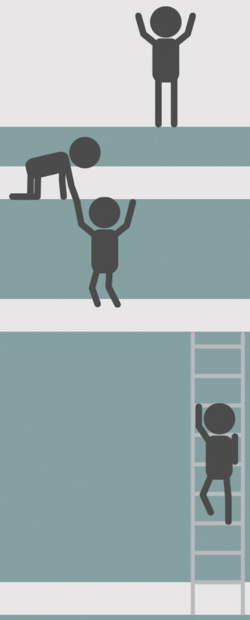

Patient-Reported Outcomes in Orthopedic Surgery: Enhancing care quality and reducing socioeconomic inequality

Joshua Michael Bonsel

Patient-Reported Outcomes in Orthopedic Surgery:
Enhancing care quality and reducing socioeconomic inequality

Joshua Michael Bonsel

The research in this thesis was primarily conducted at the department of Orthopedic Surgery & Sports Medicine at the Erasmus Medical Center, Rotterdam, The Netherlands. The research in Chapter 5 was conducted at the department of Population Health Science & Policy at the Mount Sinaï Hospital, New York, United States of America.

This research would not have been possible without the financial and intellectual support of the EuroQol Research Foundation and its enthusiastic and dedicated members.

Financial support for the printing of this thesis was generously provided by the department of Orthopedic Surgery & Sports Medicine at the Erasmus MC, the Erasmus University Rotterdam, Nederlandse Orthopaedische Vereniging (NOV), ABN AMRO, ChipSoft, Oudshoorn Chirurgische Techniek, Oxaigen, VREST, and Bauerfeind Benelux.

ISBN: 978-94-6506-804-6

Provided by thesis specialist Ridderprint, ridderprint.nl
Printing: Ridderprint
Layout and design: Tara Schollema, persoonlijkproefschrift.nl
Cover design: Joshua Bonsel

Patiënt-gerapporteerde uitkomsten in orthopedische chirurgie: verbeteren van zorgkwaliteit en verminderen van sociaal-economische ongelijkheid

Patient-Reported Outcomes in Orthopedic Surgery:
Enhancing care quality and reducing socioeconomic inequality

Proefschrift

ter verkrijging van de graad van doctor aan de

Erasmus Universiteit Rotterdam

op gezag van de

rector magnificus

Prof.dr.ir. A.J. Schuit

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

dinsdag 17 juni 2025 om 15.30 uur

door

Joshua Michael Bonsel

geboren te Rotterdam.

Erasmus University Rotterdam

Promotiecommissie

Promotor:	Prof.dr. J.A.N. Verhaar
Overige leden:	Prof.dr. D. Eygendaal
	Prof.dr. N. Gutacker
	Dr. J.A. Haagsma
Copromotor:	Dr. M. Reijman

TABLE OF CONTENTS

Chapter 1

General introduction and outline of this thesis

CONTEXT OF THIS THESIS

Patient-Reported Outcome Measures (PROMs) are self-reported questionnaires assessing the experienced burden of disease and quality of life. Over the years, PROMs have become crucial in evaluating the effectiveness of orthopedic procedures helping to quantify whether a procedure achieved the intended benefit and to what extent. For example, patients undergoing Total Hip or Knee Arthroplasty (THA, TKA) are typically asked to complete PROMs before surgery and during the one-year follow-up. This thesis demonstrates how routinely collected PROM data can be utilized to enhance the quality of care. Moreover, it aims to address methodological challenges in this context, including the optimal choice of PROM when measuring the impact on quality of life from adolescence to adulthood.

The remainder of the introduction is organized as follows. The first section provides an overview of the orthopedic procedures addressed in this thesis. The subsequent section explores the measurement and improvement of quality of care, particularly through the use of PROMs. The final section summarizes the overarching aim of this thesis and outlines the research questions.

BACKGROUND

Orthopedic surgery: evolution and current scope

The field of orthopedics has developed differently across countries. In the Netherlands, the Dutch Orthopedic Association (Dutch abbreviation: NOV) was founded in 1898[1]. Historical records indicate that early 'orthopedists' primarily focused on non-surgical treatments for malformations of the musculoskeletal system. During World War I, Dutch orthopedists expanded their scope to include surgical treatments for traumatic injuries, aligning with practice in other countries. Since then, the field of orthopedics has evolved significantly in both the conditions treated and the treatments themselves. Today, orthopedic surgery encompasses both surgical and non-surgical treatments for a wide range of musculoskeletal diseases, including acute conditions (e.g., traumatic, infectious) and chronic disorders (e.g., degenerative, oncological, congenital).

The three orthopedic treatments and procedures addressed in this thesis

This thesis focuses on three distinct areas within contemporary orthopedics, each representing a significant patient population.

Bracing therapy in Adolescent Idiopathic Scoliosis

Adolescent Idiopathic Scoliosis (AIS) is a growth disorder of the spine, that often presents during childhood or adolescence[2]. It is typically characterized by an abnormal lateral curvature of the spine, which is measured using the Cobb angle on anterior-posterior radiographs[3]. Bracing therapy is a commonly used treatment for moderate curvatures (20-45 degrees) to slow down the progression and prevent the need for surgery[4]. AIS patients, including those treated with a brace, experience a significant impact on various aspects of their lives, including pain, mental health, self-image issues, and overall quality of life[5]. The burden of this condition and its treatment extends into adulthood.

Primary THA and TKA in end-stage osteoarthritis and other conditions

THA and TKA are among the most commonly performed orthopedic procedures, primarily used to treat osteoarthritis (OA). OA, a degenerative musculoskeletal disease, significantly impacts patients' mobility, pain levels, and quality of life[6, 7]. While it mainly affects older populations, it can also affect younger individuals[8]. THA and TKA are typically performed in patients with end-stage OA that do not respond to non-surgical treatments, providing significant long-term pain relief and improved mobility[9]. Although OA is the main indication for these procedures, they are also used to treat conditions such as femoral neck fractures, posttraumatic sequelae, and osteonecrosis. The demand for THA and TKA is gradually increasing due to factors such as aging populations, rising obesity rates, and lower surgical thresholds for older patients with comorbidities.

Perioperative use of Peripheral Nerve Blocks in THA and TKA

Perioperative analgesia for THA and TKA is a crucial component of the surgical procedure. Standard multimodal protocols are available, combining oral medications and local anesthetic infiltration. Peripheral nerve block (PNB) analgesia is currently an optional addition in the Netherlands and many European countries. Its presumed benefits are pain reduction[10] and fewer post-surgical complications (e.g., venous thrombosis)[11]. In 2021, the US-based ICAROS group issued a guideline recommending the standard use of PNB in THA/TKA pain management, unless contra-indications are present. However, PNB application varies significantly across US healthcare institutions due to various factors such as clinician expertise[12]. In many European countries, including the Netherlands, rehabilitation for patients after THA and TKA follows the principles of Fast Track Surgery[13]. Multimodal pain protocols are also used, and on top of this, patients are encouraged to get out of bed and walk under the supervision of a physiotherapist on the day of surgery. Routine PNB application is not considered beneficial, as some evidence suggests that it is associated with increased falls and loss of motor control during the initial days following surgery[14, 15].

THE MEASUREMENT AND IMPROVEMENT OF QUALITY OF CARE

PROMs as reflection of the quality of care

All medical professionals strive for high-quality care, defined as care that positively impacts patients' health outcomes and that is delivered through a patient-centered process. Achieving this goal requires continuous performance evaluation and corrective actions based on suboptimal outcome metrics. Since the 1970s, the medical field has used the Donabedian framework to conceptualize quality of care, its measurement and subsequent management. This framework consists of three components: structure, process, and outcome, which are hierarchically ordered[16]. Structure refers to the healthcare delivery settings, such as provider qualifications and institutional resources. Process denotes the components of care delivered and all underlying operational and logistical procedures, such as the delivery chain of materials. Outcome refers to the observable effects of healthcare interventions on patient health status, e.g., clinical outcomes and quality of life measures. The framework assumes the structure to support the process, and the process to underpin the outcome. Donabedian assumed some structures and processes by virtue of their set-up were superior, including connected to better outcomes. Another way of thinking initiated by Cochrane gained popularity during this period, emphasizing evidence-based practice as the fundamental driver of quality improvement[17]. Cochrane also introduced economic indicators in addition to clinical indicators as important quality metrics, broadening the scope of healthcare quality assessment. However, after years of experience, these models were found to lack certain features necessary to truly improve quality of care. The Donabedian model, for example, does not acknowledge potential complex interactions between the three components and lacks opportunity for comparison at national and international levels[18]. The Cochrane model, allows evaluation of new treatments or technologies, but it offered no clear guidance as how to improve the current (standard) practice.

For these reasons, around 2000, the World Health Organization (WHO) drifted away from the Donabedian framework. The WHO introduced a quality of care model that could be universally used, regardless of the political system and the economic development of a country. Their model, described in the report ’Health Systems Performance’ starts and ends with the patient’s perspective[19]. In this model, what matters most are the patient’s experiences with the processes of the health services/providers (coined ‘responsiveness of the health system’) and the observed outcomes such as mortality and self-reported health. One may recognize the influence of the ‘patient- or person-centered care’ movement. In the following decades, the questionnaires designed to capture the patient experiences with the services/providers were coined Patient-Reported Experience Measures (PREMs), while the questionnaires to capture health outcomes were the already introduced ’PROMs’.

Important international bodies recognize this way of thinking on quality of care[20]. The framework by the Organization for Economic Cooperation and Development (OECD) is shown in Figure 1[21]. Five dimensions are deemed to represent a healthcare system’s performance: **effectiveness**, **safety** and **responsiveness**, **access to healthcare**, and **costs**. The OECD framework fully incorporates the point of view that performance is about characteristics that can be observed and/or experienced by the patient. For this thesis, it is important to add that the WHO/OECD framework recognizes the **equity** of health outcomes as a key component. In short, it addresses health differences between particular patients and population groups that are considered morally wrong and unjustified. This criterion of quality is described in more detail in the section 'PROMs in empirical research'.

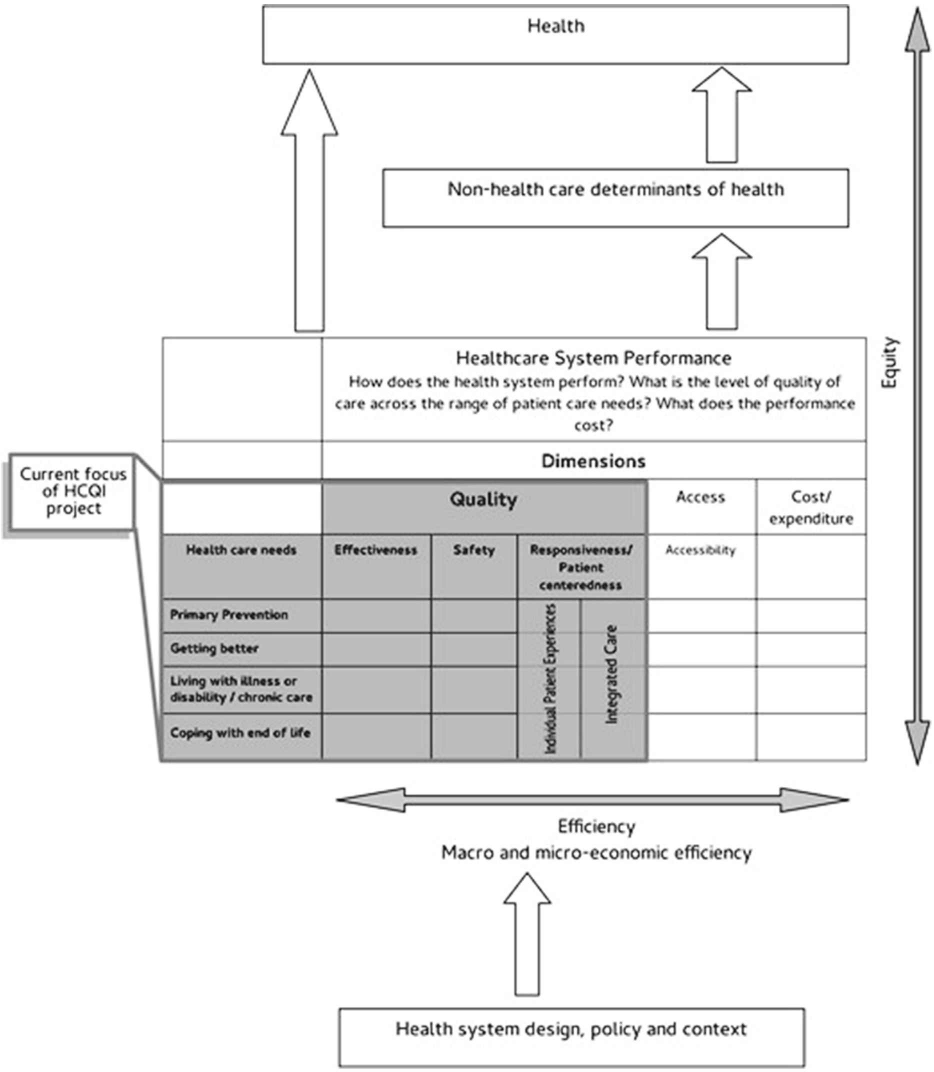

Figure 1: Framework for the evaluation of quality of care, reproduced with permission from Carinci, F. et al, on Behalf of The OECD Health Care Quality Indicators Expert Group (2015)

PROM: self-reported experience of health

A PROM is a questionnaire designed to quantify the level of health. While preferably self-reported, most PROMs allow for proxy respondents (e.g., parents, teachers), which may be necessary if the patient is too ill, too young, or cognitively/physically unable to respond to questionnaires[22]. Two types of PROMs can be (roughly) discerned. Firstly, generic PROMs, which measure health outcomes from an overarching perspective, allowing comparison between different diseases and a general judgment on severity. These measures are often multi-dimensional. Examples include measures of overall quality of life (e.g., EQ-5D[23]) or measures of well-being (e.g., WHO-5[24]). Secondly, there are disease-specific PROMs, which aim to measure health concepts, symptom burden, and functional status (e.g., Oxford Hip and Knee Score (OHS/OKS)[25]) associated with a specific disease or group of diseases.

PROMs have played a significant role in generating evidence on interventions, with generic PROMs being particularly instrumental in the exponential growth of economic evaluation. Economic evaluation data should be generalizable, as they are typically used to compare quantified benefits across diseases and conditions when prioritizing new treatments. The EQ-5D is frequently used for this purpose, with a substantial body of evidence supporting its validity and reliability in various clinical conditions. Nowadays, the WHO/OECD advises incorporating PROMs throughout health systems. The annual OECD report ‘Health at a Glance’ states, ‘A deeper understanding of quality of care requires measuring what matters to people.’ [26]. Over the past decades, the adoption of PROMs across most medical domains has been profound[27], including in the Netherlands.

PROMs in orthopedics

Over the years, professional ambition, as well as economic pressure from payers (government, insurers), has prompted orthopedic surgeons to analyze their results and quantify the impact of the treatments on patients in greater detail. Typical clinical outcome measures are adverse outcome measures, such as revision rates of joint implants, re-surgery, and mortality rates[28-30]. PROMs offer a valuable additional perspective on the impact of orthopedic treatments on important health domains relevant to the procedure, such as pain relief, improved mobility, and (joint) functioning. As joint prostheses are considered medical technologies, an economic evaluation is typically also required parallel with effectiveness measurement; hence, there is an important role for generic PROMs in this field. Also in orthopedic surgery, this role is often fulfilled by the EQ-5D, which is collected in 61% of orthopedic quality registries (QRs) (discussed below)[31]. In view of the long tradition of measuring and using PROMs, the OECD regards orthopedics as a frontrunner in this regard[26].

Collecting PROMs data: quality registries

QRs are an essential tool for collecting the abovementioned adverse event and PROM data, and increasingly, PREM data. Often initiated and managed by national medical specialty organizations, QRs collect detailed case data on patients with specific conditions or interventions. Data registration is often voluntary, but sometimes it is required as a quality indicator by the organization. Historically, the collected data consisted of specific information related to the procedure. When comprehensive, QRs serve as powerful tools for quality improvement across various levels of healthcare systems. As QRs evolve, there is a growing trend toward cross-border collaboration and alignment. These international collaborations are challenging as data processing across nations requires strict adherence to privacy regulations such as the General Data Protection Regulation (GDPR). QRs are recognized as an important quality control tool, as evidenced by upcoming legislation that provides a framework for their use and mandates their implementation in specific healthcare areas[32].

Orthopedic quality registries

Orthopedic QRs were among the first and are considered pioneers. Notable examples are the Swedish Knee and Hip Arthroplasty Register, established in 1975 and 1979. These registries had the primary aim of tracking the performance of different types of prostheses[33, 34]. Over time, these QRs have significantly improved the quality of care by identifying and discontinuing poorly performing implants.

Our thesis focuses on the Dutch Arthroplasty Register (Dutch abbreviation: LROI). The LROI is a joint implant registry founded in 2007 by the NOV. Initially, it focused on registering the most commonly performed arthroplasties, specifically THA and TKA[35]. It registers important prosthesis characteristics and a basic set of patient- and surgery-related variables, such as age, body mass index, and surgical approach. Mandatory registration has led to a coverage of >95% for these procedures. The LROI has expanded its registration to include various other joints (e.g., ankle, wrist) and procedures beyond prostheses (e.g., treatment of clubfoot).

In 2014, the LROI began collecting generic (EQ-5D[23]) and disease-specific PROMs, including pain scores, the OHS/OKS[25], the Hip dysfunction, and Osteoarthritis Outcome Score Physical Short-Form (HOOS-PS)[36], the Knee Injury and Osteoarthritis Outcome Score Physical Short-Form (KOOS-PS)[37]), as well as satisfaction and other change-related questionnaires. By doing so, it followed the example of various other orthopedic QRs[38]. Although PROM collection is not mandatory, the NOV strongly advises surgeons to collect these measures, resulting in an average nationwide response rate of 40% for PROMs before and after surgery.

The practical use of orthopedic QR data can be subdivided in a. **routine clinical care** and b. **empirical research applications**. The potential of PROMs in these applications is large. When combined with other data, PROMs offer a unique perspective, as described in the section ‘PROM: self-reported experience of health’.

Leveraging routinely collected PROMs to enhance quality of care processes

How to use outcome data to improve quality of care

The following section focuses on the use of outcome data, in particular PROMs, in orthopedics. Practically, quality can be improved through two mechanisms: a. raising the standard in all, and b. reducing practice variation (per institution, per professional). Whatever the target, we assume adequate systems and data processing such as case-mix adjustment.

From signal to action

It is practical to distinguish between health system 'levels' at which PROMs may impact quality of care. Here we cite the Alberta PROMs and EQ-5D Research and Support Unit (APERSU, Canada) framework which categorizes the potential impact of PROMs at three levels[39]. Note that this framework also acknowledges other outcomes than PROMs.

a) Micro-level: this refers to using PROMs to improve health at the individual patient level. Examples include providing feedback of PROM results to patients and/or providers to improve communication, tools to facilitate shared decision-making, monitoring treatment success, and alerting providers to symptoms that require treatment. The LROI, for example, has recently launched a shared decision-making tool ('Patients Like Me') that provides patients with insights into their expected results following arthroplasty[40].

b) Meso-level: at this level, aggregate PROM data could be used within a surgical unit or hospital to continuously evaluate the outcomes of patient groups and providers. Examples include identifying groups with suboptimal outcomes and comparing outcomes within a surgical unit to identify areas for improvement.

c) Macro-level: This level is similar to the meso-level but operates at the hospital, regional, or national level. A well-known example is using PROM data for 'benchmarking', where aggregate PROM results are used to solicit feedback, analyze contributing factors, and drive subsequent improvement. The National Health Service's PROMs program aimed to promote the use of PROM data for this purpose in several surgical fields, including orthopedics[41]. Another example is using PROMs data in a plan-do-check-act (PDCA) cycle. A PDCA cycle is a framework to guide the continuous improvement of healthcare and services provided[42].

Several systematic reviews aimed to determine the effectiveness of these interventions, however, most are of older date. In general, evidence at the micro-level supports its use in terms of effectiveness; however, at both the meso- and macro-levels, evidence is scarce and has not shown a beneficial impact[43-50]. Also, it should be noted that the effective and efficient translation of suboptimal PROM results into actions by the healthcare actors is still in its infancy. This thesis reviews contemporary literature and searches for novel applications in orthopedics and other medical specialties within this framework.

PROMs in empirical research

This section describes the use of PROMs along other data to address contemporary clinical issues in THA and TKA patients. Three examples show the utility of these data for quality improvement, through an underlying data analysis specifically developed for the issue (yet seemingly unsuitable for routine use). These examples are: socioeconomic inequalities in PROMs, inequalities in access to PNBs in perioperative care, and the impact of COVID-19 on PROMs.

Socioeconomic inequalities in PROMs in THA and TKA

QR data is vital for identifying which patients benefit most from procedures compared to those who do not, based on patient and surgical variables. A strong focus has been placed on the so-called justifiable variables within the LROI and beyond. 'Justifiable' variables are typically associated with differences in outcomes, but these differences are not viewed as unjust as they can often be explained. For example, a recent study highlights the impact of smoking on the risk of revision and mortality following THA or TKA[51]. Other examples include body mass index and comorbidities.

'Unjustifiable' variables are also often associated with differences in outcomes, even after adjusting for 'justifiable' variables. These differences are viewed as unjust because there is typically no sufficient (biological) explanation to account for the observed disparity[52]. These differences are referred to as 'inequalities'. The PROGRESS-plus variables are a fundamental set of social determinants considered driving forces of inequalities[53]. The core set covers place of residence, race/ethnicity, occupation, gender/sex, religion, education, socioeconomic status, and social capital. Over time, personal characteristics associated with discrimination (e.g., age) and features of relationships (e.g., smoking parents) have been added. For certain variables, determining whether differences are unjustified requires careful investigation, as biological explanations may play a relatively stronger role (e.g., age). Inequalities based on these variables pose a significant challenge in most medical and public health domains across all countries. Besides being considered unjust in general, they also hinder economic development[26, 54].

The Commission on Social Determinants of Health (CSDH) of the WHO has developed a framework illustrating how socioeconomic determinants impact health outcomes (Figure 2)[55]. This framework highlights the complex relationship between PROGRESS-plus variables, intermediary determinants (e.g., health behaviors), the healthcare system, and finally, health outcomes.

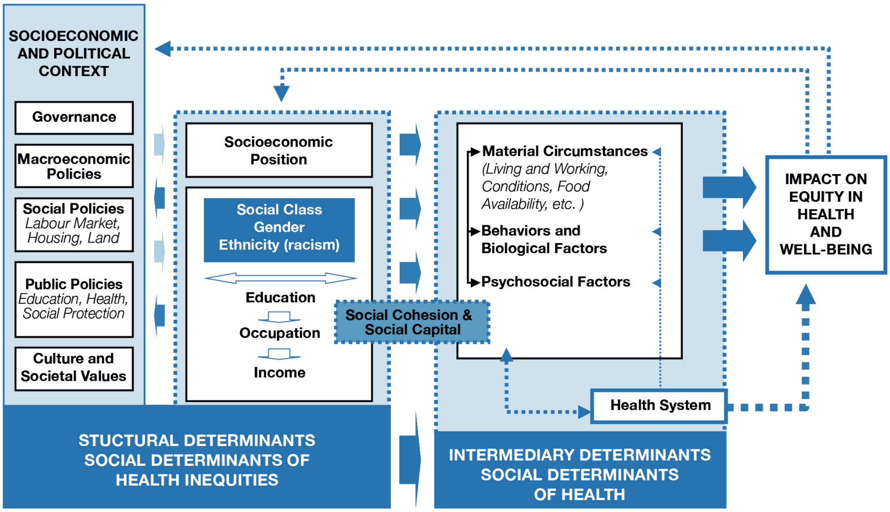

Figure 2: Framework of the WHO on social determinants and their impact on health, reproduced with permission from Solar, O. et al. (2010)

Contemporary evidence in orthopedics has convincingly demonstrated the relation between place of residence, race/ethnicity, and socioeconomic status (SES), and the effectiveness of procedures[56-58]. For example, patients with lower education, lower socioeconomic status, or belonging to a minority group are readmitted and experience complications more often following orthopedic procedures. Older age and male sex are also known associates with poorer outcomes following arthroplasty[59]. Less focus has been placed on the other PROGRESS-plus variables. Including PROMs in QRs has created opportunities to investigate inequalities in the burden of disease and quality of life, providing additional insights into underlying mechanisms. Therefore, this thesis utilizes LROI data to study the presence of socioeconomic inequalities in PROMs of THA and TKA patients.

Our understanding of intermediary determinants and the role of the healthcare system in these inequalities is limited. The LROI (and most other orthopedic QRs) lacks information on determinants such as treatment advice, comorbidities, waiting times, traveling distance, and health literacy. Of course, registering each procedure is labor-intensive for the surgeon, and filling out PROMs places a high demand on patients. As such, the burden of collecting additional information should be weighed against the expected benefits from collecting it.

Practice variation in the perioperative use of PNB in THA and TKA

Evidence in orthopedics has also consistently shown the relationship between patients' place of residence, race/ethnicity, and SES and the access to and utilization of surgical procedures[56-58]. For example, studies have shown differential rates of THA and TKA across socioeconomic groups, with lower utilization observed in more socioeconomically deprived populations. Female patients have been found to receive fewer THA and TKA relative to their estimated need compared to male patients[60]. We may assume this reflects a disadvantage for females, as the effectiveness of THA and TKA is undisputed.

At a deeper level, such practice variation may also be present in the use of PNBs in THA and TKA. US-based private insurance databases have shown that deprived and minority groups receive significantly fewer PNBs after adjusting for potential confounders[15, 61]. If PNBs are effective, the differential use can be considered a health inequality. This thesis investigated PNB practice variation within a public national insurance program (US-based Medicare (insurance) data), trying to unravel whether patient rather than hospital or provider factors were the drivers for variation. Subsequently, we investigated whether PNB use was associated with improved outcomes, which would provide additional support for the statement that practice variation is an inequality.

Impact of the COVID-19 pandemic on PROMs in THA and TKA patients

Finally, PROMs data offer the opportunity to study the impact of public health phenomena or policy implementations, such as the COVID-19 pandemic. The COVID-19 pandemic was an unprecedented challenge to all healthcare systems, with far-reaching consequences[62]. Large shifts in healthcare resources were needed. Elective care, including orthopedic procedures, was reduced to a minimum, particularly in times of high COVID-19 infections and the number of hospital beds occupied[63]. The loss of quality-adjusted life years (QALYs) due to the postponement of primary and revision THA and TKA was estimated to be large[64]. Other mechanisms may be at play in addition to loss of QALYs due to postponing care. For example, physiotherapists were also obligated to close during lockdowns, and hence, it is conceivable that recovery after THA or TKA during the COVID-19 pandemic was worse compared to pre-pandemic periods. This thesis utilizes QR data (LROI) to investigate this phenomenon. The findings can guide future policy decisions, particularly in balancing public health measures with the continued need for essential elective procedures.

Challenges of PROMs

Although PROMs have changed the landscape of modern orthopedics, numerous methodological challenges have come to light over the years. This thesis addresses several of these:

a) Lifespan consistency: Attention has been drawn to measuring the quality of life consistently over the entire lifespan. The EuroQol Research Foundation has developed a child-specific variant of the EQ-5D (EQ-5D-Y) with more child-specific questions and answers[65, 66]. While the use of EQ-5D-Y is advised for ages 8-11, it's unclear whether the EQ-5D or EQ-5D-Y is more appropriate for adolescents aged 12-17[67]. This issue plays a role in orthopedic diseases that develop during childhood or adolescence, with an impact on quality of life persisting into adulthood, such as Adolescent Idiopathic Scoliosis (AIS).

b) Selection bias: Patients who complete PROMs may differ systematically from those who don't, typically resulting in decreased observed differences[68].

c) Reporting heterogeneity: PROMs, like all other types of measures, are subject to measurement error. One specific type of measurement error is reporting heterogeneity, which refers to the systematic differences in reporting based on factors unrelated to the treatment itself. For example, well-educated patients have been shown to report a lower quality of life, on average, than their actual or 'true' level of quality of life[69].

Other methodological considerations exist both within and outside of orthopedics:

d) Sensitivity to change: PROMs may struggle to detect meaningful changes in a patient's condition.

e) Response shift: Patients' internal standards or perceptions of their health may change over time, affecting the interpretation of longitudinal data.

f) Mode of administration: The method of PROM collection (e.g., paper-based vs. electronic) can affect response rates and data quality.

g) Timing of collection: Timing of PROMs collection may influence results.

AIMS AND RESEARCH QUESTIONS

This thesis aims to identify opportunities for improved clinical practice in orthopedics using existing PROM data while developing our knowledge of the use of PROMs within this field.

Chapter 2 provides an overview and synthesizes all available evidence of PROMs applications in routine clinical care according to the framework by APERSU[39]. **Chapter 3** uses LROI data to study the presence of SES inequalities in PROMs in THA and TKA patients. **Chapter 4** uses the same data; however, it zooms in on whether the magnitude of inequalities differs across various PROMs and if specific domains are more affected. **Chapter 5** uses Medicare insurance data to study inequalities in the perioperative use of PNBs in THA and TKA patients. **Chapter 6** uses LROI data to study the impact of COVID-19 on PROMs in THA and TKA patients. **Chapter 7** addresses an important methodological consideration of measuring quality of life from adolescence to adulthood. In **Chapter 8**, we will discuss the findings from **Chapters 2 to 7** and identify future directions.

The following research questions were defined for **Chapters 2 to 7**:

1) How are PROMs effectively used to improve the quality of care, and which characteristics of PROMs-like interventions determine their effectiveness? (**Chapter 2**)

2) Is SES associated with PROMs in THA and TKA, and via which mechanisms? (**Chapter 3**)

3) How do SES inequalities vary across different PROMs, and which domains are most affected in THA and TKA patients? (**Chapter 4**)

4) What is the role of socioeconomic, patient, and hospital characteristics in the utilization of PNBs in THA and TKA patients? (**Chapter 5**)

5) Has the COVID-19 pandemic impacted PROMs in THA and TKA, and via which mechanisms? (**Chapter 6**)

6) Is a child-specific or an adult quality of life instrument more suitable for assessing an adolescent population with adolescent idiopathic scoliosis? (**Chapter 7**)

REFERENCES

1. Dutch Orthopedic Society (NOV), available at: https://www.orthopeden.org/media/svof4hio/jubileumboek_125_jaar_nov_sp_3.pdf, accessed 20-07-2024.
2. JEH P. School screening for scoliosisproefschrift. Utrecht: Universiteit Utrecht. 1996.
3. Weinstein SL, Ponseti IV. Curve progression in idiopathic scoliosis. J Bone Joint Surg Am 1983; 65: 447-455.
4. Sanders JO, Newton PO, Browne RH, Katz DE, Birch JG, Herring JA. Bracing for idiopathic scoliosis: how many patients require treatment to prevent one surgery? J Bone Joint Surg Am 2014; 96: 649-653.
5. Tones M, Moss N, Polly DW, Jr. A review of quality of life and psychosocial issues in scoliosis. Spine (Phila Pa 1976) 2006; 31: 3027-3038.
6. Yan H, Guo J, Zhou W, Dong C, Liu J. Health-related quality of life in osteoarthritis patients: a systematic review and meta-analysis. Psychol Health Med 2022; 27: 1859-1874.
7. Zhang W, Nuki G, Moskowitz RW, Abramson S, Altman RD, Arden NK, et al. OARSI recommendations for the management of hip and knee osteoarthritis: part III: Changes in evidence following systematic cumulative update of research published through January 2009. Osteoarthritis Cartilage 2010; 18: 476-499.
8. Ackerman IN, Kemp JL, Crossley KM, Culvenor AG, Hinman RS. Hip and Knee Osteoarthritis Affects Younger People, Too. J Orthop Sports Phys Ther 2017; 47: 67-79.
9. Kamaruzaman H, Kinghorn P, Oppong R. Cost-effectiveness of surgical interventions for the management of osteoarthritis: a systematic review of the literature. BMC Musculoskelet Disord 2017; 18: 183.
10. Richman JM, Liu SS, Courpas G, Wong R, Rowlingson AJ, McGready J, et al. Does Continuous Peripheral Nerve Block Provide Superior Pain Control to Opioids? A Meta-Analysis. Anesthesia & Analgesia 2006; 102.
11. Memtsoudis SG, Cozowicz C, Bekeris J, Bekere D, Liu J, Soffin EM, et al. Peripheral nerve block anesthesia/analgesia for patients undergoing primary hip and knee arthroplasty: recommendations from the International Consensus on Anesthesia-Related Outcomes after Surgery (ICAROS) group based on a systematic review and meta-analysis of current literature. Reg Anesth Pain Med 2021; 46: 971-985.
12. Cozowicz C, Poeran J, Zubizarreta N, Mazumdar M, Memtsoudis SG. Trends in the Use of Regional Anesthesia: Neuraxial and Peripheral Nerve Blocks. Reg Anesth Pain Med 2016; 41: 43-49.
13. Berg U, A WD, Rolfson O, Nauclér E, Sundberg M, Nilsdotter A. Influence of fast-track programs on patient-reported outcomes in total hip and knee replacement (THR/TKR) at Swedish hospitals 2011-2015: an observational study including 51,169 THR and 8,393 TKR operations. Acta Orthop 2020; 91: 306-312.
14. Ilfeld BM, Duke KB, Donohue MC. The association between lower extremity continuous peripheral nerve blocks and patient falls after knee and hip arthroplasty. Anesth Analg 2010; 111: 1552-1554.
15. Memtsoudis SG, Danninger T, Rasul R, Poeran J, Gerner P, Stundner O, et al. Inpatient falls after total knee arthroplasty: the role of anesthesia type and peripheral nerve blocks. Anesthesiology 2014; 120: 551-563.
16. Donabedian A. Evaluating the quality of medical care. 1966. Milbank Q 2005; 83: 691-729.
17. Cochrane AL. Effectiveness and Efficiency: Random Reflections on Health Services, Nuffield Provincial Hospitals Trust 1972.

18. (2024) O. Rethinking Health System Performance Assessment: A Renewed Framework, OECD Health Policy Studies, OECD Publishing, Paris, https://doi.org/10.1787/107182c8-en.
19. Christopher J.L. Murray DBE. Health Systems Performance, comparing health outcomes in about 100 countries, World Health Organization 2003.
20. Institute of Medicine Committee on Quality of Health Care in A. Crossing the Quality Chasm: A New Health System for the 21st Century. 2001.
21. Carinci F, Van Gool K, Mainz J, Veillard J, Pichora EC, Januel JM, et al. Towards actionable international comparisons of health system performance: expert revision of the OECD framework and quality indicators. International Journal for Quality in Health Care 2015; 27: 137-146.
22. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. BJA Education 2017; 17: 137-144.
23. Brooks R. EuroQol: the current state of play. Health Policy 1996; 37: 53-72.
24. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: A systematic review of the literature. Psychotherapy and Psychosomatics 2015; 84: 167-176.
25. Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, Dawson J. The use of the Oxford hip and knee scores. J Bone Joint Surg Br 2007; 89: 1010-1014.
26. OECD. Health at a Glance2023.
27. Black N. Patient reported outcome measures could help transform healthcare. Bmj 2013; 346: f167.
28. Berstock JR, Beswick AD, López-López JA, Whitehouse MR, Blom AW. Mortality After Total Knee Arthroplasty: A Systematic Review of Incidence, Temporal Trends, and Risk Factors. J Bone Joint Surg Am 2018; 100: 1064-1070.
29. Labek G, Thaler M, Janda W, Agreiter M, Stöckl B. Revision rates after total joint replacement: cumulative results from worldwide joint register datasets. J Bone Joint Surg Br 2011; 93: 293-297.
30. Berstock JR, Beswick AD, Lenguerrand E, Whitehouse MR, Blom AW. Mortality after total hip replacement surgery: A systematic review. Bone Joint Res 2014; 3: 175-182.
31. Rolfson O, Bohm E, Franklin P, Lyman S, Denissen G, Dawson J, 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 Suppl 1: 9-23.
32. https://www.eerstekamer.nl/wetsvoorstel/36278_wet_kwaliteitsregistraties.
33. Kärrholm J. The Swedish Hip Arthroplasty Register (www.shpr.se). Acta Orthop 2010; 81: 3-4.
34. Robertsson O, Ranstam J, Sundberg M, A WD, Lidgren L. The Swedish Knee Arthroplasty Register: a review. Bone Joint Res 2014; 3: 217-222.
35. van Steenbergen LN, Denissen GA, Spooren A, van Rooden SM, van Oosterhout FJ, Morrenhof JW, Nelissen RG. More than 95% completeness of reported procedures in the population-based Dutch Arthroplasty Register. Acta Orthop 2015; 86: 498-505.
36. Nilsdotter AK, Lohmander LS, Klässbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS)--validity and responsiveness in total hip replacement. BMC Musculoskelet Disord 2003; 4: 10.
37. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998; 28: 88-96.

38. Wilson I, Bohm E, Lübbeke A, Lyman S, Overgaard S, Rolfson O, et al. Orthopaedic registries with patient-reported outcome measures. EFORT Open Rev 2019; 4: 357-367.
39. Al Sayah F, Lahtinen M, Bonsel GJ, Ohinmaa A, Johnson JA. A multi-level approach for the use of routinely collected patient-reported outcome measures (PROMs) data in healthcare systems. Journal of Patient-Reported Outcomes 2021; 5: 98.
40. Patients like me [Patienten zoals ik], Dutch Orthopedic Society (NOV) and Dutch Arthroplasty Registry (LROI), available at: https://www.zorgvoorbeweging.nl/patienten-zoals-ik, accessed 10-07-2024.
41. Boyce MB, Browne JP. The effectiveness of providing peer benchmarked feedback to hip replacement surgeons based on patient-reported outcome measures--results from the PROFILE (Patient-Reported Outcomes: Feedback Interpretation and Learning Experiment) trial: a cluster randomised controlled study. BMJ Open 2015; 5: e008325.
42. Evidence-based practice manual: Research and outcome measures in health and human services. New York, NY, US, Oxford University Press 2004.
43. Boyce MB, Browne JP. Does providing feedback on patient-reported outcomes to healthcare professionals result in better outcomes for patients? A systematic review. Qual Life Res 2013; 22: 2265-2278.
44. Dorr MC, van Hof KS, Jelsma JGM, Dronkers EAC, Baatenburg de Jong RJ, Offerman MPJ, de Bruijne MC. Quality improvements of healthcare trajectories by learning from aggregated patient-reported outcomes: a mixed-methods systematic literature review. Health Res Policy Syst 2022; 20: 90.
45. Espallargues M, Valderas JM, Alonso J. Provision of feedback on perceived health status to health care professionals: a systematic review of its impact. Med Care 2000; 38: 175-186.
46. Gibbons C, Porter I, Gonçalves-Bradley DC, Stoilov S, Ricci-Cabello I, Tsangaris E, et al. Routine provision of feedback from patient-reported outcome measurements to healthcare providers and patients in clinical practice. Cochrane Database Syst Rev 2021; 10: CD011589.
47. Greenhalgh J, Gooding K, Gibbons E, Dalkin S, Wright J, Valderas J, Black N. How do patient reported outcome measures (PROMs) support clinician-patient communication and patient care? A realist synthesis. J Patient Rep Outcomes 2018; 2: 42.
48. Ishaque S, Karnon J, Chen G, Nair R, Salter AB. A systematic review of randomised controlled trials evaluating the use of patient-reported outcome measures (PROMs). Qual Life Res 2019; 28: 567-592.
49. Luckett T, Butow PN, King MT. Improving patient outcomes through the routine use of patient-reported data in cancer clinics: future directions. Psychooncology 2009; 18: 1129-1138.
50. Valderas JM, Kotzeva A, Espallargues M, Guyatt G, Ferrans CE, Halyard MY, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. Qual Life Res 2008; 17: 179-193.
51. Bongers J, Belt M, Spekenbrink-Spooren A, Smulders K, Schreurs BW, Koeter S. Smoking is associated with higher short-term risk of revision and mortality following primary hip or knee arthroplasty: a cohort study of 272,640 patients from the Dutch Arthroplasty Registry. Acta Orthop 2024; 95: 114-120.
52. Phelan JC, Link BG, Tehranifar P. Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. J Health Soc Behav 2010; 51 Suppl: S28-40.
53. Oliver S, Kavanagh J, Caird J, Lorenc T, Oliver K, Harden A, et al. Health Promotion, Inequalities and Young People’s health: a Systematic Review of Research. 2008.
54. Marmot M. Social determinants of health inequalities. The Lancet 2005; 365: 1099-1104.

55. World Health O. A conceptual framework for action on the social determinants of health. Geneva: World Health Organization 2010.
56. Alvarez PM, McKeon JF, Spitzer AI, Krueger CA, Pigott M, Li M, Vajapey SP. Socioeconomic factors affecting outcomes in total knee and hip arthroplasty: a systematic review on healthcare disparities. Arthroplasty 2022; 4: 36.
57. Alvarez PM, McKeon JF, Spitzer AI, Krueger CA, Pigott M, Li M, Vajapey SP. Race, Utilization, and Outcomes in Total Hip and Knee Arthroplasty: A Systematic Review on Health-Care Disparities. JBJS Rev 2022; 10.
58. Karimi A, Burkhart RJ, Hecht CJ, 2nd, Acuña AJ, Kamath AF. Is Social Deprivation Associated With Usage, Adverse Events, and Patient-reported Outcome Measures in Total Joint Arthroplasty? A Systematic Review. Clin Orthop Relat Res 2023; 481: 239-250.
59. Santaguida PL, Hawker GA, Hudak PL, Glazier R, Mahomed NN, Kreder HJ, et al. Patient characteristics affecting the prognosis of total hip and knee joint arthroplasty: a systematic review. Can J Surg 2008; 51: 428-436.
60. Gillian AH, James GW, Peter CC, Williams JI, Bart H, Richard G, Elizabeth MB. Differences between Men and Women in the Rate of Use of Hip and Knee Arthroplasty. New England Journal of Medicine 2000; 342: 1016-1022.
61. Memtsoudis SG, Poeran J, Zubizarreta N, Rasul R, Opperer M, Mazumdar M. Anesthetic Care for Orthopedic Patients: Is There a Potential for Differences in Care? Anesthesiology 2016; 124: 608-623.
62. Santomauro DF, Mantilla Herrera AM, Shadid J, Zheng P, Ashbaugh C, Pigott DM, et al. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021; 398: 1700-1712.
63. Dutch Coöperating Quality Registries (SKR) Impact Report 2021. Available at: https://skr-zorg.nl/impact-report/.
64. Latijnhouwers D, Pedersen A, Kristiansen E, Cannegieter S, Schreurs BW, van den Hout W, et al. No time to waste; the impact of the COVID-19 pandemic on hip, knee, and shoulder arthroplasty surgeries in the Netherlands and Denmark. Bone Jt Open 2022; 3: 977-990.
65. Ravens-Sieberer U, Wille N, Badia X, Bonsel G, Burström K, Cavrini G, et al. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. Qual Life Res 2010; 19: 887-897.
66. Kreimeier S, Åström M, Burström K, Egmar AC, Gusi N, Herdman M, et al. EQ-5D-Y-5L: developing a revised EQ-5D-Y with increased response categories. Qual Life Res 2019; 28: 1951-1961.
67. EuroQol Research Foundation. https://euroqol.org/publications/user-guides/, accessed November 2023.
68. van Schie P, van Bodegom-Vos L, Zijdeman TM, Gosens T, Nelissen R, Marang-van de Mheen PJ, Group IQJS. Linking Patient-Reported Outcome Measure Scores to Adverse Event Data to Gain Insight into Overestimation of Postoperative Patient-Reported Outcome Measure Improvement After Total Hip Arthroplasty and Total Knee Arthroplasty Due to Selective Nonresponse. J Arthroplasty 2024.
69. d'Uva TB, O'Donnell O, van Doorslaer E. Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans. Int J Epidemiol 2008; 37: 1375-1383.

Chapter 2

The use of patient-reported outcome measures to improve patient-related outcomes – a systematic review

Authors

Joshua M. Bonsel[1], MD

Ademola J. Itiola[2], MPH, MSc

The abovementioned 2 authors shared equal first authorship.

Anouk S. Huberts[3], MD

Gouke J. Bonsel[4], MD, PhD, emeritus professor

Hannah Penton[5], MSc, PhD

Affiliations

[1] Department of Orthopaedics and Sports Medicine, Erasmus MC, University Medical Center, Rotterdam, The Netherlands

[2] School of Public Health, University of Alberta, Edmonton, Alberta, Canada

[3] Department of Quality and Patientcare, Erasmus MC, University Medical Center, Rotterdam, The Netherlands

[4] EuroQol Research Foundation, Rotterdam, The Netherlands

[5] OPEN Health Evidence & Access, Rotterdam, The Netherlands

Status

Published.

Citation: Bonsel JM, Itiola AJ, Huberts AS, Bonsel GJ, Penton H. The use of patient-reported outcome measures to improve patient-related outcomes - a systematic review. Health Qual Life Outcomes. 2024;22(1):101.

ABSTRACT

Background

Patient-reported outcome measures (PROMs) provide invaluable information on patients' health outcomes and can be used to improve patient-related outcomes at the individual, organizational and policy levels. This systematic review aimed to a) identify contemporary applications and synthesize all evidence on the use of PROMs in these contexts and b) to determine characteristics of interventions associated with increased effectiveness.

Methods

Five databases were searched for studies providing quantitative evidence of the impact of PROM interventions. Any study design was permitted. An overall benefit (worsening) in outcome was defined as a statistically significant improvement (deterioration) in either a PROM, patient-reported experience measure or clinical outcome. Study quality was assessed using the Effective Public Healthcare Panacea Project's Quality Assessment Tool for Quantitative Studies. A narrative synthesis was conducted.

Results

Seventy-six studies of the 11,121 articles identified met the inclusion criteria. At the individual level, 10 (43%) of 23 studies that fed back PROMs to the patient or healthcare provider showed an improvement in outcome. This percentage increased in studies which used PROMs to monitor disease symptoms and linked these to care-pathways: 17 (68%) of 25 studies using this mechanism showed an improvement. Ten (71%) of 14 studies using PROMs to screen for disease found a benefit. The monitoring and screening approach was most effective using PROMs covering cancer-related, depression and gastro-intestinal symptoms. Three studies found that the mere collection of PROMs resulted in improved outcomes. Another three studies used PROMs in decision aids and found improved decision quality. At the organizational/policy level, none of the 4 studies that used PROMs for benchmarking found a benefit. The three studies that used PROMs for in-depth performance analyses and 1 study in a plan-do-study-act (PDCA) cycle found an improvement in outcome. Studies employing disease-specific PROMs tended to observe improved outcomes more often. There are concerns regarding the validity of findings, as studies varied from weak to moderate quality.

Conclusions

The use of PROMs at the individual level has matured considerably. Monitoring/screening applications seem promising particularly for diseases for which treatment algorithms rely on the experienced symptom burden by patients. Organizational/policy-level application is in its infancy, and performance evaluation via in-depth analyses and PDCA-cycles may be useful. The findings of this review may aid stakeholders in the development and implementation of PROM-interventions which truly impact patient outcomes.

BACKGROUND

Patient-reported outcomes measures (PROMs) are considered an invaluable tool to capture information on patients' health outcomes, including expectations and values. Two types of PROMs exist, namely generic and disease-specific PROMs[1]. Generic PROMs aim to measure a health outcome from an overarching perspective, allowing for comparison between different diseases and a general judgement on the severity. These measures are often multi-dimensional; examples include measures of overall Quality of Life (e.g., EQ-5D) or well-being (e.g., WHO-5) [2, 3]. Disease-specific PROMs aim to measure these concepts, the symptom burden and functional status associated with a disease or a group of diseases[4].

PROMs were introduced to complement clinical outcome measures in studies assessing the (cost-)effectiveness of new clinical interventions. However, their application has broadened, including the role as outcome indicator in clinical practice alongside traditional indicators such as mortality and prevalence/incidence[5]. This movement is adopted by medical science and leading institutions like the Organisation for Economic Co-operation and Development, which conform to the principle that assessing health system performance starts by assessment of patient-related outcomes[6]. It is pragmatic to distinguish three levels of intended use: the individual (micro-), organizational (meso-) and policy (macro-) level[7].

At the micro-level, PROMs are used at the patient-encounter level. Several systematic reviews revealed evidence that using PROMs at the micro-level has a modest beneficial impact on patient-related outcomes[8-15]. The key idea is that a patient fills out a PROM once or multiple times, and the results are fed back to the patient or clinician[15]. Greenhalgh et al. has outlined the underlying theory how PROMs may be useful at this level: the feedback of PROMs may alter the decision-making process, and initiate a change to clinical practice[16]. Several examples exist: firstly, the feedback of PROMs to patient and provider can aid in communicating symptoms which may otherwise remain unnoticed[17, 18]. Another example are novel digital patient-decision systems using PROMs, which develop rapidly parallel to digital technology (e.g., apps, e-portals, and dashboards) [19].

Aggregated PROMs can be used to inform the healthcare system at the organizational (meso-) and health system (macro-) level, respectively. Evidence of the impact of PROMs use at the meso-/macro-level is scarce, and a recent review did not find a clear impact on patient outcomes[8, 20]. The key idea at this level is that aggregated PROMs can guide the (continuous) improvement of healthcare provided by a group of clinicians, hospital or even country[21]. Their role in orthopedic surgery may illustrate their potential. At the meso-level, an orthopedic surgery unit in a hospital may use PROMs to improve local policy on eligibility criteria for surgical treatment, to rationalize pain killing strategies, or to compare performance across surgeons on a monthly basis[22]. At the macro-level, PROMs results according to hospital, region, nation, or otherwise may be presented in a standardized form (both in epidemiological and graphical meaning), inviting for a process of feedback, analysis of drivers, and if possible subsequent improvement[21]. This mechanism is often referred to as benchmarking and is thought to demonstrate performance differences among providers, facilitate more in-depth clinical audits, and inform decision-making, and is a potentially effective method to improve the quality of care[23, 24]. An example which aimed to encourage benchmarking is the NHS-programme in the UK on certain surgical procedures. This program publicly published PROMs for varicose vein, groin hernia, and hip and knee arthroplasty surgery; as of 2017 PROMs are only collected for hip/knee surgery[25]. This program also aimed to incentivize patients to select the assumed best provider, however, available evidence does not support this pathway[21, 26].

We think a contemporary review is warranted because it remains unknown why certain PROMs-interventions are more effective than others[8, 11]. Certain mechanisms underpinning the interventions may contribute to increased effectiveness. For example, a critical step to transform a suboptimal PROM level, i.e. a patient value below a particular threshold, into an improved outcome may be to link this observation to a care pathway. The doctor may receive an alert inviting her/him to check the situation. This approach seems promising in disease areas where symptom monitoring along with treatment tailoring is common practice, e.g., gastroenterology, rheumatology, and oncology[27, 28].

In this systematic review, we aim to identify contemporary evidence of the impact of the use of PROMs at the micro-, meso- and macro-level on patient outcomes. Our second aim is to identify and describe characteristics of the intervention and PROMs used which may contribute to an increased chance for success.

METHODS

The present systematic review was registered in PROSPERO under record 2022 CRD42022333400. This review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (2020) when applicable[29].

Data sources and search strategy

The following databases were searched: MEDLINE, Embase, Web of Science Core Collection, Cochrane CENTRAL Register of trials, and Google Scholar from database inception to August 24, 2023 for studies that reported the use of PROMs to improve quality of care. The final search was developed and refined through an iterative process and consisted of 3 blocks, namely: (a) various terms for PROMs, (b) various terms for quality, effectiveness and outcomes, and (c) mechanisms through which PROMs may be used to benefit healthcare (e.g., feedback, monitoring, dashboards and plan-do-check-act (PDCA) cycles) (Supplemental

File 1). A PDCA-cycle is a commonly used framework to guide the continuous improvement of healthcare and services provided[30]. Additional studies were identified by screening the references of included articles.

Study selection

Studies were eligible that (a) provided evidence on the impact of an intervention, (b) using a previously validated PROM, (c) which reported at least one quantitative outcome per the definition described below. Any study design was permitted. Studies were excluded if (a) the full-text could not be retrieved and/or only a conference abstract was available; (b) the study was conducted as a pilot; (c) there was no comparator or pre-intervention comparison; (d) the PROM was used to select patients for another type of intervention; (e) the article was not available in English. Two reviewers (JB and AI) independently screened all titles and abstracts obtained from the search and applied the inclusion criteria to eligible studies. Any disagreements regarding the inclusion of studies was discussed between the two reviewers and were resolved by consensus.

Outcome definition

We defined the potential impact of a PROM-intervention on patient-related outcomes using the Donabedian framework[31]. To evaluate the quality of healthcare or impact of an intervention, contemporary guidelines place emphasis on outcome measures which reflect the impact on the health status of patients[32]. Typically, these outcomes are of quantitative nature and are collected at the patient-level. We discerned three types of outcomes measured based on previous reviews, namely (1) PROMs, (2) patient reported experiences measures (PREMs) and (3) clinical outcomes. Outcome measures were categorized according to the dimensions/items into overarching groups based on the identified studies, e.g., Health-Related Quality of Life (HRQoL), physical functioning, mental functioning, and symptom burden. Similarly, this was done for PREMs (e.g., satisfaction) and clinical outcomes (e.g., readmissions).

A study was judged to have found an overall benefit (or a detriment/harm) if any of the above-mentioned outcomes improved (worsened) up to statistical significance. As patient-related outcomes may be specific to the intended use and medical domain, we did not attribute weight to a specific type of outcome. Studies often contained multiple comparisons through analysis of dimensions or even items separately. This approach inflates testing, increasing the potential of a type I error. Therefore, we required at least 2 subdomain/single-items to reach statistical significance to qualify the impact as a benefit or detriment, unless outcomes were defined as primary outcome a priori.

In accordance with previous reviews, process of care measures (e.g., number of symptoms discussed) were extracted, but were considered to mediate outcomes described above[14].

Data extraction and quality assessment

The following data were extracted from eligible studies by one of the reviewers (JB or AI): authors, country, setting, study design, sample, PROMs used, description of intervention using PROMs, co-interventions, training offered on the intervention and/or interpretation of PROM, all primary and secondary outcome measures and their quantification.

Two reviewers (JB and AH) independently assessed the methodological quality of included studies using the Effective Public Healthcare Panacea Project's Quality Assessment Tool for Quantitative Studies[33]. The tool was considered the most appropriate for this systematic review as it covers various study designs and public health interventions. Domains assessed using the tool included selection bias, study design, confounders, blinding, data collection methods, and withdrawals and drop-outs. Each domain was rated as 1 (strong), 2 (moderate) or 3 (weak). A global score was calculated, in which strong = no weak ratings, moderate = 1 weak rating, and weak = two or more weak ratings.

Data synthesis

A narrative synthesis was conducted as a formal meta-analysis appeared not possible at an early stage due to the heterogeneity of study designs and outcomes reported. Overall, the synthesis was split up by the micro- and meso-/macro-level. The impact of PROMs interventions was assessed by four possible determinants for increased effectiveness. The applications were categorized into mechanisms applied based on commonalities between PROMs interventions. Subsequently, we captured a broader perspective by determining the impact of PROMs interventions by the medical domain, the type of PROM used in the intervention, and by the separate outcome dimensions used to measure the effect of the intervention. For the latter, we decided to only present those which were measured in at least 3 studies. We discerned studies which used the same PROM outcome as in the intervention from studies which (only) used different outcomes. Finally, for each determinant and outcome dimension, the average quality of studies was calculated.

RESULTS

The PRISMA diagram depicting the selection process is presented in Figure 1. A total of 18,652 records were identified. After removing duplicates, 11,121 records were screened at title-abstract level, of which 159 were screened at full-text; 57 records were found to be eligible for inclusion[17, 19, 28, 34-88]. Through reference tracking another 21 records were identified[17, 89-108], leading to a total of 78 included studies. Two studies presented outcomes in two separate publications; these were combined resulting in 76 unique studies[17, 74, 75, 87].

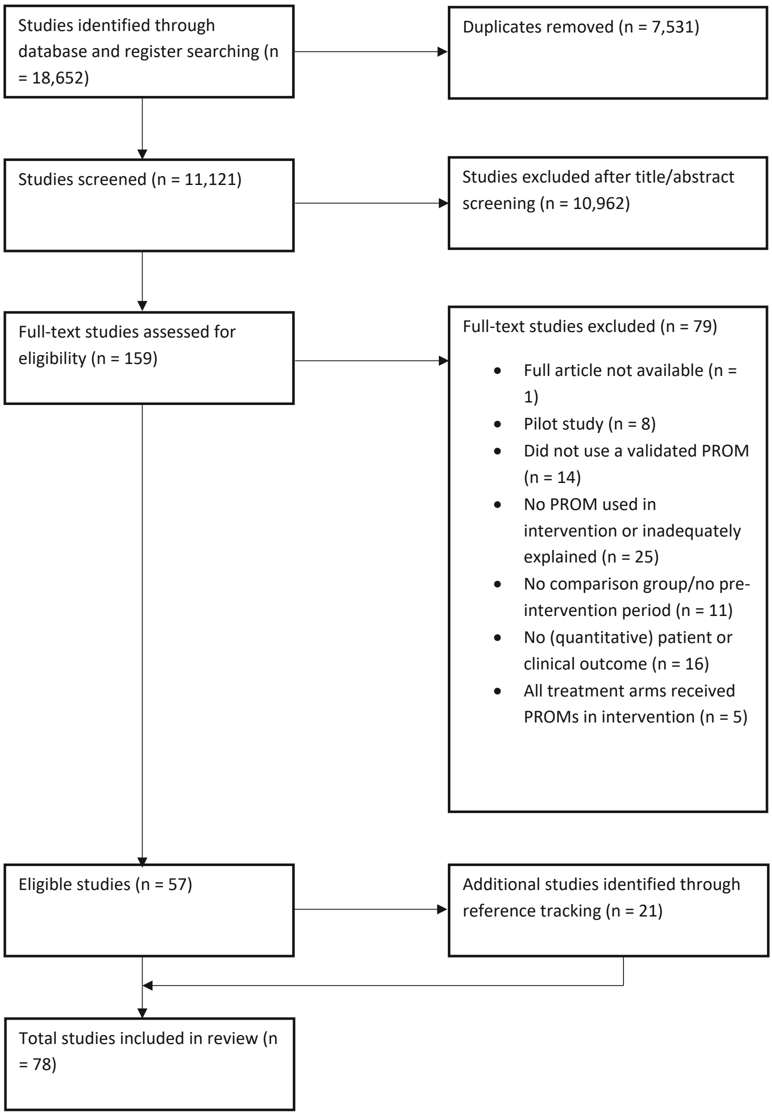

Figure 1: PRISMA flowchart of study selection

Study characteristics

An overview of study characteristics, PROMs used, overall study impact and quality is presented in Table 1 (micro-level) and Table 2 (meso-/macro-level). Below we shortly describe the included studies: for a more detailed description of study characteristics refer to Supplemental File 2, and for extended tables of study characteristics, quality assessment and outcomes extracted refer to Supplemental File 3.

Table 1: Study characteristics at the micro-level (sorted by medical domain)

Author (year)	Domain	Patients	Number of patients (I = Intervention, C = Control)	PROM(s) used (GEN = generic, DS = disease-specific)	Overall impact	Quality of study
Decision-aid						
Jayakumar (2021)	Orthopedics	Considering knee replacement for OA	I:69; C:60	KOOS-JR (DS), PROMIS Global 10 (GEN)	+	Moderate
Bansback (2022)	Orthopedics	Considering knee replacement for OA	I:82; C:81	EQ-5D-5L (GEN)	+	Weak
Volkmann (2015)	Rheumatology	Considering knee replacement for OA	I:111; C:NA	WOMAC (DS)	+	Moderate
Feedback to patient						
Ngo (2022)	Gyneacology	Pregnant women	I:89; C:103	PUQE (DS)	~	Moderate
Rogers (2021)	Oncology	Head-neck cancer, curatively treated	I:140; C:148	PCI Head Neck Cancer (DS)	+	Moderate
Steele Gray (2021)	Primary care	Elderly	I:23; C:21	PROMIS Global Health Scale (GEN), Pain Interference (DS), HAQ (GEN)	~	Weak
Berdal (2023)	Rehabilitation	Rheumatic or musculoskeletal diseases	I:168;C:206	PSFS (DS), EQ-5D-5L (GEN), EQ VAS (GEN), 30 second sit-to-stand test	~	Moderate
Gossec (2016)	Rheumatology	Reumatoid arthritis	I:159; C:161	RAPID-3/HAQ2 (GEN), RAID (DS), symptoms as free text	~	Moderate
Feedback to provider						
Hadjistavropoulos (2009)	Community care	Elderly, with complex medical problems	I:88; C:56	Pain Assessment Battery (DS), GPM (DS), GDS short form (DS), pain drawing	~	Weak
Almario (2016)	Gastro-enterology	Gastrointestinal disease	I:217; C:154	PROMIS Gastrointestinal (DS)	~	Moderate
Kjaer (2016)	Oncology	Head-neck cancer, finished treatment	I:132; C:134	HADS (DS), symptoms relevant to head-neck cancer survivors	+	Weak
Velikova (2004)	Oncology	Any cancer, commencing treatment	I:144, 70; C:72	EORTC-QLQ-C30 (DS), and HADS (DS)	+	Weak
Detmar (2002)	Oncology	Any cancer, undergoing palliative chemotherapy	I:114; C:100	EORTC-QLQ-C30 (DS)	~	Weak

Table 1: Continued

Author (year)	Domain	Patients	Number of patients (I = Intervention, C = Control)	PROM(s) used (GEN = generic, DS = disease-specific)	Overall impact	Quality of study
Rosenbloom (2007)	Oncology	Advanced breast, lung or colorectal cancer	I:71, 73; C:69	FACT-G (DS), including relevant cancer-type subscale	~	Strong
Hilarius (2008)	Oncology	Any cancer, undergoing adjuvant or palliative chemotherapy	I:111; C:108	EORTC-QLQ-C30 (DS), including relevant cancer-type (breast (QLQ-BR23), colorectal (QLQ- CR38), or lung cancer (QLQ-LC13))	~	Strong
Barbera (2015)	Oncology	Stage I–III breast cancer, received adjuvant chemotherapy	I:2541; C:5818	ESAS (DS)	+	Strong
Patel (2022)	Oncology	Advanced cancer	I:64; C:64	ESAS (DS)	+	Strong
Skovlund (2021)	Oncology	Metastic melanoma	I:137; C:142	EORTC-QLQ-C30 (DS), HADS (DS)	+	Weak
Ackermans (2017)	Orthopedics	Hip or knee OA	I:72; C:70	HOOS-PS/KOOS-PS (DS), NRS Pain (DS)	+	Strong
Holm (2020)	Orthopedics	Hip or knee OA	I:6245; C:N/A	NRS pain (DS), HOOS/KOOS (DS), EQ-5D (GEN), PSFS (DS), physical activity (DS), OA-QI (DS), ASES (DS)	+	Strong
De Wit (2008)	Pediatrics	Diabetes	I:46; C:45	PedsQL (GEN), Generic and Diabetes module (DS)	+	Moderate
Fihn (2004)	Primary care	Veterans	I:5801; C:3218	SF-36 (GEN), Seattle Outpatient Satisfaction Questionnaire (PREM), 1 of 6 disease-specific questionnaires (Seattle Angina, Seattle Obstructive Lung Disease, Drinking Practices, Seattle Diabetes, Seattle Hypertension, HSCL)	~	Weak
Kroenke (2018)	Primary care	Visiting for any reason	I:151; C:149	SPADE symptoms (GEN)	~	Moderate
Reiber (2004)	Primary care	Veterans	I:3701; C:2020	Seattle Diabetes Questionnaire, SF-36 (GEN), SOSQ (PREM)	~	Weak
Richardson (2008)	Primary care	Elderly	I:134; C:131	Self Report Task Modification and Disability Scale (DS), HUI (GEN), SF-36 (GEN), several physical and functional performance measures	+	Moderate
Santana (2010)	Pulmonary medicine	Pre- and post-lung transplant	I:108; C:105	HUI2/3 (GEN)	~	Moderate

Table 1: Continued

Author (year)	Domain	Patients	Number of patients (I = Intervention, C = Control)	PROM(s) used (GEN = generic, DS = disease-specific)	Overall impact	Quality of study
Monitoring						
Davidson (2010)	Cardiology	Myocardial infarction	I:80; C:77	PHQ-9 (DS)	+	Moderate
de Jong (2017)	Gastro-enterology	IBS	I:465; C:444	Monitor IBD At Home (DS)	+	Moderate
Berinstein (2022)	Gastro-enterology	IBS	I:100; C:105	CD-PRO (DS), UC-PRO (DS)	+	Strong
Cross (2019)	Gastro-enterology	IBS	I:115, 116; C:117	HBI (DS), SSCAI (DS)	+	Moderate
Pooni (2023)	General surgery	Colorectal surgery	I:128; C:125	QoR-15 (DS)	+	Strong
Girgis (2009)	Oncology	Non-localized breast and colorectal cancer	I:120, 119; C:117	HADS (DS), EORTC-QLQ-C30 (DS), SCNS short form (DS), selected items from the NA-ACP (DS)	~	Moderate
Cooley (2022)	Oncology	Lung cancer	I:89; C:91	PHQ-9 (DS), SDS (DS), FACT-G (DS), AUDIT (DS), MSAS (DS)	~	Weak
Price (2023)	Oncology	Cancer patients with depression	I:165; C:NA	PHQ-9 (DS) and GAD-2 (DS).	+	Weak
Livanainen (2023)	Oncology	Colorectal cancer, undergoing chemotherapy	I:36; C:35	NCTCAE (DS)	~	Weak
Basch (2016)	Oncology	Metastatic breast, genitourinary, gynecologic or lung cancer	I:441; C:325	NCTCAE (DS)	+	Moderate
Sharpe (2014)	Oncology	Cancer patients with depression	I:253; C:247	PHQ-9 (DS)	+	Weak
Patel (2020)	Oncology	Hematologic or advanced stage cancer	I:186; C:102	Edmonton Symptom Assessment Scale (DS)	+	Strong
Maguire (2021)	Oncology	Non-metastatic breast, colorectal, (non-)Hodgkin's cancer, undergoing chemotherapy	I:415; C:414	DCTAQ (DS)	+	Strong

Table 1: Continued

Author (year)	Domain	Patients	Number of patients (I = Intervention, C = Control)	PROM(s) used (GEN = generic, DS = disease-specific)	Overall impact	Quality of study
Epstein (2007)	Pediatrics	ADHD	I:162; C:215	CPRS (DS), CTRS (DS)	~	Moderate
Dobscha (2006)	Primary care	Moderate to severe depression	I:189; C:186	PHQ-9 (DS)	+	Strong
Balestrieri (2020)	Primary care	Depression	I:66; C:32	PHQ-9 (DS) and IDS-SR (DS)	+	Moderate
Dhingra (2021)	Primary care	Chronic pain	I:256; C:272	FPS (DS), SBIRT (DS), PHQ-2 (DS), BPI short form (DS), PROMIS Pain Interference short form (DS)	~	Weak
Katon (1996)	Primary care	Depression	I:77; C:76	BDI short form (DS)	+	Moderate
Katzelnick (2000)	Primary care	Depression	I:218; C:189	HAM-D (DS)	+	Moderate
Unützer (2002)	Primary care	Depression	I:218; C:189	PHQ-9 (DS)	+	Weak
Simon (2011)	Primary care	Depression	I:106; C:102	PHQ-9 (DS)	+	Weak
Katon (2004)	Primary care	Diabetes patients, with depression	I:164; C:165	PHQ-9 (DS)	+	Weak
Carola Pèrez (2021)	Psychiatry	Depression	I:84; C:83	PHQ-9 (DS)	~	Moderate
Tirelli (2020)	Reumatology	Juvenile Idiopathic Arthritis, subtype Enthesitis-Related Arthritis	I:54; C:98, 51	JADAS (DS), PROMIS physical function (DS), NRS pain (DS)	~	Moderate
Buckley (2020)	Rheumatology	Juvenile Idiopathic Arthritis	I:97; C:NA	JADAS (DS), PROMIS physical function (DS), NRS pain (DS)	+	Moderate
No feedback						
Baker (2023)	Dermatology	Eczema	I:147; C:149	POEM (DS)	+	Moderate
McCambridge (2007)	General public	University students	I:217; C:204	AUDIT (DS)	+	Moderate
Jaensson (2017)	General surgery	Various types of day surgery	I:513; C:514	SwQoR (DS)	+	Strong

2

Table 1: Continued

Author (year)	**Domain**	**Patients**	**Number of patients (I = Intervention, C = Control)**	**PROM(s) used (GEN = generic, DS = disease-specific)**	**Overall impact**	**Quality of study**
Screening						
Frasure-Smith (1997)	Cardiology	Myocardial infarction	I:692; C:684	GHQ (DS)	~	Strong
Kronish (2020)	Cardiology	Myocardial infarction	I:499, 501; C:500	PHQ-8 (DS)	~	Weak
Allen (2011)	Community care	War veterans	I:97; C:NA	DASS (DS), AUDIT (DS)	+	Weak
van der Zee-van den Berg (2017)	Midwife care	Mothers who just gave birth	I:1843; C:1246	EPDS (DS)	+	Moderate
Ferrell (2021)	Oncology	Palliative cancer, undergoing a therapeutic clinical trial	I:239; C:240	Distress Thermometer (DS), FACT-G (DS)	+	Weak
Shyu (2013)	Orthopedic and trauma surgery	Hip fracture patients	I:99; C:101, 99	GDS short form (DS)	+	Strong
Mallen (2017)	Primary care	Hip or knee osteoarthritis patients	I:1339; C:703	GAD-2 (DS), PHQ-2 (DS), NRS pain (DS)	-	Moderate
Fortmann (2020)	Primary care	Diabetes patients	I:236;C:239	PHQ-9 (DS)	+	Moderate
Rollman (2021)	Cardiology	Heart failure	I:251; C:252, 126	PHQ-2 and -9 (DS)	+	Moderate
Regueiro (2019)	Gastro-enterology	IBS	I:322; C:NA	Harvey-Bradshaw Index (DS) for CD or Ulcerative Colitis Activity Index for UC, Short Inflammatory Bowel Disease questionnaire (DS), PHQ-9 (DS), GAD-7 (DS)	+	Moderate
Howell (2020)	Oncology	Various types of cancer	I:13260, 10875; C:57594, 48068	ESAS (DS), DART (DS), BPI (DS), CFS (DS), GAD (DS), PHQ (DS), occasionally the CPC (DS) or SDI (DS)	+	Strong
Wylde (2022)	Orthopedics	Received knee replacement surgery	I:242; C:121	BPI (DS), HADS (DS), PainDETECT (DS), Douleur Neuropathique 4 (DS)	+	Weak
Buxton (2012)	Primary care	Low income and uninsured	I:36; C:81	SF-12 (GEN), AUDIT (DS), PHQ-9 (DS)	~	Moderate
Wu (2018)	Primary care	Diabetes type 2	I:366, 380; C:341	PHQ -2 and/or -9 (DS)	+	Strong

Abbreviations: ADHD = Attention Deficit Hyperactivity Disorder; ASES = Arthritis Self-Efficacy Scale; AUDIT = Alcohol Use Disorders Identification Test; BDI = Beck Depression Inventory; BPI = Brief Pain Inventory; CD-PRO = Crohn's Disease PRO; CFS = Cancer Fatigue Scale; COOP = Primary Care Cooperative Information Project; CPC = Canadian Problem Checklist; CPRS = Conners Parent Rating Scale; CTRS = Conners Teachers Rating Scale; DART = Distress Assessment Response Tool; DASS = Depression Anxiety Stress Scales; DCTAQ = Daily Chemotherapy Toxicity Self-Assessment Questionnaire; EORTC-QLQ-C30 = The European Organization for Research and Treatment of Cancer quality of life questionnaire; EPDS = Edinburgh Postnatal Depression Scale; EQ-5D-5L = Five-level version of the EQ-5D; ESAS = Edmonton Symptom Assessment System; FACT-G = Functional Assessment of Cancer Therapy - General; FPS = Faces Pain Scale; GAD = Generalized Anxiety Disorder; GDS = Geriatric Depression Scale; GHQ = General Health Questionnaire; GPM = Geriatric Pain Measure; HADS = Hospital Anxiety and Depression Scale; HAM-D = Hamilton Depression Rating Scale; HAQ = Health Assessment Questionnaire; HBI = Harvey-Bradshaw Index; HOOS-PS = Hip disability and Osteoarthritis Outcome Score – Physical Function; HSCL = Hopkins Symptom Checklist; HUI = Health Utilities Index Mark; IBS = Inflammatory Bowel Disease; IDS-SR = Inventory of Depressive Symptomatology – Self Rated; JADAS = Clinical Juvenile Arthritis Disease Activity Score; KOOS-JR = Knee Injury and Osteoarthritis Outcome Score for Joint Replacement; KOOS-PS = Knee disability and Osteoarthritis Outcome Score – Physical Function; MSAS = Memorial Symptom Assessment Scale; NA-ACP = Needs Assessment for Advanced Cancer Patient Questionnaire; NCTCAE = National Cancer Institute's Common Terminology Criteria for Adverse Events; NRS = Numerical Rating Scale; OA-QI = OsteoArthritis Quality Indicator Questionnaire; OHS = Oxford Hip Score; OKS = Oxford Knee Score; PCI = Patient Concerns Inventory; PedsQL = Pediatric Quality of Life Inventory; PHQ = Patient Health Questionnaire; POEM = Patient Oriented Eczema Measure; PREMs = patient-reported experience measures; PROMs = patient-reported outcome measures; PROMIS = Patient-Reported Outcomes Measurement Information System; PSFS = Patient Specific Functional Scale; PUQE = Pregnancy Unique Quantification of Emesis; QoR-15 = Quality of Recovery questionnaire; RAID = Reumatoid Arhritis Impact of Disease; RAPID-3/HAQ2 = RAPID3 Health Assessment Questionnaire; SBIRT = Screening, Brief Intervention, and Referral to Treatment questionnaire; SCNS = Supportive Care Needs Survey; SDI = Social Difficulties Inventory; SDS = Symptom Distress Scale; SF-36 = Short Form Health Survey; SOSQ = Seattle Outpatient Satisfaction Questionnaire; SPADE = Sleep, pain, anxiety, depression, and low energy/fatigue; SSCAI = Simple Clinical Colitis Activity Index; SwQoR = Swedish Quality of Recovery Scale; UU-PRO = Ulcerative Colitis PRO; VAS = Visual Analogue Scale; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index

Table 2: Study characteristics at the meso-/macro-level (sorted by medical domain)

Author (year)	Domain	Patients	Number of patients	PROM(s) used (GEN = generic, DS = disease-specific)	Overall impact	Quality of study
Benchmarking						
Boyce (2015)	Orthopedics	Receiving primary hip arthroplasty	I:230; C:228	OHS (DS)	~	Weak
Varagunam (2014)	Orthopedics and general surgery	Receiving hip or knee arthroplasty, varicose vein or inguinal hernia surgery	I: 7k-30k; C:NA	EQ-5D-5L (GEN), EQ VAS (GEN), 1 of 3 disease-specific questionnaires (OHS, OKS, AVVQ)	~	Moderate
Weingarten (2000)	Primary care	Elderly	I:541; C:543	Dartmouth COOP (GEN)	~	Moderate
Kumar (2021)	Urology	Undergoing prostate surgery for cancer	I:212; C:210	Selected items (continence, sexual function) from the EPIC (DS)	-	Weak
In-depth analysis of data						
Haller (2011)	Internal and surgical departments	Wards	I:1237; C:1113	IPO questionnaire (DS), NRS pain	+	Moderate
Zaslansky (2019)	Surgical departments	Wards	I:? C:?	IPO questionnaire (DS), NRS pain	+	Weak
Garduño-López (2021)	Surgical departments	Wards	I:? C:?	IPO questionnaire (DS), NRS pain	+	Weak
PDCA-cycle						
Partridge (2016)	Orthopedics	Patient receiving total knee arthroplasty	I:827, C:441	OKS (DS), EQ-5D-3L (GEN)	+	Moderate

Abbreviations: AVVQ = Aberdeen Varicose Vein Questionnaire; COOP = Primary Care Cooperative Information Project; EPIC = Expanded Prostate Cancer Index Composite; EQ-5D-3L = Three-level version of the EQ-5D; EQ-5D-5L = Five-level version of the EQ-5D; IPO = International Pain Outcomes; NRS = Numerical Rating Scale; Questionnaire; OHS = Oxford Hip Score; OKS = Oxford Knee Score; VAS = Visual Analogue Scale

Micro-level

Sixty-eight out of 76 studies provided evidence on the use of PROMs at the micro-level[17, 19, 28, 34-36, 38-44, 46-48, 50-56, 58, 59, 62-71, 73-93, 95-108]. Most studies were conducted in the United States (n=32), and were in the medical domains primary care (n=17), oncology (n=19), gastroenterology (n=5) and orthopedic (trauma) surgery (n=6). Fifty-five studies used a disease-specific instrument in their intervention, 3 used a generic instrument and 10 a combination. Sixteen studies were of strong quality, 31 were of moderate quality and 21 were of weak quality.

Macro-level

Eight out of 76 studies provided evidence of the use of PROMs at the macro-level[37, 45, 49, 57, 60, 61, 72, 94], and no studies were found at the meso-level. Studies were conducted in various countries. Most studies were conducted in surgical fields (n=7), of which 3 in both non-surgical and surgical fields; the eighth study was conducted in primary care. Five studies used a disease-specific PROM, 1 used a generic PROM, and 2 used a combination. Four studies were rated as moderate quality, while the other 4 were rated as weak quality.

Impact by determinants and outcome dimensions

Outcome of PROMs interventions by determinants are summarized in Table 3 (micro-level) and 4 (meso-/macro-level). Table 5 shows the impact by outcome dimensions. The quality of studies for each determinant generally indicated "moderate" quality, both at the micro- and meso-/macro-level; the exception is highlighted. Six mechanisms were identified at the micro-level, and 3 at the meso-/macro-level.

Table 3: Overall impact by determinants at the micro-level

		Number of studies	Improvement (%)	Quality of studies (average)
Mechanism	Feedback to patient	5	1 (20)	1.8
	Feedback to provider	18	9 (50)	1.9
	Screening	14	10 (71)*	2.0
	Monitoring	25	18 (72)	1.9
	No feedback	3	3 (100)	2.3
	Decision-aid	3	3 (100)	1.7
Medical domain	Cardiology	4	2 (50)	2.0
	Community care	2	1 (50)	1.0
	Dermatology	1	1 (100)	2.0
	Gastroenterology	5	4 (80)	2.2
	General public	1	1 (100)	2.0
	General surgery	2	2 (100)	3.0
	Gyneacology	1	0 (0)	2.0
	Midwife care	1	1 (100)	2.0
	Oncology	19	13 (68)	1.9

Table 3: Continued

		Number of studies	Improvement (%)	Quality of studies (average)
	Orthopedics/trauma surgery	6	6 (100)	2.2
	Pediatrics	2	1 (50)	2.0
	Primary care	17	10 (59)*	1.7
	Psychiatry	1	0 (0)	2.0
	Pulmonary medicine	1	0 (0)	2.0
	Rehabilitation	1	0 (0)	2.0
	Rheumatology	4	2 (50)	2.0
Type of PROM	Disease-specific	55	39 (71)*	2.0
	Generic	4	1 (25)	1.5
	Combination	9	4 (44)	1.9

*One study showed a deterioration.
Abbreviations: PROM = Patient-Reported Outcome Measure.

Table 4: Overall impact by determinants at the meso-/macro-level

		Number of studies	Improvement (%)	Quality of studies (average)
Mechanism	Benchmarking	4	0 (0)*	1.3
	In-depth analysis of data	3	3 (100)	1.7
	PDCA-cycle	1	1 (100)	2.0
Medical domain	Orthopedics	2	1 (50)	1.7
	Primary care	1	0 (0)	1.0
	Urology	1	0 (0)*	2.0
	Various internal and surgical departments	4	3 (75)	1.5
Type of PROM	Disease-specific	5	3 (60)*	1.4
	Generic	1	0 (0)	1.0
	Combination	2	1 (50)	2.0

*One study showed a deterioration.
Abbreviations: PROM = Patient-Reported Outcome Measure.

Table 5: Impact by outcome dimensions

	Micro-level			Meso-/macro-level		
	Number of studies	Improvement (%)	Quality of studies (average)	Number of studies	Improvement (%)	Quality of studies (average)
PROMs						
Functioning						
Physical	29	12 (41)	1.9	2	0	1.0
Mental	25	8 (32)	1.8	1	0	1.0
Social	16	6 (38)	1.8	1	0	1.0
HRQoL	29	11 (38)	1.8	5	2 (40)	1.6
Role limitations						
Physical	5	1 (17)	2.0	-	-	-
Emotional	5	0	2.0	-	-	-
General health perceptions	8	6 (75)	2.4	1	0	1.0
Symptoms combined	46	26 (57)**	1.9	5	3 (60)**	1.6
Depression	25	14 (56)	1.8	2	0	2.5
Anxiety	14	6 (43)	2.0	2	1 (50)	2.5
Alcohol use/disorder	3	1 (33)	1.7	-	-	-
Pain	17	7 (42)**	1.9	5	2 (40)	1.6
Vitality/fatigue	5	3 (60)	1.8	1	0	3.0
Nausea	4	0	2.3	2	1 (50)	1.0
Decision-conflict and readiness	4	3 (75)		-	-	-
PREMs						
Satisfaction	23	10 (43)	2.0	-	-	-
Patient-physician relationship	5	1 (20)	2.0	-	-	-
Experience with care	7	4 (57)	2.0	-	-	-
Supportive needs	3	1 (33)	2.7	-	-	-

Table 5: Continued

	Micro-level			Meso-/macro-level		
	Number of studies	Improvement (%)	Quality of studies (average)	Number of studies	Improvement (%)	Quality of studies (average)
Patient-activation	7	4 (57)	2.1	-	-	-
Physician awareness of HRQoL	2	0	2.0	-	-	-
Clinical outcomes						
Complications*	8	1 (13)	1.8	2	0	1.5
(Re)admissions	17	5 (29)	2.2	-	-	-
Emergency department visits	12	7 (58)	2.4	-	-	-
Survival	5	0	2.0	-	-	-
Lab values	4	2 (50)	2.0	-	-	-
Outcome same as PROM used in intervention	32	18 (56)**	1.8	7	3 (42)**	1.4
Outcome not the same as PROM used in intervention	36	26 (72)	2.0	1	1 (100)	2.0

*Complications also vary by domain and intervention, e.g., a bleed in myocardial infarction patients.
**One study showed a deterioration.
Abbreviations: PROM = Patient-Reported Outcome Measure; PREM = Patient-Reported Experience Measure, HRQoL = Health-Related Quality of Life

Impact by mechanism

Micro-level

Feedback of PROMs to patient

One of 5 studies employing feedback of PROMs to patients fed back (raw) scores directly [54], 3 included a graphical display of PROMs scores[55, 78, 85], and 1 combined a narrative report with a graphical display[43]. Studies were conducted in various domains. One (20%) study conducted in head-cancer patients fed back data from a comprehensive inventory of disease-related symptoms and found an improved overall outcome, driven by improved symptoms (pain and activity), mental and physical functioning[54].

Feedback of PROMs to provider

Two of the 18 studies employing feedback of PROMs to providers used (raw) scores in their report[79, 90], 4 included a narrative report[52, 53, 73, 93], 8 included a graphical display[17, 36, 44, 47, 48, 84, 91, 92], and 3 combined a narrative report with a graphical display[34, 41, 89]. Overall, nine (53%) studies found an improvement in outcome[17, 34, 47, 53, 73, 84, 89, 90].

When looking at the information collected, 14 of 18 studies fed back PROMs to patients which covered disease-specific information such as hip functioning, cancer-related, or gastrointestinal symptoms[17, 34, 36, 41, 47, 53, 73, 79, 84, 89-93]. Of these 14 studies, 9 (64%) found an improvement in outcome[17, 34, 47, 53, 73, 79, 84, 89, 90]. Most studies pertained to cancer-related symptoms (n=8) of which 5 (63%) reported an improvement via various outcome dimensions, including reduced emergency department (ED) visits or readmissions (n=2), improved physical, mental and social functioning (n=1), symptoms (depression and cancer-related) (n=1) or experience with care (n=1) [17, 47, 79, 84, 89]. The remaining 4 studies fed back PROMs to the provider pertaining to general HRQoL and/or pain, and found no improvement in outcome[44, 48, 52, 55].

Using PROMs to screen for disease or symptoms

Seven studies out of 14 used PROMs to screen for depression[28, 35, 50, 56, 71, 98, 102], and 1 study for oncological symptoms [70], to initiate treatment or a care pathway. Of these, five (63%) studies observed an improved outcome driven by improved symptoms (depression, stress or anxiety) (n=4), improved mental (n=2), social (n=2), and physical functioning (n=1), and reduced ED visits and readmissions (n=1) [28, 35, 56, 70, 71]. One study found an outcome deterioration via worsened pain symptoms [50].

Six studies combined the screening for depression with follow-up monitoring to evaluate whether the treatment works, and potentially adjust if treatment was ineffective[38, 59, 74, 83, 88, 105]. Of these, three also incorporated disease-specific information: knee functioning [88], cancer-related [74], and gastro-intestinal symptoms [105]. Five (83%) out of 6 studies found improved outcome particularly via improved symptoms (depression and anxiety) (n=4) and reduced ED visits (n=2) [59, 74, 83, 88, 105]. Two of three disease-specific symptoms also improved, except for oncological symptoms [74].

Using PROMs to monitor symptoms

Twelve out of 25 studies used PROMs to identify patients under treatment exceeding predefined thresholds of symptoms and linked these to treatment changes, increased monitoring or care pathways[39, 63, 66, 67, 81, 86, 95, 97, 100, 103, 107, 108]; 10 (83%) found an improved outcome[39, 63, 66, 81, 95, 97, 100, 103, 107, 108]. Seven studies also used PROMs monitor treatment but did not explicitly mention the use of predefined algorithms[40, 42, 69, 82, 99, 101, 104]; 4 (57%) reported an improvement[82, 99, 101, 104]. Six studies incorporated PROMs into the clinical pathway and sent out alerts upon exceeding a threshold without specific guidance to the provider[64, 68, 76, 80, 96, 106], 1 of these also used PROMs to monitor treatment response [106]; three (50%) found an improved outcome[64, 96, 106].

When looking at the information collected, 13 out of 25 studies used PROMs to monitor existing depression symptoms[42, 63, 68, 69, 80, 82, 97, 99-101, 106-108]. Of these, 10 (77%) found an improved outcome, mostly driven by improved depression symptoms (n=9) and satisfaction (n=5) [63, 69, 82, 97, 99-101, 106-108]. Five studies used PROMs to monitor cancer-related symptoms[64, 67, 76, 103, 104], of which 3 (60%) found various improved outcomes including HRQoL, physical and mental functioning, and satisfaction[64, 103, 104]. Three studies monitored gastro-intestinal symptoms in patients with inflammatory bowel disease and all (100%) found reduced readmissions (n=2) and improved HRQoL (n=1) [39, 66, 96]. The remaining 4 studies were conducted in various domains[40, 81, 86, 95], of which two showed improved outcomes. The first monitored surgical recovery in colorectal surgery patients and found improved perception of general health, anxiety and satisfaction. The other used PROMs to guide treatment in children with juvenile idiopathic arthritis and found reduced pain and arthritis activity[81, 95].

No feedback: filling out effect of PROMs

One of 3 studies tested the hypothesis of whether merely filling out alcohol abuse PROMs would reduce alcohol use by a direct measurement effect[51]. Similarly, another study collected PROMs weekly in patients with eczema without any additional interventions[62]. The third study collected PROMs daily after surgery via an app; patients could always contact their provider via the e-portal[46]. All (100%) studies reported improved outcome due to improved symptoms (depression and alcohol dependency) (n=2) and improved HRQoL (n=1).

PROMs in decision-aids

In three studies a one-time PROM was used in a decision-aid along an education component to help with treatment choice (surgical vs. conservative) in patients with knee osteoarthritis[19, 58, 77]. All studies (100%) found an improvement in shared-decision making, while 1 of these only found this effect in females[58].

Meso-/macro-level

PROMs in benchmarking

Three benchmarking studies used case-mix adjusted PROM scores[37, 49, 57], while the fourth used unadjusted scores[94]. Three studies presented performance reports to the provider, which included PROM scores and how they compared to peer providers[37, 49, 94]; in 2 studies complication rates were also presented[37, 49]. The other study evaluated a nationwide PROMs collection program, which provided both patients and providers the option to check providers' PROMs outcomes[57]. All studies were of weak quality, and did not find an improvement in outcome; 1 study even reported a potential worsening[49].

PROMs in in-depth analysis of data

Three studies used PROM data in combination with guidelines, teaching and protocols to improve pain management in various surgical and non-surgical departments[45, 60, 72]. One of these studies also used a feedback loop by a department representative to evaluate and provide advice on the implemented initiatives[45]. The two other studies pertained to the same quality initiative aimed to reduce the pain of patients admitted to hospitals but were conducted in different developing countries/departments[60, 72]. All 3 (100%) studies found an improvement in outcome due to reduced pain (n=3) and nausea (n=2) symptoms in particular.

PROMs in PDCA-cycles

One study conducted a PDCA-cycle where they introduced an improved total knee implant and changed their surgical technique, guided by and evaluated with PROMs scores[61]: an overall improvement in outcome (HRQoL) was observed.

Impact by medical domain

Micro-level

At the micro-level, the medical domains in which PROM interventions were conducted which seemed to be consistently associated with improved outcome were orthopedic (trauma) surgery (n=6 studies, 100% effective), gastroenterology (n=5, 80%), oncology (n=19, 68%), and primary care (n=17, 59%). Less effective seemed cardiology (n=4, 50%) and rheumatology (n=4, 50%). Limited evidence was available for other domains.

Meso-/macro-level

Interventions conducted in orthopedics, primary care, and urology were not found to be related to improved outcome. Four studies covered various internal and surgical departments, of which 3 (75%) showed improved outcome.

Impact by type of PROM used in intervention

Micro-level

Most studies used a disease-specific PROM, which showed the highest percentage of improved outcomes (n=55 studies, 71% effective). Generic PROMs or a combination of both showed an overall lower percentage (n=13, 38%). While disease-specific PROMs were used in all mechanisms, generic PROMs were used in studies employing the "feedback" mechanism (n=10), "decision-aids" (n=2), and once (combined with a disease-specific PROM) in "screening".

Meso-/macro-level

According to the type of PROM (disease-specific vs. generic) no specific pattern was observed.

Impact by outcome dimensions

Micro-level

In this section, we describe the impact of the PROMs-interventions on the outcome dimensions (PROMs, PREMs or clinical outcomes), regardless of the mechanism or other determinants.

Regarding PROMs, studies often showed an improvement in general health perceptions (n=8 studies, 75% effective), decision-readiness and conflict (n=4, 75%) and symptoms overall (n=46, 57%). Particularly depression was evaluated often (n=25), and improved in 57% of studies. The percentage decreased for HRQoL (n=29, 38%) and physical and mental functioning domains.

Regarding PREMs, satisfaction was most often studied (n=23), and improved in less than half of studies (43%). Patient-activation and experience with care tended to improve slightly more often (n=7, 57%, for both outcomes).

As for clinical outcomes, twelve studies analyzed emergency department visits, of which 58% found an improvement. Fewer studies observed a positive effect on complications (n=8, 13%) and (re)admissions (n=17, 29%), and no studies observed an effect on survival (n=5, 0%).

Studies which used a different outcome than the PROM in the intervention more often had an improved overall outcome (n=36, 72%), compared to those which did not (n=32, 56%).

Meso-/macro-level

With regard to PROMs, symptoms showed improved most often, which mostly pertained to pain (n=5, 60%). HRQoL was also measured in 5 studies, however, improved in less studies

(40%). Other domains and outcomes were studied in only a few studies, and showed no improvement.

DISCUSSION

In this systematic review, evidence on the use of PROMs to improve patient-related outcomes at the micro- (68 studies) and meso-/macro- (8) levels was collected and analyzed. Moreover, determinants for increased effectiveness were elucidated.

At the micro-level, 44% of studies employing direct feedback of PROMs to the provider and/or patient resulted in improved patient outcomes, which is in line with previous reviews[8-15]. A contemporary development was to use PROMs to screen for disease or to monitor existing disease. These studies linked the PROMs scores to care pathways or treatment adaptations, and approximately 70% of studies found improved outcomes. This approach was particularly effective for depression, oncological and gastroenterological disease. A novel application was to use PROMs to inform patients considering knee arthroplasty, which generally resulted in improved decision-quality. At the meso-/macro-level, current evidence does not support using PROMs in benchmarking. The scarce evidence available suggests, however, that PROMs might be of value in an in-depth analysis of the performance of departments and hospitals and PDCA-cycles. At both the micro- and meso-/macro-level, studies more often employed disease-specific PROMs, which – in comparison with studies which employed generic PROMs – found improved outcomes more often.

The evidence at all levels was of moderate quality at best, which raises concerns regarding the validity of the findings.

Micro-level

Providing feedback on the PROM scores to patients or providers is generally thought to benefit outcomes via improved patient-healthcare professional communication and identification of problematic symptoms[16]. This application is often used in patients with chronic disease who have multiple visits to their doctor, which in our review included diabetes, gastrointestinal disease, oncology, orthopedics, transplantation care; most evidence was available for oncology[8, 27]. For example, two studies applied a tailored symptom inventory for head-neck cancer patients and found a positive impact on PROMs[47, 54]. The effectiveness may be because this group presumably experiences a number of severe physical symptoms (e.g., problems with swallowing) which, if timely detected, are sensitive to treatment.

The application of PROMs to improve patient outcomes seems particularly effective if a deviation from the acceptable threshold occurs and can be linked to a recognizable action by the clinician, such as referral or treatment adaptation. This mechanism was effective in several studies in the medical domains, including depression, oncology and gastrointestinal care. For example, monitoring patients with diagnosed diseases such as inflammatory bowel disease or screening for disease with an expected high burden in the studied population such as post-partum depression may be beneficial[28, 39]. The purpose and goal of the tool may be clearer for both patient and provider, which could increase its effectiveness.

Various reasons may underlie decreased effectiveness of PROM-interventions. Firstly, a general trend was observed that studies utilizing generic PROMs found less positive effect overall, and these studies mostly did not link a generic PROM to a care pathway (such as "screening" or "monitoring"). Generic PROMs may provide insufficient insight into treatable or modifiable factors related to the studied population. However, it should be noted, one of the identified decision-aids successfully employed only a generic measure in patients considering knee arthroplasty. Combined, we believe this underlines the fact that the choice of PROM in the intervention should be driven by the intended use. Secondly, the measured outcome may play a role: PROM interventions tended to have a more pronounced impact on general health perceptions and symptom burden, but less so on certain outcomes such as HRQoL in general or survival. Other reasons for failure may include patients' resistance to discussing symptoms, time constraints in clinical practice and lack of provider continuity, and implementation hurdles through lack of knowledge[16].

The evaluation of interventions based on systematic PROM feedback appears to be a challenge. Firstly, the definition of 'control' treatment: about a third of the studies collected PROMs in the control group, unconnected to feedback or another intervention. This may decrease the difference as the collection of PROMs itself may induce beneficial effects as observed in 3 studies[46, 51, 62]. These findings suggest a Hawthorne-like effect through the completion of PROMs alone[51, 109]. The patient's self-knowledge and awareness are increased, and filling out the questionnaire may increase their empowerment to take a more active role in their healthcare[34]. We expected this effect to be relatively limited, as approximately half of studies used a different outcome measure than the PROM in the intervention and generally found an improvement. Secondly, most studies did not measure intervention compliance making it impossible to know to what extent (and how) patients or providers used the PROM interventions. Thirdly, PROMs are generally part of a more complex intervention with multiple facets (e.g., patient education), and it is impossible to isolate the exact role of the PROM in the intervention. However, we believe this is also one of the key roles of PROMs in contemporary medicine; they can enhance interventions by offering important insight into patient outcomes.

Meso-/macro-level

The 4 studies which evaluated PROM benchmarking did not find a benefit. Multiple reasons for the intervention not being successful have been suggested. Boyce et al. noted that

PROMs have not been developed nor validated as performance measures, and the choice of PROM may play a role in the usability of the provided feedback[37]. It is possible that inter-provider comparisons do not inherently motivate professionals to initiate additional audits and research activities or professionals may lack the knowledge to undertake such initiatives. The included studies do not describe how the data was (or wasn't) used in a feedback process of change. Kumar et al. suggested that further improvement might be prevented when the quality of care is already high[49]. The quality of the benchmarking process is also dependent on adequate case-mix variable selection, which is time-consuming and costly[110, 111]. A lack of educational support could also play a role, and it may be useful to provide examples of successes and failures with using PROMs data[112]. Finally, aggregated PROMs are used extensively in research aimed at improving quality care through, e.g. identifying subgroups at risk for poorer outcomes. These studies presumably have a large impact on national clinical guidelines, however, to our knowledge, the impact is hardly reported in peer-reviewed literature. The same applies to quality benchmarking under the supervision of professional organisations: this information is discussed with hospital groups and individuals but is generally not published.

Some examples, however, were found for in-depth analysis and PDCA-cycles with the aim to initiate quality improvements. A PDCA-cycle provides a structured and iterative approach to test changes aimed at improving the quality of systems[113]. Four studies were found that exploited these types of methods using PROMs data, all finding a benefit on patient outcomes. Zaslansky et al. suggested that the success could be attributable to the relatively low starting performance of partaking departments[60]. A commonality among these studies is the clear definition of the goal, an action plan, and feedback on the intervention along the way; all potential items which might facilitate the success of a quality improvement initiative, also highlighted by a Cochrane review[114].

Strengths and limitations

The major strength of this review is the broad search strategy, including the added value of PROMs at the micro-, meso- and macro-level. Several limitations must be acknowledged. Non-peer-reviewed literature (e.g., registry reports), which may be an important source of information on the use of PROMs as quality improvement tool, was excluded. However, this was not deemed feasible because these documents are often published in non-English languages and generally do not report clear evidence of an impact, such as a before-after comparison. Meta-analysis and estimating the effect sizes were not possible due to the heterogeneity of outcomes. PROM scores were variably reported as total score and/or by dimension, limiting the synthesis on the impact of PROMs-interventions by outcome dimensions.

CONCLUSION

This systematic review provides a comprehensive overview of novel applications of PROMs which aim improve patient outcomes, and determinants for increased effectiveness. The effectiveness appears to relate to the underlying mechanism, type of PROM used and outcome studied. At the micro-level, for example, PROMs feedback to patient or provider was positively associated with patient outcomes in approximately half of studies. Contemporary studies went a step further and linked PROMs scores to care pathways in for example depression, oncological and gastrointestinal care, which resulted in improved outcomes in a higher percentage of studies. At the meso-/macro-level evidence was limited, and evidence did not suggest a benefit of using PROMs for benchmarking. Promising applications included in-depth analysis and PDCA-cycles using PROMs data. With the increasing use of PROMs in routine clinical care, these findings may help in designing applications which truly impact patient outcomes. As the quality of studies was moderate at best raising concerns regarding the validity of findings, rigorously designed studies should be conducted on testing these applications.

REFERENCES

1. Brooks R. EuroQol: the current state of play. Health Policy 1996; 37: 53-72.
2. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015; 84: 167-176.
3. Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: Past, Present and Future. Appl Health Econ Health Policy 2017; 15: 127-137.
4. Meadows KA. Patient-reported outcome measures: an overview. Br J Community Nurs 2011; 16: 146-151.
5. Murray CJ, Frenk J. A framework for assessing the performance of health systems. Bull World Health Organ 2000; 78: 717-731.
6. The Organisation for Economic Co-operation and Development (OECD) Health at a Glance 2019.
7. Alberta PROMs and EQ-5D Research and Support Unit (APERSU). Enhancing the Use of Patient-reported Outcome Measures (PROMs) in the Healthcare System in Alberta. Link: https://apersu.ca/wp-content/uploads/2020/09/APERSU-PROMs-White-Paper.pdf. 2020.
8. Boyce MB, Browne JP. Does providing feedback on patient-reported outcomes to healthcare professionals result in better outcomes for patients? A systematic review. Qual Life Res 2013; 22: 2265-2278.
9. Espallargues M, Valderas JM, Alonso J. Provision of feedback on perceived health status to health care professionals: a systematic review of its impact. Med Care 2000; 38: 175-186.
10. Gilbody SM, House AO, Sheldon T. Routine administration of Health Related Quality of Life (HRQoL) and needs assessment instruments to improve psychological outcome--a systematic review. Psychol Med 2002; 32: 1345-1356.
11. Gilbody SM, House AO, Sheldon TA. Routinely administered questionnaires for depression and anxiety: systematic review. Bmj 2001; 322: 406-409.
12. Greenhalgh J, Meadows K. The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: a literature review. J Eval Clin Pract 1999; 5: 401-416.
13. Ishaque S, Karnon J, Chen G, Nair R, Salter AB. A systematic review of randomised controlled trials evaluating the use of patient-reported outcome measures (PROMs). Qual Life Res 2019; 28: 567-592.
14. Valderas JM, Kotzeva A, Espallargues M, Guyatt G, Ferrans CE, Halyard MY, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. Qual Life Res 2008; 17: 179-193.
15. Gibbons C, Porter I, Gonçalves-Bradley DC, Stoilov S, Ricci-Cabello I, Tsangaris E, et al. Routine provision of feedback from patient-reported outcome measurements to healthcare providers and patients in clinical practice. Cochrane Database Syst Rev 2021; 10: CD011589.
16. Greenhalgh J, Gooding K, Gibbons E, Dalkin S, Wright J, Valderas J, Black N. How do patient reported outcome measures (PROMs) support clinician-patient communication and patient care? A realist synthesis. J Patient Rep Outcomes 2018; 2: 42.
17. Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, Selby PJ. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. J Clin Oncol 2004; 22: 714-724.

18. Watson L, Link C, Qi S, DeIure A, Chmielewski L, Hildebrand A, et al. Testing a modified electronic version of the Edmonton symptom assessment system-revised for remote online completion with ambulatory cancer patients in Alberta, Canada. Digit Health 2023; 9: 20552076231190998.
19. Bansback N, Trenaman L, MacDonald KV, Durand D, Hawker G, Johnson JA, et al. An online individualised patient decision aid improves the quality of decisions in patients considering total knee arthroplasty in routine care: A randomized controlled trial. Osteoarthr Cartil Open 2022; 4: 100286.
20. Dorr MC, van Hof KS, Jelsma JGM, Dronkers EAC, Baatenburg de Jong RJ, Offerman MPJ, de Bruijne MC. Quality improvements of healthcare trajectories by learning from aggregated patient-reported outcomes: a mixed-methods systematic literature review. Health Res Policy Syst 2022; 20: 90.
21. Greenhalgh J, Dalkin S, Gibbons E, Wright J, Valderas JM, Meads D, Black N. How do aggregated patient-reported outcome measures data stimulate health care improvement? A realist synthesis. J Health Serv Res Policy 2018; 23: 57-65.
22. Al Sayah F, Lahtinen M, Bonsel GJ, Ohinmaa A, Johnson JA. A multi-level approach for the use of routinely collected patient-reported outcome measures (PROMs) data in healthcare systems. J Patient Rep Outcomes 2021; 5: 98.
23. Ellis J. All inclusive benchmarking. J Nurs Manag 2006; 14: 377-383.
24. Willmington C, Belardi P, Murante AM, Vainieri M. The contribution of benchmarking to quality improvement in healthcare. A systematic literature review. BMC Health Serv Res 2022; 22: 139.
25. Valderas JM, Fitzpatrick R, Roland M. Using health status to measure NHS performance: another step into the dark for the health reform in England. BMJ Qual Saf 2012; 21: 352-353.
26. Totten AM, Wagner J, Tiwari A, O'Haire C, Griffin J, Walker M. Closing the quality gap: revisiting the state of the science (vol. 5: public reporting as a quality improvement strategy). Evid Rep Technol Assess (Full Rep) 2012: 1-645.
27. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. BMC Health Serv Res 2013; 13: 211.
28. van der Zee-van den Berg AI, Boere-Boonekamp MM, Groothuis-Oudshoorn CGM, MJ IJ, Haasnoot-Smallegange RME, Reijneveld SA. Post-Up Study: Postpartum Depression Screening in Well-Child Care and Maternal Outcomes. Pediatrics 2017; 140.
29. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Bmj 2021; 372: n71.
30. Roberts A YK, editors. Evidence-based practice manual: research and outcome measures in health and human services, New York, NY: Oxford University Press 2004.
31. Donabedian A. Evaluating the quality of medical care. 1966. Milbank Q 2005; 83: 691-729.
32. Agency for Healthcare Research and Quality R, MD. Six Domains of Healthcare Quality. Content last reviewed December 2022. .
33. Armijo-Olivo S, Stiles CR, Hagen NA, Biondo PD, Cummings GG. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. J Eval Clin Pract 2012; 18: 12-18.
34. Ackermans L, Hageman MG, Bos AH, Haverkamp D, Scholtes VAB, Poolman RW. Feedback to Patients About Patient-reported Outcomes Does Not Improve Empowerment or Satisfaction. Clin Orthop Relat Res 2018; 476: 716-722.

35. Allen J, Annells M, Nunn R, Petrie E, Clark E, Lang L, Robins A. Evaluation of effectiveness and satisfaction outcomes of a mental health screening and referral clinical pathway for community nursing care. J Psychiatr Ment Health Nurs 2011; 18: 375-385.

36. Almario CV, Chey WD, Khanna D, Mosadeghi S, Ahmed S, Afghani E, et al. Impact of National Institutes of Health Gastrointestinal PROMIS Measures in Clinical Practice: Results of a Multicenter Controlled Trial. Am J Gastroenterol 2016; 111: 1546-1556.

37. Boyce MB, Browne JP. The effectiveness of providing peer benchmarked feedback to hip replacement surgeons based on patient-reported outcome measures--results from the PROFILE (Patient-Reported Outcomes: Feedback Interpretation and Learning Experiment) trial: a cluster randomised controlled study. BMJ Open 2015; 5: e008325.

38. Buxton JA, Chandler-Altendorf A, Puente AE. A novel collaborative practice model for treatment of mental illness in indigent and uninsured patients. Am J Health Syst Pharm 2012; 69: 1054-1062.

39. de Jong MJ, van der Meulen-de Jong AE, Romberg-Camps MJ, Becx MC, Maljaars JP, Cilissen M, et al. Telemedicine for management of inflammatory bowel disease (myIBDcoach): a pragmatic, multicentre, randomised controlled trial. Lancet 2017; 390: 959-968.

40. Epstein JN, Rabiner D, Johnson DE, Fitzgerald DP, Chrisman A, Erkanli A, et al. Improving attention-deficit/hyperactivity disorder treatment outcomes through use of a collaborative consultation treatment service by community-based pediatricians: a cluster randomized trial. Arch Pediatr Adolesc Med 2007; 161: 835-840.

41. Fihn SD, McDonell MB, Diehr P, Anderson SM, Bradley KA, Au DH, et al. Effects of sustained audit/feedback on self-reported health status of primary care patients. Am J Med 2004; 116: 241-248.

42. Girgis A, Breen S, Stacey F, Lecathelinais C. Impact of two supportive care interventions on anxiety, depression, quality of life, and unmet needs in patients with nonlocalized breast and colorectal cancers. J Clin Oncol 2009; 27: 6180-6190.

43. Gossec L, Cantagrel A, Soubrier M, Berthelot JM, Joubert JM, Combe B, et al. An e-health interactive self-assessment website (Sanoia(®)) in rheumatoid arthritis. A 12-month randomized controlled trial in 320 patients. Joint Bone Spine 2018; 85: 709-714.

44. Hadjistavropoulos T, MacNab Y, Lints-Martindale A, Martin R, Hadjistavropoulos H. Does routine pain assessment result in better care? Pain Res Manag 2009; 14: 211-216.

45. Haller G, Agoritsas T, Luthy C, Piguet V, Griesser AC, Perneger T. Collaborative quality improvement to manage pain in acute care hospitals. Pain Med 2011; 12: 138-147.

46. Jaensson M, Dahlberg K, Eriksson M, Nilsson U. Evaluation of postoperative recovery in day surgery patients using a mobile phone application: a multicentre randomized trial. Br J Anaesth 2017; 119: 1030-1038.

47. Kjaer T, Dalton SO, Andersen E, Karlsen R, Nielsen AL, Hansen MK, et al. A controlled study of use of patient-reported outcomes to improve assessment of late effects after treatment for head-and-neck cancer. Radiother Oncol 2016; 119: 221-228.

48. Kroenke K, Talib TL, Stump TE, Kean J, Haggstrom DA, DeChant P, et al. Incorporating PROMIS Symptom Measures into Primary Care Practice-a Randomized Clinical Trial. J Gen Intern Med 2018; 33: 1245-1252.

49. Kumar RM, Fergusson DA, Lavallée LT, Cagiannos I, Morash C, Horrigan M, et al. Performance Feedback May Not Improve Radical Prostatectomy Outcomes: The Surgical Report Card (SuRep) Study. J Urol 2021; 206: 346-353.

50. Mallen CD, Nicholl BI, Lewis M, Bartlam B, Green D, Jowett S, et al. The effects of implementing a point-of-care electronic template to prompt routine anxiety and depression screening in patients consulting for osteoarthritis (the Primary Care Osteoarthritis Trial): A cluster randomised trial in primary care. PLoS Med 2017; 14: e1002273.
51. McCambridge J, Day M. Randomized controlled trial of the effects of completing the Alcohol Use Disorders Identification Test questionnaire on self-reported hazardous drinking. Addiction 2008; 103: 241-248.
52. Reiber GE, Au D, McDonell M, Fihn SD. Diabetes quality improvement in Department of Veterans Affairs Ambulatory Care Clinics: a group-randomized clinical trial. Diabetes Care 2004; 27 Suppl 2: B61-68.
53. Richardson J, Chan D, Risdon K, Giles C, Mulveney S, Cripps D. Does monitoring change in function in community-dwelling older adults alter outcome? A randomized controlled trial. Clin Rehabil 2008; 22: 1061-1070.
54. Rogers SN, Allmark C, Bekiroglu F, Edwards RT, Fabbroni G, Flavel R, et al. Improving quality of life through the routine use of the patient concerns inventory for head and neck cancer patients: main results of a cluster preference randomised controlled trial. Eur Arch Otorhinolaryngol 2021; 278: 3435-3449.
55. Santana MJ, Feeny D, Johnson JA, McAlister FA, Kim D, Weinkauf J, Lien DC. Assessing the use of health-related quality of life measures in the routine clinical care of lung-transplant patients. Qual Life Res 2010; 19: 371-379.
56. Shyu YI, Liang J, Tseng MY, Li HJ, Wu CC, Cheng HS, et al. Comprehensive and subacute care interventions improve health-related quality of life for older patients after surgery for hip fracture: a randomised controlled trial. Int J Nurs Stud 2013; 50: 1013-1024.
57. Varagunam M, Hutchings A, Neuburger J, Black N. Impact on hospital performance of introducing routine patient reported outcome measures in surgery. J Health Serv Res Policy 2014; 19: 77-84.
58. Volkmann ER, FitzGerald JD. Reducing gender disparities in post-total knee arthroplasty expectations through a decision aid. BMC Musculoskelet Disord 2015; 16: 16.
59. Wu S, Ell K, Jin H, Vidyanti I, Chou CP, Lee PJ, et al. Comparative Effectiveness of a Technology-Facilitated Depression Care Management Model in Safety-Net Primary Care Patients With Type 2 Diabetes: 6-Month Outcomes of a Large Clinical Trial. J Med Internet Res 2018; 20: e147.
60. Zaslansky R, Chapman CR, Baumbach P, Bytyqi A, Castro Lopes JM, Chetty S, et al. Improving perioperative pain management: a preintervention and postintervention study in 7 developing countries. Pain Rep 2019; 4: e705.
61. Partridge T, Carluke I, Emmerson K, Partington P, Reed M. Improving patient reported outcome measures (PROMs) in total knee replacement by changing implant and preserving the infrapatella fatpad: a quality improvement project. BMJ Qual Improv Rep 2016; 5.
62. Baker A, Mitchell EJ, Partlett C, Thomas KS. Evaluating the effect of weekly patient-reported symptom monitoring on trial outcomes: results of the Eczema Monitoring Online randomized controlled trial. Br J Dermatol 2023; 189: 180-187.
63. Balestrieri M, Sisti D, Rocchi M, Rucci P, Simon G, Araya R, de Girolamo G. Effectiveness of clinical decision support systems and telemedicine on outcomes of depression: a cluster randomized trial in general practice. Fam Pract 2020; 37: 731-737.
64. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom Monitoring With Patient-Reported Outcomes During Routine Cancer Treatment: A Randomized Controlled Trial. J Clin Oncol 2016; 34: 557-565.

65. Berdal G, Sand-Svartrud AL, Linge AD, Aasvold AM, Tennebø K, Eppeland SG, et al. Bridging gaps across levels of care in rehabilitation of patients with rheumatic and musculoskeletal diseases: Results from a stepped-wedge cluster randomized controlled trial. Clin Rehabil 2023; 37: 1153-1177.
66. Berinstein JA, Cohen-Mekelburg SA, Greenberg GM, Wray D, Berry SK, Saini SD, et al. A Care Coordination Intervention Improves Symptoms But Not Charges in High-Risk Patients With Inflammatory Bowel Disease. Clin Gastroenterol Hepatol 2022; 20: 1029-1038 e1029.
67. Cooley ME, Mazzola E, Xiong N, Hong F, Lobach DF, Braun IM, et al. Clinical Decision Support for Symptom Management in Lung Cancer Patients: A Group RCT. J Pain Symptom Manage 2022; 63: 572-580.
68. Dhingra L, Schiller R, Teets R, Nosal S, Dieckmann NF, Ginzburg R, et al. Pain Management in Primary Care: A Randomized Controlled Trial of a Computerized Decision Support Tool. Am J Med 2021; 134: 1546-1554.
69. Dobscha SK, Corson K, Hickam DH, Perrin NA, Kraemer DF, Gerrity MS. Depression decision support in primary care: a cluster randomized trial. Ann Intern Med 2006; 145: 477-487.
70. Ferrell B, Chung V, Hughes MT, Koczywas M, Azad NS, Ruel NH, et al. A Palliative Care Intervention for Patients on Phase 1 Studies. J Palliat Med 2021; 24: 846-856.
71. Fortmann AL, Walker C, Barger K, Robacker M, Morrisey R, Ortwine K, et al. Care Team Integration in Primary Care Improves One-Year Clinical and Financial Outcomes in Diabetes: A Case for Value-Based Care. Popul Health Manag 2020; 23: 467-475.
72. Garduño-López AL, Acosta Nava VM, Castro Garcés L, Rascón-Martínez DM, Cuellar-Guzmán LF, Flores-Villanueva ME, et al. Towards Better Perioperative Pain Management in Mexico: A Study in a Network of Hospitals Using Quality Improvement Methods from PAIN OUT. J Pain Res 2021; 14: 415-430.
73. Holm I, Pripp AH, Risberg MA. The Active with OsteoArthritis (AktivA) Physiotherapy Implementation Model: A Patient Education, Supervised Exercise and Self-Management Program for Patients with Mild to Moderate Osteoarthritis of the Knee or Hip Joint. A National Register Study with a Two-Year Follow-Up. J Clin Med 2020; 9.
74. Howell D, Li M, Sutradhar R, Gu S, Iqbal J, O'Brien MA, et al. Integration of patient-reported outcomes (PROs) for personalized symptom management in "real-world" oncology practices: a population-based cohort comparison study of impact on healthcare utilization. Support Care Cancer 2020; 28: 4933-4942.
75. Howell D, Rosberger Z, Mayer C, Faria R, Hamel M, Snider A, et al. Personalized symptom management: a quality improvement collaborative for implementation of patient reported outcomes (PROs) in 'real-world' oncology multisite practices. J Patient Rep Outcomes 2020; 4: 47.
76. Livanainen S, Ravichandra R, Jekunen A, Arokoski R, Mentu S, Lang L, et al. ePRO symptom follow-up of colorectal cancer patients receiving oxaliplatin-based adjuvant chemotherapy is feasible and enhances the quality of patient care: a prospective multicenter study. J Cancer Res Clin Oncol 2023; 149: 6875-6882.
77. Jayakumar P, Moore MG, Furlough KA, Uhler LM, Andrawis JP, Koenig KM, et al. Comparison of an Artificial Intelligence-Enabled Patient Decision Aid vs Educational Material on Decision Quality, Shared Decision-Making, Patient Experience, and Functional Outcomes in Adults With Knee Osteoarthritis: A Randomized Clinical Trial. JAMA Netw Open 2021; 4: e2037107.
78. Ngo E, Truong MB, Wright D, Nordeng H. Impact of a Mobile Application for Tracking Nausea and Vomiting During Pregnancy (NVP) on NVP Symptoms, Quality of Life, and Decisional Conflict Regarding NVP Treatments: MinSafeStart Randomized Controlled Trial. JMIR Mhealth Uhealth 2022; 10: e36226.

79. Patel MI, Kapphahn K, Dewland M, Aguilar V, Sanchez B, Sisay E, et al. Effect of a Community Health Worker Intervention on Acute Care Use, Advance Care Planning, and Patient-Reported Outcomes Among Adults With Advanced Stages of Cancer: A Randomized Clinical Trial. JAMA Oncol 2022; 8: 1139-1148.
80. Pérez JC, Fernández O, Cáceres C, Carrasco Á E, Moessner M, Bauer S, et al. An Adjunctive Internet-Based Intervention to Enhance Treatment for Depression in Adults: Randomized Controlled Trial. JMIR Ment Health 2021; 8: e26814.
81. Pooni A, Brar MS, Anpalagan T, Schmocker S, Rashid S, Goldstein R, et al. Home to Stay: A Randomized Controlled Trial Evaluating the Effect of a Postdischarge Mobile App to Reduce 30-Day Readmission Following Elective Colorectal Surgery. Ann Surg 2023; 277: e1056-e1062.
82. Price S, Hamann HA, Halaby L, Trejo J, Rogers FC, Weihs K. Collaborative depression care sensitive to the needs of underserved patients with cancer: Feasibility, acceptability and outcomes. J Psychosoc Oncol 2023: 1-23.
83. Rollman BL, Anderson AM, Rothenberger SD, Abebe KZ, Ramani R, Muldoon MF, et al. Efficacy of Blended Collaborative Care for Patients With Heart Failure and Comorbid Depression: A Randomized Clinical Trial. JAMA Intern Med 2021; 181: 1369-1380.
84. Skovlund PC, Vind Thaysen H, Schmidt H, Alsner J, Hjollund NH, Lomborg K, Nielsen BK. Effect of patient-reported outcomes as a dialogue-based tool in cancer consultations on patient self-management and health-related quality of life: a clinical, controlled trial. Acta Oncol 2021; 60: 1668-1677.
85. Steele Gray C, Chau E, Tahsin F, Harvey S, Loganathan M, McKinstry B, et al. Assessing the Implementation and Effectiveness of the Electronic Patient-Reported Outcome Tool for Older Adults With Complex Care Needs: Mixed Methods Study. J Med Internet Res 2021; 23: e29071.
86. Tirelli F, Xiao R, Brandon TG, Burnham JM, Chang JC, Weiss PF. Determinants of disease activity change over time in Enthesitis related arthritis: effect of structured outcome monitoring and clinical decision support. Pediatr Rheumatol Online J 2020; 18: 79.
87. Velikova G, Keding A, Harley C, Cocks K, Booth L, Smith AB, et al. Patients report improvements in continuity of care when quality of life assessments are used routinely in oncology practice: secondary outcomes of a randomised controlled trial. Eur J Cancer 2010; 46: 2381-2388.
88. Wylde V, Bertram W, Sanderson E, Noble S, Howells N, Peters TJ, et al. The STAR care pathway for patients with pain at 3 months after total knee replacement: a multicentre, pragmatic, randomised, controlled trial. Lancet Rheumatol 2022; 4: e188-e197.
89. Barbera L, Sutradhar R, Howell D, Sussman J, Seow H, Dudgeon D, et al. Does routine symptom screening with ESAS decrease ED visits in breast cancer patients undergoing adjuvant chemotherapy? Support Care Cancer 2015; 23: 3025-3032.
90. de Wit M, Delemarre-van de Waal HA, Bokma JA, Haasnoot K, Houdijk MC, Gemke RJ, Snoek FJ. Monitoring and discussing health-related quality of life in adolescents with type 1 diabetes improve psychosocial well-being: a randomized controlled trial. Diabetes Care 2008; 31: 1521-1526.
91. Detmar SB, Muller MJ, Schornagel JH, Wever LD, Aaronson NK. Health-related quality-of-life assessments and patient-physician communication: a randomized controlled trial. Jama 2002; 288: 3027-3034.
92. Hilarius DL, Kloeg PH, Gundy CM, Aaronson NK. Use of health-related quality-of-life assessments in daily clinical oncology nursing practice: a community hospital-based intervention study. Cancer 2008; 113: 628-637.
93. Rosenbloom SK, Victorson DE, Hahn EA, Peterman AH, Cella D. Assessment is not enough: a randomized controlled trial of the effects of HRQL assessment on quality of life and satisfaction in oncology clinical practice. Psychooncology 2007; 16: 1069-1079.

94. Weingarten SR, Kim CS, Stone EG, Kristopaitis RJ, Pelter M, Sandhu M. Can peer-comparison feedback improve patient functional status? Am J Manag Care 2000; 6: 35-39.

95. Buckley L, Ware E, Kreher G, Wiater L, Mehta J, Burnham JM. Outcome Monitoring and Clinical Decision Support in Polyarticular Juvenile Idiopathic Arthritis. J Rheumatol 2020; 47: 273-281.

96. Cross RK, Langenberg P, Regueiro M, Schwartz DA, Tracy JK, Collins JF, et al. A Randomized Controlled Trial of TELEmedicine for Patients with Inflammatory Bowel Disease (TELE-IBD). Am J Gastroenterol 2019; 114: 472-482.

97. Davidson KW, Rieckmann N, Clemow L, Schwartz JE, Shimbo D, Medina V, et al. Enhanced depression care for patients with acute coronary syndrome and persistent depressive symptoms: coronary psychosocial evaluation studies randomized controlled trial. Arch Intern Med 2010; 170: 600-608.

98. Frasure-Smith N, Lespérance F, Prince RH, Verrier P, Garber RA, Juneau M, et al. Randomised trial of home-based psychosocial nursing intervention for patients recovering from myocardial infarction. Lancet 1997; 350: 473-479.

99. Katon W, Robinson P, Von Korff M, Lin E, Bush T, Ludman E, et al. A multifaceted intervention to improve treatment of depression in primary care. Arch Gen Psychiatry 1996; 53: 924-932.

100. Katon WJ, Von Korff M, Lin EH, Simon G, Ludman E, Russo J, et al. The Pathways Study: a randomized trial of collaborative care in patients with diabetes and depression. Arch Gen Psychiatry 2004; 61: 1042-1049.

101. Katzelnick DJ, Simon GE, Pearson SD, Manning WG, Helstad CP, Henk HJ, et al. Randomized trial of a depression management program in high utilizers of medical care. Arch Fam Med 2000; 9: 345-351.

102. Kronish IM, Moise N, Cheung YK, Clarke GN, Dolor RJ, Duer-Hefele J, et al. Effect of Depression Screening After Acute Coronary Syndromes on Quality of Life: The CODIACS-QoL Randomized Clinical Trial. JAMA Intern Med 2020; 180: 45-53.

103. Maguire R, McCann L, Kotronoulas G, Kearney N, Ream E, Armes J, et al. Real time remote symptom monitoring during chemotherapy for cancer: European multicentre randomised controlled trial (eSMART). Bmj 2021; 374: n1647.

104. Patel MI, Ramirez D, Agajanian R, Agajanian H, Bhattacharya J, Bundorf KM. Lay Health Worker-Led Cancer Symptom Screening Intervention and the Effect on Patient-Reported Satisfaction, Health Status, Health Care Use, and Total Costs: Results From a Tri-Part Collaboration. JCO Oncol Pract 2020; 16: e19-e28.

105. Regueiro M, Click B, Anderson A, Shrank W, Kogan J, McAnallen S, Szigethy E. Reduced Unplanned Care and Disease Activity and Increased Quality of Life After Patient Enrollment in an Inflammatory Bowel Disease Medical Home. Clin Gastroenterol Hepatol 2018; 16: 1777-1785.

106. Sharpe M, Walker J, Holm Hansen C, Martin P, Symeonides S, Gourley C, et al. Integrated collaborative care for comorbid major depression in patients with cancer (SMaRT Oncology-2): a multicentre randomised controlled effectiveness trial. Lancet 2014; 384: 1099-1108.

107. Simon GE, Ralston JD, Savarino J, Pabiniak C, Wentzel C, Operskalski BH. Randomized trial of depression follow-up care by online messaging. J Gen Intern Med 2011; 26: 698-704.

108. Unützer J, Katon W, Callahan CM, Williams JW, Jr., Hunkeler E, Harpole L, et al. Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. Jama 2002; 288: 2836-2845.

109. Sandberg T, Conner M. A mere measurement effect for anticipated regret: impacts on cervical screening attendance. Br J Soc Psychol 2009; 48: 221-236.

110. Anne-Margreet van D, Hester FL, Johan PM, Ewout WS. Random variation and rankability of hospitals using outcome indicators. BMJ Quality & Safety 2011; 20: 869-874.

111. Goldstein H, Spiegelhalter DJ. League Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance. Journal of the Royal Statistical Society Series A: Statistics in Society 2018; 159: 385-409.

112. Marshall DA, Jin X, Pittman LB, Smith CJ. The use of patient-reported outcome measures in hip and knee arthroplasty in Alberta. Journal of Patient-Reported Outcomes 2021; 5: 87.

113. Taylor MJ, McNicholas C, Nicolay C, Darzi A, Bell D, Reed JE. Systematic review of the application of the plan-do-study-act method to improve quality in healthcare. BMJ Qual Saf 2014; 23: 290-298.

114. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev 2012: Cd000259.

DECLARATIONS

Ethics: Given the study design (systematic review) ethics approval was not required nor sought. This study was registered prospectively in PROSPERO under record 2022 CRD42022333400.

Consent for publication: Not applicable

Availability of data and materials: All data generated or analysed during this study are included in this published article

Competing interests: The authors declare that they have no competing interests

Funding/Support: This work was funded by a PhD grant (PHD-287) provided by the EuroQol Research Foundation. The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Author Contributions: Concept and design: JB, GB, HP; Literature search: JB, AI; Quality assessment: JB, AH; Interpretation of results: JB, AI, AH; Drafting of the manuscript: JB; Critical revision of the paper for important intellectual content: all authors; Obtaining funding: JB, HP, GB; Supervision: HP, GB

Acknowledgement: We would like to thank dr. Wichor Bramer (biomedical information specialist at Erasmus Medical Center) for his help in developing the search strategy.

Supplementary data: Supplementary data associated with this article can be found in the online version at doi: 10.1186/s12955-024-02312-4.

Chapter 3

Socioeconomic inequalities in patient-reported outcome measures of Dutch primary hip and knee arthroplasty patients for osteoarthritis

Authors
Joshua M. Bonsel[1] MD
Max Reijman[1], MSc, PhD
Jan A.N. Verhaar[1], MD, PhD, emeritus professor
Liza N. van Steenbergen[2], MSc, PhD
Mathieu F. Janssen[3], MSc, PhD
Gouke J. Bonsel[4], MD, PhD, emeritus professor

Affiliations
[1] Department of Orthopaedics and Sports Medicine, Erasmus MC, University Medical Center, Rotterdam, The Netherlands
[2] Dutch Arthroplasty Register (Landelijke Registratie Orthopedische Interventies), ’s Hertogenbosch, The Netherlands
[3] Department of Medical Psychology and Psychotherapy, Department of Psychiatry, Erasmus MC, University Medical Center, Rotterdam, The Netherlands
[4] EuroQol Research Foundation, Rotterdam, the Netherlands

Status
Published.
Citation: Bonsel JM, Reijman M, Verhaar JAN, van Steenbergen LN, Janssen MF, Bonsel GJ. Socioeconomic inequalities in patient-reported outcome measures of Dutch primary hip and knee arthroplasty patients for osteoarthritis. Osteoarthritis Cartilage. 2024;32(2):200-9.

ABSTRACT

Objective

To study socio-economic inequalities in patient-reported outcomes in primary hip and knee arthroplasty (THA/TKA) patients for osteoarthritis, using two analytical techniques.

Methods

We obtained data from 44,732 THA and 30,756 TKA patients with preoperative and 12-month follow-up PROMs between 2014 and 2020 from the Dutch Arthroplasty Registry. A deprivation indicator based on neighborhood income, unemployment rate, and education level was linked and categorized into quintiles. The primary outcome measures were the EQ-5D-3L index and Oxford Hip/Knee Score (OHS/OKS) preoperative, at 12-month follow-up, and the calculated change score between these measurements. We contrasted the most and least deprived quintiles using multivariable linear regression, adjusting for patient characteristics. Concurrently, we calculated concentration indices as a non-arbitrary tool to quantify inequalities.

Results

Compared to the least deprived, the most deprived THA patients had poorer preoperative (EQ-5D –0.03 (95%CI –0.02, –0.04), OHS –1.26 (–0.99, –1.52)) and 12-month follow-up health (EQ-5D –0.02 (–0.01, –0.02), OHS –0.42 (–0.19, –0.65)), yet higher mean change (EQ-5D 0.02 (0.01, 0.03), OHS 0.84 (0.52, 1.16)). The most deprived TKA patients had similar results. The higher mean change among the deprived resulted from lower preoperative health in this group (confounding). After accounting for this, the most deprived patients had a lower mean change. The concentration indices showed similar inequality effects and provided information on the magnitude of inequalities over the entire socio-economic range.

Conclusion

The most deprived THA and TKA patients have worse preoperative health, which persisted after surgery. The concentration indices allow comparison of inequalities across different outcomes (e.g., revision risk).

INTRODUCTION

Health inequalities are unfair and potentially avoidable health differences between population groups. Socioeconomic status (SES), a composite measure typically covering education, income, and occupation, is recognized as a universal driving force of health inequalities[1, 2]. Unequal distribution of health is considered a human right infringement, but it also hampers economic prosperity and social development[3, 4]. Traditionally, clinical outcomes such as mortality, are used to measure health inequalities. However, a useful alternative may be patient-reported outcome measures (PROMs), which are standardized surveys used to quantify generic or disease-specific health status from the patient's perspective[5].

Hip and knee arthroplasty (THA, TKA) are commonly performed surgical procedures, resulting in pain relief and improved mobility in individuals with end-stage osteoarthritis[6, 7]. Despite being standardized procedures, striking socio-economic inequalities have been demonstrated in these patients. Deprived patients report inferior PROMs after THA or TKA, even after adjusting for patient and surgical characteristics[8-16]. Some studies observed worse PROMs in deprived patients prior to surgery due to patient or provider selection[8, 9, 11, 15]. Surgical recovery can be expressed as the change score, which is the difference between post- and preoperative PROMs. When analyzing this score, studies commonly control for preoperative PROMs because of their significant association with the change score[11]. However, adjustment obscures our understanding of how the effect of deprivation prior to surgery might influence the effect of deprivation during recovery. One exploratory study recognized this issue and stratified patients by preoperative health, in order to expose deprivation effects on the mean change score[9].

Furthermore, to analyze inequalities, outcomes of the top 20% (most affluent) and bottom 20% (most deprived) patients are usually contrasted[17, 18]. However, this approach does not provide insight into the amount of inequality in the overall population and does not allow for comparison of the size of inequalities across different outcomes. Additionally, it is impossible to determine the role of attributable factors (e.g., patient characteristics), which could inform efforts aimed at reducing inequalities. For this purpose, economists developed inequality measures such as the concentration index, which can be transferred to public health sciences[19]. This index, closely related to the Gini coefficient, provides a non-arbitrary estimate of the 'volume' of health inequality in a population according to a deprivation indicator[18].

In this study, we aimed to analyze socio-economic health inequalities in THA and TKA patients using Dutch registry data, unraveling the effect of deprivation on pre- and postoperative PROMs. Additionally, we used the concentration index to express possible inequalities and assess the impact of deprivation amidst other potentially influential factors.

METHODS

An observational cohort study was performed using anonymized prospectively collected data from the Dutch Arthroplasty Registry (LROI). Ethical approval is not required by Dutch law for this type of study. We adhered to the STROBE guideline for observational studies.

Data and inclusion criteria

In the Netherlands, clinical registration of data from THA and TKA patients is professionally endorsed. The LROI has a completion rate of over 95%[20]. Variables include patient and surgical characteristics and outcomes. The collection of internationally validated general health and disease-specific PROMs has been strongly recommended since 2014. Since then, the collection of PROMs from patients undergoing surgery increased notably, and stabilized in 2017, with approximately 40% of primary THA and TKA patients having preoperative and 12-month follow-up outcomes[21].

All patients who had primary THA or TKA for osteoarthritis from 2014 to 2020, and of whom preoperative and 12-month follow-up PROMs were available were included. We selected only patients who received arthroplasty for osteoarthritis, as this is the largest and most homogeneous group of elective orthopedic patients in which PROMs are collected. 11% of THA patients and 12% of TKA patients also underwent contralateral joint replacement during this period; the second record was ignored to avoid data dependencies.

Variables

The following patient- and surgical data was extracted from the registry: age, biological sex, body mass index (BMI), Charnley score[22], American Society of Anesthesiologists (ASA) score[23], previous surgery of the replaced joint, smoking status, type of hospital, fixation method, and surgical approach. The Charnley score represents the extent of osteoarthritis disease, ranging from “A” (one joint affected) to “C” (multiple joints affected or severely impaired quality of life due to the disease). The hospital type is categorized into private, general, or university hospitals. Fixation methods were categorized into cemented, uncemented, or hybrid. Approach methods for THA were categorized into anterior, anterolateral, posterolateral, direct lateral, and other; for TKA into medial para-patellar and other.

Deprivation indicator

As a deprivation indicator, a small neighborhood SES score was linked to the data using the 4-digit postal code of the home address of each patient[24]. The Netherlands Institute for

Social Research, a government organization, calculates this SES score for each 4-digit postal code area with a minimum of 100 inhabitants (mean 4300 inhabitants). This standardized continuous score is derived from the mean income per household, % households with a low income, % unemployed inhabitants, and % households with a low education per postal code area. We categorized this score into quintiles, using cumulative z-distribution thresholds for the 20^{th}, 40^{th}, 60^{th}, and 80^{th} percentile. For the descriptive analysis, we regrouped the 5 quintiles into 3 groups: most deprived (quintile 1), medium deprived (quintile 2–4), and least deprived (quintile 5), following guidelines of the World Health Organisation[25].

Outcomes (PROMs)

The Dutch Orthopaedic Association advises to collect PROMs at three time points: a maximum of 6 months before surgery, postoperatively at 3-months follow-up (range 2–4 months) for THA patients, at 6-months follow-up (5–7 months) for TKA patients, and at 12-months follow-up (11–13 months) for both procedures. Preoperative PROMs are generally completed at the outpatient clinic, whilst postoperative PROMs are completed electronically after invitation per email or with pen and paper.

The LROI contains multiple PROMs; we selected the EQ-5D-3L, the Oxford Hip Score (OHS), and the Oxford Knee Score (OKS). The EQ-5D-3L questionnaire has 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, on which patients score their general health. With the scores on the 5 dimensions, an index value was calculated using the Dutch National Value set[26]. This value ranges from –0.329 to 1.0, indicating the poorest and full health, respectively. The OHS and OKS are 12-item questionnaires which assess functioning and pain in patients with osteoarthritis of the hip or knee, respectively. The OHS and OKS range from 0 to 48, indicating most and no disability, respectively.

Missing data

Missing Charnley scores were conservatively estimated as "A", as the indication for joint replacement was osteoarthritis. Ages < 10 years and > 105 years, and BMI values < 10 and > 70, were considered erroneous and were recoded as 'missing' conform to the guidelines of the LROI. Missing values in one or more of the extracted variables were present in 4% of THA and TKA patients with complete preoperative and 12-month follow-up PROMs. Visual plots indicated random missingness (Supplemental Figures 1 and 2). Therefore, we excluded patients with missing data and did a complete case analysis.

Statistical analysis

For the descriptive analysis of the association between deprivation and patient/surgical characteristics, we used the Analysis of Variance (ANOVA (t-test)) and chi-squared test.

Multivariable linear regression models were used to estimate the association of deprivation with preoperative, 12-month follow-up, and change score between these measurements, expressed with the EQ-5D index and OHS/OKS. The least deprived quintile was used as the reference category. The coefficients of the change score analysis reflect differences in mean change scores by deprivation quintiles, which we refer to as the "mean change" henceforth. Analyses were repeated with adjustments for sex, age, BMI, ASA score, Charnley score, and type of hospital. Age and sex were regarded as potential confounders, as their relation with arthroplasty outcomes has been found consistently, and they were unlikely to act as intermediate variables in the studied association[27, 28]. BMI, ASA score, Charnley score, and the type of hospital may act as confounders and intermediate variables[29, 30]. To help assess clinical relevance, the coefficients were compared with coefficients of other included variables in the analyses. The full models are included in Supplemental File 1, Tables 1-4.

To analyze the effect of the preoperative score on the mean change, separately from the role of deprivation, patients were stratified according to the preoperative score in quintiles. Subsequently, we present the mean change for each preoperative quintile according to the deprivation indicator. Confounding is regarded to be present if deprivation is associated with the preoperative score and if, within the separate preoperative strata, deprivation influences the mean change.

We calculated concentration indices, a bivariate quantity which relates the inequality in outcomes (here: PROMs) to the ranking of the individual's SES in the entire population[31]. As the polarity of the outcome and its scale impacts the computed concentration index, we employed Erreyger's standardization for convenience of interpretation and to enable comparisons between results obtained from different PROMs[32, 33]. The index range is –1.0 to +1.0, where a positive value indicates that better health is concentrated among higher-ranked individuals. Zero represents perfect equality. More details on the used formulas are included in Supplemental File 2.

The concentration index was 'decomposed' to identify what percentage particular variables contribute to the observed inequality[32, 34]. This standard technique uses a multivariable linear regression model to determine the sensitivity of each variable to the outcome. The relative (percentage) contribution of each variable to the concentration index is the product of its sensitivity to the outcome and the degree of socio-economic inequality in the variable of interest. Similarly, relative positive and negative percentages reflect an increase and decrease respectively in overall inequalities. Zero means no contribution to inequality. The same variables were used as in the regression analysis; of particular interest were the variables which might act as confounders and intermediates.

All statistical analyses were performed in Stata (version 17.0). Coefficients and indices are presented with 95% confidence intervals (95% CIs). A p-value < 0.05 was considered significant. We explored the potential presence of response selection by comparing differences in patient characteristics by deprivation when including non-responders to PROMs. We also checked results in several patient respondent groups: respondents with preoperative and 3-/6-month PROMs, respondents with preoperative PROMs regardless of having other measurements, and respondents with 12-month follow-up PROMs regardless of having other measurements. As results were nearly identical to primary analyses, these were not separately presented (Supplemental File 1, Tables 1–5).

RESULTS

Descriptive analysis

Data from 44,732 THA and 30,756 TKA arthroplasty patients met our inclusion criteria (Figure 1). In both THA and TKA patients, 89% had complete EQ-5D and OHS/OKS response, and for 11%, either was missing.

Most arthroplasty patients were female (THA 63%, TKA 61%). The most deprived THA patients were on average 1 year older than the least deprived patients (69.7, SD 9.5 vs 68.6, SD 9.5); this difference was not present in TKA patients (68.7, SD 8.7 vs 69.0, SD 8.4). In both THA and TKA patients, deprivation was associated with higher BMI and ASA scores and undergoing surgery in non-private hospitals. In THA patients, the deprived patients received no cementation compared to cementation more often than affluent patients (most deprived: 63% vs 28%; least deprived: 75% vs 18%). Similarly, surgery was mainly performed via the posterolateral compared to the anterior approach (most deprived: 57% vs 31%, least deprived: 45% vs 44%) (Tables 1A and Table 1B).

Table 1A: Demographics of hip arthroplasty patients with PROMs by socio-economic status

	Q1 (most depr.)	**Q2-4 (med. depr.)a**	**Q5 (least depr.)**	**p-value**
Total	8,522 (19)	28,409 (64)	7,801 (17)	
Age, mean (SD)	69.7 (9.5)	68.9 (9.3)	68.6 (9.5)	<0.001
BMI, mean (SD)	27.8 (4.5)	27.4 (4.4)	26.7 (4.2)	<0.001
Sex, n (%)				
Female	5,560 (65)	17,844 (63)	4,927 (63)	<0.001
ASA, n (%)				
I	1,248 (15)	5,399 (19)	1,699 (22)	<0.001
II	5,491 (64)	18,313 (65)	4,952 (64)	
III-IV	1,783 (21)	4,697 (17)	1,150 (15)	

Table 1A: Continued

	Q1 (most depr.)	Q2-4 (med. depr.)a	Q5 (least depr.)	p-value
Charnley, n (%)				
A	4,168 (49)	13,652 (48)	3,930 (50)	<0.001
B1	2,804 (33)	9,851 (35)	2,640 (39)	
B2	1,253 (15)	4,010 (14)	1,077 (14)	
C	297 (4)	896 (3)	154 (2)	
Smoking, n (%)				
Yes	883 (10)	2,593 (9)	602 (8)	<0.001
Previous surgery of the joint, n (%)				
Yes	155 (2)	458 (2)	145 (2)	0.201
Type of hospital, n (%)				
Private Hospital	463 (5)	2,503 (9)	830 (11)	<0.001
General Hospital	7,805 (92)	25,503 (90)	6,887 (88)	
University Hospital	254 (3)	403 (1)	84 (1)	
Fixation, n (%)				
Uncemented	5,349 (63)	20,096 (71)	5,864 (75)	<0.001
Cemented	2,387 (28)	5,779 (20)	1,393 (18)	
Hybrid	786 (9)	2,534 (9)	544 (7)	
Approach, n (%)				
Anterior	2,648 (31)	10,041 (35)	3,420 (44)	<0.001
Anterolateral	409 (5)	1,223 (4)	271 (4)	
Direct Lateral	556 (7)	1,676 (6)	520 (7)	
Posterolateral	4,837 (57)	15,122 (53)	3,517 (45)	
Other	72 (1)	347 (1)	73 (1)	
EQ-5D index, mean (SD)				
Preoperative	0.54 (0.28)	0.57 (0.27)	0.59 (0.26)	
12-month follow-up	0.85 (0.20)	0.87 (0.18)	0.88 (0.17)	
Mean change	0.30 (0.30)	0.29 (0.28)	0.28 (0.28)	
OHS/OKS, mean (SD)				
Preoperative	22.19 (8.62)	23.19 (8.32)	24.12 (8.35)	
12-month follow-up	41.54 (7.37)	42.29 (7.18)	42.55 (7.25)	
Mean change	19.39 (9.76)	19.06 (9.69)	18.42 (9.83)	

Only patients with complete patient and surgical characteristics were included. ANOVA (t-test) and chi-squared tests were performed for continuous and categorical variables respectively.

[a] For convenience purposes the middle 3 deprivation quintiles [Q2, 3, 4] were collapsed under medium socio-economic status (SES).

Abbreviations: BMI = Body Mass Index; ASA = American Society of Anesthesiology score

Table 1B: Demographics of knee arthroplasty patients with PROMs by socio-economic status

	Q1 (most depr.)	**Q2-4 (med. depr.)a**	**Q5 (least depr.)**	**p-value**
Total	6,497 (21)	19,689 (64)	4,570 (15)	
Age, mean (SD)	68.7 (8.7)	68.3 (8.5)	69.0 (8.4)	<0.001
BMI, mean (SD)	30.1 (5.0)	29.4 (4.8)	28.9 (4.6)	<0.001
Sex, n (%)				
Female	4,096 (63)	11,878 (60)	2,769 (61)	<0.001
ASA, n (%)				
I	655 (10)	2,716 (14)	682 (15)	<0.001
II	4,288 (66)	13,187 (67)	3,066 (67)	
III-IV	1,554 (24)	3,786 (19)	822 (18)	
Charnley, n (%)				
A	3,138 (48)	9,386 (48)	2,151 (47)	0.072
B1	2,370 (36)	7,391 (38)	1,756 (38)	
B2	761 (12)	2,174 (11)	525 (12)	
C	228 (4)	738 (4)	138 (3)	
Smoking, n (%)				
Yes	594 (9)	1,498 (8)	284 (6)	<0.001
Previous surgery of the joint, n (%)				
Yes	1,910 (29)	6,267 (32)	1,328 (29)	<0.001
Type of hospital, n (%)				
Private Hospital	417 (6)	2,011 (10)	490 (11)	<0.001
General Hospital	5,914 (91)	17,473 (89)	4,030 (88)	
University Hospital	166 (3)	205 (1)	50 (1)	
Fixation, n (%)				
Uncemented	210 (3)	503 (3)	312 (7)	<0.001
Cemented	6,056 (93)	18,725 (95)	4,123 (90)	
Hybrid	231 (4)	461 (2)	135 (3)	
Approach, n (%)				
Medial parapatellar	6,365 (98)	19,065 (97)	4,378 (96)	<0.001
Other	132 (2)	624 (3)	192 (4)	
EQ-5D index, mean (SD)				
Preoperative	0.58 (0.27)	0.61 (0.25)	0.64 (0.24)	
12-month follow-up	0.82 (0.21)	0.84 (0.19)	0.85 (0.18)	
Mean change	0.24 (0.28)	0.23 (0.27)	0.22 (0.26)	
OHS/OKS, mean (SD)				
Preoperative	22.29 (7.49)	23.50 (7.31)	24.28 (7.50)	
12-month follow-up	37.94 (8.60)	39.12 (7.89)	39.57 (7.63)	
Mean change	15.56 (9.07)	15.60 (8.86)	15.29 (8.90)	

The most deprived patients had slightly lower PRO response rates than medium (quintile 2–4) and least (quintile 5) deprived patients, implying that the complete LROI dataset shows a slightly higher prevalence of THA and TKA among the deprived. The abovementioned deprivation patterns among responders did not change if non-responders to PROMs were added (Supplemental File 1, Table 6).

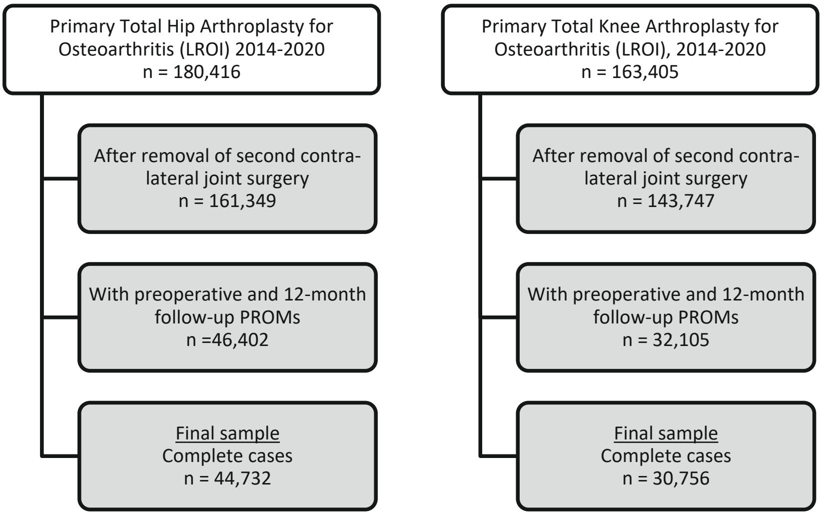

Figure 1: Inclusion flowchart of patients with PROMs

Epidemiological analysis of deprivation

The outcome contrast between most and least deprived quintiles was the largest preoperative, in both THA and TKA patients, with similar magnitude. These differences were slightly attenuated after adjustment for patient characteristics. The most deprived THA patients had 0.03 (95% CI: 0.02, 0.04) lower EQ-5D index values and 1.26 (95% CI: 0.99, 1.52) lower OHS scores. In TKA patients, this difference was 0.04 (95% CI: 0.03, 0.05) on the EQ-5D and 1.38 (95% CI: 1.09, 1.66) on the OKS. The difference between most and least deprived quintiles halved at 12 months follow-up. The most deprived THA patients had 0.02 (95% CI: 0.01, 0.02) lower EQ-5D values and 0.42 (95% CI: 0.19, 0.65) lower OHS scores. Differences in EQ-5D (0.02 (95% CI: 0.01, 0.03)) were similar in TKA patients but were larger on the OKS (1.14 (95% CI: 0.83, 1.46)) (Table 2).

Table 2: Socio-economic gradient in EQ-5D index and OHS/OKS scores

	Preoperative		12-month follow-up		Mean change[a]	
	Unadjusted Difference (95% CI)	**Adjusted Difference (95% CI)**	**Unadjusted Difference (95% CI)**	**Adjusted Difference (95% CI)**	**Unadjusted Difference (95% CI)**	**Adjusted Difference (95% CI)**
THA, EQ-5D index	*n = 39,744*					
5 (least deprived)	*ref*	*ref*	*ref*	*ref*	*ref*	*ref*
4	-0.01 (-0.02, 0.00)	-0.01 (-0.02, 0.00)	0.00 (-0.01, 0.00)	0.00 (-0.01, 0.00)	0.01 (0.00, 0.02)	0.01 (0.00, 0.02)
3	-0.02 (-0.03, -0.01)	-0.02 (-0.02, -0.01)	-0.01 (-0.02, 0.00)	-0.01 (-0.01, 0.00)	0.01 (0.00, 0.02)	0.01 (0.00, 0.02)
2	-0.03 (-0.04, -0.02)	-0.02 (-0.03, -0.01)	-0.01 (-0.02, -0.01)	-0.01 (-0.01, 0.00)	0.02 (0.01, 0.03)	0.01 (0.01, 0.02)
1 (most deprived)	-0.05 (-0.06, -0.04)	-0.03 (-0.04, -0.02)	-0.03 (-0.03, -0.02)	-0.02 (-0.02, -0.01)	0.02 (0.01, 0.03)	0.02 (0.01, 0.03)
OHS	*n = 39,991*					
5 (least deprived)	*ref*	*ref*	*ref*	*ref*	*ref*	*ref*
4	-0.47 (-0.74, -0.20)	-0.38 (-0.64, -0.12)	-0.05 (-0.29, 0.18)	0.03 (-0.20, 0.26)	0.42 (0.10, 0.74)	0.41 (0.09, 0.73)
3	-0.86 (-1.13, -0.60)	-0.64 (-0.9, -0.39)	-0.28 (-0.51, -0.05)	-0.09 (-0.32, 0.13)	0.58 (0.28, 0.89)	0.55 (0.24, 0.86)
2	-1.26 (-1.52, -1.01)	-0.92 (-1.17, -0.67)	-0.39 (-0.62, -0.17)	-0.10 (-0.32, 0.11)	0.87 (0.57, 1.18)	0.82 (0.51, 1.12)
1 (most deprived)	-1.95 (-2.23, -1.68)	-1.26 (-1.52, -0.99)	-0.98 (-1.21, -0.74)	-0.42 (-0.65, -0.19)	0.98 (0.66, 1.30)	0.84 (0.52, 1.16)
TKA, EQ-5D index	*n = 27,391*					
5 (least deprived)	*ref*	*ref*	*ref*	*ref*	*ref*	*ref*
4	-0.01 (-0.02, 0.00)	-0.01 (-0.02, 0.00)	0.00 (-0.01, 0.01)	0.00 (-0.01, 0.01)	0.01 (0.00, 0.02)	0.01 (0.00, 0.02)
3	-0.02 (-0.03, -0.01)	-0.02 (-0.03, -0.01)	-0.01 (-0.01, 0.00)	-0.01 (-0.01, 0.00)	0.02 (0.01, 0.03)	0.02 (0.00, 0.03)
2	-0.03 (-0.04, -0.02)	-0.03 (-0.04, -0.02)	-0.01 (-0.02, -0.01)	-0.01 (-0.02, 0.00)	0.02 (0.01, 0.03)	0.02 (0.01, 0.03)
1 (most deprived)	-0.06 (-0.07, -0.05)	-0.04 (-0.05, -0.03)	-0.03 (-0.04, -0.02)	-0.02 (-0.03, -0.01)	0.03 (0.02, 0.04)	0.02 (0.01, 0.03)
OKS	*n = 27,418*					
5 (least deprived)	*ref*	*ref*	*ref*	*ref*	*ref*	*ref*
4	-0.32 (-0.63, -0.01)	-0.27 (-0.57, 0.02)	-0.17 (-0.51, 0.17)	-0.15 (-0.48, 0.17)	0.15 (-0.22, 0.53)	0.12 (-0.25, 0.49)
3	-0.78 (-1.07, -0.48)	-0.66 (-0.94, -0.38)	-0.33 (-0.65, -0.01)	-0.24 (-0.55, 0.07)	0.45 (0.09, 0.80)	0.42 (0.06, 0.77)

Table 2: Continued

	Preoperative		12-month follow-up		Mean change[a]	
	Unadjusted Difference (95% CI)	Adjusted Difference (95% CI)	Unadjusted Difference (95% CI)	Adjusted Difference (95% CI)	Unadjusted Difference (95% CI)	Adjusted Difference (95% CI)
2	-1.08 (-1.37, -0.79)	-0.81 (-1.09, -0.54)	-0.78 (-1.09, -0.47)	-0.54 (-0.84, -0.24)	0.30 (-0.05, 0.65)	0.27 (-0.07, 0.62)
1 (most deprived)	-1.94 (-2.24, -1.65)	-1.38 (-1.66, -1.09)	-1.67 (-1.99, -1.34)	-1.14 (-1.46, -0.83)	0.27 (-0.09, 0.64)	0.23 (-0.13, 0.60)

Linear regression was used to analyse difference between the socio-economic status quintiles in patients with preoperative and 12-month follow-up PROMs. The least deprived category was used as reference. Adjustments were made for sex, age, BMI, American Society of Anesthesiology score, Charnley score, and type of hospital. Coefficients are reported with the respective 95% CIs.

[a] The mean change is based on the difference (change) score between the 12-month follow-up and preoperative score.

The mean change was higher in the most deprived patients. Adjustments for patient characteristics did not affect this. The most deprived THA patients had 0.02 (95% CI: 0.01, 0.03) higher EQ-5D and 0.84 (95% CI: 0.52, 1.16) higher OHS mean change scores. Differences in the EQ-5D were similar in TKA patients, however, the OKS showed no significant deprivation effect on the mean change. The most deprived TKA patients had 0.02 (95% CI: 0.01, 0.03) higher EQ-5D and 0.23 (95% CI: -0.13, 0.60) higher OKS mean change.

The preoperative, 12-month follow-up and mean change regression coefficients assigned to the most deprived group generally exceeded the coefficients assigned to Charnley score "C" (Supplemental File 1, Tables 1–4). This was not consistently the case for other variables included in the analyses.

Confounding effects of deprivation

Through stratification, we evaluated whether the lower preoperative scores in the deprived patients confounded the finding that deprived patients experience a larger mean change. We stratified patients according to the preoperative score (in quintiles) and deprivation. Subsequently, within each preoperative stratum, we calculated the mean change for each deprivation quintile. Independent from deprivation, a lower preoperative state in the case of THA and TKA led to a larger mean change. The EQ-5D and OHS/OKS mean change were lower in the most compared to the least deprived quintile of patients for all preoperative scores. With increasing preoperative scores, this difference in mean change of both outcomes decreased. While the pattern was identical, differences were larger in TKA compared to THA patients (Figure 2).

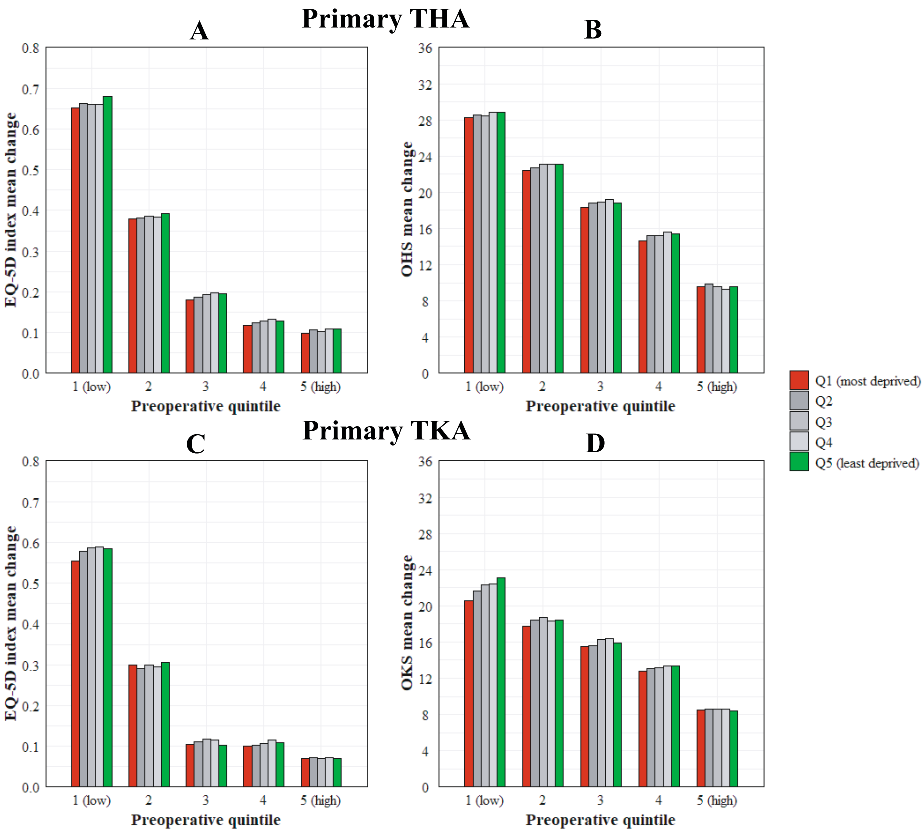

Figure 2: Stratification of mean change scores according to the deprivation quintiles, preoperative EQ-5D index, and OHS/OKS quintiles in hip (A, B) and knee (C, D) patients

Economic inequality analysis

The sign of the concentration index indicates the direction of the association between SES and outcomes, whilst the magnitude reflects both the strength of the relationship and variability in the outcome. Positive (negative) indices indicate that outcomes are more centered among the affluent (deprived). The concentration indices also indicated outcome inequalities: cross-sectional outcomes were more centered among the affluent, while the mean change generally was centered among the deprived. In THA and TKA patients, preoperative concentration indices calculated on the EQ-5D and OHS/OKS were +0.03, and they decreased slightly at 12-month follow-up to +0.02, except for the OKS which remained at +0.03. The concentration indices of the mean change were –0.01 in all outcomes except for the OKS, which was not significantly different from zero. The 95% CIs were narrow for all indices, with a maximum range of 0.01 (Table 3).

Table 3: Size of inequalities measured with concentration indices in EQ-5D index and OHS/OKS scores

	Concentration indices
Primary THA	
Preoperative	
EQ-5D index	0.03 (0.03, 0.03)
OHS	0.03 (0.03, 0.04)
12-month follow-up	
EQ-5D index	0.02 (0.01, 0.02)
OHS	0.02 (0.01, 0.02)
Mean change[a]	
EQ-5D index	-0.01 (-0.01, 0.00)
OHS	-0.01 (-0.01, -0.01)
Primary TKA	
Preoperative	
EQ-5D-3L index	0.03 (0.03, 0.04)
OKS	0.03 (0.03, 0.04)
12-month follow-up	
EQ-5D index	0.02 (0.02, 0.02)
OKS	0.03 (0.02, 0.03)
Mean change[a]	
EQ-5D index	-0.01 (-0.01, -0.01)
OKS	0.00 (-0.01, 0.00)

Concentration indices were calculated in patients with preoperative and 12-month follow-up PROMs. Indices are reported with the respective 95% CIs.
[a] The mean change is based on the difference (change) score between the 12-month follow-up and preoperative score.

Economic decomposition analysis

Subsequently, we decomposed the concentration indices to identify the relative contribution of each variable on the observed inequality and what remaining inequalities can be attributed to SES. The relative contributions of entered variables were similar for the EQ-5D index and Oxford set (Supplemental File 1, Tables 6–10). In THA and TKA patients, the entered variables explained 24–45% of the observed inequality in the cross-sectional (preoperative, 12-month follow-up) EQ-5D index, while the rest was attributed to SES. Variables with relatively large contributions were ASA III–IV (9–16%) and the lowest BMI quintile (8–9%). The other potential confounding/intermediate variables (type of hospital, Charnley score) had negligible contributions. The remaining SES inequalities were explained mainly by the most deprived (41–63%) and 2nd most deprived quintiles (11–19%).

As with the linear regression model, the explanatory variables accounted for a small part of the observed inequality (15–23%) on the mean change. The second-largest contributor was the lowest BMI quintile, explaining 9% in THA and TKA patients. Other variables had negligible contributions. The remaining SES inequalities were mostly explained by the most deprived (67–77%) and 2nd most deprived quintiles (25–30%) (Figure 3).

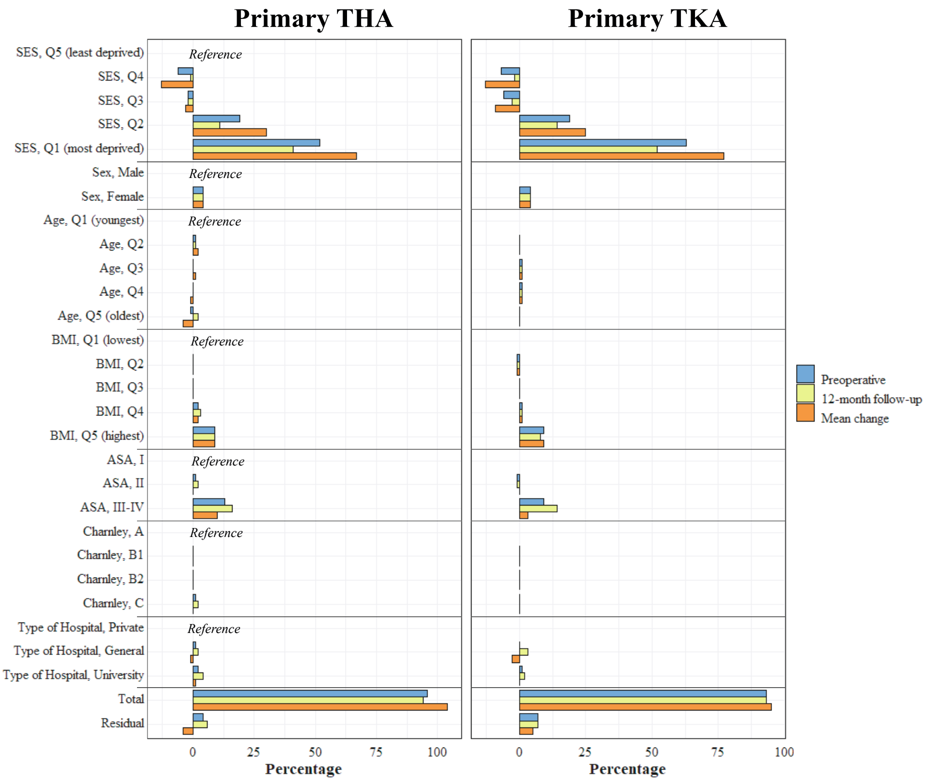

Figure 3: Decomposition of inequalities in the EQ-5D index. Percentages represent the relative contribution of the entered variable on the total amount of inequalities measured with the concentration indices. The category with the least expected contribution was used as the reference category. The residual is the amount of socioeconomic inequality not explained by the model.

DISCUSSION

In this study, we established the presence of systematic socio-economic health inequalities in Dutch primary THA and TKA patients using two complementary techniques from epidemiological and economic origin. Overall, results among THA and TKA patients were strikingly similar. Deprivation was strongly associated with poorer health prior to and, although to a lesser extent, after surgery. After accounting for the confounding effect of the preoperative score, deprivation was also associated with a lower mean change. The observed inequalities were only in part explained by patient and surgical characteristics in epidemiological and economic analysis. Interestingly, in the epidemiological analysis, the differences between the lowest and highest deprivation quintile superseded the Charnley score both before and after surgery, which are routinely taken into account (Supplemental File 1, Tables 1-4). Within the medical domain, no examples currently exist as a reference for the size of the inequalities measured with the concentration index; our results are the first and may therefore serve as a benchmark.

The association of deprivation with worse PROMs before and after arthroplasty is aligned with previous studies[8-16]. However, to our knowledge, only Clement et al. [8] revealed the double mechanism of deprivation in affecting preoperative health and health change. This study however only had a small sample size and collected data from a single medical center, resulting in limitations in terms of generalizability.

Preoperative inequalities

Preoperative outcome inequality may depend on several independent mechanisms. The first mechanism is selection: different thresholds of pain and dysfunction for surgery may be used by the provider and patient[35]. Surgeons may apply more stringent criteria in deprived patients to compensate for the higher prevalence of patient characteristics impairing postoperative outcomes or causing complications. Selecting the most severe cases leads to a higher mean change, as our results confirm. Provider selection may also already occur at the referral stage by the general practitioner[36, 37]. Furthermore, deprived patients may delay seeking help, as they are more accepting of chronic pain and functional limitations compared to non-deprived patients[38]. Neuburger et al. [15] confirmed this and reported more and a longer duration of symptoms at the time of surgery in deprived THA and TKA patients in the United Kingdom.

Additionally, deprived patients receive fewer arthroplasties relative to the number of patients in need[39, 40]. This could potentially be explained by access barriers posed by high deductibles, which also applies to the Netherlands. For example, high deductibles are associated with decreased use of psychotherapy in low-income adults[41]. Far less research has been conducted on other factors that may influence access, such as medical knowledge, cultural beliefs, travel distance, avoidance of seeking care, and access stops when hospitals run out of planned production volume[42].

Postoperative inequalities

The recovery disadvantage in the deprived, whilst accounting for preoperative health status, must depend on other mechanisms. On the patient side, we think of risk factors such as higher BMI and ASA scores, which are more prevalent in the most deprived patients[43, 44]. The decomposition analysis showed that these factors explained part of the observed health disadvantage. However, after adjusting for these factors, the majority of the deprivation effect persisted in both epidemiological and economic analyses, which has also been observed in other studies[8, 11, 15, 16].

Several clinical pathways may be responsible for this effect, and could be modifiable in order to reduce health inequality. A specific physician-related factor may be the surgical strategy. We observed that the cementation strategy (cemented vs. uncemented) and surgical approach (posterolateral vs. anterior) in THA patients was different according to deprivation status. At the time of drafting the present study, Dutch orthopedic guidelines consider both approaches and cementation strategies equal[45]. An explanation may be that the least deprived actively search for surgeons using potentially superior surgical techniques, such as the anterior approach. However, it seems unlikely that this mechanism alone is responsible for the reported difference. We cannot rule out physician-induced inequalities given the substantial surgical approach differences and the strong deprivation association. Unfortunately, data to analyze at the level of regions, hospitals or physician groups was currently unavailable which may have revealed the role of being treated in certain hospitals.

Next, hospital quality variations may occur. This was observed at Dutch obstetrics departments, where considerable inter-hospital variation in inequality was observed[46]. A possible explanation is that some hospitals adhere more strictly to protocols, which in orthopedics may include Enhance Recovery After Surgery programs[47]. Also, hospitals providing services to a relatively high proportion of deprived patients may provide more equitable care.

Finally, we found slightly larger inequalities in TKA compared to THA patients during follow-up with OKS/OHS that were not found with EQ-5D. We assume the EQ-5D and OKS/OHS have a different 'target' area: the OKS/OHS are more sensitive to capture problems in relatively healthier patients[19]. A previous study similarly found lower improvements for education level on role-physical functioning, general health (SF-36) and the Physical Component Summary scale in TKA but not in THA patients[14]. This might be because TKA patients require more rehabilitation than THA patients, which is subject to lower performance as it depends on instruction and self-care.

Besides the 'true' effects of deprivation on reported health, we should be aware of differences in response attitude. Previously, more educated, older Europeans were more likely to rate a given health state negatively[48]. This emphasizes the potential of systematically lower cross-sectional PROM scores by the more affluent and in return, attenuation of inequality effects. Response shift, which is a change in response attitude over time, could influence the change scores[49]. It is possible that a life-changing intervention, such as arthroplasty, has a different impact on values and priorities in deprived compared to affluent patients, which may result in a different self-evaluation of health at follow-up.

Concentration indices

Overall, the epidemiological measures and the concentration indices showed similar deprivation effects. The concentration indices emphasized that inequalities were present over the entire socioeconomic range and provide an indication of the magnitude of inequalities. No obvious comparator currently exists within the medical domain. Adding such an economic measure could be valuable for registry purposes, as it facilitates the comparison of inequalities in different outcomes and between providers and health systems.

Strengths and limitations

A strength of the present study is the use of national registry data, which provides generalizable results for Dutch THA and TKA patients. Furthermore, longitudinal data was employed (before surgery to 12 months follow-up). The data was analyzed using regression models in combination with concentration indices, of which the latter has not been applied to a medical population before. A limitation of this study is the exclusion of non-responders, those with only preoperative or 12-month follow-up data, and patients with missing data. However, several sensitivity analyses were conducted, leading to findings similar to the main analyses. A second limitation is that socio-economic status was measured with area-based information, which may not always accurately reflect SES at the individual level. However, area-based information generally leads to an underestimation of the associations between individual information of SES and outcomes[50, 51]. We did not have hospital data and were therefore unable to analyze their role in the observed inequalities. Lastly, the LROI does not collect data on ethnicity, which often is an important determinant of health inequalities, and might associate with observed differences according to SES.

CONCLUSION

The most deprived quintile of THA and TKA patients for the indication osteoarthritis had the worst EQ-5D index and OHS/OKS scores before and after surgery, which persisted into recovery. The observed inequalities were similar in size compared to having Charnley “C”, i.e. multiple joints affected by osteoarthritis or severely impaired quality of life due to the disease. Concentration index analysis showed that the inequality pattern exists across the full range of SES. The addition of this analysis may provide a benchmark of the magnitude of inequalities in orthopedic surgery and allows comparison of inequalities across different outcomes such as revision risk and mortality. These analytical tools, especially if hospital characteristics are added, may guide quality improvement initiatives aimed at reducing inequalities.

REFERENCES

1. Mackenbach J. Health Inequalities: Europe in Profile. 2006.
2. O'Neill J, Tabish H, Welch V, Petticrew M, Pottie K, Clarke M, et al. Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. Journal of Clinical Epidemiology 2014; 67: 56-64.
3. Marmot M. Social determinants of health inequalities. The Lancet 2005; 365: 1099-1104.
4. Phelan JC, Link BG, Tehranifar P. Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. J Health Soc Behav 2010; 51 Suppl: S28-40.
5. Spronk I, Haagsma JA, Lubetkin EI, Polinder S, Janssen MF, Bonsel GJ. Health Inequality Analysis in Europe: Exploring the Potential of the EQ-5D as Outcome. Front Public Health 2021; 9: 744405.
6. The Organisation for Economic Co-operation and Development (OECD) Health at a Glance 2019.
7. Kamaruzaman H, Kinghorn P, Oppong R. Cost-effectiveness of surgical interventions for the management of osteoarthritis: a systematic review of the literature. BMC Musculoskelet Disord 2017; 18: 183.
8. Clement ND, Jenkins PJ, MacDonald D, Nie YX, Patton JT, Breusch SJ, et al. Socioeconomic status affects the Oxford knee score and short-form 12 score following total knee replacement. Bone Joint J 2013; 95-B: 52-58.
9. Clement ND, Muzammil A, Macdonald D, Howie CR, Biant LC. Socioeconomic status affects the early outcome of total hip replacement. J Bone Joint Surg Br 2011; 93: 464-469.
10. Dowsey MM, Nikpour M, Choong PF. Outcomes following large joint arthroplasty: does socio-economic status matter? BMC Musculoskelet Disord 2014; 15: 148.
11. Gonzalez AI, Nguyen U-SDT, Franklin P, Barea C, Hannouche D, Lübbeke A. 1-year trajectories of patients undergoing primary total hip arthroplasty: Patient reported outcomes and resource needs according to education level. BMC Musculoskeletal Disorders 2022; 23: 84.
12. Goodman SM, Mandl LA, Parks ML, Zhang M, McHugh KR, Lee YY, et al. Disparities in TKA Outcomes: Census Tract Data Show Interactions Between Race and Poverty. Clin Orthop Relat Res 2016; 474: 1986-1995.
13. Jenkins PJ, Perry PR, Yew Ng C, Ballantyne JA. Deprivation influences the functional outcome from total hip arthroplasty. Surgeon 2009; 7: 351-356.
14. Keurentjes JC, Blane D, Bartley M, Keurentjes JJ, Fiocco M, Nelissen RG. Socio-economic position has no effect on improvement in health-related quality of life and patient satisfaction in total hip and knee replacement: a cohort study. PLoS One 2013; 8: e56785.
15. Neuburger J, Hutchings A, Allwood D, Black N, van der Meulen JH. Sociodemographic differences in the severity and duration of disease amongst patients undergoing hip or knee replacement surgery. J Public Health (Oxf) 2012; 34: 421-429.
16. Neuburger J, Hutchings A, Black N, van der Meulen JH. Socioeconomic differences in patient-reported outcomes after a hip or knee replacement in the English National Health Service. Journal of Public Health 2012; 35: 115-124.
17. Mackenbach JP, Kunst AE. Measuring the magnitude of socio-economic inequalities in health: an overview of available measures illustrated with two examples from Europe. Soc Sci Med 1997; 44: 757-771.
18. Wagstaff A, Paci P, van Doorslaer E. On the measurement of inequalities in health. Social Science & Medicine 1991; 33: 545-557.

19. Szende A JB, Cabases J, editors. Self-Reported Population Health: An International Perspective based on EQ-5D. 2014.
20. van Steenbergen LN, Denissen GAW, Spooren A, van Rooden SM, van Oosterhout FJ, Morrenhof JW, et al. More than 95% completeness of reported procedures in the population-based Dutch Arthroplasty Register. Acta orthopaedica 2015; 86: 498-505.
21. Dutch Orthopaedic National Quality Registry (LROI) Annual report 2021. Available at: www.lroi-rapportage.nl.
22. Charnley J. The long-term results of low-friction arthroplasty of the hip performed as a primary intervention. J Bone Joint Surg Br 1972; 54: 61-76.
23. Dripps RD. New classification of physical status. Anesthesiology 1963; 24: 111.
24. Dutch Institute of Social Research (Sociaal Cultureel Planbureau (SCP)) Socioenomic status score per postal code area, accessible by request, retrieved October 20th 2020, https://bronnen.zorggegevens.nl/Bron?naam=Sociaal-Economische-Status-per-postcodegebied.
25. World Health Organisation (WHO) Handbook on health inequality monitoring: with a special focus on low- and middle-income countries. 2013. Available at: https://apps.who.int/iris/bitstream/handle/10665/85345/9789241548632_eng.pdf?sequence=1.
26. Lamers LM, McDonnell J, Stalmeier PF, Krabbe PF, Busschbach JJ. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Econ 2006; 15: 1121-1132.
27. Peters RM, van Steenbergen LN, Stewart RE, Stevens M, Rijk PC, Bulstra SK, et al. Which patients improve most after total hip arthroplasty? Influence of patient characteristics on patient-reported outcome measures of 22,357 total hip arthroplasties in the Dutch Arthroplasty Register. Hip Int 2021; 31: 593-602.
28. Tolk JJ, Waarsing JEH, Janssen RPA, van Steenbergen LN, Bierma-Zeinstra SMA, Reijman M. Development of Preoperative Prediction Models for Pain and Functional Outcome After Total Knee Arthroplasty Using The Dutch Arthroplasty Register Data. J Arthroplasty 2020; 35: 690-698 e692.
29. Feldman CH, Dong Y, Katz JN, Donnell-Fink LA, Losina E. Association between socioeconomic status and pain, function and pain catastrophizing at presentation for total knee arthroplasty. BMC Musculoskelet Disord 2015; 16: 18.
30. Hoelen T-CA, Schotanus M, van Kuijk S, Bastiaenen C, Boonen B, Most J. The relation between socioeconomic status and patient symptoms before and one year after lower extremity arthroplasty. Journal of Orthopaedics 2023; 39: 11-17.
31. Kakwani N, Wagstaff A, van Doorslaer E. Socioeconomic inequalities in health: Measurement, computation, and statistical inference. Journal of Econometrics 1997; 77: 87-103.
32. Erreygers G. Correcting the concentration index. J Health Econ 2009; 28: 504-515.
33. Wagstaff A. The bounds of the concentration index when the variable of interest is binary, with an application to immunization inequality. Health Econ 2005; 14: 429-432.
34. Wagstaff A, van Doorslaer E, Watanabe N. On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam. Journal of Econometrics 2003; 112: 207-223.
35. Gademan MGJ, Hofstede SN, Vliet Vlieland TPM, Nelissen RGHH, Marang-van de Mheen PJ. Indication criteria for total hip or knee arthroplasty in osteoarthritis: a state-of-the-science overview. BMC musculoskeletal disorders 2016; 17: 463-463.
36. Dreinhöfer KE, Dieppe P, Stürmer T, Gröber-Grätz D, Flören M, Günther KP, et al. Indications for total hip replacement: comparison of assessments of orthopaedic surgeons and referring physicians. Ann Rheum Dis 2006; 65: 1346-1350.

37. Jinks C, Vohora K, Young J, Handy J, Porcheret M, Jordan KP. Inequalities in primary care management of knee pain and disability in older adults: an observational cohort study. Rheumatology (Oxford) 2011; 50: 1869-1878.
38. Delpierre C, Lauwers-Cances V, Datta GD, Lang T, Berkman L. Using self-rated health for analysing social inequalities in health: a risk for underestimating the gap between socioeconomic groups? J Epidemiol Community Health 2009; 63: 426-432.
39. Edwards NM, Varnum C, Overgaard S, Pedersen AB. The impact of socioeconomic status on the utilization of total hip arthroplasty during 1995-2017: 104,055 THA cases and 520,275 population controls from national databases in Denmark. Acta Orthop 2021; 92: 29-35.
40. Judge A, Welton NJ, Sandhu J, Ben-Shlomo Y. Equity in access to total joint replacement of the hip and knee in England: cross sectional study. Bmj 2010; 341: c4092.
41. Lopes FV, Riumallo Herl CJ, Mackenbach JP, Van Ourti T. Patient cost-sharing, mental health care and inequalities: A population-based natural experiment at the transition to adulthood. Social Science & Medicine 2022; 296: 114741.
42. Oliver A, Mossialos E. Equity of access to health care: outlining the foundations for action. J Epidemiol Community Health 2004; 58: 655-658.
43. Boyce L, Prasad A, Barrett M, Dawson-Bowling S, Millington S, Hanna SA, et al. The outcomes of total knee arthroplasty in morbidly obese patients: a systematic review of the literature. Arch Orthop Trauma Surg 2019; 139: 553-560.
44. Hooper GJ, Rothwell AG, Hooper NM, Frampton C. The relationship between the American Society Of Anesthesiologists physical rating and outcome following total hip and knee arthroplasty: an analysis of theNew Zealand Joint Registry. J Bone Joint Surg Am 2012; 94: 1065-1070.
45. Dutch Orthopedic Society (NOV) 2022. Available at: https://www.orthopeden.org/kwaliteit/kwaliteitsbeleid/richtlijnen. 2022.
46. Poeran J, Borsboom GJ, de Graaf JP, Birnie E, Steegers EA, Bonsel GJ. Population attributable risks of patient, child and organizational risk factors for perinatal mortality in hospital births. Matern Child Health J 2015; 19: 764-775.
47. Frassanito L, Vergari A, Nestorini R, Cerulli G, Placella G, Pace V, et al. Enhanced recovery after surgery (ERAS) in hip and knee replacement surgery: description of a multidisciplinary program to improve management of the patients undergoing major orthopedic surgery. Musculoskelet Surg 2020; 104: 87-92.
48. d’Uva TB, O’Donnell O, van Doorslaer E. Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans. Int J Epidemiol 2008; 37: 1375-1383.
49. Razmjou H, Yee A, Ford M, Finkelstein JA. Response shift in outcome assessment in patients undergoing total knee arthroplasty. J Bone Joint Surg Am 2006; 88: 2590-2595.
50. Mesterton J, Willers C, Dahlström T, Rolfson O. Comparison of individual and neighbourhood socioeconomic status in case mix adjustment of hospital performance in primary total hip replacement in Sweden: a register-based study. BMC Health Serv Res 2020; 20: 645.
51. Subramanian SV, Chen JT, Rehkopf DH, Waterman PD, Krieger N. Comparing individual- and area-based socioeconomic measures for the surveillance of health disparities: A multilevel analysis of Massachusetts births, 1989-1991. Am J Epidemiol 2006; 164: 823-834.

DECLARATIONS

Acknowledgements: None.

CRediT authorship contribution statement: Study Conception and design: JB, MR, JV, GB. Analysis and interpretation of the data: JB, MR, JV, MJ, GB. Drafting of the article: JB, GB. Critical revision of the article for important intellectual content: JB, MR, JV, LS, MJ, GB. Final approval of the article: JB, MR, JV, LS, MJ, GB. Statistical expertize: JB, MJ, and GB. Obtaining funding: JB, MR, JV, GB. Collection and assembly of data: JB. Data acquisition: JB, MR, JV, LS, MJ, and GB; JB and GB take responsibility for the integrity of the work as a whole.

Funding: This study was in part funded by the EuroQol Research Foundation. The funder did not play any role in the study design, collection, analysis and interpretation of data, in the writing of the manuscript, or in the decision to submit the manuscript for publication. The views expressed by the authors do not necessarily reflect the views of the EuroQol Research Foundation.

3

Declaration of competing interest: JB, MR, JV, and MJ declare they had financial support from the EuroQol Research Foundation for the submitted work. LV and GB declare to have no conflicts of interest.

Data availability: By the terms of data usage from the Dutch Arthroplasty Registry, no additional data can be made available by the authors.

Supplementary data: Supplementary data associated with this article can be found in the online version at doi: 10.1016/j.joca.2023.07.004.

Chapter 4

Socioeconomic Inequalities in Patient-Reported Outcome Measures among Total Hip and Knee Arthroplasty Patients: A Comprehensive Analysis of Instruments and Domains

Authors

Joshua M. Bonsel[1], MD

Max Reijman[1], MSc, PhD

Erin M. Macri[1], MSc, PhD

Jan A.N. Verhaar[1], MD, PhD, emeritus professor

Liza N. van Steenbergen[2], MSc, PhD

Gouke J. Bonsel[3], MD, PhD, emeritus professor

Affiliations

[1] Department of Orthopaedics and Sports Medicine, Erasmus MC, Rotterdam, The Netherlands

[2] Dutch Arthroplasty Register (Landelijke Registratie Orthopedische Interventies), ’s Hertogenbosch, The Netherlands

[3] EuroQol Research Foundation, Rotterdam, The Netherlands

Status

Submitted.

ABSTRACT

Background

Prior to total hip and knee arthroplasty (THA/TKA), patients with low socioeconomic status (SES) report worse Patient-Reported Outcome Measures (PROMs), persisting postoperatively. This study explores which self-reported PROMs and their specific domains are most involved.

Methods

We obtained data from the Dutch Arthroplasty Registry (2014-2022), including over 100,000 THA/TKA patients with complete preoperative and 12-month follow-up PROMs. The EQ-5D-3L and EQ-5D-5L, EQ Visual Analogue Scale (VAS), Oxford Hip and Oxford Knee Score (OHS/OKS), and Numerical Rating Scales (NRS) for pain and satisfaction (TKA only, at 12-month follow-up) were obtained. PROMs were transformed to a 0-100 scale for direct comparison. We categorized a SES-indicator based on neighborhood income, unemployment rate, and education level into quintiles. Through linear regression we contrasted the most vs. least deprived SES quintiles, adjusting for patient and surgical characteristics. The contribution (percentage) of each domain to the overall inequalities was estimated for the EQ-5D's and the OHS/OKS.

Results

Preoperatively, the most vs. the least deprived TKA patients had lower EQ-5D-3L: -2.1 [95% confidence interval -2.6, -1.6]. At 12-month follow-up, differences were smaller: EQ-5D-3L 1.3 [-1.9, -0.7]). Analogous differences were present in OKS scores (preoperatively: -4.3 [-5.3, -3.4]; 12-month: -1.8 [-2.5, -1.2]). The differences in EQ VAS scores were smaller (preoperatively: -0.8 [-1.5, -0.1]; 12-month: -0.5 [-1.2, 0.1]). The differences in NRS pain (in rest) were comparable to those in EQ-5D-3L and OKS (preoperatively: -4.5 [-5.4, -3.5]; 12-month: -2.7 [-3.5, -1.9]), while NRS satisfaction showed no inequality at 12 months. For EQ-5D-3L, the domain usual activities accounted for up to 46% of outcome inequalities, while anxiety/depression played a limited role (up to 17%). For OHS/OKS, functioning contributed most in THA (up to 61%) and pain contributed most in TKA (up to 68%). Findings were similar in THA patients, and for TKA and THA patients measured using the EQ-5D-5L.

Conclusions

Deprived THA/TKA patients have poorer pre- and postoperative health, which was primarily related to worse functioning and pain, providing potential opportunities for amendment. These inequalities did not translate into worse overall health (EQ VAS) or into higher dissatisfaction among deprived patients. Future research should investigate whether the latter two questionnaires reflect true inequality-related effects or whether they were disproportionately affected by reporting heterogeneity, e.g., due to differences in the interpretation of wording.

BACKGROUND

Socioeconomic status (SES), a composite measure including education, income and occupation, is a significant driver of health inequalities in most medical fields[1, 2]. Total hip and knee arthroplasty (THA/TKA) are two commonly performed procedures that typically alleviate pain and symptoms in patients suffering from end-stage osteoarthritis of the hip and knee[3]. Within THA/TKA, SES-related health inequalities have convincingly been demonstrated, with deprived patients experiencing poorer clinical outcomes such as increased mortality and complication rates[4]. The health disadvantage associated with low SES may arise from several mechanisms, including increased risk exposure, less healthy occupational environment, and reduced access to medical care.

Patient-Reported Outcome Measures (PROMs) provide a valuable alternative to above-mentioned clinical metrics for measuring health inequalities. PROMs are standardized, validated, mostly self-reported questionnaires that assess patients' generic and condition-specific health status, and are widely used in orthopedics, including THA/TKA. For example, previous studies have found that low SES associates with poorer preoperative[5-10] and postoperative PROMs[6, 8-15], which have illuminated that inequalities occur in the starting position before surgery and during recovery after THA/TKA. The cited studies employed various PROMs to explore health inequalities by SES, and reported varying strengths of associations. Multiple instrument-specific factors and design features may be responsible, as may be differing populations, hampering our understanding of how inequalities manifest across these PROM instruments.

First, the instrument construct, including the health domains covered, matters[16]. For example, the Oxford Hip and Knee Score (OHS/OKS) instruments cover similar constructs, with emphasis on joint-related function and pain. In comparison, the EQ-5D covers five generic quality of life domains: mobility, self-care, usual activities, pain/discomfort (not specified any further), and anxiety/depression. As it turns out, the OHS and EQ-5D provide similar results in THA patients, with the exception of anxiety/depression, a domain with independent relevance[17]. In our context, this construct difference might impact measured health inequality, if deprived patients disproportionately experience mental issues following surgery.

Other differences in instrument characteristics such as the scale used may also matter, which usually affects reliability and responsiveness[16]. Generally, multi-item measures such as the OHS/OKS and EQ-5D psychometrically perform better compared to single-item measures such as the EQ Visual Analogue Scale (VAS) and Numerical Rating Scale (NRS) for pain in THA and TKA patients[18-21]. Consequently, single-item measures such as the EQ VAS and Numerical Rating Scale (NRS) for pain may show less SES-related inequalities compared to the multi-item OHS/OKS and EQ-5D.

A third factor potentially affecting the size of SES-related inequalities is reporting heterogeneity, which refers to the phenomenon where a respondent systematically interprets questions or response options differently. A simple example is the avoidance of extreme answers, more often observed in the aged[22]. In our context, it is known that poor individuals express greater satisfaction with health care than their more affluent counterparts, even when they show inferior outcomes measured by other metrics[23]. Reporting heterogeneity may therefore induce underestimation of 'true' health inequalities. While all PROMs are susceptible to this phenomenon, some may be more susceptible than others, such as the EQ VAS and satisfaction questions in general. We believe this to be the case due to the phrasing used in these instruments, i.e., "How would you rate your health today?" and "How satisfied are you with the treatment received?", respectively. Such phrasing may invite patients to respond using their own reference of the quality of care received, more so compared to, e.g., the OHS/OKS or EQ-5D. While accounting for reporting heterogeneity is complex, it should be considered as an explanans when comparing PROMs, for example in assessments of health inequality magnitude[24].

This study uses Dutch Arthroplasty Registry (LROI) data of THA and TKA patients to explore SES-related inequalities using a range of PROMs, and to relate differences in size to their domain structure and type of scale. By design, differences of methodology or patient population do not play a role. Our study departs from the assumption that, among all types of measures, health measured using the EQ-5D and OHS/OKS best reflect 'true' health in the studied population. We propose the following hypotheses that we believe may explain different PROMs scores in different categories of SES:

Hypothesis 1: SES-related inequalities will be more pronounced when measured by the EQ-5D compared to the OHS/OKS, because the EQ-5D includes the domain of anxiety/depression.

Hypothesis 2: SES-related inequalities will be more pronounced in multi-item scales (e.g., EQ-5D or OHS/OKS) compared to single-item scales (e.g., EQ VAS, NRS pain and satisfaction), due to the poorer instrument characteristics of single-item scores in general.

Hypothesis 3: SES-related inequalities will be most pronounced when measured with the EQ-5D and OHS/OKS compared to the EQ VAS and NRS satisfaction, which may suggest the presence of reporting heterogeneity.

METHODS

Data source

This observational cohort study used anonymized, prospectively collected clinical registry data (www.lroi.nl). This registry is under responsibility of the Netherlands Orthopedic Association (NOV). Patients undergoing surgery may opt out for sharing their data with the LROI. Studies using LROI data are subject to technical and ethical judgment by the registry holder, and Dutch Law does not require additional institutional ethical judgment by a Medical Ethical Review Board. Our study followed the STROBE guideline for observational studies and when appropriate the COSMIN Study Design checklist[25, 26].

The LROI captures over 95% of THA and TKA surgeries performed in the Netherlands since 2009[27]. Variables encompass patient and surgical characteristics, and outcomes. Since 2014, a set of internationally validated generic and disease-specific PROMs has been included in the LROI, supported by the NOV. PROMs are collected at three time points: preoperatively (maximal 6 months before surgery), 3 months after surgery (range 2–4 months) in THA, 6 months after surgery (range 5–7 months) in TKA, and 12-months after surgery (range 11–13 months) for both procedures. Response rates since 2017 are approximately 40%, depending largely on hospital participation[28].

Inclusion criteria

We selected primary THA and TKA patients between 2014 to 2022 with complete preoperative and 12-month follow-up PROMs. Further selection was through diagnosis of osteoarthritis, which is the largest and most homogeneous group. Records of contralateral joint replacements during this period were also obtained. From 2014 until 2020, the EQ-5D-3L was used. In 2021, it was replaced by the EQ-5D-5L. Since scores of the -3L and -5L version are not directly interchangeable on the individual level (see ‘Outcomes: PROMs’ below) [29], the dataset was split accordingly. Patients who completed a preoperative EQ-5D-3L and responded to a 12-month follow-up EQ-5D-5L during the transition period were excluded.

Variables

Patient and surgical data included age, biological sex, body mass index (BMI), Charnley score[30], American Society of Anesthesiologists (ASA) score[31], previous surgery of the replaced joint, smoking status, type of hospital, prothesis fixation method, and surgical approach. The Charnley score represents the extent of osteoarthritis disease, and ranges from “A” (one joint affected) to “C” (multiple joints affected or quality of life severely impaired due to the disease). Hospital type was categorized as private, general, or university. Fixation methods were categorized as cemented, cementless, or hybrid. Surgical approach for THA was categorized as anterior, anterolateral, posterolateral, straight lateral, and other.

For TKA, surgical approach was categorized as lateral parapatellar, medial parapatellar, and mid- or sub-vastus.

Exposure: socioeconomic status

A neighborhood SES score was linked to patients using four-digit postal codes[32]. This standardized score is calculated by two government institutions (Statistics Netherlands and the Netherlands Institute for Social Research) from the mean income per household, percentage of households with low income, percentage of unemployed inhabitants, and percentage of households with low education per postal code area. The SES score is only calculated for postal code areas with a minimum of 100 inhabitants (mean 4300 inhabitants per postal code). We used the 2017 scoring because it was the mid-point of our primary cohort. Moreover, neighborhood deprivation is known to be very stable over several years[33]. Based on customary practice and guidelines, SES was divided into quintiles[34].

Outcomes: PROMs

For our research questions, we obtained data from EQ-5D-3L, EQ-5D-5L, EQ VAS, OHS/OKS, NRS for pain at rest, NRS for pain during activity, and NRS for satisfaction with the undergone procedure (the latter was only available in TKA at 12-month follow-up). All PROMs were self-reported. The EQ-5D-3L and EQ-5D-5L have 5 domains (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) on which patients report their perceived general health[35]. For each domain the -3L version has 3 response options ranging from 'extreme' to 'no complaints'. The EQ-5D-5L has 5 response options, which has been shown to increase sensitivity and reducing ceiling[29]. Typically, the scores for the 5 domains can be linked to a 'value set', which transforms the 5 domain scores into an overall 'utility' value for this health state. When the purpose of EQ-5D data is non-economic, as in this study, the EuroQol Research Foundation advises to sum the domain scores directly into a level-sum-score (LSS) [36]. The LSS of the EQ-5D-3L ranges from 5 to 15, and for the EQ-5D-5L it ranges from 5 to 25, with lower scores indicating better health. The EQ VAS rates own health on a visual analogue scale from 0 ('The worst health you can imagine') to 100 ('The best health you can imagine'). The OHS/OKS consists of 12 items on two domains, namely (physical) functioning and pain, in patients with osteoarthritis of the hip/knee[37]. In the OHS each domain is covered with 6 items, while in the OKS the function and pain domains are covered by 5 and 7 items, respectively. The OHS/OKS scores range from 0 to 48, with 48 indicating no disability. The NRS pain outcomes are rated on a scale of 0 to 10, with 10 reflecting severe pain. For the NRS satisfaction, however, a score of 10 reflects the highest degree of satisfaction with the result. To facilitate direct comparison, all PROM scores were transformed to a 0 to 100-scale, with 100 representing the best attainable outcome.

Missing data

Missing Charnley scores were conservatively estimated as 'A', because only patients with osteoarthritis were selected. Ages under 10 years or over 105 years, and BMI values below 10 or above 70, were recoded as 'missing' in accordance with LROI guidelines[38]. As missing data in one or more variables was present in only 4% of patients with EQ-5D data, a complete case analysis was conducted. Among patients with EQ-5D data, other PROMs were missing in up to 9% of patients per cohort; therefore, a complete case analysis was also conducted for these outcomes.

Statistical analysis

Analyses were performed for THA and TKA separately, and for the primary (EQ-5D-3L) and secondary (EQ-5D-5L) cohorts.

For continuous variables we calculated medians with interquartile ranges (IQR), and for categorical variables percentages. To quantify 'ceiling' we calculated the proportion of patients reporting best health, i.e., a score of 100, for each total PROM score. Based on existing reporting practice, a percentage greater than 15% was considered indicative of a ceiling. The ANOVA (t-test) and chi-squared tests were used to compare the most and least deprived quintile when appropriate.

First, we explored the size of SES-related inequalities across different PROMs. Linear regression (LR) was used for this purpose, testing and quantifying the association between SES quintiles and total PROMs scores, separately for the preoperative and 12-month follow-up measurement. The least deprived SES quintile was used as reference; we expected negative regression coefficients for less affluent SES categories. Models were adjusted for sex, age, BMI, ASA score, Charnley score, and type of hospital, which were considered potential confounders[10]. The regression models for 12-month follow-up outcomes were also adjusted for the preoperative score of the respective PROM[39]. To facilitate meaningful interpretation of coefficients, the preoperative score was entered as a categorical variable, grouped into tertiles (low, medium, high) of approximately equal size using the 'santoku' package[40]. Cut-off values are included in Supplemental File, Table 1.

Previously, we found that the relation between SES and postoperative PROMs differed according to preoperative score[10]. Therefore, we also stratified the models according to the tertiles of the preoperative score of the respective PROM, rather than adjusting for the preoperative score. No preoperative measurement is available for the NRS satisfaction, which should be regarded as a concept both covering current health and health change. As satisfaction following arthroplasty is known to be associated with preoperative health we may expect stronger associations between SES and satisfaction for worse preoperative health[41]. Thus, NRS satisfaction was also stratified according to the preoperative EQ-5D LSS.

A three-step procedure was conducted to estimate the contribution of separate domains of the PROMs (EQ-5D and OHS/OKS) to the health inequalities. First, we removed the domain for which the contribution was to be calculated from the total score of that PROM[42]. The model was re-run with this 'total minus one' score. Second, the SES coefficients of the total and 'total minus one' model were compared, and the difference in coefficients was calculated as a percentage. Third, the percentages over the SES quintiles were averaged, resulting in a percentage-wise expression of how much a domain contributed to the overall association between SES and the total score. We repeated this procedure for each domain of the PROM. For the OKS, if each domain contributed equally, the functional domain would account for around 42% (5/12 items) and the pain domain would account for 58% (7/12 items) of the inequality found with the total OKS score. The OHS and EQ-5D have a balanced number of items across their domains, so equal contributions would result in percentages of equal size.

Because our dataset contained contralateral procedures, we evaluated whether we needed to account for nesting of outcomes at the patient level (i.e., using hierarchical modeling). We compared empty (null) models with and without random intercepts for the patient. As model fit did not improve, we reported the results of regular regression models.

We compared coefficients between PROMs, presenting them with 95% confidence intervals [95% CIs] to determine the degree of certainty in differences in inequalities observed. We accepted/refuted our hypotheses based on whether the difference between coefficients exceeded the uncertainty interval (i.e., showed no overlap) in the expected direction, following standard practice for evaluating hypotheses in the validation of PROM instruments. In our previous publication, no evidence of bias was observed in the inequality patterns due to non-response; hence, this was not assessed again[10]. A p-value of <0.05 was considered statistically significant. All analyses were performed in R version 4.1.2 [43].

RESULTS

After removing incomplete cases, the primary (EQ-5D-3L) cohorts had 45,822 THA and 32,734 TKA procedures (Figure 1). The secondary (EQ-5D-5L) cohorts had 14,388 THA and 9,191 TKA procedures.

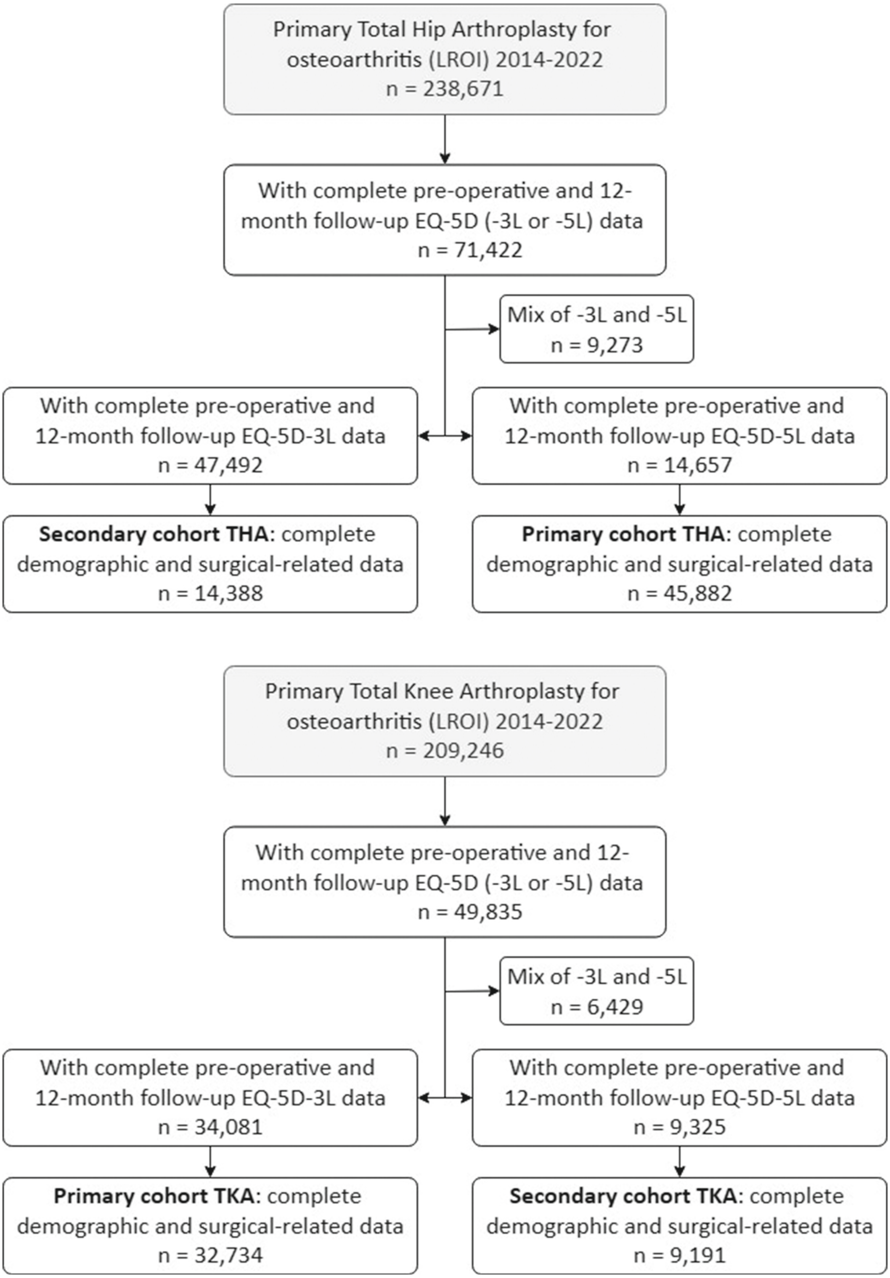

Figure 1: Flowchart depicting case-selection (THA/TKA)

Primary (EQ-5D-3L) cohorts

Descriptive analysis

THA and TKA patients were about 69 years old and were more likely to be female. Inequalities by SES followed expected patterns, e.g., those who were more deprived tended to be less healthy measured with ASA score (Supplemental File, Table 2–3).

Preoperatively, median values for the OHS/OKS and NRS pain scores were lower than those for the EQ-5D-3L and EQ VAS scores, and were closer to the scales' midrange score of 50 (i.e., the center of the 0-100 scale) for all SES quintiles. OHS/OKS and NRS pain scores also showed greater distribution (larger IQRs). At 12-month follow-up, median values and distributions for all SES quintiles were similar across all PROMs.

Preoperatively, there was no ceiling in any of the PROMs across deprivation quintiles. At 12-month follow-up, all PROMs exhibited ceiling to some extent. The highest was observed for the NRS pain outcomes, while the lowest was observed for EQ VAS. The most deprived patients responded best health about 5% less often for all PROMs.

Linear regression models

The full adjusted models can be found in Supplemental File, Table 4–7.

Hypothesis 1: SES-related inequalities will be more pronounced when measured using the EQ-5D compared to the OHS/OKS

Preoperatively, the most deprived patients had lower EQ-5D-3L LSS (THA: -1.6 [95% CI -2.0, -1.2]; TKA: -2.1 [95% CI -2.6, -1.6]) scores than the least deprived group (Table 1). The differences on OHS/OKS compared to EQ-5D-3L scores was similar in THA, but larger in TKA patients with non-overlapping CIs (THA: -2.4 [95% CI -3.0, -1.9]); TKA: -4.3 [95% CI -5.3, -3.4]). At 12-month follow-up, differences were smaller (Table 2). The most deprived patients reported lower EQ-5D-3L LSS (THA: -1.1 [95% CI -1.6, -0.7]; TKA: -1.3 [95% CI -1.9, -0.7]) and OHS/OKS (THA: -0.7 [95% CI -1.1, -0.2]; TKA: -1.8 [95% CI -2.5, -1.2]) scores. The results indicate that this hypothesis was <u>not met</u>.

Table 1: Association between socioeconomic status and preoperative health status

	EQ-5D-3L LSS	EQ VAS	OHS/OKS	NRS Pain in rest	NRS Pain during activity
Variables	**Beta [95% CI]**	**Beta [95% CI]**	**Beta [95% CI]**	**Beta [95% CI]**	**Beta [95% CI]**
Total Hip Arthroplasty					
Intercept	61.5 [60.6, 62.4]	67.4 [66.2, 68.7]	50.5 [50.0, 50.9]	45.1 [43.4, 46.8]	25.5 [24.2, 26.9]
SES					
Q1 [least depr.]	ref	ref	ref	ref	ref
Q2	-0.5 [-1.0, -0.1]	-0.8 [-1.4, -0.3]	-0.7 [-1.2, -0.1]	-1.0 [-1.8, -0.3]	-0.7 [-1.3, -0.1]
Q3	-1.1 [-1.5, -0.7]	-0.8 [-1.3, -0.2]	-1.4 [-1.9, -0.8]	-1.7 [-2.5, -1.0]	-1.2 [-1.8, -0.6]
Q4	-1.4 [-1.8, -1.0]	-0.6 [-1.1, 0.0]	-1.8 [-2.4, -1.3]	-2.4 [-3.1, -1.6]	-2.1 [-2.7, -1.5]
Q5 [most depr.]	-1.6 [-2.0, -1.2]	-1.0 [-1.6, -0.4]	-2.4 [-3.0, -1.9]	-3.4 [-4.2, -2.7]	-1.5 [-2.2, -0.9]
R-squared	0.05	0.04	0.08	0.03	0.03
Total Knee Arthroplasty					
Intercept	65.4 [64.1, 66.8]	71.4 [69.4, 73.4]	51.2 [49.0, 53.3]	45.7 [43.0, 48.5]	26.0 [23.9, 28.2]
SES					
Q1 [least depr.]	ref	ref	ref	ref	ref
Q2	-0.5 [-1.0, 0.0]	0.6 [-0.1, 1.3]	-1.4 [-2.4, -0.5]	-0.8 [-1.8, 0.2]	-0.7 [-1.4, 0.1]
Q3	-1.2 [-1.7, -0.7]	0.3 [-0.4, 0.9]	-2.0 [-2.9, -1.1]	-2.5 [-3.5, -1.6]	-1.8 [-2.5, -1.0]
Q4	-1.5 [-2.0, -1.1]	0.2 [-0.5, 0.8]	-1.9 [-2.8, -1.0]	-3.4 [-4.3, -2.5]	-2.0 [-2.7, -1.3]
Q5 [most depr.]	-2.1 [-2.6, -1.6]	-0.8 [-1.5, -0.1]	-4.3 [-5.3, -3.4]	-4.5 [-5.4, -3.5]	-1.9 [-2.6, -1.2]
R-squared	0.05	0.05	0.09	0.03	0.03

The table depicts data from the EQ-5D-3L cohort. Multivariable linear regression models were used. The LSS, OHS, OKS and NRS outcomes were transformed to 0-100 where 100 is the best score possible. The regression models were adjusted for sex, age, BMI, ASA score, Charnley score, and type of hospital.
Abbreviations: LSS = level sum score; VAS = Visual Analogue Scale; OHS = Oxford Hip Score; OKS = Oxford Knee Score; BMI = Body Mass Index; ASA = American Society of Anesthesiology score; NRS = Numerical Rating Scale

Table 2: Association between socioeconomic status and 12-month follow-up health status

	EQ-5D-3L LSS	EQ VAS	OHS/OKS	NRS Pain in rest	NRS Pain during activity
Variables	**Beta [95% CI]**	**Beta [95% CI]**	**Beta [95% CI]**	**Beta [95% CI]**	**Beta [95% CI]**
Total Hip Arthroplasty					
Intercept	87.4 [86.3, 88.4]	79.9 [78.7, 81.1]	89.7 [88.7, 90.8]	89.5 [88.3, 90.7]	83.3 [81.8, 84.8]
SES					
Q1 [least depr.]	ref	ref	ref	ref	ref
Q2	0.2 [-0.3, 0.6]	0.4 [-0.2, 0.9]	0.3 [-0.2, 0.7]	0.0 [-0.6, 0.6]	-0.2 [-0.8, 0.5]
Q3	-0.5 [-0.9, 0.0]	-0.1 [-0.7, 0.4]	0.1 [-0.4, 0.6]	-0.5 [-1.1, 0.0]	-0.8 [-1.5, -0.2]
Q4	-0.5 [-0.9, 0.0]	-0.3 [-0.8, 0.2]	0.0 [-0.5, 0.4]	-0.8 [-1.4, -0.3]	-1.1 [-1.7, -0.4]
Q5 [most depr.]	-1.1 [-1.6, -0.7]	-0.8 [-1.4, -0.3]	-0.7 [-1.2, -0.2]	-1.8 [-2.4, -1.3]	-2.0 [-2.7, -1.4]
R-squared	0.10	0.10	0.09	0.03	0.02
Total Knee Arthroplasty					
Intercept	81.5 [79.8, 83.1]	74.9 [73.0, 76.7]	79.3 [77.5, 81.1]	83.3 [80.9, 85.7]	73.0 [70.2, 75.8]
SES					
Q1 [least depr.]					
Q2	0.0 [-0.6, 0.6]	0.1 [-0.6, 0.8]	-0.3 [-1.0, 0.3]	-0.9 [-1.8, -0.1]	-0.9 [-1.9, 0.1]
Q3	-0.4 [-0.9, 0.2]	-0.3 [-1.0, 0.3]	-0.3 [-0.9, 0.3]	-0.8 [-1.6, 0.0]	-0.5 [-1.5, 0.5]
Q4	-0.7 [-1.2, -0.1]	-0.6 [-1.2, 0.0]	-0.7 [-1.3, -0.1]	-1.4 [-2.2, -0.7]	-1.6 [-2.5, -0.7]
Q5 [most depr.]	-1.3 [-1.9, -0.7]	-0.5 [-1.2, 0.1]	-1.8 [-2.5, -1.2]	-2.7 [-3.5, -1.9]	-2.8 [-3.8, -1.9]
R-squared	0.12	0.11	0.12	0.05	0.03

The table depicts data from the EQ-5D-3L cohort. Multivariable linear regression models were used. The LSS, OHS, OKS, and NRS outcomes were transformed to 0-100 where 100 is the best score possible. All models were adjusted for sex, age, BMI, ASA score, Charnley score, and type of hospital, and the preoperative score of the respective PROM. Abbreviations: LSS = level sum score; VAS = Visual Analogue Scale; OHS = Oxford Hip Score; OKS = Oxford Knee Score; BMI = Body Mass Index; ASA = American Society of Anesthesiology score; NRS = Numerical Rating Scale

Hypothesis 2: SES-related inequalities will be less pronounced in single-item compared to multi-item scales

Preoperatively, the difference between the most vs. least deprived quintile on the single-item EQ VAS (THA: -1.0 [95% CI -1.6, -0.4]; TKA: -0.8 [95% CI -1.5, -0.1]) was about half the size of the multi-item EQ-5D-3L LSS, and even smaller compared to the multi-item OHS/OKS. For most of these comparisons, CIs did not overlap. The differences on the single-item NRS pain scores (NRS pain in rest; THA: -3.4 [95% CI -4.2, -2.7]; TKA: -4.5 [95% CI -5.4, -3.5]) were similar to the EQ-5D-3L and OHS/OKS scores. At 12-month follow-up these patterns persisted. Moreover, at 12-month follow-up, the most vs. least deprived TKA patients did not report different satisfaction levels (-0.0 [95% CI -0.9, 0.8]) (Table 3). The results indicate that this hypothesis was not met.

Table 3: Association between socioeconomic status and 12-month follow-up NRS satisfaction in TKA patients

	NRS Satisfaction*
Variable	**Beta [95% CI]**
Intercept	76.5 [74.1, 79.0]
SES	
Q1 [least depr.]	ref
Q2	0.2 [-0.7, 1.1]
Q3	0.7 [-0.1, 1.6]
Q4	0.3 [-0.6, 1.1]
Q5 [most depr.]	0.0 [-0.9, 0.8]
R-squared	0.01

The table depicts data from the EQ-5D-3L cohort. Multivariable linear regression models were used, with adjustment for sex, age, BMI, ASA score, Charnley score, and type of hospital. The NRS for satisfaction was transformed to 0-100 where 100 is the best score possible.
Abbreviations: NRS = Numerical Rating Scale

Hypothesis 3: SES-inequalities will be least pronounced measured with EQ VAS and NRS satisfaction compared to EQ-5D and OHS/OKS

The results described under hypothesis 2 indicate that this hypothesis was met.

Impact of SES on 12-month follow-up PROMs, stratified by preoperative health

The point-estimates of inequalities on all PROMs between most and least deprived groups tended to be larger for a low preoperative score (Figure 2A and 2B). This pattern seemed less pronounced for EQ VAS, particularly in TKA patients. Points-estimates of SES-related differences in NRS satisfaction in TKA patients, after stratifying according to preoperative EQ-5D-3L LSS, behaved similar to EQ VAS. Overall, these findings illustrate that the patterns described under the hypotheses seem more evident in a low vs. high preoperative score.

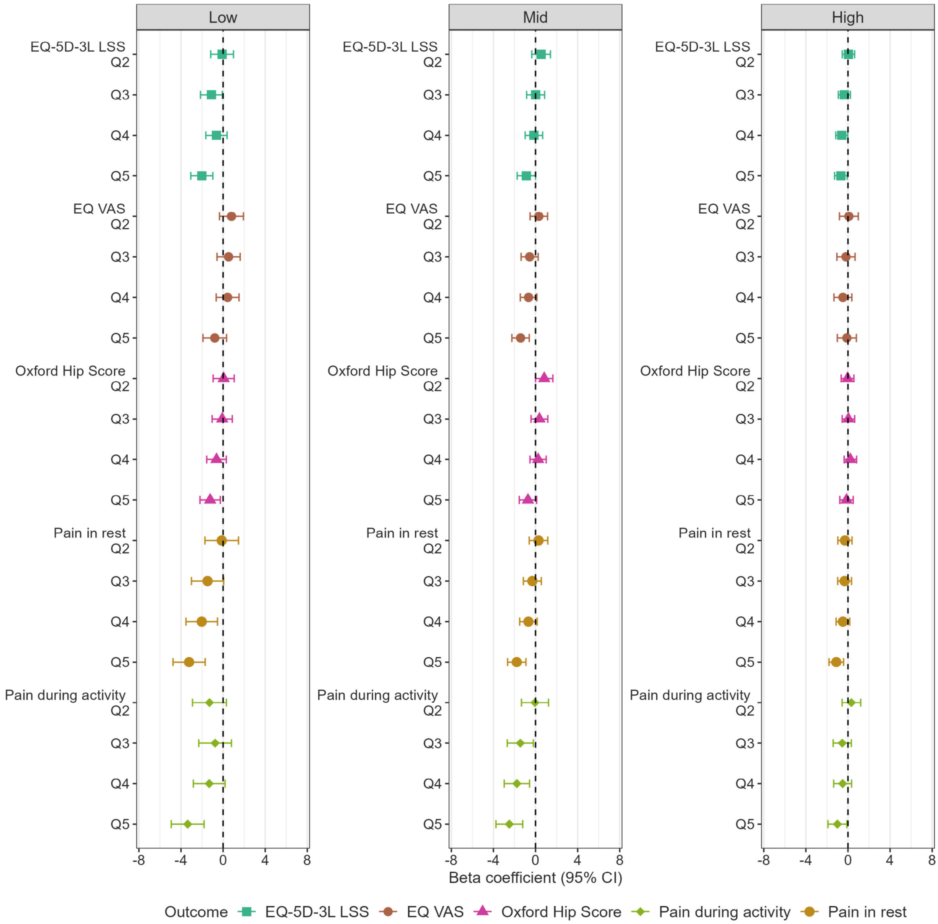

Figure 2A: Impact of SES on 12-month follow-up outcome, stratified by preoperative outcome (THA)

Abbreviations: LSS = level sum score; VAS = Visual Analogue Scale

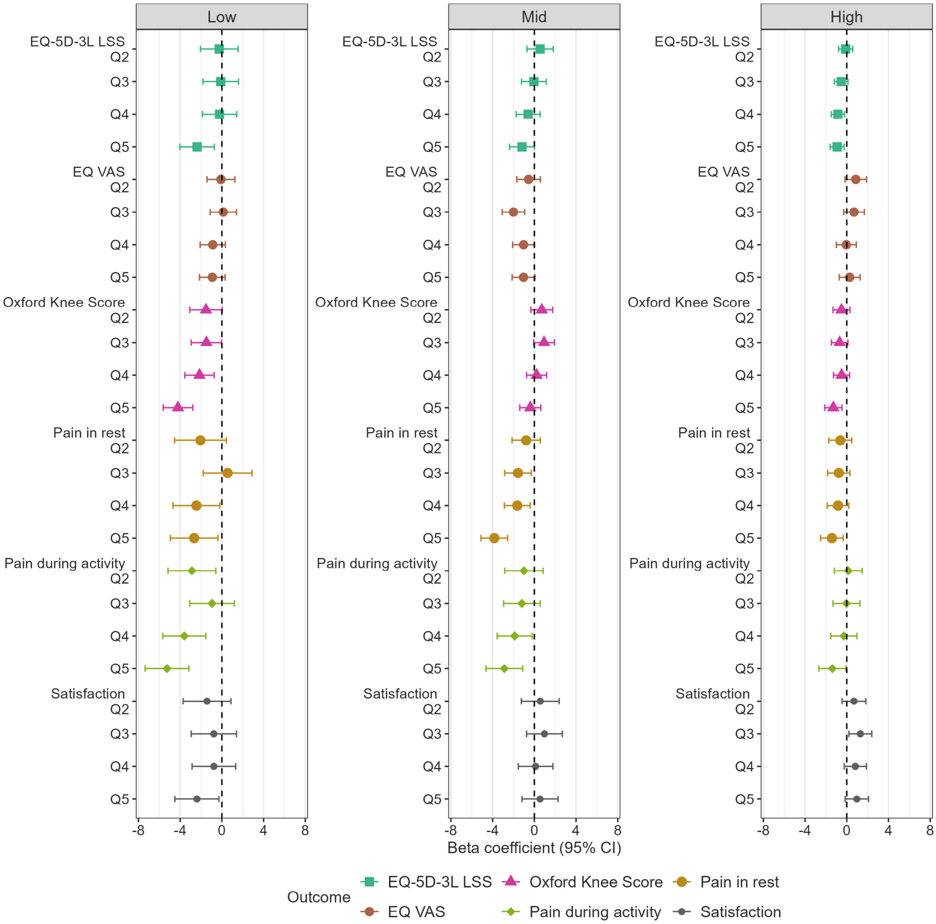

Figure 2B: Impact of SES on 12-month follow-up outcome, stratified by preoperative outcome (TKA)

Domain contribution

For the EQ-5D-3L, at each measurement point and in both THA and TKA patients, the 'usual activities' domain played a larger role in the SES-related PROM score differences (explained 25-46%) (Table 4). 'Mobility' had a limited role preoperatively (5%), which increased at 12-month follow-up (21-24%). 'Anxiety/depression' had a limited role in general (9-17%). Diverging patterns were seen for the other domains. For the OHS/OKS, function and pain had similar contributions to the observed differences preoperatively. At the 12-month follow-up, function (61%) was slightly more important for THA patients, while pain (68%) was more important for TKA patients.

Table 4: Percentage of inequality explained by each EQ-5D-3L and OHS/OKS dimension

	THA		TKA	
	Preoperative	12-month follow-up	Preoperative	12-month follow-up
EQ-5D-3L				
Mobility	5	21	5	24
Self-care	26	36	19	13
Usual activities	37	25	31	46
Pain/discomfort	18	7	28	12
Anxiety/depression	13	13	17	9
OHS/OKS				
Function	46	61	38	24
Pain	54	35	62	68

The table depicts data from the EQ-5D-3L cohort. The multivariable linear regression models were used to calculate the percentage of the overall coefficient explained by each domain, as described in the methods section. Abbreviations: OHS = Oxford Hip Score; OKS = Oxford Knee Score

Secondary (EQ-5D-5L) cohorts

Findings regarding demographics (Supplemental File, Table 8–9), preoperative outcomes (Supplemental File, Table 10–11) and outcomes at 12-month follow-up (Supplemental File, Table 12–13) were similar between the EQ-5D-5L and EQ-5D-3L cohorts. Two exceptions were noted. First, preoperatively, inequalities between most and least deprived for the EQ-5D-5L LSS in TKA patients were smaller and insignificant (-0.78 [95% CI -1.76, 0.20]). Second, at 12-month follow-up, inequalities were smaller across all PROMs compared to the EQ-5D-3L cohort, except for OHS in THA. In the EQ-5D-5L cohort, domain contributions were roughly similar compared to the EQ-5D-3L cohort, although the earlier described pattern observed on OHS/OKS was less evident (Supplemental File, Table 14).

DISCUSSION

Main findings

This study demonstrated the presence of SES-related health inequalities in Dutch THA and TKA patients, using a set of acknowledged condition-specific and generic PROMs. Higher deprivation levels were invariably associated with worse health status, both preoperatively and at the 12-month follow-up. The differences were relatively small, bringing into question the clinical relevance. To put the findings into perspective, the average difference between lowest and highest deprivation quintile ranged from 25 to 50% of the minimal important difference (an individual-based relevance criterion) of the instruments[44, 45].

Contrary to our first hypothesis, the SES-related inequalities expressed with the condition-specific OHS/OKS were similar to those in the generic EQ-5D. The second hypothesis was also rejected, as the single-item NRS pain instrument showed similar inequalities compared to the EQ-5D and OHS/OKS. This is most likely due to pain being the hallmark of the osteoarthritic condition, and relatively high responsiveness compensates for the limited domain coverage of the NRS for pain[21]. The single-item EQ VAS showed smaller to no inequalities. We found evidence supporting the third hypothesis, i.e., the presence of reporting heterogeneity of PROMs influencing inequality size: e.g., NRS satisfaction did not show inequalities in TKA patients at 12-month follow-up, despite inequalities being present measured with PROMs assumed to exhibit less reporting heterogeneity. We are aware this evidence is not conclusive and dedicated research is needed to separate a true health difference (i.e., a similar clinical outcome in the deprived) from reporting heterogeneity (i.e., overrating by deprived patients).

Comparison with the literature

The OHS/OKS showed a similar magnitude of outcome inequalities by SES compared to the generic EQ-5D, both pre- and postoperatively. This could be an indirect effect of deprivation not affecting the anxiety/depression domain of the EQ-5D. It contradicts a study conducted in the United States (US) reporting poorer mental health (PROMIS General Health Short Form) among deprived TKA patients[46]. While the anxiety/depression domain is considered sensitive[47], we presume this to be a true effect. Among the Dutch patient population, mental health problems are perhaps less likely to vary according to SES compared to the US, given the SES-indicator is categorized into quintiles. We assume the increased mental health issues in the US could be the result of steeper economic inequalities the Netherlands. Overall, we think it is reasonable to assume that the EQ-5D and OHS/OKS are measuring a similar deprivation effect, due to the considerable overlap of the domains/items that are relevant for TKA/THA patients.

Preoperative and postoperative SES pathways

To understand the sources of SES-related inequalities, we should separate preoperative from postoperative outcome differences by SES. The greater preoperative pain and function among the deprived suggests more severe osteoarthritis when surgery is prescribed and that different thresholds for prescribing surgery are applied. This may indicate the presence of (presumably unintentional) selection, either by the patient or the provider. A US study found that marginalized groups expressed less preference for undergoing surgery[48]. For example, Black patients perceived THA or TKA half as likely to be beneficial compared to White patients. Cost-sharing and deductibles are common features of the Dutch insurance system, which may contribute to self-selection. For example, these financial factors associate with reduced use of mental healthcare resources, potentially influencing who seeks treatment[49]. Deprived patients may also be selected at a later stage for surgery. The general practitioner at the referral stage or the surgeon during the selection procedure may show more reluctance to advise the surgery given the increased presence of comorbidity (e.g., higher BMI) [50]. Finally, communication barriers may lead to an underestimation of the pain and dysfunction levels in deprived patients at all stages[51]. In particular, more affluent patients may be able to make a stronger case for surgery. While the fact is clear-cut, the attribution of SES-related inequalities to specific underlying pathways is in its infancy, hampering reduction.

Postoperative SES-related outcome differences may rest on other factors, apart from a direct relation between pre- and postoperative health. TKA typically requires a more intensive recovery trajectory compared to THA. For that reason, physiotherapy is often advised for TKA patients. The first 20 sessions of physiotherapy following TKA/THA are, however, not covered by the compulsory health insurance in the Netherlands. Financial barriers can be expected for the deprived, and the pressure to restart work or household roles may be higher. Additionally, health literacy issues or language barriers may limit full understanding of the instructions during recovery.

Potential reporting heterogeneity

Separating ‘true’ inequality effects from reporting heterogeneity, a potential source of measurement error, is challenging. We assumed that inequalities in EQ-5D and OHS/OKS by SES, given their wide use in THA and TKA research, best represent ‘true’ health disadvantages. The SES-related inequality was reduced/absent when measured with EQ VAS and NRS satisfaction, compared to the EQ-5D and OHS/OKS. The first explanation states that deprived patients ‘simply’ rate their overall health state to be better or are more satisfied with the result than less deprived patients, which would explain similar scores using the EQ VAS and NRS satisfaction. This explanation implies that overall health or satisfaction has the same meaning across deprivation groups. If this is the case, these measures warrant careful interpretation in the context of SES-related inequalities. Given the broad scope of these questionnaires and the lack of information on which factors influence these measurements[52], we cannot conclude that similar scores across SES categories indicate better THA/TKA care. In other words, without contextual information, these measures provide limited insight into important domains pertinent to the quality of THA/TKA, such as aspects of pain, physical and mental functioning. Although the EQ-5D and OHS/OKS lack important domains such as social functioning, they provide an explicit assessment of the abovementioned domains[53].

Another explanation for differences in the magnitude of SES-inequalities is that both EQ VAS and NRS satisfaction show reporting heterogeneity[54]. In other words, the EQ VAS and NRS satisfaction measures have subtle differences in meaning and severity grading across deprivation groups. This phenomenon may turn into what has been described as the ‘disability paradox’, and entails a discrepancy between the perceived quality of life by external observers and the reported quality of life by individuals with disabilities[55]. As described earlier, phrasing may matter, and the EQ VAS and satisfaction could invite patients to respond using their own reference of the quality of care; these measures may also cover more than health care and health status alone. This internal judgment scale may be affected by prior expectations and experiences. As the EQ-5D and OHS/OKS show relatively similar inequality effects, we expect less bias in these measures. ‘Response shift’, which in essence is a change of response style over time, may also play a role. This phenomenon is described as occurring when patients undergo a process of adaptation over time, leading to a change in their internal standards, values, or conceptualization of their quality of life[56]. It is a desirable human capacity per se, enabling to live up with changed conditions, but it interferes with objective measurement. In our study, the arthroplasty procedure could be an evoking factor, and VAS/satisfaction measures could be more sensitive to this adaptation process. The best way to identify these reporting heterogeneity phenomena would be to conduct a study using vignettes (external anchors) [24].

Strengths and limitations

We included a diverse set of PROMs, and the quality of the registry data was excellent. Moreover, the large national sample and the inclusion of both THA and TKA patients enhanced the generalizability of our findings. A key limitation is the limited information on factors that could explain SES-related inequalities. We relied on postal codes to link an area-based SES indicator. Despite its proven reputation as an explanatory factor, as in our study, more specific variables such as individual-level education are also needed to delineate pathways to reduce SES-related inequalities. A second limitation is our reliance on an indirect method to suggest the presence of reporting heterogeneity. Finally, while the primary and secondary cohorts yielded largely consistent findings, some discrepancies were observed. The starting year of the secondary cohort coincided with the onset of the COVID-19 pandemic, which significantly affected orthopedic care. The PROMs collected during the pandemic differed from those collected in the pre-pandemic periods, potentially accounting for the subtle differences between cohorts[57].

CONCLUSIONS

This study demonstrated that SES-related inequalities among Dutch THA and TKA patients differ significantly across PROM instruments, although the clinical relevance of these differences remains questionable. The most significant SES-related inequalities were observed for functioning and pain (OHS/OKS, NRS pain, EQ-5D), providing potential opportunities for amendment. Notably, the anxiety/depression domain of the EQ-5D had a limited role. Overall, these findings did not translate into deprived patients rating their overall general health (EQ VAS) worse or expressing dissatisfaction (NRS satisfaction). Caution should be exercised when interpreting these latter two measures, as they may lead to overly simplistic interpretations of differences in health based on SES. Future research should focus on further identifying the drivers of inequalities and assessing whether PROMs, in particular EQ VAS and satisfaction, reflect ‘true’ inequalities or are disproportionally affected by reporting heterogeneity.

REFERENCES

1. Marmot M. Social determinants of health inequalities. The Lancet 2005; 365: 1099-1104.
2. Phelan JC, Link BG, Tehranifar P. Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. J Health Soc Behav 2010; 51 Suppl: S28-40.
3. Kamaruzaman H, Kinghorn P, Oppong R. Cost-effectiveness of surgical interventions for the management of osteoarthritis: a systematic review of the literature. BMC Musculoskelet Disord 2017; 18: 183.
4. Mahomed NN, Barrett J, Katz JN, Baron JA, Wright J, Losina E. Epidemiology of total knee replacement in the United States Medicare population. J Bone Joint Surg Am 2005; 87: 1222-1228.
5. Gonzalez AI, Nguyen U-SDT, Franklin P, Barea C, Hannouche D, Lübbeke A. 1-year trajectories of patients undergoing primary total hip arthroplasty: Patient reported outcomes and resource needs according to education level. BMC Musculoskeletal Disorders 2022; 23: 84.
6. Hoelen T-CA, Schotanus M, van Kuijk S, Bastiaenen C, Boonen B, Most J. The relation between socioeconomic status and patient symptoms before and one year after lower extremity arthroplasty. Journal of Orthopaedics 2023; 39: 11-17.
7. Neuburger J, Hutchings A, Allwood D, Black N, van der Meulen JH. Sociodemographic differences in the severity and duration of disease amongst patients undergoing hip or knee replacement surgery. J Public Health (Oxf) 2012; 34: 421-429.
8. Clement ND, Muzammil A, Macdonald D, Howie CR, Biant LC. Socioeconomic status affects the early outcome of total hip replacement. J Bone Joint Surg Br 2011; 93: 464-469.
9. Clement ND, Jenkins PJ, MacDonald D, Nie YX, Patton JT, Breusch SJ, et al. Socioeconomic status affects the Oxford knee score and short-form 12 score following total knee replacement. Bone Joint J 2013; 95-B: 52-58.
10. Bonsel JM, Reijman M, Verhaar JAN, van Steenbergen LN, Janssen MF, Bonsel GJ. Socioeconomic inequalities in patient-reported outcome measures of Dutch primary hip and knee arthroplasty patients for osteoarthritis. Osteoarthritis and Cartilage 2024; 32: 200-209.
11. Jenkins PJ, Perry PR, Yew Ng C, Ballantyne JA. Deprivation influences the functional outcome from total hip arthroplasty. Surgeon 2009; 7: 351-356.
12. Goodman SM, Mandl LA, Parks ML, Zhang M, McHugh KR, Lee YY, et al. Disparities in TKA Outcomes: Census Tract Data Show Interactions Between Race and Poverty. Clin Orthop Relat Res 2016; 474: 1986-1995.
13. Dowsey MM, Nikpour M, Choong PF. Outcomes following large joint arthroplasty: does socio-economic status matter? BMC Musculoskelet Disord 2014; 15: 148.
14. Keurentjes JC, Blane D, Bartley M, Keurentjes JJ, Fiocco M, Nelissen RG. Socio-economic position has no effect on improvement in health-related quality of life and patient satisfaction in total hip and knee replacement: a cohort study. PLoS One 2013; 8: e56785.
15. Neuburger J, Hutchings A, Black N, van der Meulen JH. Socioeconomic differences in patient-reported outcomes after a hip or knee replacement in the English National Health Service. Journal of Public Health 2012; 35: 115-124.
16. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018; 27: 1147-1157.
17. Oppe M, Devlin N, Black N. Comparison of the Underlying Constructs of the EQ-5D and Oxford Hip Score: Implications for Mapping. Value in Health 2011; 14: 884-891.

18. Lin DY, Cheok TS, Samson AJ, Kaambwa B, Brown B, Wilson C, et al. A longitudinal validation of the EQ-5D-5L and EQ-VAS stand-alone component utilising the Oxford Hip Score in the Australian hip arthroplasty population. J Patient Rep Outcomes 2022; 6: 71.
19. Lin DY, Cheok TS, Kaambwa B, Samson AJ, Morrison C, Chan T, et al. Evaluation of the EQ-5D-5L, EQ-VAS stand-alone component and Oxford knee score in the Australian knee arthroplasty population utilising minimally important difference, concurrent validity, predictive validity and responsiveness. Health and Quality of Life Outcomes 2023; 21: 41.
20. Diamantopoulos A, Sarstedt M, Fuchs C, Wilczynski P, Kaiser S. Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. Journal of the Academy of Marketing Science 2012; 40: 434-449.
21. Ruyssen-Witrand A, Fernandez-Lopez CJ, Gossec L, Anract P, Courpied JP, Dougados M. Psychometric properties of the OARSI/OMERACT osteoarthritis pain and functional impairment scales: ICOAP, KOOS-PS and HOOS-PS. Clin Exp Rheumatol 2011; 29: 231-237.
22. Penton H, Dayson C, Hulme C, Young T. An Investigation of Age-Related Differential Item Functioning in the EQ-5D-5L Using Item Response Theory and Logistic Regression. Value in Health 2022; 25: 1566-1574.
23. Christopher J.L. Murray DBE. Health Systems Performance, comparing health outcomes in about 100 countries, World Health Organization 2003.
24. d'Uva TB, O'Donnell O, van Doorslaer E. Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans. Int J Epidemiol 2008; 37: 1375-1383.
25. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. BMC Medical Research Methodology 2010; 10: 22.
26. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. The Lancet 2007; 370: 1453-1457.
27. van Steenbergen LN, Denissen GAW, Spooren A, van Rooden SM, van Oosterhout FJ, Morrenhof JW, Nelissen RGHH. More than 95% completeness of reported procedures in the population-based Dutch Arthroplasty Register. Acta orthopaedica 2015; 86: 498-505.
28. Dutch Orthopaedic National Quality Registry (LROI) Annual report 2023. Available at: www.lroi-rapportage.nl.
29. Janssen MF, Bonsel GJ, Luo N. Is EQ-5D-5L Better Than EQ-5D-3L? A Head-to-Head Comparison of Descriptive Systems and Value Sets from Seven Countries. Pharmacoeconomics 2018; 36: 675-697.
30. Charnley J. The long-term results of low-friction arthroplasty of the hip performed as a primary intervention. J Bone Joint Surg Br 1972; 54: 61-76.
31. Dripps RD. New classification of physical status. Anesthesiology 1963; 24: 111.
32. Dutch Institute of Social Research (Sociaal Cultureel Planbureau (SCP)) Socioenomic status score per postal code area, accessible by request, retrieved October 20th 2020, https://bronnen.zorggegevens.nl/Bron?naam=Sociaal-Economische-Status-per-postcodegebied.
33. Ralston K, Dundas R, Leyland AH. A comparison of the Scottish Index of Multiple Deprivation (SIMD) 2004 with the 2009 + 1 SIMD: does choice of measure affect the interpretation of inequality in mortality? Int J Health Geogr 2014; 13: 27.
34. World Health Organisation (WHO) Handbook on health inequality monitoring: with a special focus on low- and middle-income countries. 2013. Available at: https://apps.who.int/iris/bitstream/handle/10665/85345/9789241548632_eng.pdf?sequence=1.

35. Brooks R. EuroQol: the current state of play. Health Policy 1996; 37: 53-72.
36. Devlin N, Parkin D, Janssen B. An Introduction to EQ-5D Instruments and Their Applications. In: Methods for Analysing and Reporting EQ-5D Data Cham: Springer International Publishing 2020:1-22.
37. Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, Dawson J. The use of the Oxford hip and knee scores. J Bone Joint Surg Br 2007; 89: 1010-1014.
38. https://www.lroi.nl/downloads/181/methodologie-voor-analyse-van-lroi-data.pdf. Dutch Arthroplasty Registry (Landelijke Registratie Orthopedische Implantaten (LROI)) 2020 methodological report. Accessed May 2022.
39. Tennant PWG, Arnold KF, Ellison GTH, Gilthorpe MS. Analyses of 'change scores' do not estimate causal effects in observational data. Int J Epidemiol 2022; 51: 1604-1615.
40. Hugh-Jones D (2024). santoku: A Versatile Cutting Tool. R package version 0.10.0, https://github.com/hughjonesd/santoku.
41. Pronk Y, Peters MCWM, Brinkman J-M. Is Patient Satisfaction After Total Knee Arthroplasty Predictable Using Patient Characteristics and Preoperative Patient-Reported Outcomes? The Journal of Arthroplasty 2021; 36: 2458-2465.
42. von dem Knesebeck O, Bickel H, Fuchs A, Gensichen J, Höfels S, Riedel-Heller SG, et al. Social inequalities in patient-reported outcomes among older multimorbid patients – results of the MultiCare cohort study. International Journal for Equity in Health 2015; 14: 17.
43. R Core Team (2024). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
44. Beard DJ, Harris K, Dawson J, Doll H, Murray DW, Carr AJ, Price AJ. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. J Clin Epidemiol 2015; 68: 73-79.
45. Luo N, Johnson JA, Coons SJ. Using Instrument-Defined Health State Transitions to Estimate Minimally Important Differences for Four Preference-Based Health-Related Quality of Life Instruments. Medical Care 2010; 48.
46. Gulley ML, Carender CN, Glass NA, Bedard NA. What is the Impact of Social Deprivation on Mental and Physical Health Before and After Primary Total Knee Arthroplasty? Arthroplast Today 2023; 22: 101156.
47. Franklin M, Enrique A, Palacios J, Richards D. Psychometric assessment of EQ-5D-5L and ReQoL measures in patients with anxiety and depression: construct validity and responsiveness. Quality of Life Research 2021; 30: 2633-2647.
48. Ang DC, Monahan PO, Cronan TA. Understanding ethnic disparities in the use of total joint arthroplasty: application of the health belief model. Arthritis Rheum 2008; 59: 102-108.
49. Lopes FV, Riumallo Herl CJ, Mackenbach JP, Van Ourti T. Patient cost-sharing, mental health care and inequalities: A population-based natural experiment at the transition to adulthood. Social Science & Medicine 2022; 296: 114741.
50. Algren MH, Bak CK, Berg-Beckhoff G, Andersen PT. Health-Risk Behaviour in Deprived Neighbourhoods Compared with Non-Deprived Neighbourhoods: A Systematic Literature Review of Quantitative Observational Studies. PLoS One 2015; 10: e0139297.
51. De Ruddere L, Goubert L, Prkachin KM, Louis Stevens MA, Van Ryckeghem DML, Crombez G. When you dislike patients, pain is taken less seriously. PAIN 2011; 152.
52. Feng Y, Parkin D, Devlin NJ. Assessing the performance of the EQ-VAS in the NHS PROMs programme. Qual Life Res 2014; 23: 977-989.

53. Holmenlund C, Overgaard S, Bilberg R, Varnum C. Evaluation of the Oxford Hip Score: Does it still have content validity? Interviews of total hip arthroplasty patients. Health Qual Life Outcomes 2021; 19: 237.
54. Elster J. Sour Grapes: Studies in the Subversion of Rationality, Cambridge University Press 1983.
55. Albrecht GL, Devlieger PJ. The disability paradox: high quality of life against all odds. Soc Sci Med 1999; 48: 977-988.
56. Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. Soc Sci Med 1999; 48: 1507-1515.
57. Bonsel JM, Groot L, Cohen A, Verhaar JAN, Gademan MGJ, Spekenbrink-Spooren A, et al. Impact of the COVID-19 lockdown on patient-reported outcome measures in Dutch hip and knee arthroplasty patients. Acta Orthop 2022; 93: 808-818.

DECLARATIONS

Ethics approval: This study used anonymized prospectively collected clinical registry data from the Dutch Arthroplasty Registry (LROI, www.lroi.nl). Patients undergoing surgery were given the option to op-out of sharing their data with the LROI. Studies using LROI data are subject to technical and ethical judgment by the registry holder. If permission is granted, Dutch Law does not require additional institutional ethical judgment by a Medical Ethical Review Board.

Consent for publication: not applicable

Consent to participate: not applicable

Availability of data and materials: per the agreement with the LROI, the authors cannot share any data used in this study. Codes used to conduct the analyses, however, are obtainable from the corresponding author.

Competing interests: the authors declare that they have no competing interests.

Funding: this work was funded by a PhD grant (PHD-287) provided by the EuroQol Research Foundation. The funder had no role in the design and conduct of the study, or approval of the manuscript, nor the decision to submit the manuscript for publication. The views expressed are those of the individual authors and do not necessarily reflect the views of the EuroQol Research Foundation.

Authors' contributions: Joshua Bonsel, Max Reijman, Erin Macri, Jan Verhaar and Gouke Bonsel contributed to the study conception and design. Funding acquisition was performed by Joshua Bonsel, Max Reijman, Jan Verhaar and Gouke Bonsel. Data collection and analysis were performed by Joshua Bonsel. The first draft of the manuscript was written by Joshua Bonsel, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Acknowledgements: not applicable.

Supplementary data: Supplementary data associated with this article can be found at the end of this chapter.

SUPPLEMENTAL FILE, TABLES 1–14

- Exposure: SES, which was categorized into quintiles.
- Outcome: all PROM outcomes were transformed into a 0-100 scale where 100 represents the best attainable outcome.
- Table 1 presents cut-off values of preoperative PROM scores (tertiles).
- Table 2–3 presents descriptive statistics of demographics and outcomes of the primary EQ-5D-3L cohorts.
- Table 4–7 presents the full adjusted linear regression models studying the association between SES and respective PROM outcomes for the primary EQ-5D-3L cohorts. The regression models were adjusted for sex, age, BMI, ASA score, Charnley score, and type of hospital.
- Table 8–9 presents descriptive statistics of demographics and outcomes of the secondary EQ-5D-5L cohorts.
- Table 10–13 presents the full adjusted linear regression models studying the association between SES and respective PROM outcomes for the secondary EQ-5D-5L cohorts. The regression models were adjusted for sex, age, BMI, ASA score, Charnley score, and type of hospital.
- Table 14 presents the percentage of inequality explained by each EQ-5D-5L and OHS/OKS dimension, using data from the secondary EQ-5D-5L cohort.
- Abbreviations: PROM = patient-reported outcome measure; THA = total hip arthroplasty; TKA = total knee arthroplasty; SES = socioeconomic status; BMI = Body Mass Index; ASA = American Society of Anesthesiology score; LSS = level-sum-score; VAS = Visual Analogue Scale; OHS = Oxford Hip Score; OKS = Oxford Knee Score; NRS = Numerical Rating Scale; IQR = Interquartile Range

Table 1: Cut-off points of tertiles of preoperative scores

	EQ-5D-3L cohort				EQ-5D-5L cohort			
	THA		TKA		THA		TKA	
	Median (range)	N	Median (range)	N	Median (range)	N	Median (range)	N
Pre-op score LSS								
lowest	50 (0, 50)	13392	40 (0, 45)	6269	50 (0, 50)	3639	45 (0, 50)	2216
middle	60 (60, 60)	13100	55 (50, 60)	8042	60 (60, 60)	5067	60 (55, 65)	3410
highest	70 (70, 100)	19290	70 (65, 100)	18423	70 (70, 100)	5682	75 (70, 100)	3565
Pre-op score OHS/OKS								
lowest	29 (0, 40)	13893	33 (0, 40)	8863	29 (0, 38)	4523	31 (0, 40)	2831
middle	48 (42, 54)	12459	48 (42, 54)	10478	48 (40, 54)	4614	48 (42, 54)	3080
highest	65 (56, 100)	15948	65 (56, 100)	10456	65 (56, 100)	4872	65 (56, 100)	2970
Pre-op score EQ VAS								
lowest	49 (0, 59)	13392	50 (0, 63)	10622	45 (0, 59)	4655	50 (0, 60)	2957
middle	70 (60, 77)	16781	70 (64, 79)	9951	70 (60, 75)	4851	70 (61, 79)	2919
highest	85 (78, 100)	15304	86 (80, 100)	11852	84 (76, 100)	4793	85 (80, 100)	3245
Pre-op score NRS pain in rest								
lowest	20 (0, 20)	9108	20 (0, 20)	6259	20 (0, 20)	2983	20 (0, 20)	1805
middle	40 (30, 50)	19284	40 (30, 50)	13116	40 (30, 50)	6263	40 (30, 50)	4002
highest	70 (60, 100)	16816	70 (60, 100)	11355	70 (60, 100)	5116	80 (60, 100)	3334
Pre-op score NRS pain during activity								
lowest	10 (0, 10)	11392	10 (0, 10)	8305	10 (0, 10)	3882	10 (0, 10)	2471
middle	20 (20, 20)	13465	20 (20, 20)	9288	20 (20, 20)	4447	20 (20, 20)	2887
highest	40 (30, 100)	20391	40 (30, 100)	13125	40 (30, 100)	6016	40 (30, 100)	3769

Table 2: Demographics and outcomes of THA patients (EQ-5D-3L cohort)

	Q1 (least depr.)	Q2	Q3	Q4	Q5 (most depr.)	p-value
Total	7913	8508	9804	10796	8861	
SES Z-score, median [range]	1.3 [0.9, 2.8]	0.6 [0.4, 0.9]	0.2 [-0.1, 0.4]	-0.4 [-0.8, -0.1]	-1.4 [-6.3, -0.8]	
Demographics						
Age, median [IQR]	69.0 [63.0, 75.0]	69.0 [63.0, 75.0]	70.0 [63.0, 75.0]	70.0 [63.0, 75.0]	70.0 [64.0, 76.0]	<0.001
<50	271 (3)	258 (3)	262 (3)	283 (3)	226 (3)	<0.001
50-69	3739 (47)	4108 (48)	4631 (47)	5088 (47)	3930 (44)	
>70	3903 (49)	4142 (49)	4911 (50)	5425 (50)	4705 (53)	
BMI, median [IQR]	26.1 [23.9, 29.0]	26.6 [24.1, 29.7]	26.9 [24.3, 29.8]	27.0 [24.4, 30.0]	27.3 [24.7, 30.4]	<0.001
<25	3400 (43)	3278 (39)	3570 (36)	3793 (35)	2876 (32)	<0.001
25-30	3162 (40)	3496 (41)	4173 (43)	4600 (43)	3817 (43)	
>30	1351 (17)	1734 (20)	2061 (21)	2403 (22)	2168 (24)	
Male	5020 (63)	5329 (63)	6155 (63)	6964 (65)	5817 (66)	0.003
ASA						<0.001
I	1749 (22)	1744 (20)	1960 (20)	1874 (17)	1303 (15)	
II	4997 (63)	5489 (65)	6260 (64)	7078 (66)	5776 (65)	
III-IV	1167 (15)	1275 (15)	1584 (16)	1844 (17)	1782 (20)	
Charnley						<0.001
A	3619 (46)	3714 (44)	4371 (45)	4584 (42)	3907 (44)	
B1	2497 (32)	2695 (32)	2987 (30)	3530 (33)	2693 (30)	
B2	1634 (21)	1871 (22)	2118 (22)	2324 (22)	1942 (22)	
C	163 (2)	228 (3)	328 (3)	358 (3)	319 (4)	
Smoking (yes)	600 (8)	766 (9)	928 (9)	986 (9)	896 (10)	<0.001
Previous surgery of the joint (yes)	135 (2)	130 (2)	134 (1)	183 (2)	162 (2)	0.589
Year of surgery						<0.001
2014	426 (5)	372 (4)	433 (4)	426 (4)	452 (5)	

Table 2: Continued

	Q1 (least depr.)	Q2	Q3	Q4	Q5 (most depr.)	p-value
2015	1034 (13)	1026 (12)	1142 (12)	1119 (10)	941 (11)	
2016	1338 (17)	1492 (18)	1710 (17)	1776 (16)	1433 (16)	
2017	1546 (20)	1685 (20)	1986 (20)	2164 (20)	1811 (20)	
2018	1748 (22)	1884 (22)	2202 (22)	2420 (22)	1955 (22)	
2019	1468 (19)	1686 (20)	1974 (20)	2300 (21)	1736 (20)	
2020	243 (3)	233 (3)	222 (2)	323 (3)	343 (4)	
2021	77 (1)	82 (1)	102 (1)	193 (2)	145 (2)	
2022	33 (0)	48 (1)	33 (0)	75 (1)	45 (1)	
Type of hospital						<0.001
General Hospital	7013 (89)	7559 (89)	8822 (90)	9836 (91)	8099 (91)	
Private Hospital	819 (10)	817 (10)	852 (9)	817 (8)	493 (6)	
University Medical Center	81 (1)	132 (2)	130 (1)	143 (1)	269 (3)	
Fixation						<0.001
Cemented	1441 (18)	1605 (19)	1927 (20)	2338 (22)	2470 (28)	
Cementless	5918 (75)	6152 (72)	6882 (70)	7549 (70)	5510 (62)	
Hybrid	554 (7)	751 (9)	995 (10)	909 (8)	881 (10)	
Approach						<0.001
Anterior	3458 (44)	3393 (40)	3462 (35)	3546 (33)	2830 (32)	
Anterolateral	235 (3)	341 (4)	414 (4)	445 (4)	418 (5)	
Other	78 (1)	64 (1)	114 (1)	193 (2)	105 (1)	
Posterolateral	3608 (46)	4121 (48)	5262 (54)	5997 (56)	4945 (56)	
Straight lateral	534 (7)	589 (7)	552 (6)	615 (6)	563 (6)	
Contralateral procedure (yes)	760 (10)	879 (10)	1024 (10)	1152 (11)	900 (10)	0.242

Table 2: Continued

	Q1 (least depr.)	Q2	Q3	Q4	Q5 (most depr.)	p-value
Outcomes						
Preoperative outcomes, median [IQR]						
EQ-5D-3L LSS	60.0 [50.0, 70.0]	60.0 [50.0, 70.0]	60.0 [50.0, 70.0]	60.0 [50.0, 70.0]	60.0 [50.0, 70.0]	
OHS	52.1 [37.5, 62.5]	50.0 [37.5, 62.5]	50.0 [35.4, 60.4]	47.9 [35.4, 60.4]	47.9 [33.3, 60.4]	
EQ VAS	70.0 [56.0, 80.0]	70.0 [54.0, 80.0]	70.0 [53.0, 80.0]	70.0 [53.0, 80.0]	70.0 [51.0, 80.0]	
NRS Pain in rest	50.0 [30.0, 70.0]	50.0 [30.0, 70.0]	50.0 [30.0, 70.0]	40.0 [30.0, 70.0]	40.0 [30.0, 60.0]	
NRS Pain during activity	20.0 [20.0, 40.0]	20.0 [20.0, 40.0]	20.0 [10.0, 40.0]	20.0 [10.0, 40.0]	20.0 [10.0, 40.0]	
12-month follow-up outcomes, median [IQR]						
EQ-5D-3L LSS	100.0 [80.0, 100.0]	90.0 [80.0, 100.0]	90.0 [80.0, 100.0]	90.0 [80.0, 100.0]	90.0 [70.0, 100.0]	
OHS	93.8 [83.3, 97.9]	93.8 [83.3, 97.9]	93.8 [83.3, 97.9]	93.8 [83.3, 97.9]	91.7 [81.2, 97.9]	
EQ VAS	80.0 [70.0, 90.0]	80.0 [70.0, 90.0]	80.0 [70.0, 90.0]	80.0 [70.0, 90.0]	80.0 [70.0, 90.0]	
NRS Pain in rest	100.0 [90.0, 100.0]	100.0 [90.0, 100.0]	100.0 [90.0, 100.0]	100.0 [90.0, 100.0]	100.0 [90.0, 100.0]	
NRS Pain during activity	100.0 [80.0, 100.0]	100.0 [80.0, 100.0]	100.0 [80.0, 100.0]	100.0 [80.0, 100.0]	90.0 [80.0, 100.0]	
Preoperative outcomes, n (%) ceiling						
EQ-5D-3L LSS	68 (0.9)	79 (0.9)	66 (0.7)	54 (0.5)	68 (0.8)	
OHS	9 (0.1)	3 (0.0)	2 (0.0)	3 (0.0)	3 (0.0)	
EQ VAS	151 (1.9)	124 (1.5)	173 (1.8)	185 (1.7)	167 (1.9)	
NRS Pain in rest	432 (5.5)	414 (5.0)	468 (4.9)	500 (4.7)	446 (5.1)	
NRS Pain during activity	71 (0.9)	76 (0.9)	84 (0.9)	90 (0.8)	95 (1.1)	
12-month follow-up outcomes, n (%) ceiling						
EQ-5D-3L LSS	3978 (50.3)	4227 (49.7)	4705 (48.0)	5023 (46.5)	3924 (44.3)	
OHS	1728 (24.7)	1824 (23.6)	2104 (23.6)	2150 (21.7)	1545 (20.0)	
EQ VAS	480 (6.1)	570 (6.7)	718 (7.4)	715 (6.7)	618 (7.0)	
NRS Pain in rest	5426 (70.4)	5743 (68.9)	6591 (68.6)	7192 (67.4)	5679 (65.1)	
NRS Pain during activity	4190 (54.4)	4440 (53.2)	5053 (52.5)	5460 (51.2)	4346 (49.7)	

Table 3: Demographics and outcomes of TKA patients (EQ-5D-3L cohort)

	Q1 (least depr.)	Q2	Q3	Q4	Q5 (most depr.)	p-value
Total	4860	5589	6990	8335	6960	
SES Z-score, median [range]	1.3 [0.9, 2.7]	0.6 [0.4, 0.9]	0.2 [-0.1, 0.4]	-0.4 [-0.8, -0.1]	-1.4 [-6.2, -0.8]	
Demographics						
Age, median [IQR]	69.0 [63.0, 75.0]	69.0 [63.0, 74.0]	69.0 [63.0, 74.0]	69.0 [63.0, 74.0]	69.0 [63.0, 75.0]	0.373
<50	49 (1)	80 (1)	93 (1)	119 (1)	93 (1)	0.105
50-69	2386 (49)	2915 (52)	3706 (53)	4355 (52)	3495 (50)	
>70	2425 (50)	2594 (46)	3191 (46)	3861 (46)	3372 (48)	
BMI, median [IQR]	28.4 [25.7, 31.7]	28.7 [25.9, 32.0]	28.8 [26.0, 32.2]	29.0 [26.2, 32.7]	29.4 [26.4, 33.2]	<0.001
<25	1101 (23)	1184 (21)	1444 (21)	1585 (19)	1233 (18)	<0.001
25-30	2114 (43)	2456 (44)	2928 (42)	3508 (42)	2777 (40)	
>30	1645 (34)	1949 (35)	2618 (37)	3242 (39)	2950 (42)	
Male						0.009
ASA						<0.001
I	738 (15)	826 (15)	1042 (15)	1085 (13)	705 (10)	
II	3284 (68)	3794 (68)	4697 (67)	5598 (67)	4657 (67)	
III-IV	838 (17)	969 (17)	1251 (18)	1652 (20)	1598 (23)	
Charnley						0.154
A	2053 (42)	2415 (43)	2967 (42)	3479 (42)	2948 (42)	
B1	1682 (35)	1851 (33)	2329 (33)	2798 (34)	2306 (33)	
B2	991 (20)	1114 (20)	1442 (21)	1728 (21)	1479 (21)	
C	134 (3)	209 (4)	252 (4)	330 (4)	227 (3)	
Smoking (yes)	288 (6)	381 (7)	518 (7)	646 (8)	614 (9)	<0.001
Previous surgery of the joint (yes)	1370 (28)	1687 (30)	2146 (31)	2595 (31)	1929 (28)	0.586
Year of surgery						<0.001
2014	78 (2)	64 (1)	66 (1)	54 (1)	54 (1)	

Table 3: Continued

	Q1 (least depr.)	Q2	Q3	Q4	Q5 (most depr.)	p-value
2015	329 (7)	307 (5)	336 (5)	417 (5)	325 (5)	
2016	863 (18)	1045 (19)	1367 (20)	1567 (19)	1313 (19)	
2017	1254 (26)	1357 (24)	1761 (25)	2077 (25)	1768 (25)	
2018	1124 (23)	1337 (24)	1664 (24)	2052 (25)	1634 (23)	
2019	948 (20)	1212 (22)	1455 (21)	1698 (20)	1462 (21)	
2020	172 (4)	155 (3)	203 (3)	269 (3)	258 (4)	
2021	61 (1)	89 (2)	111 (2)	146 (2)	105 (2)	
2022	31 (1)	23 (0)	27 (0)	55 (1)	41 (1)	
Type of hospital						<0.001
General Hospital	4202 (86)	4820 (86)	6146 (88)	7506 (90)	6316 (91)	
Private Hospital	602 (12)	708 (13)	784 (11)	757 (9)	484 (7)	
University Medical Center	56 (1)	61 (1)	60 (1)	72 (1)	160 (2)	
Fixation						<0.001
Cemented	4338 (89)	5292 (95)	6644 (95)	7806 (94)	6473 (93)	
Cementless	346 (7)	119 (2)	211 (3)	276 (3)	234 (3)	
Hybrid	176 (4)	178 (3)	135 (2)	253 (3)	253 (4)	
Approach						<0.001
Lateral parapatellar	35 (1)	31 (1)	34 (0)	39 (0)	34 (0)	
Medial parapatellar	4697 (97)	5426 (97)	6784 (97)	8153 (98)	6834 (98)	
Other	3 (0)	5 (0)	4 (0)	9 (0)	6 (0)	
Vastus (mid/sub)	125 (3)	127 (2)	168 (2)	134 (2)	86 (1)	
Contralateral procedure (yes)	565 (12)	694 (12)	892 (13)	1101 (13)	890 (13)	0.062

Table 3: Continued

	Q1 (least depr.)	Q2	Q3	Q4	Q5 (most depr.)	p-value
Outcomes						
Preoperative outcomes, median [IQR]						
EQ-5D-3L LSS	70.0 [60.0, 70.0]	70.0 [60.0, 70.0]	70.0 [60.0, 70.0]	70.0 [60.0, 70.0]	70.0 [60.0, 70.0]	
OKS	52.1 [39.6, 62.5]	50.0 [39.6, 60.4]	50.0 [39.6, 60.4]	47.9 [37.5, 58.3]	47.9 [35.4, 58.3]	
EQ VAS	72.0 [60.0, 82.0]	73.0 [60.0, 83.0]	72.0 [60.0, 81.0]	71.0 [60.0, 80.0]	70.0 [56.0, 80.0]	
NRS Pain in rest	50.0 [30.0, 70.0]	50.0 [30.0, 70.0]	50.0 [30.0, 70.0]	40.0 [30.0, 70.0]	40.0 [30.0, 60.0]	
NRS Pain during activity	20.0 [20.0, 40.0]	20.0 [20.0, 40.0]	20.0 [10.0, 30.0]	20.0 [10.0, 30.0]	20.0 [10.0, 30.0]	
12-month follow-up outcomes, median [IQR]						
EQ-5D-3L LSS	90.0 [70.0, 100.0]	90.0 [70.0, 100.0]	90.0 [70.0, 100.0]	90.0 [70.0, 100.0]	90.0 [70.0, 100.0]	
OKS	87.5 [75.0, 95.8]	87.5 [75.0, 93.8]	87.5 [75.0, 93.8]	85.4 [72.9, 93.8]	83.3 [68.8, 93.8]	
EQ VAS	80.0 [70.0, 90.0]	80.0 [70.0, 90.0]	80.0 [70.0, 90.0]	80.0 [70.0, 90.0]	80.0 [67.0, 90.0]	
NRS Pain in rest	100.0 [80.0, 100.0]	100.0 [80.0, 100.0]	100.0 [80.0, 100.0]	90.0 [80.0, 100.0]	90.0 [70.0, 100.0]	
NRS Pain during activity	90.0 [70.0, 100.0]	90.0 [60.0, 100.0]	90.0 [60.0, 100.0]	80.0 [60.0, 100.0]	80.0 [60.0, 100.0]	
NRS Satisfaction	80.0 [70.0, 100.0]	80.0 [70.0, 100.0]	80.0 [70.0, 100.0]	80.0 [70.0, 90.0]	80.0 [70.0, 100.0]	
Preoperative outcomes, n (%) ceiling						
EQ-5D-3L LSS	64 (1.3)	75 (1.3)	73 (1.0)	72 (0.9)	65 (0.9)	
OKS	2 (0.0)	3 (0.1)	4 (0.1)	2 (0.0)	0 (0.0)	
EQ VAS	92 (1.9)	119 (2.2)	136 (2.0)	173 (2.1)	148 (2.1)	
NRS Pain in rest	307 (6.8)	294 (5.6)	345 (5.3)	398 (5.1)	355 (5.3)	
NRS Pain during activity	39 (0.9)	38 (0.7)	49 (0.8)	56 (0.7)	80 (1.2)	
12-month follow-up outcomes, n (%) ceiling						
EQ-5D-3L LSS	2069 (42.6)	2346 (42.0)	2852 (40.8)	3314 (39.8)	2578 (37.0)	
OKS	344 (8.0)	358 (7.1)	440 (6.9)	535 (7.0)	370 (5.9)	
EQ VAS	214 (4.4)	252 (4.6)	326 (4.7)	383 (4.6)	307 (4.5)	

Table 3: Continued

	Q1 (least depr.)	Q2	Q3	Q4	Q5 (most depr.)	p-value
NRS Pain in rest	2498 (54.6)	2784 (51.7)	3454 (51.5)	4029 (49.8)	3282 (48.1)	
NRS Pain during activity	1586 (34.6)	1802 (33.4)	2238 (33.3)	2617 (32.3)	2146 (31.5)	
NRS Satisfaction	1075 (26.2)	1284 (25.6)	1575 (25.2)	1877 (24.9)	1540 (25.0)	

Table 4: Association between socioeconomic status and preoperative health status of THA patients (EQ-5D-3L cohort)

	LSS	VAS	OHS	NRS Pain in rest	NRS Pain during activity
Variables	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**
Intercept	61.48 (60.56, 62.40)	67.43 (66.15, 68.72)	50.52 (49.33, 51.71)	45.07 (43.38, 46.75)	25.54 (24.20, 26.89)
SES					
Q1, least deprived	*reference*				
Q2	-0.53 (-0.96, -0.11)	-0.84 (-1.42, -0.25)	-0.65 (-1.20, -0.11)	-1.04 (-1.81, -0.26)	-0.73 (-1.34, -0.11)
Q3	-1.07 (-1.48, -0.67)	-0.77 (-1.34, -0.20)	-1.35 (-1.88, -0.83)	-1.72 (-2.46, -0.97)	-1.16 (-1.76, -0.56)
Q4	-1.40 (-1.80, -1.00)	-0.56 (-1.12, 0.00)	-1.84 (-2.35, -1.33)	-2.35 (-3.08, -1.62)	-2.13 (-2.71, -1.54)
Q5, most deprived	-1.62 (-2.04, -1.20)	-0.97 (-1.56, -0.39)	-2.42 (-2.96, -1.88)	-3.42 (-4.19, -2.65)	-1.54 (-2.16, -0.93)
Male (vs. female)	2.93 (2.67, 3.20)	3.71 (3.34, 4.08)	5.42 (5.08, 5.76)	5.30 (4.82, 5.79)	4.06 (3.68, 4.45)
Age					
<50	*reference*				
50-69	4.26 (3.49, 5.04)	4.62 (3.54, 5.71)	3.74 (2.73, 4.74)	5.14 (3.71, 6.56)	3.93 (2.79, 5.06)
>70	4.33 (3.54, 5.11)	6.72 (5.62, 7.81)	3.68 (2.67, 4.70)	8.78 (7.34, 10.22)	5.55 (4.40, 6.70)
BMI					
<25	*reference*				
25-30	-1.20 (-1.48, -0.91)	-0.96 (-1.37, -0.56)	-2.62 (-2.99, -2.25)	-2.37 (-2.90, -1.85)	-1.93 (-2.35, -1.51)
>30	-3.32 (-3.68, -2.96)	-3.36 (-3.86, -2.86)	-6.42 (-6.87, -5.96)	-4.32 (-4.98, -3.67)	-3.97 (-4.49, -3.45)
ASA					
I					
II	-2.29 (-2.64, -1.95)	-3.66 (-4.15, -3.18)	-2.87 (-3.31, -2.43)	-2.19 (-2.82, -1.56)	-2.46 (-2.96, -1.95)
III-IV	-6.89 (-7.35, -6.43)	-9.26 (-9.90, -8.62)	-8.22 (-8.81, -7.63)	-4.50 (-5.34, -3.66)	-5.70 (-6.37, -5.03)
Charnley					
A	*reference*				
B1	0.00 (-0.30, 0.29)	0.23 (-0.18, 0.64)	0.57 (0.19, 0.95)	-0.06 (-0.60, 0.48)	0.17 (-0.26, 0.60)
B2	1.29 (0.96, 1.62)	0.73 (0.27, 1.20)	1.78 (1.36, 2.21)	-0.34 (-0.95, 0.27)	1.47 (0.99, 1.96)
C	-1.45 (-2.21, -0.70)	-1.17 (-2.23, -0.12)	-1.82 (-2.77, -0.88)	-3.68 (-5.05, -2.30)	-0.67 (-1.77, 0.43)

Table 4: Continued

	LSS	VAS	OHS	NRS Pain in rest	NRS Pain during activity
Variables	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**
Hospital					
Private Hospital	*reference*				
General Hospital	-0.98 (-1.45, -0.50)	-2.45 (-3.11, -1.79)	-0.99 (-1.58, -0.39)	0.76 (-0.10, 1.62)	1.74 (1.06, 2.43)
University Medical Center	-5.71 (-6.80, -4.63)	-5.45 (-6.97, -3.94)	-9.74 (-11.18, -8.30)	-2.03 (-4.04, -0.02)	-2.03 (-3.63, -0.42)
R-squared	0.053	0.040	0.078	0.026	0.027

4

Table 5: Association between socioeconomic status and preoperative health status of TKA patients (EQ-5D-3L cohort)

	LSS	VAS	OKS	NRS Pain in rest	NRS Pain during activity
Variables	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**
Intercept	65.44 (64.07, 66.81)	71.39 (69.44, 73.35)	51.16 (49.00, 53.32)	45.72 (42.97, 48.48)	26.00 (23.86, 28.15)
SES					
Q1, least deprived	*reference*				
Q2	-0.51 (-1.00, -0.01)	0.58 (-0.13, 1.28)	-1.44 (-2.37, -0.52)	-0.81 (-1.81, 0.19)	-0.67 (-1.44, 0.11)
Q3	-1.22 (-1.69, -0.74)	0.26 (-0.41, 0.94)	-1.96 (-2.87, -1.06)	-2.54 (-3.50, -1.59)	-1.77 (-2.52, -1.03)
Q4	-1.52 (-1.97, -1.06)	0.15 (-0.50, 0.80)	-1.94 (-2.84, -1.04)	-3.37 (-4.29, -2.45)	-2.02 (-2.74, -1.31)
Q5, most deprived	-2.07 (-2.55, -1.60)	-0.78 (-1.46, -0.10)	-4.33 (-5.28, -3.39)	-4.47 (-5.42, -3.51)	-1.89 (-2.63, -1.15)
Male (vs. female)	3.07 (2.78, 3.36)	3.62 (3.20, 4.03)	4.80 (4.20, 5.39)	5.01 (4.43, 5.60)	3.68 (3.22, 4.13)
Age					
<50	*reference*				
50-69	4.10 (2.87, 5.34)	4.56 (2.81, 6.32)	4.05 (2.06, 6.03)	4.19 (1.72, 6.66)	3.17 (1.24, 5.09)
>70	5.01 (3.77, 6.25)	6.41 (4.64, 8.17)	4.20 (2.20, 6.21)	9.26 (6.78, 11.75)	6.07 (4.13, 8.00)
BMI					
<25	*reference*				
25-30	-0.93 (-1.31, -0.55)	-0.56 (-1.10, -0.02)	-3.16 (-3.81, -2.52)	-2.08 (-2.85, -1.31)	-1.39 (-1.98, -0.79)
>30	-2.70 (-3.10, -2.30)	-2.74 (-3.31, -2.16)	-8.24 (-9.05, -7.42)	-4.93 (-5.74, -4.12)	-3.52 (-4.15, -2.89)
ASA					
I	*reference*				
II	-2.14 (-2.58, -1.70)	-4.24 (-4.87, -3.62)	-3.09 (-3.91, -2.27)	-0.81 (-1.70, 0.07)	-1.62 (-2.31, -0.93)
III-IV	-5.70 (-6.24, -5.16)	-9.78 (-10.54, -9.01)	-8.03 (-9.07, -6.98)	-2.74 (-3.82, -1.66)	-4.12 (-4.97, -3.28)
Charnley					
A	*reference*				
B1	-0.14 (-0.47, 0.18)	-0.40 (-0.86, 0.06)	0.65 (-0.02, 1.31)	0.22 (-0.44, 0.87)	-0.61 (-1.12, -0.10)
B2	1.48 (1.11, 1.86)	1.25 (0.71, 1.79)	1.89 (1.14, 2.63)	0.76 (0.00, 1.52)	1.63 (1.04, 2.23)
C	-2.20 (-2.98, -1.42)	-0.64 (-1.74, 0.47)	-1.52 (-3.20, 0.17)	-1.43 (-3.01, 0.15)	-0.84 (-2.07, 0.39)

Table 5: Continued

	LSS	VAS	OKS	NRS Pain in rest	NRS Pain during activity
Variables	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**
Hospital					
Private Hospital	*reference*				
General Hospital	-1.24 (-1.73, -0.76)	-3.32 (-4.01, -2.63)	-2.09 (-2.80, -1.38)	0.49 (-0.51, 1.50)	0.41 (-0.38, 1.19)
University Medical Center	-3.51 (-4.85, -2.18)	-4.03 (-5.95, -2.11)	-5.02 (-9.78, -0.25)	-0.20 (-2.96, 2.56)	0.22 (-1.93, 2.36)
R-squared	0.053	0.051	0.089	0.033	0.028

Table 6: Association between socioeconomic status and 12-month follow-up health status of THA patients (EQ-5D-3L cohort)

	LSS	VAS	OHS	NRS Pain in rest	NRS Pain during activity
Variables	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**
Intercept	87.38 (86.33, 88.44)	79.93 (78.71, 81.14)	89.74 (88.71, 90.77)	89.50 (88.26, 90.74)	83.28 (81.76, 84.80)
SES					
Q1, least deprived	*reference*				
Q2	0.16 (-0.31, 0.63)	0.37 (-0.18, 0.91)	0.25 (-0.21, 0.71)	0.00 (-0.55, 0.55)	-0.15 (-0.83, 0.52)
Q3	-0.46 (-0.92, -0.01)	-0.14 (-0.67, 0.38)	0.10 (-0.35, 0.55)	-0.52 (-1.06, 0.01)	-0.84 (-1.50, -0.18)
Q4	-0.45 (-0.90, 0.00)	-0.31 (-0.82, 0.21)	-0.02 (-0.46, 0.41)	-0.84 (-1.36, -0.32)	-1.05 (-1.69, -0.41)
Q5, most deprived	-1.14 (-1.61, -0.67)	-0.83 (-1.37, -0.29)	-0.69 (-1.15, -0.22)	-1.83 (-2.38, -1.29)	-2.03 (-2.71, -1.35)
Pre-op score in 3 categories					
lowest					
middle	5.06 (4.69, 5.43)	4.34 (3.94, 4.75)	3.64 (3.29, 4.00)	2.91 (2.47, 3.36)	2.11 (1.56, 2.66)
highest	8.44 (8.10, 8.79)	9.88 (9.46, 10.30)	6.20 (5.86, 6.54)	7.03 (6.57, 7.49)	4.68 (4.17, 5.19)
Male (vs. female)	2.70 (2.40, 3.00)	2.01 (1.67, 2.36)	1.99 (1.70, 2.29)	1.13 (0.79, 1.48)	1.95 (1.52, 2.38)
Age					
<50	*reference*				
50-69	0.87 (0.01, 1.74)	-0.09 (-1.08, 0.91)	0.65 (-0.20, 1.51)	2.11 (1.10, 3.12)	3.73 (2.49, 4.98)
>70	-0.42 (-1.30, 0.46)	-2.25 (-3.26, -1.24)	-0.96 (-1.83, -0.10)	1.56 (0.54, 2.58)	4.24 (2.97, 5.50)
BMI					
<25	*reference*				
25-30	-1.31 (-1.63, -0.99)	-1.10 (-1.47, -0.73)	-1.34 (-1.65, -1.02)	-0.76 (-1.14, -0.39)	-1.53 (-2.00, -1.07)
>30	-3.28 (-3.68, -2.88)	-2.79 (-3.25, -2.33)	-3.41 (-3.81, -3.02)	-1.59 (-2.05, -1.12)	-2.86 (-3.44, -2.29)
ASA					
I	*reference*				
II	-2.54 (-2.92, -2.15)	-3.35 (-3.79, -2.90)	-1.95 (-2.33, -1.58)	-1.78 (-2.23, -1.33)	-2.16 (-2.72, -1.60)
III-IV	-6.32 (-6.83, -5.80)	-8.17 (-8.77, -7.57)	-5.44 (-5.94, -4.93)	-3.23 (-3.83, -2.64)	-4.11 (-4.85, -3.37)

Table 6: Continued

	LSS	VAS	OHS	NRS Pain in rest	NRS Pain during activity
Variables	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**
Charnley					
A	*reference*				
B1	-0.49 (-0.82, -0.16)	-0.19 (-0.57, 0.19)	-0.46 (-0.78, -0.13)	0.07 (-0.31, 0.45)	0.28 (-0.19, 0.76)
B2	-1.13 (-1.50, -0.76)	-0.47 (-0.90, -0.04)	-0.58 (-0.95, -0.22)	0.39 (-0.04, 0.82)	0.87 (0.33, 1.40)
C	-2.63 (-3.48, -1.79)	-2.28 (-3.26, -1.30)	-1.54 (-2.35, -0.73)	-0.71 (-1.69, 0.27)	-0.86 (-2.07, 0.35)
Hospital					
Private Hospital	*reference*				
General Hospital	-2.06 (-2.58, -1.53)	-2.03 (-2.64, -1.42)	-2.06 (-2.56, -1.55)	-1.68 (-2.29, -1.08)	-1.48 (-2.23, -0.72)
University Medical Center	-5.96 (-7.17, -4.74)	-3.76 (-5.16, -2.36)	-5.14 (-6.36, -3.92)	-1.29 (-2.73, 0.15)	-2.79 (-4.58, -1.01)
R-squared	0.103	0.095	0.085	0.034	0.021

Table 7: Association between socioeconomic status and 12-month follow-up health status of TKA patients (EQ-5D-3L cohort)

	LSS	VAS	OKS	NRS Pain in rest	NRS Pain during activity	Satisfaction
Variables	Coefficient (95% CI)	Coefficient (95% CI)	Coefficient (95% CI)	Coefficient (95% CI)	Coefficient (95% CI)	Coefficient (95% CI)
Intercept	81.46 (79.78, 83.13)	74.85 (72.97, 76.72)	79.31 (77.53, 81.09)	83.28 (80.87, 85.69)	73.02 (70.22, 75.82)	76.52 (74.06, 78.97)
SES						
Q1, least deprived	*reference*					
Q2	0.00 (-0.60, 0.59)	0.11 (-0.56, 0.78)	-0.34 (-0.98, 0.30)	-0.91 (-1.76, -0.05)	-0.91 (-1.91, 0.09)	0.20 (-0.70, 1.10)
Q3	-0.35 (-0.92, 0.21)	-0.34 (-0.98, 0.30)	-0.27 (-0.89, 0.34)	-0.78 (-1.60, 0.03)	-0.50 (-1.46, 0.45)	0.71 (-0.14, 1.57)
Q4	-0.66 (-1.21, -0.12)	-0.62 (-1.24, 0.00)	-0.67 (-1.27, -0.08)	-1.44 (-2.23, -0.65)	-1.58 (-2.50, -0.66)	0.25 (-0.58, 1.08)
Q5, most deprived	-1.30 (-1.87, -0.73)	-0.53 (-1.17, 0.12)	-1.84 (-2.46, -1.22)	-2.67 (-3.49, -1.86)	-2.82 (-3.77, -1.87)	-0.03 (-0.89, 0.83)
Pre-op score in 3 categories						
lowest	*reference*					
middle	7.03 (6.52, 7.54)	5.69 (5.22, 6.17)	6.74 (6.28, 7.19)	5.25 (4.60, 5.89)	3.88 (3.13, 4.62)	
highest	11.79 (11.34, 12.25)	11.96 (11.49, 12.43)	11.08 (10.61, 11.55)	11.44 (10.77, 12.11)	7.79 (7.09, 8.49)	
Male (vs. female)	2.06 (1.71, 2.40)	1.60 (1.20, 1.99)	2.63 (2.25, 3.01)	1.25 (0.75, 1.75)	2.76 (2.17, 3.34)	1.64 (1.12, 2.16)
Age						
<50	*reference*					
50-69	0.58 (-0.90, 2.05)	0.70 (-0.97, 2.36)	0.93 (-0.65, 2.50)	1.37 (-0.74, 3.48)	3.96 (1.49, 6.42)	1.32 (-0.87, 3.51)
>70	-0.07 (-1.55, 1.42)	-1.05 (-2.73, 0.63)	0.78 (-0.81, 2.37)	1.43 (-0.70, 3.56)	5.85 (3.37, 8.34)	1.80 (-0.40, 4.00)
BMI						
<25	*reference*					
25-30	-0.83 (-1.29, -0.38)	-0.65 (-1.17, -0.14)	-1.13 (-1.62, -0.63)	-1.46 (-2.11, -0.80)	-1.33 (-2.10, -0.57)	-0.43 (-1.11, 0.25)
>30	-2.18 (-2.66, -1.70)	-1.64 (-2.19, -1.10)	-2.73 (-3.26, -2.21)	-2.27 (-2.96, -1.58)	-2.42 (-3.23, -1.61)	-0.33 (-1.05, 0.39)
ASA						
I	*reference*					
II	-1.61 (-2.14, -1.09)	-2.28 (-2.88, -1.69)	-1.51 (-2.07, -0.95)	-2.13 (-2.89, -1.37)	-2.22 (-3.10, -1.33)	-1.85 (-2.63, -1.06)
III-IV	-5.34 (-5.99, -4.69)	-6.74 (-7.48, -6.01)	-5.19 (-5.89, -4.50)	-4.50 (-5.42, -3.58)	-5.18 (-6.26, -4.10)	-4.57 (-5.53, -3.61)

Table 7: Continued

	LSS	VAS	OKS	NRS Pain in rest	NRS Pain during activity	Satisfaction
Variables	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**
Charnley						
A	*reference*					
B1	-0.08 (-0.47, 0.31)	-0.26 (-0.70, 0.18)	-0.24 (-0.66, 0.18)	0.02 (-0.54, 0.58)	0.35 (-0.30, 1.00)	0.02 (-0.56, 0.60)
B2	0.13 (-0.32, 0.58)	0.35 (-0.16, 0.86)	-0.33 (-0.81, 0.16)	0.56 (-0.09, 1.21)	0.89 (0.13, 1.65)	1.95 (1.27, 2.62)
C	-1.21 (-2.14, -0.28)	-1.28 (-2.33, -0.23)	-0.71 (-1.70, 0.28)	0.48 (-0.86, 1.83)	0.62 (-0.95, 2.19)	1.51 (0.13, 2.88)
Hospital						
Private Hospital	*reference*					
General Hospital	-2.31 (-2.89, -1.74)	-1.46 (-2.12, -0.81)	-2.02 (-2.65, -1.40)	-1.92 (-2.81, -1.04)	-1.92 (-2.95, -0.89)	2.57 (1.68, 3.45)
University Medical Center	-4.30 (-5.90, -2.71)	-1.21 (-3.05, 0.63)	-3.37 (-5.12, -1.62)	-3.29 (-5.67, -0.90)	-2.19 (-4.98, 0.59)	2.13 (-0.25, 4.51)
R-squared	0.119	0.113	0.124	0.053	0.033	0.006

Table 8: Demographics and outcomes of THA patients (EQ-5D-5L cohort)

	Q1 (least depr.)	**Q2**	**Q3**	**Q4**	**Q5 (most depr.)**	**p-value**
Total	2628	2801	3050	3250	2659	
SES Z-score, median [range]	1.3 [0.9, 2.8]	0.6 [0.4, 0.9]	0.2 [-0.1, 0.4]	-0.4 [-0.8, -0.1]	-1.37 [-6.13, -0.78]	
Demographics						
Age, median [IQR]	70.0 [63.0, 75.0]	70.0 [63.0, 75.0]	70.0 [63.0, 75.0]	70.0 [64.0, 76.0]	71.0 [63.5, 76.0]	<0.001
<50	51 (2)	80 (3)	67 (2)	52 (2)	58 (2)	0.012
50-69	1244 (47)	1310 (47)	1401 (46)	1471 (45)	1151 (43)	
>70	1333 (51)	1411 (50)	1582 (52)	1727 (53)	1450 (55)	
BMI, median [IQR]	26.0 [23.7, 29.0]	26.4 [24.0, 29.4]	26.9 [24.2, 30.0]	26.9 [24.2, 29.8]	27.2 [24.7, 30.3]	<0.001
<25	1143 (43)	1126 (40)	1099 (36)	1172 (36)	880 (33)	<0.001
25-30	1066 (41)	1134 (40)	1296 (42)	1386 (43)	1134 (43)	
>30	419 (16)	541 (19)	655 (21)	692 (21)	645 (24)	
Male	1688 (64)	1730 (62)	1946 (64)	2085 (64)	1779 (67)	0.044
ASA						<0.001
I	494 (19)	505 (18)	561 (18)	484 (15)	332 (12)	
II	1687 (64)	1760 (63)	1876 (62)	2022 (62)	1712 (64)	
III-IV	447 (17)	536 (19)	613 (20)	744 (23)	615 (23)	
Charnley						0.03
A	1169 (44)	1205 (43)	1298 (43)	1344 (41)	1139 (43)	
B1	849 (32)	878 (31)	973 (32)	1064 (33)	843 (32)	
B2	562 (21)	626 (22)	677 (22)	717 (22)	598 (22)	
C	48 (2)	92 (3)	102 (3)	125 (4)	79 (3)	
Smoking (yes)	172 (7)	212 (8)	210 (7)	252 (8)	240 (9)	0.001
Previous surgery of the joint (yes)	31 (1)	26 (1)	38 (1)	28 (1)	26 (1)	0.564
Year of surgery						N/A
2014	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	

Table 8: Continued

	Q1 (least depr.)	Q2	Q3	Q4	Q5 (most depr.)	p-value
2015	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	
2016	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	
2017	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	
2018	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	
2019	1 (0)	0 (0)	0 (0)	0 (0)	0 (0)	
2020	232 (9)	221 (8)	212 (7)	205 (6)	191 (7)	
2021	1612 (61)	1737 (62)	1923 (63)	1969 (61)	1546 (58)	
2022	783 (30)	843 (30)	915 (30)	1076 (33)	922 (35)	
Type of hospital						<0.001
General Hospital	1733 (66)	2036 (73)	2323 (76)	2613 (80)	2160 (81)	
Private Hospital	889 (34)	758 (27)	720 (24)	620 (19)	483 (18)	
University Medical Center	6 (0)	7 (0)	7 (0)	17 (1)	16 (1)	
Fixation						<0.001
Cemented	322 (12)	340 (12)	499 (16)	526 (16)	602 (23)	
Cementless	2133 (81)	2235 (80)	2303 (76)	2434 (75)	1751 (66)	
Hybrid	173 (7)	226 (8)	248 (8)	290 (9)	306 (12)	
Approach						<0.001
Anterior	1729 (66)	1598 (57)	1482 (49)	1481 (46)	1172 (44)	
Anterolateral	53 (2)	70 (2)	70 (2)	60 (2)	30 (1)	
Other	24 (1)	30 (1)	99 (3)	60 (2)	55 (2)	
Posterolateral	769 (29)	1058 (38)	1352 (44)	1580 (49)	1332 (50)	
Straight lateral	53 (2)	45 (2)	47 (2)	69 (2)	70 (3)	
Contralateral procedure (yes)	339 (13)	399 (14)	423 (14)	453 (14)	374 (14)	0.23

Table 8: Continued

	Q1 (least depr.)	Q2	Q3	Q4	Q5 (most depr.)	p-value
Outcomes						
Preoperative outcomes, median [IQR]						
EQ-5D-5L LSS	60.0 [50.0, 70.0]	60.0 [50.0, 70.0]	60.0 [45.0, 70.0]	60.0 [45.0, 70.0]	55.0 [45.0, 65.0]	
OHS	50.0 [37.5, 62.5]	47.9 [35.4, 60.4]	47.9 [35.4, 60.4]	47.9 [33.3, 60.4]	43.8 [31.2, 58.3]	
EQ VAS	70.0 [51.0, 80.0]	70.0 [51.0, 80.0]	70.0 [50.0, 80.0]	70.0 [50.0, 80.0]	67.0 [50.0, 80.0]	
NRS Pain in rest	50.0 [30.0, 70.0]	50.0 [30.0, 70.0]	50.0 [30.0, 70.0]	40.0 [30.0, 60.0]	40.0 [20.0, 60.0]	
NRS Pain during activity	20.0 [20.0, 40.0]	20.0 [10.0, 40.0]	20.0 [10.0, 30.0]	20.0 [10.0, 30.0]	20.0 [10.0, 30.0]	
12-month follow-up outcomes, median [IQR]						
EQ-5D-5L LSS	95.0 [85.0, 100.0]	95.0 [85.0, 100.0]	95.0 [85.0, 100.0]	95.0 [80.0, 100.0]	90.0 [80.0, 100.0]	
OHS	93.8 [85.4, 100.0]	93.8 [85.4, 100.0]	93.8 [83.3, 97.9]	93.8 [83.3, 97.9]	91.7 [79.2, 97.9]	
EQ VAS	81.0 [71.0, 90.0]	81.0 [71.0, 90.0]	80.0 [70.0, 90.0]	80.0 [70.0, 90.0]	80.0 [70.0, 90.0]	
NRS Pain in rest	100.0 [90.0, 100.0]	100.0 [90.0, 100.0]	100.0 [90.0, 100.0]	100.0 [90.0, 100.0]	100.0 [90.0, 100.0]	
NRS Pain during activity	100.0 [80.0, 100.0]	100.0 [80.0, 100.0]	100.0 [80.0, 100.0]	100.0 [80.0, 100.0]	100.0 [70.0, 100.0]	
Preoperative outcomes, n (%) ceiling						
EQ-5D-5L LSS	14 (0.5)	5 (0.2)	8 (0.3)	10 (0.3)	8 (0.3)	
OHS	2 (0.1)	2 (0.1)	1 (0.0)	0 (0.0)	1 (0.0)	
EQ VAS	47 (1.8)	55 (2.0)	64 (2.1)	61 (1.9)	36 (1.4)	
NRS Pain in rest	142 (5.4)	153 (5.5)	131 (4.3)	148 (4.6)	114 (4.3)	
NRS Pain during activity	18 (0.7)	17 (0.6)	13 (0.4)	25 (0.8)	17 (0.6)	
12-month follow-up outcomes, n (%) ceiling						
EQ-5D-5L LSS	1155 (43.9)	1222 (43.6)	1256 (41.2)	1301 (40.0)	995 (37.4)	
OHS	700 (27.1)	745 (27.2)	735 (24.6)	721 (22.7)	571 (22.0)	
EQ VAS	162 (6.2)	206 (7.4)	202 (6.7)	220 (6.9)	159 (6.1)	
NRS Pain in rest	1815 (69.2)	1938 (69.4)	2050 (67.5)	2141 (66.0)	1737 (65.5)	
NRS Pain during activity	1410 (53.8)	1529 (54.8)	1575 (51.8)	1656 (51.1)	1327 (50.1)	

Table 9: Demographics and outcomes of TKA patients (EQ-5D-5L cohort)

	Q1 (least depr.)	Q2	Q3	Q4	Q5 (most depr.)	p-value
Total	1312	1665	1971	2314	1929	
SES Z-score, median [range]	1.2 [0.9, 2.8]	0.6 [0.4, 0.9]	0.2 [-0.1, 0.4]	-0.36 [-0.78, -0.07]	-1.4 [-7.8, -0.8]	
Demographics						
Age, median [IQR]	70.0 [65.0, 75.0]	70.0 [63.0, 75.0]	69.0 [63.0, 75.0]	70.0 [63.0, 75.0]	70.0 [63.0, 75.0]	<0.001
<50	9 (1)	7 (0)	15 (1)	17 (1)	20 (1)	0.042
50-69	586 (45)	815 (49)	975 (49)	1120 (48)	937 (49)	
>70	717 (55)	843 (51)	981 (50)	1177 (51)	972 (50)	
BMI, median [IQR]	27.8 [25.2, 31.1]	28.3 [25.7, 31.6]	28.7 [26.0, 32.2]	29.0 [26.0, 32.3]	29.3 [26.4, 33.0]	<0.001
<25	351 (27)	381 (23)	405 (21)	488 (21)	345 (18)	<0.001
25-30	576 (44)	724 (43)	825 (42)	950 (41)	804 (42)	
>30	385 (29)	560 (34)	741 (38)	876 (38)	780 (40)	
Male	811 (62)	1027 (62)	1187 (60)	1370 (59)	1219 (63)	0.447
ASA						0.001
I	164 (12)	208 (12)	246 (12)	258 (11)	175 (9)	
II	850 (65)	1092 (66)	1251 (63)	1530 (66)	1237 (64)	
III-IV	298 (23)	365 (22)	474 (24)	526 (23)	517 (27)	
Charnley						0.014
A	541 (41)	690 (41)	781 (40)	929 (40)	805 (42)	
B1	399 (30)	561 (34)	661 (34)	802 (35)	662 (34)	
B2	319 (24)	351 (21)	457 (23)	500 (22)	407 (21)	
C	53 (4)	63 (4)	72 (4)	83 (4)	55 (3)	
Smoking (yes)	77 (6)	85 (5)	112 (6)	160 (7)	158 (8)	0.015
Previous surgery of the joint (yes)	270 (21)	391 (23)	510 (26)	566 (24)	482 (25)	0.004
Year of surgery						NaN
2014	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	

Table 9: Continued

	Q1 (least depr.)	Q2	Q3	Q4	Q5 (most depr.)	p-value
2015	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	
2016	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	
2017	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	
2018	1 (0)	0 (0)	0 (0)	0 (0)	0 (0)	
2019	0 (0)	2 (0)	0 (0)	0 (0)	0 (0)	
2020	108 (8)	126 (8)	148 (8)	149 (6)	108 (6)	
2021	829 (63)	1020 (61)	1259 (64)	1455 (63)	1199 (62)	
2022	374 (29)	517 (31)	564 (29)	710 (31)	622 (32)	
Type of hospital						0.033
General Hospital	1035 (79)	1333 (80)	1551 (79)	1879 (81)	1567 (81)	
Private Hospital	274 (21)	324 (19)	412 (21)	424 (18)	349 (18)	
University Medical Center	3 (0)	8 (0)	8 (0)	11 (0)	13 (1)	
Fixation						0.002
Cemented	1214 (93)	1543 (93)	1816 (92)	2107 (91)	1712 (89)	
Cementless	62 (5)	89 (5)	118 (6)	164 (7)	142 (7)	
Hybrid	36 (3)	33 (2)	37 (2)	43 (2)	75 (4)	
Approach						0.184
Lateral parapatellar	8 (1)	6 (0)	17 (1)	13 (1)	18 (1)	
Medial parapatellar	1273 (97)	1627 (98)	1906 (97)	2253 (97)	1881 (98)	
Other	6 (0)	4 (0)	4 (0)	6 (0)	3 (0)	
Vastus (mid/sub)	25 (2)	28 (2)	44 (2)	42 (2)	27 (1)	
Contralateral procedure (yes)	225 (17)	245 (15)	338 (17)	360 (16)	291 (15)	0.127
Outcomes						
Preoperative outcomes, median [IQR]						
EQ-5D-5L LSS	65.0 [55.0, 70.0]	65.0 [50.0, 70.0]	65.0 [55.0, 70.0]	65.0 [55.0, 70.0]	60.0 [50.0, 70.0]	

Table 9: Continued

	Q1 (least depr.)	Q2	Q3	Q4	Q5 (most depr.)	p-value
OHS	50.0 [39.6, 60.4]	47.9 [39.6, 60.4]	47.9 [37.5, 60.4]	47.9 [37.5, 58.3]	45.8 [33.3, 56.2]	
EQ VAS	70.0 [58.0, 80.0]	70.0 [59.0, 81.0]	70.0 [60.0, 80.0]	70.0 [59.0, 81.0]	70.0 [55.0, 80.0]	
NRS Pain in rest	50.0 [30.0, 70.0]	50.0 [30.0, 70.0]	50.0 [30.0, 70.0]	40.0 [30.0, 70.0]	40.0 [30.0, 60.0]	
NRS Pain during activity	20.0 [10.0, 40.0]	20.0 [20.0, 40.0]	20.0 [10.0, 30.0]	20.0 [10.0, 30.0]	20.0 [10.0, 30.0]	
12-month follow-up outcomes, median [IQR]						
EQ-5D-5L LSS	90.0 [80.0, 100.0]	90.0 [80.0, 100.0]	90.0 [80.0, 100.0]	90.0 [80.0, 100.0]	90.0 [80.0, 100.0]	
OHS	87.5 [75.0, 93.8]	87.5 [75.0, 93.8]	87.5 [75.0, 93.8]	85.4 [70.8, 93.8]	85.4 [68.8, 93.8]	
EQ VAS	80.0 [70.0, 90.0]	80.0 [70.0, 90.0]	80.0 [70.0, 90.0]	80.0 [70.0, 90.0]	80.0 [70.0, 90.0]	
NRS Pain in rest	100.0 [80.0, 100.0]	100.0 [80.0, 100.0]	90.0 [80.0, 100.0]	90.0 [80.0, 100.0]	90.0 [70.0, 100.0]	
NRS Pain during activity	90.0 [60.0, 100.0]	90.0 [70.0, 100.0]	80.0 [60.0, 100.0]	80.0 [60.0, 100.0]	80.0 [60.0, 100.0]	
NRS Satisfaction	80.0 [70.0, 90.0]	80.0 [70.0, 90.0]	80.0 [70.0, 90.0]	80.0 [70.0, 90.0]	80.0 [70.0, 90.0]	
Preoperative outcomes, n (%) ceiling						
EQ-5D-5L LSS	3 (0.2)	9 (0.5)	3 (0.2)	10 (0.4)	12 (0.6)	
OHS	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)	
EQ VAS	24 (1.8)	34 (2.0)	34 (1.7)	51 (2.2)	52 (2.7)	
NRS Pain in rest	96 (7.4)	103 (6.2)	104 (5.3)	99 (4.3)	95 (5.0)	
NRS Pain during activity	5 (0.4)	13 (0.8)	6 (0.3)	10 (0.4)	12 (0.6)	
12-month follow-up outcomes, n (%) ceiling						
EQ-5D-5L LSS	407 (31.0)	523 (31.4)	561 (28.5)	659 (28.5)	559 (29.0)	
OHS	78 (6.1)	101 (6.3)	111 (5.9)	121 (5.4)	101 (5.4)	
EQ VAS	68 (5.2)	77 (4.7)	84 (4.3)	103 (4.5)	104 (5.5)	
NRS Pain in rest	690 (52.8)	868 (52.6)	977 (49.6)	1114 (48.4)	945 (49.2)	
NRS Pain during activity	439 (33.7)	544 (32.9)	605 (30.8)	667 (29.0)	620 (32.3)	
NRS Satisfaction	294 (23.5)	399 (24.6)	434 (22.3)	510 (22.5)	467 (24.6)	

Table 10: Association between socioeconomic status and preoperative health status of THA patients (EQ-5D-5L cohort)

	LSS	VAS	OHS	NRS Pain in rest	NRS Pain during activity
Variables	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**
Intercept	61.34 (59.49, 63.19)	64.68 (62.16, 67.19)	50.16 (48.53, 51.80)	51.78 (48.71, 54.86)	29.42 (27.09, 31.75)
SES					
Q1, least deprived	*reference*				
Q2	-1.19 (-1.98, -0.39)	0.95 (-0.13, 2.03)	-0.74 (-1.33, -0.14)	-0.15 (-1.47, 1.16)	-1.19 (-2.19, -0.19)
Q3	-2.13 (-2.91, -1.35)	-0.04 (-1.10, 1.02)	-1.50 (-2.06, -0.93)	-1.29 (-2.59, 0.00)	-1.47 (-2.45, -0.49)
Q4	-1.78 (-2.55, -1.01)	-0.06 (-1.11, 0.98)	-1.77 (-2.32, -1.22)	-2.11 (-3.39, -0.83)	-2.26 (-3.23, -1.29)
Q5, most deprived	-2.32 (-3.13, -1.52)	-1.50 (-2.60, -0.40)	-3.03 (-3.60, -2.45)	-4.18 (-5.53, -2.84)	-2.56 (-3.58, -1.54)
Male (vs. female)	3.07 (2.55, 3.58)	3.64 (2.94, 4.34)	6.48 (6.13, 6.83)	4.74 (3.88, 5.59)	3.86 (3.21, 4.50)
Age					
<50	*reference*				
50-69	4.20 (2.50, 5.90)	3.87 (1.55, 6.19)	3.76 (2.29, 5.23)	4.31 (1.48, 7.14)	3.52 (1.37, 5.66)
>70	4.60 (2.89, 6.32)	6.18 (3.84, 8.51)	4.86 (3.38, 6.34)	6.94 (4.09, 9.79)	4.73 (2.57, 6.89)
BMI					
<25	*reference*				
25-30	-1.77 (-2.32, -1.22)	-1.26 (-2.01, -0.50)	-1.78 (-2.23, -1.32)	-1.73 (-2.65, -0.81)	-2.25 (-2.95, -1.56)
>30	-5.17 (-5.87, -4.48)	-4.11 (-5.06, -3.16)	-5.26 (-5.75, -4.78)	-5.16 (-6.32, -4.00)	-5.44 (-6.32, -4.56)
ASA					
I	*reference*				
II	-2.65 (-3.35, -1.95)	-3.73 (-4.69, -2.78)	-2.59 (-3.11, -2.07)	-2.32 (-3.49, -1.15)	-1.69 (-2.58, -0.81)
III-IV	-6.65 (-7.54, -5.76)	-9.31 (-10.53, -8.09)	-7.10 (-7.74, -6.46)	-4.95 (-6.44, -3.47)	-4.65 (-5.78, -3.53)
Charnley					
A	*reference*				
B1	0.12 (-0.45, 0.69)	0.35 (-0.43, 1.12)	-0.23 (-0.62, 0.16)	0.80 (-0.14, 1.75)	0.41 (-0.30, 1.13)
B2	1.66 (1.02, 2.30)	0.40 (-0.47, 1.27)	1.45 (1.00, 1.90)	0.31 (-0.76, 1.37)	1.74 (0.94, 2.55)
C	-2.58 (-4.01, -1.14)	-1.06 (-3.02, 0.91)	-3.19 (-4.10, -2.27)	-3.97 (-6.36, -1.58)	-1.32 (-3.13, 0.49)

Table 10: Continued

	LSS	VAS	OHS	NRS Pain in rest	NRS Pain during activity
Variables	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**
Hospital					
Private Hospital	*reference*				
General Hospital	-3.51 (-4.12, -2.91)	-0.73 (-1.55, 0.10)	-0.47 (-1.05, 0.11)	-7.75 (-8.76, -6.74)	-3.99 (-4.75, -3.22)
University Medical Center	-0.34 (-4.39, 3.71)	3.56 (-1.94, 9.06)	-7.45 (-9.06, -5.83)	-6.84 (-13.77, 0.08)	-4.98 (-10.28, 0.31)
R-squared	0.074	0.039	0.105	0.049	0.046

4

Table 11: Association between socioeconomic status and preoperative health status of TKA patients (EQ-5D-5L cohort)

	LSS	VAS	OKS	NRS Pain in rest	NRS Pain during activity
Variables	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**
Intercept	64.81 (61.30, 68.33)	67.20 (62.52, 71.88)	52.68 (48.78, 56.58)	47.30 (40.99, 53.62)	29.05 (24.33, 33.77)
SES					
Q1, least deprived	*reference*				
Q2	-0.66 (-1.67, 0.35)	0.70 (-0.65, 2.05)	0.16 (-0.95, 1.27)	0.72 (-1.10, 2.54)	-0.06 (-1.42, 1.30)
Q3	-0.53 (-1.50, 0.45)	0.83 (-0.47, 2.13)	-0.16 (-1.23, 0.90)	-1.70 (-3.46, 0.06)	-1.64 (-2.95, -0.33)
Q4	-0.60 (-1.54, 0.35)	0.66 (-0.60, 1.92)	-0.58 (-1.61, 0.46)	-2.77 (-4.48, -1.06)	-1.41 (-2.69, -0.14)
Q5, most deprived	-0.78 (-1.76, 0.20)	0.63 (-0.68, 1.95)	-2.50 (-3.57, -1.42)	-4.90 (-6.66, -3.13)	-1.97 (-3.29, -0.65)
Male (vs. female)	2.74 (2.15, 3.33)	2.88 (2.09, 3.67)	6.10 (5.45, 6.75)	5.07 (4.00, 6.14)	3.55 (2.76, 4.35)
Age					
<50	*reference*				
50-69	3.19 (-0.15, 6.52)	5.29 (0.84, 9.74)	1.99 (-1.72, 5.70)	8.16 (2.16, 14.17)	2.08 (-2.40, 6.57)
>70	3.87 (0.52, 7.22)	7.13 (2.66, 11.59)	2.40 (-1.33, 6.12)	12.08 (6.05, 18.10)	4.69 (0.19, 9.19)
BMI					
<25	*reference*				
25-30	-1.80 (-2.56, -1.04)	-2.02 (-3.04, -1.01)	-3.13 (-3.96, -2.29)	-3.09 (-4.46, -1.72)	-2.51 (-3.53, -1.49)
>30	-3.64 (-4.46, -2.82)	-3.53 (-4.62, -2.43)	-5.89 (-6.79, -4.99)	-4.69 (-6.16, -3.22)	-3.96 (-5.06, -2.86)
ASA					
I	*reference*				
II	-1.99 (-2.94, -1.05)	-4.25 (-5.52, -2.99)	-2.77 (-3.81, -1.73)	-1.66 (-3.37, 0.05)	-1.39 (-2.67, -0.12)
III-IV	-5.55 (-6.68, -4.43)	-9.85 (-11.35, -8.34)	-6.79 (-8.02, -5.55)	-4.06 (-6.09, -2.03)	-3.07 (-4.59, -1.55)
Charnley					
A	*reference*				
B1	-0.54 (-1.21, 0.13)	0.30 (-0.59, 1.19)	-0.29 (-1.02, 0.44)	0.86 (-0.34, 2.07)	0.35 (-0.55, 1.25)
B2	1.18 (0.42, 1.94)	1.17 (0.15, 2.18)	1.37 (0.54, 2.21)	0.95 (-0.42, 2.32)	0.84 (-0.18, 1.86)
C	-1.59 (-3.18, -0.01)	-0.10 (-2.23, 2.02)	-0.67 (-2.41, 1.06)	1.06 (-1.80, 3.92)	1.74 (-0.40, 3.88)

Table 11: Continued

	LSS	VAS	OKS	NRS Pain in rest	NRS Pain during activity
Variables	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**
Hospital					
Private Hospital	*reference*				
General Hospital	-2.39 (-3.16, -1.62)	0.24 (-0.78, 1.27)	-1.93 (-2.77, -1.08)	-5.55 (-6.93, -4.17)	-3.01 (-4.04, -1.97)
University Medical Center	1.08 (-3.14, 5.30)	4.11 (-1.58, 9.79)	-8.95 (-13.61, -4.29)	-0.35 (-8.21, 7.52)	-1.87 (-7.67, 3.94)
R-squared	0.051	0.039	0.098	0.041	0.031

4

Table 12: Association between socioeconomic status and 12-month follow-up health status of THA patients (EQ-5D-5L cohort)

	LSS	VAS	OHS	NRS Pain in rest	NRS Pain during activity
Variables	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**
Intercept	90.22 (88.49, 91.95)	80.38 (78.17, 82.58)	92.15 (90.28, 94.02)	88.59 (86.29, 90.89)	85.58 (82.76, 88.40)
SES					
Q1, least deprived	*reference*				
Q2	0.54 (-0.18, 1.26)	0.54 (-0.39, 1.46)	0.14 (-0.64, 0.93)	0.46 (-0.49, 1.41)	0.15 (-1.03, 1.32)
Q3	0.24 (-0.47, 0.94)	0.41 (-0.50, 1.32)	-0.24 (-1.01, 0.53)	0.10 (-0.83, 1.03)	-0.13 (-1.29, 1.03)
Q4	0.16 (-0.54, 0.86)	-0.17 (-1.07, 0.73)	-0.45 (-1.21, 0.32)	0.31 (-0.61, 1.23)	0.15 (-1.00, 1.29)
Q5, most deprived	-0.47 (-1.21, 0.26)	0.03 (-0.92, 0.97)	-1.16 (-1.97, -0.36)	-0.31 (-1.28, 0.66)	-0.97 (-2.17, 0.24)
Pre-op score in 3 categories					
lowest	*reference*				
middle	4.63 (4.05, 5.21)	3.91 (3.21, 4.61)	3.26 (2.66, 3.86)	3.88 (3.10, 4.66)	1.68 (0.73, 2.64)
highest	7.61 (7.03, 8.18)	9.41 (8.69, 10.12)	5.83 (5.22, 6.44)	7.84 (7.02, 8.66)	4.44 (3.53, 5.35)
Male (vs. female)	1.44 (0.97, 1.90)	1.25 (0.65, 1.85)	1.69 (1.18, 2.20)	1.25 (0.63, 1.86)	2.08 (1.31, 2.84)
Age					
<50	*reference*				
50-69	0.15 (-1.40, 1.69)	-0.07 (-2.06, 1.92)	0.04 (-1.64, 1.72)	1.70 (-0.34, 3.74)	2.41 (-0.12, 4.94)
>70	-1.28 (-2.84, 0.28)	-1.76 (-3.77, 0.24)	-1.45 (-3.15, 0.24)	1.00 (-1.05, 3.06)	3.24 (0.69, 5.79)
BMI					
<25	*reference*				
25-30	-1.43 (-1.93, -0.92)	-1.50 (-2.15, -0.85)	-1.51 (-2.06, -0.96)	-1.25 (-1.91, -0.58)	-1.71 (-2.53, -0.88)
>30	-3.28 (-3.91, -2.64)	-3.19 (-4.00, -2.37)	-3.40 (-4.10, -2.70)	-2.04 (-2.88, -1.21)	-2.96 (-4.00, -1.92)
ASA					
I	*reference*				
II	-1.14 (-1.78, -0.50)	-2.67 (-3.49, -1.85)	-1.32 (-2.02, -0.63)	-0.84 (-1.69, 0.00)	-1.28 (-2.33, -0.24)
III-IV	-3.69 (-4.50, -2.87)	-7.21 (-8.26, -6.16)	-3.81 (-4.70, -2.92)	-1.93 (-3.01, -0.86)	-2.32 (-3.65, -0.99)

Table 12: Continued

	LSS	VAS	OHS	NRS Pain in rest	NRS Pain during activity
Variables	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**
Charnley					
A	*reference*				
B1	-0.95 (-1.47, -0.44)	-0.85 (-1.52, -0.18)	-1.44 (-2.01, -0.88)	-0.53 (-1.21, 0.15)	-0.64 (-1.48, 0.21)
B2	-1.06 (-1.64, -0.48)	-0.72 (-1.46, 0.03)	-0.96 (-1.60, -0.33)	0.40 (-0.36, 1.17)	0.37 (-0.58, 1.32)
C	-3.33 (-4.63, -2.02)	-3.20 (-4.89, -1.51)	-2.83 (-4.26, -1.40)	-1.13 (-2.85, 0.60)	-3.11 (-5.25, -0.97)
Hospital					
Private Hospital	*reference*				
General Hospital	-3.28 (-3.83, -2.73)	-2.17 (-2.88, -1.46)	-3.80 (-4.40, -3.20)	-2.83 (-3.56, -2.10)	-4.46 (-5.36, -3.56)
University Medical Center	-6.32 (-10.00, -2.64)	-5.90 (-10.61, -1.20)	-5.09 (-9.29, -0.88)	-1.82 (-6.80, 3.16)	-5.67 (-11.91, 0.56)
R-squared	0.106	0.094	0.086	0.043	0.027

Table 13: Association between socioeconomic status and 12-month follow-up health status of TKA patients (EQ-5D-5L cohort)

	LSS	VAS	OKS	NRS Pain in rest	NRS Pain during activity	NRS Satisfaction
Variables	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**
Intercept	86.84 (83.27, 90.40)	81.01 (76.61, 85.42)	81.68 (77.57, 85.78)	84.98 (79.56, 90.39)	74.87 (68.59, 81.15)	82.42 (77.31, 87.53)
SES						
Q1, least deprived	*reference*					
Q2	-0.32 (-1.33, 0.69)	-0.28 (-1.54, 0.98)	-0.12 (-1.28, 1.03)	-0.59 (-2.14, 0.95)	-0.50 (-2.30, 1.30)	0.03 (-1.46, 1.53)
Q3	-1.02 (-1.99, -0.04)	-0.96 (-2.18, 0.26)	-0.25 (-1.36, 0.87)	-1.12 (-2.61, 0.37)	-1.39 (-3.13, 0.35)	-0.06 (-1.50, 1.38)
Q4	-1.38 (-2.33, -0.44)	-1.19 (-2.37, 0.00)	-1.06 (-2.14, 0.02)	-1.60 (-3.04, -0.15)	-2.24 (-3.92, -0.55)	-0.22 (-1.62, 1.18)
Q5, most deprived	-1.72 (-2.71, -0.74)	-0.71 (-1.95, 0.52)	-1.91 (-3.03, -0.79)	-1.81 (-3.31, -0.30)	-1.92 (-3.67, -0.17)	0.04 (-1.41, 1.48)
Pre-op score in 3 categories						
lowest	*reference*					
middle	6.20 (5.45, 6.95)	4.80 (3.90, 5.70)	5.56 (4.75, 6.38)	5.56 (4.37, 6.74)	2.35 (1.02, 3.69)	
highest	9.46 (8.71, 10.22)	10.78 (9.90, 11.67)	9.80 (8.96, 10.65)	12.51 (11.28, 13.75)	7.99 (6.71, 9.26)	
Male (vs. female)	1.36 (0.77, 1.96)	1.17 (0.43, 1.92)	2.90 (2.22, 3.59)	1.25 (0.34, 2.16)	3.13 (2.07, 4.19)	2.19 (1.32, 3.06)
Age						
<50	*reference*					
50-69	-0.85 (-4.19, 2.50)	-1.91 (-6.08, 2.26)	-0.06 (-3.94, 3.82)	-1.48 (-6.58, 3.61)	3.19 (-2.74, 9.12)	1.51 (-3.35, 6.36)
>70	-1.03 (-4.38, 2.32)	-2.89 (-7.07, 1.29)	-1.03 (-4.91, 2.86)	-1.87 (-6.98, 3.25)	4.28 (-1.67, 10.23)	0.33 (-4.53, 5.20)
BMI						
<25	*reference*					
25-30	-0.19 (-0.95, 0.57)	-0.69 (-1.64, 0.27)	-0.61 (-1.48, 0.26)	0.00 (-1.16, 1.17)	0.01 (-1.35, 1.37)	-0.64 (-1.76, 0.47)
>30	-0.93 (-1.75, -0.11)	-1.50 (-2.53, -0.47)	-2.06 (-3.00, -1.11)	-0.47 (-1.72, 0.79)	-0.92 (-2.38, 0.54)	-0.55 (-1.75, 0.65)
ASA						
I	*reference*					
II	-1.82 (-2.77, -0.87)	-3.51 (-4.70, -2.32)	-1.55 (-2.64, -0.47)	-1.78 (-3.23, -0.33)	-2.09 (-3.78, -0.40)	-1.99 (-3.38, -0.60)
III-IV	-4.58 (-5.71, -3.45)	-7.95 (-9.37, -6.53)	-4.44 (-5.74, -3.15)	-4.12 (-5.84, -2.40)	-4.16 (-6.17, -2.15)	-3.34 (-4.99, -1.69)

Table 13: Continued

	LSS	VAS	OKS	NRS Pain in rest	NRS Pain during activity	NRS Satisfaction
Variables	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**	**Coefficient (95% CI)**
Charnley						
A	*reference*					
B1	0.48 (-0.19, 1.14)	-0.97 (-1.81, -0.13)	0.07 (-0.69, 0.84)	0.39 (-0.63, 1.41)	0.64 (-0.55, 1.83)	0.51 (-0.47, 1.49)
B2	1.09 (0.33, 1.85)	-0.11 (-1.06, 0.84)	0.28 (-0.59, 1.15)	0.91 (-0.25, 2.08)	1.11 (-0.25, 2.46)	2.43 (1.32, 3.55)
C	-2.12 (-3.71, -0.54)	-2.25 (-4.25, -0.26)	-2.00 (-3.82, -0.19)	0.09 (-2.33, 2.52)	-0.29 (-3.12, 2.54)	0.44 (-1.89, 2.77)
Hospital						
Private Hospital	*reference*					
General Hospital	-3.06 (-3.83, -2.29)	-2.42 (-3.38, -1.46)	-3.02 (-3.89, -2.15)	-3.67 (-4.84, -2.49)	-4.75 (-6.11, -3.38)	-2.97 (-4.09, -1.85)
University Medical Center	-6.53 (-10.75, -2.30)	-6.95 (-12.34, -1.57)	-7.95 (-12.80, -3.09)	-4.14 (-10.88, 2.60)	-8.06 (-15.82, -0.30)	-7.34 (-13.61, -1.07)
R-squared	0.104	0.104	0.116	0.063	0.039	0.012

4

Table 14: Percentage of inequality explained by each EQ-5D-5L and OHS/OKS dimension (EQ-5D-5L cohort)

	THA		TKA	
	Preoperative	**12-month follow-up**	**Preoperative**	**12-month follow-up**
EQ-5D-5L				
Mobility	15	23	8	25
Self-care	21	53	29	10
Usual activities	31	51	45	25
Pain/discomfort	19	62	13	22
Anxiety/depression	13	61	5	16
OHS/OKS				
Function	57	47	53	54
Pain	43	46	50	41

Chapter 5

Socioeconomic, patient, and hospital determinants for the utilization of peripheral nerve blocks in total joint arthroplasty

Authors
Joshua M. Bonsel[1], MD
Hanish Kodali[2], MBBS, MPH
Jashvant Poeran[2], MD, PhD
Gouke J. Bonsel[3], MD, PhD, emeritus professor

Affiliations
[1] Department of Orthopaedics and Sports Medicine, Erasmus MC, University Medical Center, Rotterdam, The Netherlands
[2] Department of Population Health and Policy, Icahn School of Medicine, Mount Sinai Hospital, New York, New York, United States of America
[3] EuroQol Research Foundation, Rotterdam, The Netherlands

Status
Published.
Citation: Bonsel JM, Kodali H, Poeran J, et al. Socioeconomic, Patient, and Hospital Determinants for the Utilization of Peripheral Nerve Blocks in Total Joint Arthroplasty. Anesth Analg. 2025; 140: 675-86.

ABSTRACT

Background

While peripheral nerve blocks (PNBs) are associated with various improved outcomes in patients undergoing total hip or knee arthroplasty (THA/TKA), disparities in PNB utilization have been reported. This study assessed the importance of socioeconomic, demographic, clinical and hospital determinants in explaining PNB utilization using the population attributable risk (PAR) framework. Subsequently, we examined the association between PNB use and 3 secondary outcomes: Centers for Medicare and Medicaid Services (CMS)-defined complications, 90-day all-cause readmissions and length of stay >3 days.

Methods

This retrospective cohort study included 52,926 THA and 94,795 TKA cases from the 5% 2012 to 2021 Medicare dataset. Mixed-effects logistic regression models measured the association between study variables and PNB utilization. Variables of interest were demographic (age, sex), clinical (outpatient setting, diagnosis, prior hospitalizations in the year before surgery, Deyo-Charlson index, obesity, (non)-opioid abuse, smoking), socioeconomic (neighborhood Social Deprivation Index, race and ethnicity) and hospital variables (beds, ownership, region, rurality, resident-to-bed ratio). The model was used for the calculation of variable-specific and variable category-specific PARs (presented in percentages), reflecting the proportion of variation in PNB use explained after eliminating variables (or groups of variables) of interest with all other factors held constant. Subsequently, regression models measured the association between PNB use and secondary outcomes. Associations are presented with odds ratios (OR) and 95% confidence intervals (95% CIs).

Results

Socioeconomic and demographic variables accounted for only a small proportion of variation in PNB use (up to 3% and 7%, respectively). Clinical (THA: 46%; TKA: 34%) and hospital variables (THA: 31%; TKA: 22%) were the primary drivers of variation. In THA, variation by clinical variables was driven by increased PNB use in the inpatient setting (OR, 1.28 (95% CI 1.07, 1.53)) and decreased use in patients with ≥2 prior hospitalizations (OR, 0.72 (95% CI 0.57, 0.90)). Moreover, non-osteoarthritis diagnoses associated with reduced PNB utilization in THA (OR, 0.64 (95% CI 0.58, 0.72)) and TKA (OR, 0.35 (95% CI 0.34, 0.37)).

In TKA, PNB use was subsequently associated with fewer complications (OR, 0.82 (0.75, 0.90)) and less prolonged length of stay (OR, 0.90 (0.86, 0.95)); no association was found for readmissions (OR, 0.98 (0.93, 1.03)). In THA, associations did not reach statistical significance.

Conclusions

Among THA and TKA patients on Medicare, large variations exist in the utilization of PNBs by clinical and hospital variables, while demographic and socioeconomic variables played a limited role. Given the consistent benefits of PNBs, particularly in TKA patients, more standardized provision may be warranted to mitigate the observed variation.

Key points

Question: What is the importance of socioeconomic, demographic, clinical and hospital determinants for the utilization of peripheral nerve blocks (PNBs) in total hip and knee arthroplasty patients (THA/TKA), and do PNBs associate with improved outcomes?

Findings: In both THA and TKA patients on Medicare, clinical (eg, indication for surgery) and hospital variables explained most variation in PNB use, while demographic and socioeconomic variables played a limited role; in TKA patients, PNBs were also associated with reduced complications and length of stay.

Meaning: Our findings emphasize substantial individual and hospital practice variation in PNB use; as PNBs are consistently associated with improved outcomes, particularly in TKA patients, the findings are a plea for more standardized provision of PNBs.

INTRODUCTION

While peripheral nerve blocks (PNBs) have been associated with improved outcomes in patients undergoing total hip or knee arthroplasty (THA/TKA)[1, 2], disparities in their utilization based on patient and hospital determinants have been reported[3-5]. Indeed, studies have shown that being younger, un(der)insured or belonging to a minority group is associated with lower odds of receiving PNBs[6, 7]. At the hospital level, a rural location and teaching status are associated with decreased PNB utilization[4]. However, size and even direction of effect are not always consistent as illustrated by a recent study including both patient- and hospital-level variables[8].

Separating the effect of factors of interest is complex in both statistical analysis and daily practice, as they are often interrelated. One should account for the so-called 'level' of their action (eg, patient-level versus hospital-level effects). Also, the impact indicator should reflect both the prevalence and the strength of the determinants. For example, even if there is a very strong association indicating Black patients receive fewer PNBs, its population-level impact will be limited in a hypothetical population with only a few Black patients. In a population with more Black patients, the population-level impact may still be limited if the strength of the race-PNB association is weaker than other or higher-level factors. The population attributable risk (PAR) concept, combined with stepwise analysis methods to account for the aforementioned "level" of action issue, provides a valuable approach for this purpose[9, 10]. The PAR assesses the impact of a determinant (or study variable of effect) in terms of the proportion of PNB use accounted for by that determinant.

Our study aimed to get a deeper understanding of the source of PNB variation. We estimated the importance of the socioeconomic background (including race/ethnicity and a proxy for socioeconomic status [SES]), demographic, clinical, and hospital determinants of the patient in explaining PNB utilization. We hypothesized a greater role of hospital versus patient variables and that within the latter PNB use would be lower in minority patients and those with a lower SES. We subsequently examined the association between PNB use and 3 important outcomes related to THA/TKA (complications, 90-day all-cause readmissions and length of stay) hypothesizing that PNB would be associated with improved outcomes, further emphasizing the importance of minimizing the hypothesized variation in PNB use.

METHODS

Data

In this retrospective cohort study, we analyzed inpatient and outpatient THAs and TKAs performed between 2012 and 2021 (all data available to our research group) as recorded in the Medicare Limited Dataset[11]. Given the deidentified nature of the data source, this study was exempt from full review by the Mount Sinai Institutional Review Board (STUDY-20-01677). This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines[12].

The Medicare database includes patient-level claims of all Medicare-insured patients in the United States. The inpatient and outpatient files contain an incomplete overview of the PNBs used; therefore, we also used the Carrier file to define PNBs, which made it necessary to use the 5% random sample. Each encounter contains information on procedural (Current Procedural Terminology [CPT]) and diagnosis-related (International Classification of Diseases ninth Revision codes [ICD]) codes. In 2015 the International Classification of Diseases tenth Revision coding system (ICD-10) was introduced. As this database mainly captures individuals aged over 65 and/or with disabilities, younger patients with or without private insurance are not included.

Inclusion and exclusion criteria of the sample

We constructed an initial cohort of 241,326 primary THA (CPT: 27130, ICD-9: 81.51, ICD-10: 0SR90, 0SRB0) and TKA (CPT: 27477, ICD-9: 81.54, ICD-10: 0SRC0, 0SRD0) patients. To define comorbidity prevalence and complication rate before and after surgery, respectively, we excluded patients who had surgery in 2012 or the last 3 months of 2021 (n = 21,609). Subsequently, we excluded patients who were not continuously enrolled in the database for at least 1 year prior to and 3 months after their joint arthroplasty (n = 41,375); patients inhabiting unincorporated territories of the United States (n = 186); aged <66 (n = 21,467); eligibility for Medicare due to end-stage renal disease (n = 347); patients having claims of both THA and TKA simultaneously (n = 711). For each patient, we only kept index arthroplasties which were 90 days apart, excluding another n = 2309 claims. Therefore, repeat procedures are not expected[13]. Patients may have had a contralateral procedure or procedure of another joint. Given the nature of the data, we were unable to determine whether patients had prior primary joint replacement before enrolling in Medicare. This process resulted in a sample of 153,322 patients (Figure).

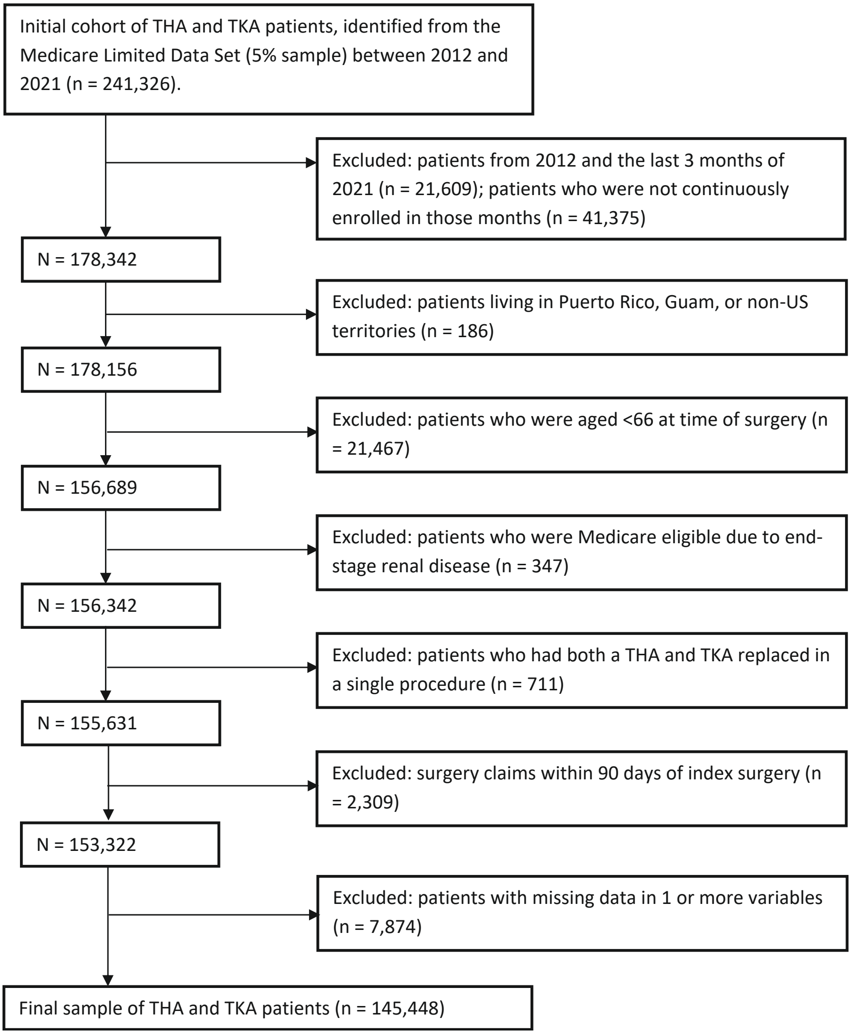

Figure: Flowchart depicting sample selection

Variables

Variables were selected based on the potential influence on PNB use either found in previous studies or based on clinical judgement[4-8, 14]. Directly obtained were age, sex, and year of surgery. Race and ethnicity was grouped as White, Black, and other (Asian, Hispanic, North American Native, other)[15]. Diagnosis was defined as either osteoarthritis or nonosteoarthritis using ICD codes (attached in Supplemental File 1). We calculated the Deyo-Charlon comorbidity index and categorized it into 0, 1, 2, and ≥3[16, 17].We identified a history of obesity, smoking, nonopioid, and opioid abuse. Prior hospitalization was defined as any acute care hospitalization in the 365 days before receiving surgery (excluding index arthroplasties) and was categorized into 0, 1, or ≥2. The effect of these variables on arthroplasty outcomes is well known, however, to our knowledge they have not been studied in the context of PNB utilization.

The 2019 neighborhood Social Deprivation Index (SDI) provides details on the place of living and was individually linked to the state-county ID. The SDI is a composite measure based on 7 characteristics, where a 1 to 100 SDI score is calculated using 5-year (2014–2019) averaged data from the American Community Survey: % living in poverty, % with <12 years of education, % single-parent households, % living in rented housing units, % living in the overcrowded housing unit, % of households without a car, and % nonemployed adults under 65 years of age[18]. We categorized it into 5 groups based on the thresholds for the 20th, 40th, 60th, and 80th percentiles[19]. Neighborhood deprivation indices are generally stable over time, and the currently applied measure is assumed valid for the included timeframe[20].

The following hospital data were derived from the 2017 Hospital Inpatient Prospective Payment System 2017 impact file: beds (0–150, 150–499 and ≥500), ownership (government, physician/proprietary, voluntary), rurality (large urban, small urban and rural), region (Northeast, South, Midwest, and West) and resident-to-bed ratio[21]. The resident-to-bed ratio is defined as the ratio of (interns + residents)/average operating beds; this ratio has previously been used as a proxy for teaching intensity. The ratio was categorized as <0.05 reflecting no teaching, 0.05 to 0.249 reflecting minor teaching, and ≥0.25 reflecting major teaching[22, 23]. Slight variations in hospital variables may occur over time: to maintain the ability to estimate the effects of these variables at the hospital level, we linked the IPPS file available in the middle of the included timeframe.

5

Missingness in any of the variables was present in 5% of THA and TKA patients, and was mainly attributable to hospitals not being recorded in the IPPS file. Missingness was unrelated to outcomes; therefore, we conducted a complete case analysis (Supplemental File 2, Figures 1–4). This resulted in a final sample of n = 147,721 patients.

Peripheral nerve block utilization, complication incidence and length of stay (outcomes)

Our primary outcome was the utilization of any form of PNB for the received joint arthroplasty, which was defined using CPT and ICD codes (Supplemental File 1, Table 1) submitted on the day of surgery, or at maximum 1 day before or 1 day after surgery. The 3 secondary outcomes of interest were Centers for Medicare and Medicaid Services (CMS)-defined complication (definition in Supplemental File 3, Table 2), 90-day all-cause readmissions and length of stay >3 days. Based on empirical evidence, CMS has defined outcome measures for commonly performed procedures, which are used to determine hospitals' performance and to adjust reimbursement rates. The measure captures negative outcomes that are likely attributable to the studied procedure, and which represent the quality of the provided care. In THA and TKA, a CMS-defined complication includes acute myocardial infarction, pneumonia, or sepsis/septicemia/shock during the admission or within 7 days from the date of admission; pulmonary embolism, surgical site bleeding or death during the admission or within 30 days from the date of admission; mechanical complications, surgical site bleeding or peri-prosthetic joint/wound infection during the admission or within 90 days from the date of admission[24].

Statistical analysis

THA and TKA patients were analyzed separately. Patient and hospital variables were compared between recipients and nonrecipients of a PNB. We used $\chi 2$ test for categorical variables.

For our first research question, we aimed to explore the relative contribution of each variable to the use of PNBs whilst accounting for potential confounding effects. Because hospital-level variables generally strongly affect PNB use, unmeasured factors may play a role. Logistic regression null models were compared with and without a random intercept for the hospital. Model fit improved drastically, with intraclass correlation coefficients (ICC) of 0.67 (THA) and 0.47 (TKA). The addition of a random intercept for the patient (3-level model) to account for contra-lateral procedures had minimal effect and was therefore not included in the final models. All variables of interest were entered as fixed effects: age, sex, outpatient setting, diagnosis, prior hospitalizations, Deyo index, obesity, (non)-opiod abuse, smoking, SDI, race and ethnicity, hospital beds, ownership, region, rurality, and resident-to-bed ratio. We assessed potential multicollinearity among variables using Spearman rank correlation indices. With all pairwise correlation indices <0.4, we determined the risk of multicollinearity to be low, as commonly accepted thresholds range from 0.5 to 0.8[25, 26]. The year of the surgery was adjusted for in the analysis, but was not considered a variable of interest with regard to our primary research question; it would have provided no opportunity for practice change. Within a categorical variable, the category with the highest amount of PNBs used was selected as a reference category. The reference category was generally the same for THA and TKA; if not, we opted to keep the same category as a reference in both THA and TKA.

To visualize the contribution of each variable to the use of PNBs we calculated PARs. Conventionally, PARs are calculated as $[Pr(O) – Pr(O \mid E)]/ Pr(O)$, where $Pr(O)$ is the probability of outcome in the study population, $Pr(O \mid E)$ is the hypothetical probability of outcome if the variable of interest were eliminated[27]. This approach fails to take confounders into account; therefore, we used abovementioned regression models to estimate confounder-adjusted PARs, currently the most efficient method available[27]. We followed a previously described approach[9]. Study population estimates of the number of PNBs used [Pr(O)] and estimates for each variable's category predicting the least amount of PNBs [Pr(O | E)] were obtained using the regression models. The PAR for each variable was calculated using the abovementioned formula (univariable PARs). The interpretation of an univariable PAR is the maximum % of PNB variation explained by eliminating that variable while controlling for confounding effects of other variables, assuming all potential confounders have been accounted for.

Sequential PARs were calculated for groups of variables: demographic (age, sex), clinical (outpatient setting, diagnosis, prior hospitalizations, Deyo index, obesity, (non)-opioid abuse, smoking), socioeconomic (SDI, race and ethnicity) and hospital (hospital beds, ownership, region, rurality, resident-to-bed ratio) variables. First, estimates for a group of variables set to the least amount of PNBs are obtained, and the PAR is calculated. Subsequently, the next group of variables is also set to the least amount of PNBs and new estimates are obtained, and the difference in PAR is calculated. This second procedure is repeated until all 4 groups are eliminated. The sequential PAR at each step represents the added variation explained by a group of variables after the preceding group(s) of variables have been eliminated. The order of eliminating groups of variables influences the sequential PARs because variables are potentially interrelated[28, 29]. Our primary order of interest was demographic > clinical > socioeconomic > hospital-level variables, based on the assumption that biological variables precede hospital variables. Reverse ordering was also assessed to evaluate the degree of interrelatedness of variables. If PAR estimates are similar for different orderings, groups of variables act independently on the utilization of PNBs. If PAR estimates differ, variable groups are at least statistically interrelated; for example, eliminating clinical variables may explain 10% of the variation in PNB. However, by eliminating demographics before clinical variables, the additional effect of clinical variables may be reduced to 5%, which would indicate interrelatedness between these groups of variables. In case of interrelatedness, univariable PARs exceed the total PAR for all risk factors combined[30].

For our second research question, mixed-effects logistic regression models with adjustment for the same variables measured the association between PNBs and the 3 secondary outcomes.

Two sensitivity analyses were conducted. Firstly, we checked whether there was a variation in PNB utilization by county of residence beyond the variation measured with the SDI. We re-ran the primary models replacing SDI by state-county ID (n = 3200) as random intercept (3-level model), and entering both simultaneously, and evaluated coefficients and model fit. Secondly, dual eligibility status (Medicare and Medicaid) was added as a variable of interest. As this variable has been collected since 2017[31], all models were re-run from using data from 2017 onwards.

No post hoc power calculations were performed given the exploratory nature of this study. Interaction between independent variables was not estimated, and the standard errors obtained from the models did not account for potential clustering. We report odds ratios (ORs) with 95% confidence intervals (95% CIs), while PARs are reported in %. Model performance was assessed using the c-statistic. ICCs reflect the variance captured by the random intercept for the hospital in each model. Although a *P*-value <0.05 was considered statistically significant, they were interpreted in combination with the strength of association. We conducted our analyses using R (version 4.2.3).

RESULTS

Descriptive analysis

We included 52,000 THAs and 93,448 TKAs with 7.9% (n = 4086) and 57.2% (n = 53,459) PNB use, respectively. The use of PNBs increased over the years. Univariable comparison (Table 1) of variables according to PNB use produced a similar pattern compared to the multivariable analysis (Table 2) and is therefore not separately discussed; the few exceptions to this pattern are mentioned.

Table 1: Descriptive statistics of Total Hip and Knee Arthroplasty patients by the utilization of any type of PNB

	Total hip arthroplasty			**Total knee arthroplasty**		
Variables	**No PNB**	**PNB**	***P*-value**	**No PNB**	**PNB**	***P*-value**
N	47,914	4,086		39,989	53,459	
Patient characteristics						
Age			0.085			<0.001
65–69	10,774 (22.5)	932 (22.8)		10,254 (25.6)	12,871 (24.1)	
70–74	14,572 (30.4)	1,243 (30.4)		13,264 (33.2)	17,842 (33.4)	
75–79	11,034 (23.0)	998 (24.4)		9,370 (23.4)	13,404 (25.1)	
80–84	7,051 (14.7)	564 (13.8)		4,956 (12.4)	6,786 (12.7)	
>84	4,483 (9.4)	349 (8.5)		2,145 (5.4)	2,556 (4.8)	
Female	30,186 (63.0)	2,622 (64.2)	0.141	25,670 (64.2)	34,046 (63.7)	0.112
Inpatient (vs outpatient)	43,726 (91.3)	3,685 (90.2)	0.022	35,562 (88.9)	44,193 (82.7)	<0.001
Diagnosis non-osteoarthritis	11,904 (24.8)	841 (20.6)	<0.001	9,257 (23.1)	6,910 (12.9)	<0.001
Prior hospitalizations			0.027			<0.001
0	39,949 (83.4)	3,434 (84.0)		34,611 (86.6)	46,984 (87.9)	
1	5,889 (12.3)	511 (12.5)		4,162 (10.4)	5,200 (9.7)	
≥2	2076 (4.3)	141 (3.5)		1,216 (3.0)	1,275 (2.4)	
Deyo index			0.834			0.835
0 (healthiest)	17,600 (36.7)	1,491 (36.5)		14,449 (36.1)	19,177 (35.9)	
1	10,531 (22.0)	905 (22.1)		9,512 (23.8)	12,747 (23.8)	
2	7,668 (16.0)	637 (15.6)		6,291 (15.7)	8,408 (15.7)	
≥3 (least healthy)	12,115 (25.3)	1,053 (25.8)		9,737 (24.3)	13,127 (24.6)	
Obese	11,132 (23.2)	958 (23.4)	0.772	12,138 (30.4)	16,888 (31.6)	<0.001
Abuse of nonopioids	2,880 (6.0)	268 (6.6)	0.169	1,485 (3.7)	2,221 (4.2)	0.001

Table 1: Continued

Variables	Total hip arthroplasty			Total knee arthroplasty		
	No PNB	PNB	*P*-value	No PNB	PNB	*P*-value
Abuse of opioids	460 (1.0)	58 (1.4)	0.006	270 (0.7)	387 (0.7)	0.399
Smoking	4,910 (10.2)	396 (9,7)	0.271	3,892 (9.7)	4,460 (8.3)	<0.001
SDI score			<0.001			<0.001
Q1 (most affluent)	10,083 (21.0)	789 (19.3)		7,598 (19.0)	10,977 (20.5)	
Q2–Q4	28,854 (60.2)	2,360 (57.8)		23,829 (59.6)	32,003 (59.9)	
Q5 (least affluent)	8,977 (18.7)	937 (22.9)		8,562 (21.4)	10,479 (19.6)	
Race and ethnicity			0.016			0.001
White	45,255 (94.5)	3,832 (93.8)		36,592 (91.5)	49,276 (92.2)	
Black	1,878 (3.9)	163 (4.0)		2,040 (5.1)	2,461 (4.6)	
Asian, Hispanic, North American native, other	781 (1.6)	91 (2.2)		1,357 (3.4)	1,722 (3.2)	
Year of surgery			<0.001			<0.001
2013	4,566 (9.5)	341 (8.3)		4,673 (11.7)	5,454 (10.2)	
2014	4,690 (9.8)	384 (9.4)		4,976 (12.4)	4,994 (9.3)	
2015	5,396 (11.3)	361 (8.8)		5,236 (13.1)	5,219 (9.8)	
2016	5,748 (12.0)	378 (9.3)		5,576 (13.9)	5,941 (11.1)	
2017	5,884 (12.3)	464 (11.4)		4,946 (12.4)	6,787 (12.7)	
2018	6,029 (12.6)	547 (13.4)		4,565 (11.4)	7,165 (13.4)	
2019	6,218 (13.0)	620 (15.2)		4,446 (11.1)	7,424 (13.9)	
2020	5,334 (11.1)	519 (12.7)		3,202 (8.0)	5,730 (10.7)	
2021	4,049 (8.5)	472 (11.6)		2,369 (5.9)	4,745 (8.9)	
Hospital characteristics						
Hospital beds			<0.001			<0.001
0–150	14,627 (30.5)	1,230 (30.1)		10,815 (27.0)	15,629 (29.2)	
150–499	25,639 (53.5)	2,086 (51.1)		21,867 (54.7)	27,345 (51.2)	
≥500	7,648 (16.0)	770 (18.8)		7,307 (18.3)	10,485 (19.6)	
Hospital ownership			<0.001			0.024
Government	4,950 (10.3)	521 (12.8)		4,274 (10.7)	5,568 (10.4)	
Physician/proprietary	7,499 (15.7)	976 (23.9)		7,349 (18.4)	10,179 (19.0)	
Voluntary	35,465 (74.0)	2,589 (63.4)		28,366 (70.9)	37,712 (70.5)	
Region			<0.001			<0.001
Northeast	9,098 (19.0)	539 (13.2)		6,334 (15.8)	8,954 (16.7)	
South	17,649 (36.8)	1,505 (36.8)		16,671 (41.7)	20,975 (39.2)	
Midwest	11,561 (24.1)	1,230 (30.1)		9,512 (23.8)	13,932 (26.1)	
West	9,606 (20.0)	812 (19.9)		7,472 (18.7)	9,598 (18.0)	
Rurality			<0.001			<0.001
Large urban	23,333 (48.7)	2,294 (56.1)		17,409 (43.5)	25,815 (48.3)	
Small urban	20,681 (43.2)	1,457 (35.7)		18,405 (46.0)	22,799 (42.6)	
Rural	3,900 (8.1)	335 (8.2)		4,175 (10.4)	4,845 (9.1)	

Table 1: Continued

	Total hip arthroplasty			Total knee arthroplasty		
Variables	**No PNB**	**PNB**	***P*-value**	**No PNB**	**PNB**	***P*-value**
Resident-to-bed ratio			0.249			<0.001
No teaching	30,173 (63.0)	2,525 (61.8)		26,866 (67.2)	36,093 (67.5)	
Minor teaching	10,861 (22.7)	941 (23.0)		9,139 (22.9)	11,035 (20.6)	
Major teaching	6,880 (14.4)	620 (15.2)		3,984 (10.0)	6,331 (11.8)	

Values are presented as n (%). *P*-values indicates differences between PNB and no PNB patients.
Abbreviations: PNB = peripheral nerve block; SDI = Social Deprivation Index

Table 2: Mixed-effects logistic regression models of patient and hospital variables on the use of PNBs

	Total Hip Arthroplasty		Total Knee Arthroplasty	
	OR (95% CI)	***P*-value**	**OR (95% CI)**	***P*-value**
Intercept	0.21 (0.10, 0.43)	<0.001	6.92 (4.57, 10.46)	<0.001
Age				
65–69	0.97 (0.86, 1.10)	0.644	0.84 (0.80, 0.88)	<0.001
70–74	0.91 (0.82, 1.02)	0.109	0.90 (0.86, 0.94)	<0.001
75–79	*ref*		*ref*	
80–84	0.86 (0.75, 0.98)	0.027	0.97 (0.91, 1.02)	0.246
>84	0.84 (0.72, 0.99)	0.038	0.81 (0.75, 0.88)	<0.001
Female	1.02 (0.93, 1.11)	0.724	0.95 (0.92, 0.99)	0.008
Inpatient (vs outpatient)	1.28 (1.07, 1.53)	0.007	0.71 (0.67, 0.76)	<0.001
Diagnosis nonosteoarthritis	0.64 (0.58, 0.72)	<0.001	0.35 (0.34, 0.37)	<0.001
Prior hospitalizations				
0	*ref*		*ref*	
1	0.97 (0.86, 1.10)	0.653	0.97 (0.92, 1.03)	0.328
≥2	0.72 (0.57, 0.90)	0.004	0.76 (0.68, 0.84)	<0.001
Deyo index				
0	0.96 (0.86, 1.08)	0.525	0.95 (0.90, 0.99)	0.019
1	1.00 (0.88, 1.12)	0.953	1.01 (0.96, 1.06)	0.705
2	0.93 (0.81, 1.06)	0.249	0.96 (0.91, 1.02)	0.178
≥3	*ref*		*ref*	
No obesity	1.01 (0.91, 1.11)	0.868	0.98 (0.95, 1.02)	0.344
No abuse of nonopioids	0.96 (0.81, 1.14)	0.654	0.98 (0.89, 1.07)	0.595
No abuse of opioids	0.58 (0.40, 0.84)	0.004	1.04 (0.85, 1.28)	0.671
No smoking	1.07 (0.92, 1.25)	0.380	0.88 (0.83, 0.94)	<0.001
SDI score				
Q1 (most affluent)	*ref*		*ref*	
Q2–Q4	0.94 (0.82, 1.07)	0.328	1.00 (0.95, 1.06)	0.942
Q5 (least affluent)	0.98 (0.83, 1.16)	0.842	0.96 (0.89, 1.03)	0.248
Race and ethnicity				
White	*ref*		*ref*	

Table 2: Continued

	Total Hip Arthroplasty		Total Knee Arthroplasty	
	OR (95% CI)	***P*-value**	**OR (95% CI)**	***P*-value**
Black	0.95 (0.76, 1.17)	0.620	0.91 (0.84, 0.99)	0.029
Asian, Hispanic, North American native, other	1.18 (0.87, 1.59)	0.285	0.94 (0.85, 1.04)	0.213
Year of surgery				
2013	0.49 (0.39, 0.60)	<0.001	0.83 (0.76, 0.91)	<0.001
2014	0.53 (0.43, 0.65)	<0.001	0.62 (0.56, 0.68)	<0.001
2015	0.42 (0.34, 0.52)	<0.001	0.55 (0.50, 0.60)	<0.001
2016	0.34 (0.28, 0.42)	<0.001	0.47 (0.43, 0.51)	<0.001
2017	0.44 (0.36, 0.54)	<0.001	0.68 (0.63, 0.75)	<0.001
2018	0.52 (0.43, 0.63)	<0.001	0.80 (0.74, 0.87)	<0.001
2019	0.64 (0.53, 0.77)	<0.001	0.85 (0.79, 0.93)	<0.001
2020	0.67 (0.56, 0.79)	<0.001	0.91 (0.83, 0.99)	0.024
2021	*ref*		*ref*	
Hospital beds				
≥500	*ref*		*ref*	
150–499	0.82 (0.60, 1.13)	0.219	0.85 (0.70, 1.02)	0.075
0–150	1.08 (0.73, 1.61)	0.696	0.98 (0.79, 1.23)	0.876
Hospital ownership				
Government	0.71 (0.45, 1.13)	0.152	0.94 (0.72, 1.23)	0.661
Physician/proprietary	*ref*		*ref*	
Voluntary	0.51 (0.36, 0.71)	<0.001	1.03 (0.85, 1.26)	0.732
Region				
Northeast	0.40 (0.26, 0.62)	<0.001	0.74 (0.58, 0.94)	0.012
South	0.87 (0.62, 1.23)	0.422	0.74 (0.61, 0.90)	0.003
Midwest	*ref*		*ref*	
West	1.09 (0.73, 1.60)	0.682	0.71 (0.56, 0.89)	0.003
Rurality				
Large urban	*ref*		*ref*	
Small urban	0.54 (0.40, 0.72)	<0.001	0.80 (0.67, 0.94)	0.007
Rural	0.74 (0.49, 1.12)	0.159	0.50 (0.40, 0.63)	<0.001
Resident-to-bed ratio				
No teaching	0.55 (0.35, 0.86)	0.010	0.96 (0.73, 1.26)	0.780
Minor teaching	0.62 (0.37, 1.04)	0.071	0.84 (0.62, 1.15)	0.282
Major teaching	*ref*		*ref*	
C-statistic	0.94		0.87	
ICC	0.67		0.49	

A random intercept was included for the hospital. Reference categories are those with the highest number of PNBs as observed in the univariable analysis. This is not always the same for hip and knee arthroplasty patients; we opted to choose the same reference category in those instances for comprehensibility. The c-statistic measured the overall model performance, and the ICC depicted the variance explained by the random intercept for the hospital. Abbreviations: CI = confidence interval; ICC = Intraclass Correlation Coefficient; OR = odds ratio; PNB = peripheral nerve block; SDI = Social Deprivation Index.

Association of patient and hospital variables with utilization of PNBs

THA

After adjustment, race/ethnicity and SDI did not significantly influence the utilization of PNBs. Lower odds for PNBs were seen for diagnoses other than osteoarthritis (OR, 0.64 (95% CI 0.58, 0.72)) and ≥2 prior hospitalizations (OR, 0.72 (95% CI 0.57, 0.90)). Contrary to the unadjusted analysis, an inpatient (vs. outpatient) setting showed higher odds for PNB use (OR, 1.28 (95% CI 1.07, 1.53)). Stronger effect estimates were observed for hospital-level variables: voluntary (OR, 0.51 (95% CI 0.36, 0.71)) and government-owned hospitals (OR, 0.71 (95% CI 0.45, 0.13)) showed lower odds for receiving PNBs compared to physician/proprietary-owned hospitals. Patients undergoing surgery in nonteaching hospitals also had lower odds of receiving a PNB (OR 0.55 (95% CI 0.35, 0.86)). Regional differences were substantial: patients inhabiting the West and Midwest had higher odds of receiving PNBs, as had large urban hospitals.

TKA

SDI did not significantly affect PNB utilization, while Black (compared to White) patients had slightly lower odds (OR 0.91 (95% CI 0.84, 0.99)). The inpatient setting showed lower odds for PNB use (OR 0.71 (95% CI 0.67, 0.76)). Similar to THA patients, diagnoses other than osteoarthritis (OR 0.35 (95% CI 0.33, 0.37)) and ≥2 prior hospitalizations (OR 0.76 (95% CI 0.68, 0.84)) were associated with lower odds of receiving PNBs. Regarding hospital variables, only region and rurality reached significance, of which associations aligned with THA results.

Population Attributable Risks for utilization of PNB

The highest univariable PAR of patient variables was that of having a diagnosis other than osteoarthritis (Table 3). In other words, if all THAs and TKAs were theoretically done for only nonosteoarthritis indications, this would result in 19% and 24% lower PNB utilization, respectively. In THA, the outpatient setting (13%) and ≥2 prior hospitalizations (18%) also played a large role. In THA, and to a lesser extent in TKA patients, hospital variables had high univariable PARs. In THA the largest PAR was observed for the region (38%), rurality (20%), and hospital ownership (10%). In TKA, these were rurality (14%), region (3%), and teaching status (3%).

Table 3: Population Attributable Risks for the utilization of PNBs

Variable	Total hip arthroplasty			Total knee arthroplasty		
	Worst category	**Predicted PNBs**	**Percentage**	**Worst category**	**Predicted PNBs**	**Percentage**
Predicted blocks (unadjusted variables)[a]		3,903			53,465	
Univariable PAR						
Demographics						
Age	>84	3,671	6	>84	51,849	3
Sex	Male	3,880	1		53,219	0

Table 3: Continued

Variable	Total hip arthroplasty			Total knee arthroplasty		
	Worst category	Predicted PNBs	Percentage	Worst category	Predicted PNBs	Percentage
Clinical						
Inpatient (vs outpatient)	Outpatient	3,395	13	Inpatient	52,788	1
Diagnosis	Nonosteo-arthritis	3,151	19	Non-osteo-arthritis	40,732	24
Prior hospitalizations	≥2	3,194	18	≥2	49,598	7
Deyo index	2	3,783	3	0	53,006	1
Obesity		3,908	0		53,383	0
Abuse of non-opioids		3,897	0		53,452	0
Abuse of opioids		3,888	0		53,469	0
Smoking	Yes	3,758	4		53,298	0
Socioeconomic						
SDI score	Q2–Q4 (medium)	3,847	1	Q5 (worst)	52,968	1
Race and ethnicity	Black	3,771	3	Black	52,266	2
Hospital variables						
Hospital beds	150–500	3,647	7	150–500	52,394	2
Hospital ownership	Voluntary	3,498	10	Government	52,343	2
Region	Northeast	2,422	38	West	51,821	3
Rurality	Small urban	3,131	20	Rural	45,779	14
Resident-to-bed ratio	No teaching	3,627	7	Minor teaching	51,923	3
Sequential PARs						
Demographics		3,649	7		51,600	3
Clinical		1,842	46		33,306	34
SES		1,736	3		31,578	3
Hospital variables		520	31		19,615	22
Total			87			62
Sequential PARs; reverse order						
Hospital variables		1,386	64		40,170	25
SES		1,304	2		38,410	3
Clinical		571	19		21,175	32
Demographics		520	1		19,615	3
Total			86			63

Predicted blocks and respective PARs reflect the number of PNBs utilized if a variable is set to the worst category. In other words, the PAR reflects how much PNBs are attributable to that variable, when all other variables are kept constant. We used a mixed-effects logistic regression model to calculate the PARs with a random intercept for the hospital.

[a] The predicted number of PNBs in the Total Hip Arthroplasty cohort does not entirely match the observed number of PNBs (4,086), while it is close for the total knee arthroplasty cohort (53,459). This is presumably due to the relative lower incidence of PNBs applied in the THA cohort, resulting in a slightly poorer predictive capability of the model.

Abbreviations: PAR = population attributable risk; PNB = peripheral nerve block; SDI = Social Deprivation Index; SES = socioeconomic status; THA = total hip arthroplasty.

The sequential PARs visualization provided the cumulative proportion of PNBs which could be explained by the regression model[9]. All variables combined explained a considerably higher percentage of PNBs in THA (87%) compared to in TKA patients (63%). Starting with demographics and ending with hospital variables, in THA, the largest contributing factors were clinical (46%), followed by hospital (31%), demographic (7%), and socioeconomic variables (3%). In TKA, the largest contributors were also clinical (34%), followed by hospital (22%), demographic (3%) and socioeconomic variables (3%). If the order of the groups of variables was reversed (starting with the hospital), hospital variables explained a larger proportion in THA, but not in TKA (THA: 64%, TKA: 25%). The effect of clinical variables reduced in THA (19%) which illustrates the statistical interrelatedness of hospital and clinical variables.

Association of PNBs with secondary outcomes

In THA, use of PNB did not significantly relate to CMS-defined complications (OR, 0.92 (95% CI 0.78, 1.10)), 90-day all-cause readmission (OR, 0.98 (95% CI 0.87, 1.10)) nor a length of stay >3 days (OR, 0.99 (95% CI 0.89, 1.11)) (Table 4). In TKA, use of PNBs was significantly associated with a reduction in CMS-defined complications (OR, 0.82 (95% CI 0.75, 0.90)) and length of stay >3 days (OR, 0.90 (95% CI 0.86, 0.95)), however, no benefit was found on 90-day all-cause readmissions (OR, 0.98 (95% CI 0.93, 1.03)). The full models with adjustment factors can be found in Supplemental File 4, Tables 3–8.

Model performance

The mixed-effects regression models on PNB utilization produced high c-statistics (THA: 0.94, TKA: 0.86) and ICCs (THA: 0.67, TKA: 0.49; Table 2). The models for the secondary outcomes had lower c-statistics, varying from 0.66 to 0.82 in both THA and TKA (Table 4). ICCs were also lower, ranging from 0.02 to 0.16.

Sensitivity analysis

The addition of state-county ID as a random effect had no improvement on model fit, and estimates of SDI did not change (Supplemental File 4, Tables 9–10). In other words, we did not find evidence of variation in PNB use by county of residence beyond the SDI measure used. The inclusion of dual eligibility had no effect on PNB utilization (Supplemental File 4, Table 11), nor did multivariable estimates of outcomes change substantially (Supplemental File 4, Tables 3–8).

Table 4: Mixed-effects logistic regression models of PNBs on outcomes

	Total hip arthroplasty				Total knee arthroplasty			
	OR (95% CI)	**P-value**	**C-statistic**	**ICC**	**OR (95% CI)**	**P-value**	**C-statistic**	**ICC**
CMS complications	0.92 (0.78, 1.10)	0.357	0.71	0.04	0.82 (0.75, 0.90)	<0.001	0.70	0.04
90-day all-cause readmissions	0.98 (0.87, 1.10)	0.738	0.67	0.01	0.98 (0.93, 1.03)	0.380	0.67	0.02
Length of stay >3 d	0.99 (0.89, 1.11)	0.904	0.83	0.13	0.90 (0.86, 0.95)	<0.001	0.78	0.14

The OR reflects the use of PNBs versus no PNB. The same set of variables that were used in the analysis of the utilization of PNBs was entered as fixed effects. A random effect for the hospital is included in the models. Length of stay is only analyzed in patients who had inpatient surgery (approximately 90% of patients). The c-statistic measures the overall model performance, and the ICC depicts the variance explained by the random intercept for the hospital.
Abbreviations: CI = confidence interval; ICC = Intraclass Correlation Coefficient; OR = odds ratio; PNB = peripheral nerve block; CMS = Centers for Medicare & Medicaid Services.

DISCUSSION

Main findings

To the best of our knowledge, this is the first study to determine patterns of use and effectiveness of PNBs in THA and TKA patients using Medicare data. Contrary to our expectations, socioeconomic background (PAR: THA: 2%–3%, TKA: 3%) played a minor role in the observed variation in PNB utilization. Most variation was explained by clinical (THA: 19%–46%, TKA: 32%–34%) and hospital variables (THA: 31%–64%, TKA: 22%–25%). The PAR for clinical variables was driven by the decreased use of PNBs in patients with a nonosteoarthritis diagnosis, and in THA also by decreased use in the outpatient setting and patients with prior hospitalizations. In all, statistical relations in TKA echo those in THA, but the relative role of hospital-related effects is larger in THA. These findings illustrate that the strongest driving force behind disparities in the utilization of PNBs is based on practice differences (provider based) in semi- and nonelective arthroplasty patients.

Our study adds to the extensive evidence base[1, 2] that the use of PNB is associated with improved clinical outcome: in TKA patients, we found fewer complications and length of stay; differences in THA patients did not reach statistical significance.

Comparison with other literature

Previous studies examined the impact of patient and hospital variables on the utilization of PNBs through standard regression techniques, which provide insufficient insight into the strength of the association. The PAR method used in this paper had additional value in this regard. Overall, our study found no clear evidence of disparities according to socioeconomic (SDI, race and ethnicity, and dual eligibility) variables. This finding diverges from a study by Keneally et al[6], which used ZIP-code-linked median income as SES indicator and found a higher income to significantly relate to increased utilization of PNBs in TKA. A reason for this discrepancy may be because our study applied a different comprehensive type of neighborhood SES indicator and at a different level of linkage. However, as we did not observe variation in PNB utilization according to the county of residence, it is unlikely variation by neighborhood indicators will be found in the current dataset. In TKA a weak association suggested Black patients (compared to White) received fewer PNBs. We do not believe this is strong evidence of an association, as the PAR analysis did not show substantial variation by race/ethnicity and this may also be the result of a type I error. In comparison, a recent study by Zhong et al[8] used a private insurance database and found nonwhite compared to White TKA patients receiving PNBs less often. The contrasting findings highlight that the effect of socioeconomic variables also may differ by the studied population and type of health coverage, that is, private versus public.

PNB utilization was less in patients receiving THA and TKA for nonosteoarthritis indications. Fracture patients typically present in a nonelective setting which could limit the timely administration of PNBs. However, nonosteoarthritis indications for THA/TKA will also include a variety of (semi-)elective diagnoses such as posttraumatic osteoarthritis, osteonecrosis, and rheumatoid arthritis[32, 33]. Especially in TKA in which the number of fracture patients is relatively small, there is a large group of (semi-)elective patients in whom the abovementioned explanation may not suffice. Additionally, in THA patients with prior hospitalizations PNBs were used less often, which highlights another potential explanation for differential use: comorbid and/or semielective patients may fall outside of protocolized care pathways with as a consequence less use of PNBs.

Regional practice variations explained a large part of the variation in PNB use. In both THA and TKA patients the Midwest region and urban hospitals are associated with increased utilization of PNBs. Strengthening the assertion that PNB utilization is largely determined by practice variations was that the addition of a random intercept for the hospital (which covers unspecified hospital effects) drastically improved model fit. Practice variations may indirectly lead to variations in use by socioeconomic or clinical variables, or vice versa. For example, different PNB utilization profiles may drive socioeconomic disparities in the background, because certain regions are inhabited by relatively less affluent and/or more Black patients, such as the South[34, 35]. This effect is probably limited, as we did not observe the interrelatedness of socioeconomic and hospital-related variables in the PAR analysis. Clinical and hospital-related variables, however, were statistically interrelated in THA patients, as the role of clinical variables reduced markedly after first accounting for hospital variables. In other words, particular patients (ie, comorbid/nonosteoarthritis) treated in hospitals/regions as reflected by the hospital-related variables received PNBs less often. We currently cannot determine the directionality of this effect. We think that survey data with targeted questions on barriers for use of PNBs per hospital/specialist group could provide valuable insights[36]. This may also reveal if the overall socioeconomic or clinical profile of patients presenting at hospitals in certain regions affects PNB utilization at the policy level.

We expect the practitioners' choice (surgeon or anesthesiologist) plays a key role in explaining these regional variations, which in turn largely depends on the training received and the experienced comfort with PNB utilization[4]. For example, one study found that PNBs were applied more often in TKA patients when a board-certified anesthesiologist was present[35]. In a study on the utilization of regional anesthesia for acute pain management among military anesthesiology residents and specialists, a potential barrier to apply PNBs was the lack of opportunities to practice during traing[37]. The practice environment may differ largely by regions, and specific hospitals (urban, teaching) may have increased opportunities to practice PNB utilization for residents.

Diverging patterns between THA and TKA patients were observed with regard to hospital ownership: THA patients undergoing surgery in physician/proprietary (for-profit) hospitals had higher odds of receiving a PNB compared to voluntary or government (nonprofit) hospitals, while this was not the case in TKA patients. For-profit hospitals have different incentives and resources available compared to nonprofit hospitals. which may result in increased (earlier) adaptation of novel treatments with a slimmer evidence base[38]. Supporting this notion is the fact that overall uptake of PNBs is far less in THA compared to TKA (8% vs. 57%, respectively).

Strengths and limitations

This study has some limitations. Due to the observational nature of this study, we can only assess associations and not causal relations. Moreover, it is possible that potential confounders in the studied association were missed; in this scenario, currently observed associations may be overestimated. At the hospital level, separating ambulatory surgical centers owned by or affiliated with teaching hospitals might have resulted in more detailed insights into the effect of the resident-to-bed ratio. Secondly, our findings are only generalizable to the Medicare population; different patterns may exist for commercially insured patients, a growing group of arthroplasty recipients[4, 8]. Thirdly, the area-based social deprivation indicator may not entirely reflect deprivation at the individual level. Finally, PAR estimates represent the maximum attainable reduction in variation of PNB utilization; it is unlikely a change in clinical practice will eliminate all variation.

CONCLUSIONS

In THA and TKA patients on Medicare, large variations exist in the utilization of PNBs by clinical (eg, indication for arthroplasty) and hospital variables, while demographic and socioeconomic variables played a limited role. These findings emphasize the substantial individual and hospital practice variation in PNB utilization. In light of the potential benefit of PNBs observed in our study and various other studies, we believe stakeholders should strive for more standardized provision of PNBs.

REFERENCES

1. Memtsoudis SG, Cozowicz C, Bekeris J, Bekere D, Liu J, Soffin EM, et al. Peripheral nerve block anesthesia/analgesia for patients undergoing primary hip and knee arthroplasty: recommendations from the International Consensus on Anesthesia-Related Outcomes after Surgery (ICAROS) group based on a systematic review and meta-analysis of current literature. Reg Anesth Pain Med 2021; 46: 971-985.
2. Memtsoudis SG, Cozowicz C, Bekeris J, Bekere D, Liu J, Soffin EM, et al. Anaesthetic care of patients undergoing primary hip and knee arthroplasty: consensus recommendations from the International Consensus on Anaesthesia-Related Outcomes after Surgery group (ICAROS) based on a systematic review and meta-analysis. Br J Anaesth 2019; 123: 269-287.
3. Cozowicz C, Poeran J, Zubizarreta N, Mazumdar M, Memtsoudis SG. Trends in the Use of Regional Anesthesia: Neuraxial and Peripheral Nerve Blocks. Reg Anesth Pain Med 2016; 41: 43-49.
4. Memtsoudis SG, Poeran J, Zubizarreta N, Rasul R, Opperer M, Mazumdar M. Anesthetic Care for Orthopedic Patients: Is There a Potential for Differences in Care? Anesthesiology 2016; 124: 608-623.
5. Memtsoudis SG, Danninger T, Rasul R, Poeran J, Gerner P, Stundner O, et al. Inpatient falls after total knee arthroplasty: the role of anesthesia type and peripheral nerve blocks. Anesthesiology 2014; 120: 551-563.
6. Keneally RJ, Mazzeffi MA, Chow JH, Heinz ER, Shinton ZA, Meyers BA, Dutton RP. Socioeconomic Disparities in Method of Anesthesia for Knee Arthroplasties in the US. J Health Care Poor Underserved 2022; 33: 1809-1820.
7. Schaar AN, Finneran JJ, Gabriel RA. Association of race and receipt of regional anesthesia for hip fracture surgery. Reg Anesth Pain Med 2023.
8. Zhong H, Poeran J, Liu J, Liguori G, Popovic M, Poultsides L, Memtsoudis SG. Disparities in the provision of regional anesthesia and analgesia in total joint arthroplasty: The role of patient and hospital level factors. J Clin Anesth 2021; 75: 110440.
9. Poeran J, Borsboom GJJM, de Graaf JP, Birnie E, Steegers EAP, Bonsel GJ. Population Attributable Risks of Patient, Child and Organizational Risk Factors for Perinatal Mortality in Hospital Births. Maternal and Child Health Journal 2015; 19: 764-775.
10. Danninger T, Rasul R, Poeran J, Stundner O, Mazumdar M, Fleischut PM, et al. Blood transfusions in total hip and knee arthroplasty: an analysis of outcomes. ScientificWorldJournal 2014; 2014: 623460.
11. MarketScan Research Databases - Databases 2021. Available at: https://www.ibm.com/products/marketscan-research-databases/databases. Accessed xxx.
12. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, Initiative S. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007; 370: 1453-1457.
13. Schwarzkopf R, Behery OA, Yu H, Suter LG, Li L, Horwitz LI. Patterns and Costs of 90-Day Readmission for Surgical and Medical Complications Following Total Hip and Knee Arthroplasty. J Arthroplasty 2019; 34: 2304-2307.
14. Cozowicz C, Poeran J, Memtsoudis SG. Epidemiology, trends, and disparities in regional anaesthesia for orthopaedic surgery. British Journal of Anaesthesia 2015; 115: ii57-ii67.
15. Barbera JP, Raymond HE, Zubizarreta N, Poeran J, Chen DD, Hayden BL, Moucha CS. Racial Differences in Manipulation Under Anesthesia Rates Following Total Knee Arthroplasty. The Journal of Arthroplasty 2022; 37: 1865-1869.

16. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987; 40: 373-383.
17. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 2005; 43: 1130-1139.
18. Social deprivation index (SDI), 2019. Robert Graham Center - Policy Studies in Family Medicine & Primary Care. Retrieved May 1, 2023, from https://www.graham-center.org/rgc/maps-data-tools/sdi/social-deprivation-index.html. .
19. Handbook on Health Inequality Monitoring, by The World Health Organization. Retrieved May 1, 2023, from https://www.who.int/publications/i/item/9789241548632.
20. Ralston K, Dundas R, Leyland AH. A comparison of the Scottish Index of Multiple Deprivation (SIMD) 2004 with the 2009 + 1 SIMD: does choice of measure affect the interpretation of inequality in mortality? International Journal of Health Geographics 2014; 13: 27.
21. Centers for Medicare & Medicaid Services (CMS) Impact File Hospital IPPS, 2017. Retrieved May 1, 2023, from https://www.nber.org/research/data/centers-medicare-medicaid-services-cms-impact-file-hospital-ipps.
22. Silber JH, Rosenbaum PR, Niknam BA, Ross RN, Reiter JG, Hill AS, et al. Comparing Outcomes and Costs of Medical Patients Treated at Major Teaching and Non-teaching Hospitals: A National Matched Analysis. J Gen Intern Med 2020; 35: 743-752.
23. Volpp KG, Rosen AK, Rosenbaum PR, Romano PS, Even-Shoshan O, Wang Y, et al. Mortality among hospitalized Medicare beneficiaries in the first 2 years following ACGME resident duty hour reform. Jama 2007; 298: 975-983.
24. Centers for Medicare & Medicaid Services (CMS) complication measures. Retrieved May 1, 2023, from https://qualitynet.cms.gov/inpatient/measures/complication/methodology.
25. Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 2013; 36: 27-46.
26. Vatcheva KP, Lee M, McCormick JB, Rahbar MH. Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. Epidemiology (Sunnyvale) 2016; 6.
27. Benichou J. A review of adjusted estimators of attributable risk. Stat Methods Med Res 2001; 10: 195-216.
28. Eide GE, Gefeller O. Sequential and average attributable fractions as aids in the selection of preventive strategies. J Clin Epidemiol 1995; 48: 645-655.
29. Gefeller O, Land M, Eide GE. Averaging Attributable Fractions in the Multifactorial Situation: Assumptions and Interpretation. Journal of Clinical Epidemiology 1998; 51: 437-441.
30. Rowe AK, Powell KE, Flanders WD. Why population attributable fractions can sum to more than one. American Journal of Preventive Medicine 2004; 26: 243-249.
31. Thirukumaran CP, Cai X, Glance LG, Kim Y, Ricciardi BF, Fiscella KA, Li Y. Geographic Variation and Disparities in Total Joint Replacement Use for Medicare Beneficiaries: 2009 to 2017. J Bone Joint Surg Am 2020; 102: 2120-2128.
32. Adhia AH, Feinglass JM, Suleiman LI. What Are the Risk Factors for 48 or More–Hour Stay and Nonhome Discharge After Total Knee Arthroplasty? Results From 151 Illinois Hospitals, 2016-2018. The Journal of Arthroplasty 2020; 35: 1466-1473.e1461.
33. Weiner JA, Adhia AH, Feinglass JM, Suleiman LI. Disparities in Hip Arthroplasty Outcomes: Results of a Statewide Hospital Registry From 2016 to 2018. The Journal of Arthroplasty 2020; 35: 1776-1783.e1771.

34. Andrews MR, Tamura K, Claudel SE, Xu S, Ceasar JN, Collins BS, et al. Geospatial Analysis of Neighborhood Deprivation Index (NDI) for the United States by County. J Maps 2020; 16: 101-112.
35. Fleischut PM, Eskreis-Winkler JM, Gaber-Baylis LK, Giambrone GP, Faggiani SL, Dutton RP, Memtsoudis SG. Variability in anesthetic care for total knee arthroplasty: an analysis from the anesthesia quality institute. Am J Med Qual 2015; 30: 172-179.
36. Posthumus AG, Borsboom GJ, Poeran J, Steegers EA, Bonsel GJ. Geographical, Ethnic and Socio-Economic Differences in Utilization of Obstetric Care in the Netherlands. PLoS One 2016; 11: e0156621.
37. Jaffe E, Patzkowski MS, Hodgson JA, Foerschler DL, Gonzalez SC, Giordano NA, et al. Practice Variation in Regional Anesthesia Utilization by Current and Former U.S. Military Anesthesiology Residents. Military Medicine 2021; 186: e98-e103.
38. Horwitz JR. Making profits and providing care: comparing nonprofit, for-profit, and government hospitals. Health Aff (Millwood) 2005; 24: 790-801.

DECLARATIONS

Acknowledgements: We would like to acknowledge Dr Brocha Stern for her help with the additional analyses performed during the review phase. Moreover, we would like to acknowledge the reviewers for their valuable and in-depth review of this work.

Funding: This work was funded by a travel grant provided by the EuroQol Research Foundation (1483-TVG) and the Erasmus Trustfonds (97030.2022.101.935/292/RB). The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Conflicts of interest: None.

Supplementary data: Supplementary data associated with this article can be found in the online version at doi: 10.1213/ANE.0000000000007107.

Chapter 6

Impact of the COVID-19 lockdown on patient-reported outcome measures in Dutch hip and knee arthroplasty patients

Authors

Joshua M. Bonsel[1], MD

Lichelle Groot[1,2], MSc

Abigael Cohen[1], MD

Jan A.N. Verhaar[1], MD, PhD, emeritus professor

Maaike G.J. Gademan[2,3], MSc, PhD

Anneke Spekenbrink-Spooren[4], MSc

Gouke J. Bonsel[5], MD, PhD, emeritus professor

Max Reijman[1], MSc, PhD

Affiliations

[1] Department of Orthopaedics and Sports Medicine, Erasmus MC, University Medical Center, Rotterdam, The Netherlands

[2] Department of Orthopaedics, Leiden University Medical Center, Leiden, The Netherlands

[3] Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

[4] Dutch Arthroplasty Register (Landelijke Registratie Orthopedische Interventies), ’s Hertogenbosch, The Netherlands

[5] Department of Public Health, Erasmus MC, University Medical Center, Rotterdam, The Netherlands

Status

Published.

Citation: Bonsel JM, Groot L, Cohen A, Verhaar JAN, Gademan MGJ, Spekenbrink-Spooren A, et al. Impact of the COVID-19 lockdown on patient-reported outcome measures in Dutch hip and knee arthroplasty patients. Acta Orthop. 2022;93:808-18.

ABSTRACT

Background and purpose

During the first COVID-19 lockdown elective surgery was greatly reduced. Prioritization of patients with greater need and expected benefit in terms of quality of life was advised. The lockdown also potentially affected follow-up outcomes. Therefore, our study compared patient-reported outcome measures (PROMs) retrieved during the lockdown of Dutch primary total hip and knee arthroplasty (THA, TKA) patients with previous years.

Patients and methods

We performed cross-sectional analyses using national data from the Dutch Orthopaedic Registry (LROI). All primary elective THA and TKA patients with preoperative or postoperative PROMs (EQ-5D-3L index, OHS/OKS) during the first COVID-19 lockdown between March and July 15, 2020 were included. Patients with PROMs during the same months in 2018 plus 2019 were used as control. Finally, 33,453 THA and 27,335 TKA patients were included. Patient characteristics were compared during versus before the lockdown. Subsequently, the lockdown effect on PROMs scores was analyzed with multivariable linear regression.

Results

During the COVID-19 lockdown, THA and TKA patients had a lower age and BMI preoperatively, and more often had surgery in private clinics. Both preoperative PROMs in THA patients, but not in TKA patients, were worse (EQ-5D: Adjusted mean difference (AMD) -0.021, $p < 0.001$) during the lockdown compared with prior years. Both postoperative PROMs in THA and TKA patients were better during the lockdown (12-month EQ-5D in THA: AMD 0.010, $p = 0.003$; and in TKA: AMD 0.013, $p < 0.001$).

Interpretation

During the COVID-19 lockdown, THA patients had slightly worse preoperative PROMs, suggesting selection of patients with greater urgency. Postoperative PROMs in both THA and TKA patients differed minimally. Overall, the observed differences were likely not clinically relevant.

INTRODUCTION

In response to the SARS-CoV-2 (COVID-19) pandemic, extreme measures have been taken to contain the virus, most notable being national lockdowns and quarantine policies[1]. Several studies noted an increase in psychological distress symptoms of e.g., feelings of anxiety, and a decrease in quality of life in the general population following the pandemic and the measures taken[2-4]. This effect increased with lower age, female sex, poor health status and low socioeconomic status (SES) [5, 6]. In most countries there were regional variations in infection rates, which were also observed in the Netherlands[7]. People inhabiting places with higher infection rates might also experience increased effects on mental and physical health[8].

In the Netherlands the first lockdown was instigated on March 16, 2020, which was relaxed on June 1, 2020. During the lockdown, large events were prohibited, and people were advised to stay at and work from home if possible. Also, non-essential shops and places including outpatient healthcare facilities such as physical therapists' premises were closed. Large shifts in healthcare resources were needed and most elective orthopedic care was reduced to a minimum[9]. The Dutch Orthopaedic Society issued a statement on May 1, 2020, to restrict elective arthroplasty to patients who have the largest need and expected benefit in terms of quality of life[10]. Also, to minimize COVID-19 infection risks, healthier patients eligible for day treatment were preferred[11].

The lockdown in combination with the pandemic also likely resulted in a lower grade of physical activity and postoperative rehabilitation, negatively influencing the postoperative patient-reported outcome measures (PROMs) of these patients[12, 13]. Reporting this impact has important implications regarding the assessment of effectiveness of interventions during a pandemic and the resulting lockdown.

This study compared PROMs retrieved during the COVID-19 lockdown of primary THA and TKA patients registered in the Dutch Orthopaedic Registry (Dutch abbreviation: LROI) with previous years. We had the following hypotheses. First, preoperative PROMs are lower due to a selection effect based on urgency assessment. Second, early postoperative PROMs in particular are lower due to the impaired rehabilitative process. Third, at all follow-up points the lockdown negatively affect PROMs, e.g., through feelings of anxiety. Specific subgroups that were possibly more affected by the COVID-19 lockdown were also analyzed.

PATIENTS AND METHODS

Dutch Orthopaedic Registry (LROI)

We used data from the LROI, which prospectively collects data on orthopedic interventions in the Netherlands. The LROI contains demographic and surgical information, and for arthroplasties additionally prosthesis characteristics. Data completeness is over 95% for primary THA and TKA patients[14]. Since 2014 the Dutch Orthopaedic Society has strongly recommended the collection of internationally validated general health and disease-specific PROMs in elective arthroplasty. PROMs are retrieved before surgery (at max. 182 days before surgery), at 3-month follow-up (63–110 days) in THA and at 6-month follow-up (154–210 days) in TKA, and at 12-month follow-up (323–407 days) in both. Preoperative PROMs are mostly completed at the outpatient clinic, whilst postoperative PROMs are completed either electronically after invitation via email, or with pen and paper. In 2018 and 2019, 63–66% of THA patients for osteoarthritis completed preoperative PROMs, whilst 34–42% completed both preoperative and postoperative PROMs. In TKA these rates were 55–61% and 30–40% respectively. During COVID-19 in 2020 response rates were in the same ranges[9].

Study design

We performed a retrospective cross-sectional study, and adhered to the STROBE guidelines. Primary THA and TKA patients for any elective indication were selected. Patients who filled out either pre- or postoperative PROMs between March 23 and July 15, 2020 were included. This window allowed for any potential impact of the lockdown to reach its full extent. THA and TKA patients were analyzed separately. This resulted in 6 COVID-19 lockdown groups (THA: preoperative, 3-, and 12-month follow-up; TKA: preoperative, 6-, and 12-month follow-up). Using the same inclusion window patients from 2018 plus 2019 were selected as control groups. COVID-19 groups were compared with the respective control groups, resulting in 6 cross-sectional comparisons.

Data

The following patient characteristics were obtained: age, sex, BMI, Charnley score, ASA score, previous surgery on the joint, indication for joint replacement (osteoarthritis or non-osteoarthritis such as post-traumatic), and type of hospital (general, academic, or private). Additionally, data on COVID-19 infection rate and SES was linked to registry data using patients' 4-digit postal codes. The Dutch Institute for Health and Milieu published COVID-19 infection rates in the Netherlands bi-weekly[15]. The Netherlands is divided into 12 "provinces". Each province was given an infection rate score of 1 through 5 calculated at 2-week intervals (1: ≤24, 2: 25–49, 3:50–74, 4: 75–99, 5: ≥100 infections per 100,000 inhabitants). This score was assigned to records based on the date the PROMs were filled in. Data on SES was obtained from the Dutch Institute of Social Research[16]. For each 4-digit postal code area with more than 100 inhabitants, a numeric SES-score was created. The

SES-score is calculated with multiple variables from a postal code area: mean income per household, % households with a low income, % unemployed inhabitants and % households with an average low education. This method to approximate the individual SES score is considered a validated technique[17]. The SES-score was categorized into 5 groups based on the quantiles. These groups were referred to as quintiles.

Patient-reported outcome measures

The EuroQol 5-Dimensions (EQ-5D-3L) questionnaire, visual analogue scale (EQ-VAS), and 2 disease-specific questionnaires were obtained[18]. For THA the latter were the Oxford Hip Score (OHS) and the short version of the Hip disability and Osteoarthritis Outcome Score (HOOS-PS), and for TKA these were the Oxford Knee Score (OKS) and the short version of the Knee disability and Osteoarthritis Outcome Score (KOOS-PS) [19-21]. We selected the EQ-5D-3L and the Oxford set as our main outcome measures. For the EQ-5D-3L an overall index score was calculated using the Dutch National Value set[22].

Statistics

Patient characteristics of the COVID-19 groups were compared with the respective control groups using the chi-square and Student's t-test. The representativeness of responders was assessed by also comparing patient characteristics of each COVID-19 group with non-responders operated during the same period based on the inclusion window. The pattern of representativeness was compared with previous years (2018 plus 2019). Duplicate cases, i.e., patients who had their contralateral joint replaced as well, made up a small number of patients (THA: 5%, TKA: 6%) in the entire cohort. Given the present study design, they were not expected to affect results, therefore they were not removed.

Subsequently, PROMs retrieved during the COVID-19 lockdown were compared with control groups using multivariable linear regression analysis. Potential confounders were included based on the theoretical association with the exposure (COVID-19 lockdown) and the known association with outcomes (PROMs) [23]. The analyses were adjusted for sex, BMI, ASA score, Charnley score, previous surgery on the joint, indication for joint replacement, type of hospital, and SES[24-27]. If a PROM was statistically significantly associated with the COVID-19 group, interaction terms between this group and specific high-risk subgroups were used to explore whether the COVID-19 lockdown had a different effect in these subgroups. The subgroups that we explored were BMI > 30, ASA ≥ 3, age > 70, non-osteoarthritis indication for joint replacement, female sex, and SES quintile ≤ 2. Similarly, we investigated the effect of inhabiting a region with ≥ 50 COVID-19 infections per 100,000 inhabitants. Each interaction term was added individually to the regression analysis, which was then assessed for fit and significance. If interaction terms were not relevant, i.e., resulted in a lower R^2 (worsened model fit) and/or did not reach statistical significance, they were removed from the analyses. Robust 95% confidence intervals (CIs) were calculated to account for heteroscedasticity of the PROMs outcomes. The differences are presented as adjusted mean differences (AMDs) with robust CIs and p-values. Clinical relevance was determined by comparing AMDs of the main analyses with currently accepted minimal clinically important differences (MCIDs). These have been reported to be 0.03 for the EQ-5D-3L index in musculoskeletal patients, 5.2 for the OKS, and 4.8 for the OKS in arthroplasty patients[28, 29]. Potential clinical relevance was confirmed if the CIs' bounds exceeded the defined MCIDs. A p-value < 0.05 was considered significant. All analyses were performed using SPSS (version 25; IBM Corp, Armonk, NY, USA).

Ethics, funding, and potential conflicts of interest

This study was based on registry data with an extensive protocol for legally conforming data access, therefore no ethical approval additional to LROI permission was required. One of the authors (JB) has received funding from EuroQol for a PhD project including this study. The views expressed by the authors in this manuscript do not necessarily reflect the views of the EuroQol group.

RESULTS

Included patients

There were 33,453 elective THA and 27,335 elective TKA patients eligible for inclusion (Figure 1). The number of patients at 6-month follow-up (TKA) and 12-month follow-up during the COVID-19 lockdown was similar to prior years. There were 35–50% fewer preoperative and 3-month follow-up (THA) patients.

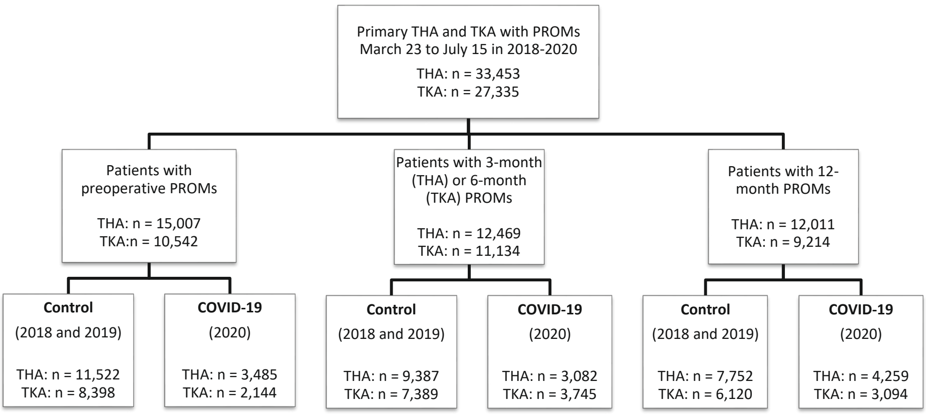

Figure 1: Flowchart of included primary THA and TKA patients with completed PROMs during defined inclusion windows. In a small number of patients with PROMs at multiple measurement points the postoperative follow-up PROM falls in an inclusion window in the subsequent year. Hence, the number of patients in the defined groups exceed the total number of patients.

Patient characteristics and representativeness

Characteristics of patients during the lockdown were compared with the control groups for THA and TKA separately (Supplemental File, Tables 1–2). Most notable differences occurred before surgery: during the lockdown both THA and TKA patients had a lower age and BMI, and more often had surgery in private clinics. THA patients slightly more often had a non-osteoarthritis indication, and TKA patients slightly more often were male and had a Charnley score of A/B1/B2 during the lockdown. At 6- and 12-month follow-up in TKA, and at 12-month follow-up in THA, patients more often had an ASA score of III–IV. Other characteristics of patients with postoperative PROMs were comparable during the lockdown.

Characteristics of responders were also compared with non-responders for THA and TKA separately at each follow-up point (Supplemental File, Tables 3–6). The comparison of patterns confirmed that elective care was reduced during the lockdown and also confirmed the above-mentioned differences in patient characteristics. Responders at all follow-up points were slightly younger, more often were male, more often had joint replacement for the indication osteoarthritis, and had better orthopedic (i.e., Charnley score) and general vitality (i.e., ASA score) scores. During the lockdown similar patterns emerged, except for type of hospital. In control years, preoperative responders had surgery in general hospitals more often, whilst during the lockdown they were more likely to have had surgery in private clinics.

PROMs

The estimated mean EQ-5D index for each comparison between PROMs retrieved during the COVID-19 lockdown and control groups is presented in Figure 2. In THA patients, the adjusted EQ-5D index and OHS were slightly worse during the COVID-19 lockdown (Supplemental File, Table 7). In TKA patients, both PROMs were unchanged during the lockdown. For adjusted mean differences in EQ-VAS and HOOS-PS/KOOS-PS see Table 8 in Supplementary File.

In THA patients, at 3-month and 12-month follow-up the EQ-5D index and OHS were slightly better during the lockdown. In TKA patients, the EQ-5D index and OKS did not differ at 6-month follow-up, whilst at 12-month follow-up they were slightly better. All identified differences and CI bounds did not exceed predefined MCIDs.

Subgroup analyses

Several interaction terms reached significance; however, they were not consistent across PROMs or follow-up points. Only in THA patients at 12-month follow-up did the term for inhabiting a region with ≥ 50 infections per 100,000 inhabitants reach statistical significance in both PROMs. Higher EQ-5D index and OHS scores (EQ-5D AMD 0.014, CI 0.007 to 0.022; OHS AMD 0.93, CI 0.60 to 1.27) were counteracted (EQ-5D interaction term –0.012, CI –0.022 to –0.001; OHS interaction term –0.58, CI –1.03 to –0.13).

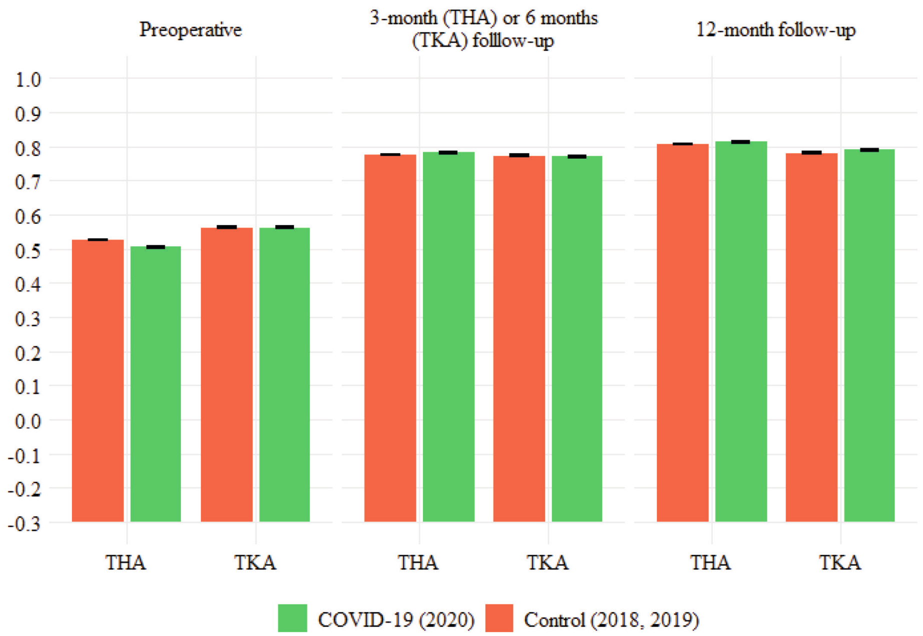

Figure 2: Estimated mean EQ-5D index scores for COVID-19 and control groups, for primary THA and primary TKA separately. A higher EQ-5D index score represents better health status. The error bars represent the 95% confidence intervals of the predicted mean; the upper and lower limit are very close to each other due to the large number of patients for each bar.

DISCUSSION

Our study demonstrated only small differences in preoperative and postoperative PROMs retrieved during the COVID-19 lockdown compared with previous years in Dutch THA and TKA patients. In THA patients most preoperative PROMs were slightly worse during the lockdown, which was not the case in TKA patients. This suggests a selection effect based on urgency assessment reflected in preoperative PROMs in THA. In both THA and TKA patients, most postoperative PROMs were slightly better during the lockdown. Contrary to our hypotheses, this suggests that the rehabilitation process remained mostly unaffected, and that an overall negative lockdown effect was not observed. Identified differences and CI bounds of all PROMs used in this study did not exceed predefined MCIDs, and are likely not clinically relevant.

This study had several strengths. 2 additional area-based variables were linked to the LROI data set, namely SES and COVID-19 infection rate. Also, multiple PROMs were used at 3 measurement moments, and an extensive set of potential confounders was included. Finally, given the study design (national registry study) the results can be considered generalizable for Dutch orthopedics.

This study had the following limitations. First, in this cross-sectional study we did not explore the effect of the lockdown on the longitudinal change scores (i.e., difference between pre- and postoperative PROMs). These analyses would provide information on a different hypothesis, i.e., whether a lockdown effect before surgery could persist into recovery. Furthermore, due to privacy laws the area-based variables were not allowed to contain patient-identifiable information. Therefore, we could not study whether regional differences in PROMs response percentage and representativeness occurred. Additionally, the response rate of PROMs is relatively low in the LROI. However, response rates remained unchanged during the lockdown. Lastly, although the lockdown in the Netherlands shared many similarities with other countries during the first COVID-19 wave, certain differences may still influence the generalizability of results to other countries.

Our study confirms the reduction in elective joint replacements during lockdown, and that a shift of orthopedic care from general hospitals to private clinics occurred during the first COVID-19 lockdown. Furthermore, younger patients with a lower BMI were selected for arthroplasty during the lockdown. Early on in the pandemic focus had shifted towards how to prioritize treatment with reduced capacity for elective surgery[30-32]. As the formal announcement of the Dutch Orthopaedic Society was made public halfway through the first lockdown, it could reflect a shift in the collective clinical opinion of Dutch orthopedic surgeons. They may have been aware of risk factors for worse COVID-19 infections such as high age and BMI, and (sub-)consciously selected candidates with a lower risk. A limitation of registry data is that we are unable to discern to what extent self-selection played a role, e.g., patients delaying the procedure themselves because of fear of becoming infected in the hospital.

Besides selection based on characteristics, we additionally found evidence of selection based on urgency. Even after adjustment, THA patients had worse preoperative EQ-5D index and disease-specific PROMs during the COVID-19 lockdown compared with previous years. In TKA patients similar PROMs scores during COVID-19 were observed compared with control years; we do not have a definite explanation for this contrast with THA patients.

We noted no difference or, rather, a slight improvement in most PROMs scores in both THA and TKA patients compared with previous years, an unexpected finding. This was already apparent in the short term (3- and 6-month for THA and TKA respectively), which indicates that rehabilitation success was unaffected. All arthroplasty patients in the Netherlands receive an unsupervised exercise schedule after surgery, and generally receive postoperative outpatient exercise therapy subsequently. It is possible that unsupervised therapy was sufficient, a notion supported by contemporary systematic reviews[12, 13]. However, this finding should be interpreted with caution. Although during the lockdown physical therapists had to close their physical practice, in approximately 25% of patients they continued via telemedicine, which is not recorded in the LROI[33]. Moreover, many patients included in the short term cross-sectional comparison will have had surgery sometime before the lockdown, and thus might have already initiated physical therapy.

A global scientific body reported severe psychological stress during the COVID-19 pandemic. This is also reflected in studies using the EQ-5D, where lower scores are found in different populations from different nationalities including Dutch, mainly driven by poorer scores for pain/discomfort and anxiety/depression[3, 4]. The slightly higher postoperative PROMs indicate there was no direct negative effect of the lockdown in this Dutch orthopedic population. In THA patients, inhabiting a region with a COVID-19 high infection rate appeared to negate the improvement in EQ-5D and OHS during the lockdown. However, this effect was minimal and resulted in approximately equal scores compared with prior to the lockdown, which we do not believe provides sufficient evidence of an effect of this interaction term. These findings are in line with a recent study on the impact of COVID-19 on PROMs in hand–wrist patients, which had a similar cross-sectional study design[34]. Combined, these results attest to the fact that the general and the orthopedic population do not necessarily experience the same impact of the COVID-19 lockdown.

A potential explanation for the discrepancy in EQ-5D index between the Dutch arthroplasty patient group and the general population is “response-scale heterogeneity,” which refers to the difference in the way individuals interpret a response scale, i.e., a PROM. For instance, if a difference in health occurs between two groups, this may reflect a true difference in health or that the groups perceive the response scale differently due to psychological mechanisms[35]. In this population, this may have been caused by feeling privileged during the lockdown: patients considered themselves lucky to have already had their hip or knee replacement before the pandemic hit.

In conclusion, we demonstrated that PROMs scores in Dutch primary THA and TKA patients during the COVID-19 lockdown were hardly affected. Orthopedic surgeons were forced to delay elective surgery due to COVID-19 and also to identify the best candidates for surgery. The observed lower preoperative PROMs scores of THA patients during the COVID-19 lockdown could indicate a (sub-)conscious selection effect based on urgency. Postoperative PROMs in both THA and TKA patients differed minimally. Overall, differences found were likely not clinically relevant.

REFERENCES

1. Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int J Antimicrob Agents 2020; 55: 105924.
2. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, Rubin GJ. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020; 395: 912-920.
3. van Ballegooijen H, Goossens L, Bruin RH, Michels R, Krol M. Concerns, quality of life, access to care and productivity of the general population during the first 8 weeks of the coronavirus lockdown in Belgium and the Netherlands. BMC Health Services Research 2021; 21: 227.
4. Hay JW, Gong CL, Jiao X, Zawadzki NK, Zawadzki RS, Pickard AS, et al. A US Population Health Survey on the Impact of COVID-19 Using the EQ-5D-5L. J Gen Intern Med 2021; 36: 1292-1301.
5. Horesh D, Kapel Lev-Ari R, Hasson-Ohayon I. Risk factors for psychological distress during the COVID-19 pandemic in Israel: Loneliness, age, gender, and health status play an important role. Br J Health Psychol 2020; 25: 925-933.
6. Ping W, Zheng J, Niu X, Guo C, Zhang J, Yang H, Shi Y. Evaluation of health-related quality of life using EQ-5D in China during the COVID-19 pandemic. PloS one 2020; 15: e0234850-e0234850.
7. Hoekman LM, Smits MMV, Koolman X. The Dutch COVID-19 approach: Regional differences in a small country. Health Policy Technol 2020; 9: 613-622.
8. Santomauro DF, Mantilla Herrera AM, Shadid J, Zheng P, Ashbaugh C, Pigott DM, et al. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021; 398: 1700-1712.
9. Dutch Orthopaedic National Quality Registry (LROI) Annual report 2021. Available at: www.lroi-rapportage.nl.
10. Nederlands Orthopaedische Vereniging (NOV) Statement on how to reinitiate orthopedic care during the COVID-19 lockdown. 1st of May, 2020. Available at: https://bit.ly/2yjf0UN.
11. Dutch Coöperating Quality Registries (SKR) Impact Report 2021. Available at: https://skr-zorg.nl/impact-report/.
12. Hansen S, Aaboe J, Mechlenburg I, Overgaard S, Mikkelsen LR. Effects of supervised exercise compared to non-supervised exercise early after total hip replacement on patient-reported function, pain, health-related quality of life and performance-based function - a systematic review and meta-analysis of randomized controlled trials. Clin Rehabil 2019; 33: 13-23.
13. Wang D, Wu T, Li Y, Jia L, Ren J, Yang L. A systematic review and meta-analysis of the effect of preoperative exercise intervention on rehabilitation after total knee arthroplasty. Ann Palliat Med 2021; 10: 10986-10996.
14. van Steenbergen LN, Denissen GAW, Spooren A, van Rooden SM, van Oosterhout FJ, Morrenhof JW, Nelissen RGHH. More than 95% completeness of reported procedures in the population-based Dutch Arthroplasty Register. Acta Orthop 2015; 86: 498-505.
15. National Institute for Public Health and the Environment (RIVM) COVID-19 actual report. Accessed October 23rd, 2020. Available at: https://www.rivm.nl/coronavirus-covid-19/actueel.
16. Dutch Institute of Social Research (SCP) Socioenomic status score per postal code area. Retrieved October 20th, 2020. Available by request at: https://bronnen.zorggegevens.nl/Bron?naam=Sociaal-Economische-Status-per-postcodegebied.

17. Subramanian SV, Chen JT, Rehkopf DH, Waterman PD, Krieger N. Comparing individual- and area-based socioeconomic measures for the surveillance of health disparities: A multilevel analysis of Massachusetts births, 1989-1991. Am J Epidemiol 2006; 164: 823-834.
18. Brooks R. EuroQol: the current state of play. Health Policy 1996; 37: 53-72.
19. Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, Dawson J. The use of the Oxford hip and knee scores. J Bone Joint Surg Br 2007; 89: 1010-1014.
20. Nilsdotter AK, Lohmander LS, Klässbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS)--validity and responsiveness in total hip replacement. BMC Musculoskelet Disord 2003; 4: 10.
21. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998; 28: 88-96.
22. Lamers LM, McDonnell J, Stalmeier PF, Krabbe PF, Busschbach JJ. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Econ 2006; 15: 1121-1132.
23. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. BMC Med Res Methodol 2008; 8: 70.
24. Garriga C, Leal J, Sánchez-Santos MT, Arden N, Price A, Prieto-Alhambra D, et al. Geographical Variation in Outcomes of Primary Hip and Knee Replacement. JAMA Netw 2019; 2: e1914325-e1914325.
25. Peters RM, van Steenbergen LN, Stewart RE, Stevens M, Rijk PC, Bulstra SK, Zijlstra WP. Which patients improve most after total hip arthroplasty? Influence of patient characteristics on patient-reported outcome measures of 22,357 total hip arthroplasties in the Dutch Arthroplasty Register. HIP International 2020; 31: 593-602.
26. Tolk JJ, Waarsing JH, Janssen RPA, van Steenbergen LN, Bierma-Zeinstra SMA, Reijman M. Development of Preoperative Prediction Models for Pain and Functional Outcome After Total Knee Arthroplasty Using The Dutch Arthroplasty Register Data. J Arthroplasty 2020; 35: 690-698.e692.
27. Neuburger J, Hutchings A, Black N, van der Meulen JH. Socioeconomic differences in patient-reported outcomes after a hip or knee replacement in the English National Health Service. J Public Health (Oxf) 2013; 35: 115-124.
28. Beard DJ, Harris K, Dawson J, Doll H, Murray DW, Carr AJ, Price AJ. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. J Clin Epidemiol 2015; 68: 73-79.
29. Coretti S, Ruggeri M, McNamee P. The minimum clinically important difference for EQ-5D index: a critical review. Expert Review of Pharmacoeconomics & Outcomes Research 2014; 14: 221-233.
30. Logishetty K, Edwards TC, Subbiah Ponniah H, Ahmed M, Liddle AD, Cobb J, Clark C. How to prioritize patients and redesign care to safely resume planned surgery during the COVID-19 pandemic. Bone & Joint Open 2021; 2: 134-140.
31. Vles GF, Ghijselings S, De Ryck I, Meyfroidt G, Sweeney NA, Oosterlinck W, et al. Returning to Elective Orthopedic Surgery During the COVID-19 Pandemic: A Multidisciplinary and Pragmatic Strategy for Initial Patient Selection. J Patient Saf 2020; 16.
32. Zorzi C, Piovan G, Screpis D, Natali S, Marocco S, Iacono V. Elective Orthopaedic Surgery During COVID-19: A Safe Way to Get Back on Track. JB JS Open Access 2020; 5.

33. Nivel. Use of paramedic care during COVID-19, consequences for first line exercise therapy. Report week 2 - 22, 2020. Available at: https://www.nivel.nl/nl/publicatie/gevolgen-corona-voor-de-eerstelijns-oefen-en-fysiotherapiepraktijk-gebruik-van#:~:text=Na%20vier%20weken%20lag%20het,%2Duitbraak%20(begin%202020).
34. Cohen A, Selles RW, De Ridder WA, Ter Stege MHP, Souer JS, Wouters RM, Hand–Wrist Study Group C. What Is the Impact of the COVID-19 Pandemic on Quality of Life and Other Patient-reported Outcomes? An Analysis of the Hand-Wrist Study Cohort. Clin Orthop Relat Res 2021; 479: 335-345.
35. Knott RJ, Black N, Hollingsworth B, Lorgelly PK. Response-Scale Heterogeneity in the EQ-5D. Health Economics 2017; 26: 387-394.

6

DECLARATIONS

Acknowledgements: The authors would like to thank Dr S Baart for her help designing the analytical strategy.

Authors contributions: All authors fulfill the ICMJE authorship criteria. JB: data curation, formal analysis, visualization, writing - original Draft, writing - review & editing, funding acquisition, project administration. LG: formal analysis, writing - original draft, writing - review & editing. AC: conceptualization, writing - review & editing. JV: writing - original draft, writing - review & editing, funding acquisition. MG: validation, writing - original draft, writing - review & editing. AS: data curation, writing - original draft. GB: conceptualization, validation, writing - original draft, writing - review & editing, funding acquisition, supervision. MR: conceptualization, validation, writing - original draft, writing - review & editing, funding acquisition, supervision.

Supplementary data: Supplementary data associated with this article can be found at the end of this chapter and in the online version at doi: 10.2340/17453674.2022.4856.

SUPPLEMENTAL FILE, TABLES 1–8

Table 1: Characteristics of primary THA patients with completed PROMs during inclusion windows

Factor	Preoperative			3-month postoperative			12-month postoperative		
	Control (n = 11,522)	COVID-19 (n = 3,485)	p-value	Control (n = 9,387)	COVID-19 (n = 3,082)	p-value	Control (n = 7,752)	COVID-19 (n = 4,259)	p-value
Age, mean (SD)	69 (10)	68 (10)	0.006	69 (9.8)	68 (10)	0.05	69 (10)	69 (10)	0.4
BMI, mean (SD)	27.4 (4.6)	27.0 (4.3)	< 0.001	27.3 (4.5)	27.2 (4.3)	0.2	27.4 (4.5)	27.4 (4.5)	1
Female sex[a]	7,415 (64)	2,225 (64)	0.6	5,873 (63)	1,922 (62)	0.9	4,992 (64)	2,698 (63)	0.3
Previous surgery on the joint	475 (4.1)	158 (4.5)	0.3	344 (3.7)	109 (3.5)	0.7	288 (3.7)	164 (3.9)	
ASA classification[a]			0.1			0.2			< 0.001
I–II	9,018 (78)	2,772 (80)		7,558 (81)	2,515 (82)		6,394 (82)	3,327 (78)	
III–IV	2,504 (22)	712 (20)		1,829 (19)	565 (18)		1,358 (18)	932 (22)	
Type of hospital[a]			< 0.001			< 0.001			< 0.001
General	10,285 (89)	2,748 (79)		8,103 (86)	2,521 (82)		6,903 (89)	3,663 (86)	
Private	1,002 (8.7)	683 (20)		1,086 (12)	499 (16)		684 (8.8)	488 (11)	
Academic	235 (2.0)	54 (1.5)		198 (2.1)	62 (2.0)		165 (2.1)	108 (2.5)	
Indication[a]			< 0.001			0.8			0.6
Osteoarthritis	10,696 (93)	3,160 (91)		8,727 (93)	2,860 (93)		7,196 (93)	3,940 (93)	
Non-osteoarthritis	818 (7.1)	319 (9.2)		658 (7.0)	221 (7.2)		556 (7.2)	317 (7.4)	
Charnley classification[a]			0.1			0.3			0.6
A/B1/B2	10,975 (95)	3,304 (95)		8,986 (96)	2,930 (95)		7,406 (96)	4,059 (95)	
C	397 (3.4)	105 (3.0)		271 (2.9)	99 (3.2)		249 (3.2)	136 (3.2)	
N/A	146 (1.3)	58 (1.7)		127 (1.4)	50 (1.6)		95 (1.2)	62 (1.5)	
Socioeconomic status score[a]			0.3			0.01			0.8
3rd–5th quintile (highest)	6,357 (55)	1,954 (56)		5,404 (58)	1,707 (55)		4,382 (57)	2,407 (57)	
1st–2nd quintile (lowest)	5,058 (44)	1,497 (43)		3,853 (41)	1,349 (44)		3,302 (43)	1,830 (43)	

Table 1: Continued

Factor	Preoperative			3-month postoperative			12-month postoperative		
	Control (n = 11,522)	COVID-19 (n = 3,485)	p-value	Control (n = 9,387)	COVID-19 (n = 3,082)	p-value	Control (n = 7,752)	COVID-19 (n = 4,259)	p-value
Regional COVID infection rate per 100.000[a]									
≤ 24	11,522 (100)	2,956 (85)	–	9,387 (100)	1,592 (52)	–	7,752 (100)	2,629 (62)	–
25 – 49	–	319 (9.2)		–	621 (20)		–	606 (14)	
50 – 74	–	80 (2.3)		–	346 (11)		–	413 (10)	
75 – 99	–	48 (1.4)		–	234 (7.6)		–	277 (6.5)	
≥ 100	–	49 (1.4)		–	256 (8.3)		–	312 (7.3)	

Data is presented as n (%), unless reported otherwise. The control groups were subsequently compared to the respective COVID-19 groups except for the variable infection rate.
[a] Numbers do not always add up due to missing or unknown values.

6

Table 2: Characteristics of primary TKA patients with completed PROMs during inclusion windows

Factor	Preoperative			6-month postoperative			12-month postoperative		
	Control (n = 8,398)	COVID-19 (n = 2,144)	p-value	Control (n = 7,389)	COVID-19 (n = 3,745)	p-value	Control (n = 6,120)	COVID-19 (n = 3,094)	p-value
Age, mean (SD)	69 (8.9)	69 (8.5)	0.02	68 (8.4)	69 (8.4)	0.1	69 (8.7)	69 (8.5)	0.9
BMI, mean (SD)	29.7 (5.0)	27.0 (4.3)	0.1	29.4 (4.8)	29.5 (4.8)	0.3	29.7 (5.0)	29.8 (4.9)	0.5
Female sex[a]	5,390 (64)	1,308 (61)	0.006	4,458 (60)	2,291 (61)	0.4	3,819 (62)	1,946 (63)	0.6
Previous surgery on the joint	2,098 (25)	527 (25)	0.2	2,049 (28)	948 (25)	< 0.001	1,682 (27)	787 (25)	0.6
ASA classification[a]			0.3			0.003			< 0.001
I–II	6,322 (75)	1,636 (76)		5,930 (80)	2,916 (78)		4,912 (80)	2,340 (76)	
III–IV	2,075 (25)	504 (24)		1,459 (20)	829 (22)		1,208 (20)	754 (24)	
Type of hospital[a]			< 0.001			< 0.001			< 0.001
General	7,399 (88)	1,715 (80)		6,312 (85)	3,064 (82)		5,274 (86)	2,563 (83)	
Private	840 (10)	412 (19)		983 (13)	617 (16)		765 (13)	470 (15)	
Academic	159 (1.9)	17 (0.8)		94 (1.3)	64 (1.7)		81 (1.3)	61 (2.0)	
Indication[a]			0.3			0.1			0.1
Osteoarthritis	8,117 (97)	2,056 (96)		7,199 (97)	3,626 (97)		5,957 (97)	2,990 (97)	
Non-osteoarthritis	275 (3.3)	79 (3.7)		190 (2.6)	114 (3.0)		163 (2.7)	98 (3.2)	
Charnley classification[a]			0.006			0.5			0.6
A/B1/B2	8,016 (95)	2,074 (97)		7,081 (96)	3,602 (96)		5,860 (96)	2,950 (95)	
C	359 (4.3)	59 (2.8)		300 (4.1)	135 (3.6)		247 (4.0)	135 (4.4)	
N/A	20 (0.2)	5 (0.2)		7 (0.1)	4 (0.1)		13 (0.2)	9 (0.3)	
Socioeconomic status score[a]			0.2			0.2			1
3rd–5th quintile (highest)	4,404 (52)	1,162 (54)		4,009 (54)	1,978 (53)		3,273 (53)	1,637 (53)	
1st–2nd quintile (lowest)	3,926 (47)	967 (45)		3,304 (45)	1,731 (46)		2,807 (46)	1,400 (45)	

Table 2: Continued

Factor	Preoperative			6-month postoperative			12-month postoperative		
	Control (n = 8,398)	COVID-19 (n = 2,144)	p-value	Control (n = 7,389)	COVID-19 (n = 3,745)	p-value	Control (n = 6,120)	COVID-19 (n = 3,094)	p-value
Regional COVID infection rate per 100.000[a]									
≤ 24	8,398 (100)	1,811 (84)	–	7,389 (100)	2,367 (63)	–	6,120 (100)	1,909 (62)	–
25 – 49	–	191 (8.9)		–	548 (15)		–	425 (14)	
50 – 74	–	50 (2.3)		–	338 (9.0)		–	284 (9.2)	
75 – 99	–	47 (2.2)		–	232 (6.2)		–	191 (6.2)	
≥ 100	–	31 (1.4)		–	225 (6.0)		–	228 (7.4)	

Data is presented as n (%), unless reported otherwise. The control groups were subsequently compared to the respective COVID-19 groups except for the variable infection rate.
[a] Numbers do not always add up due to missing or unknown values.

Table 3: Characteristics of primary THA patients who were operated in the same period based on the inclusion window in 2020 (COVID-19)

	Preoperative			3-month postoperative			12-month postoperative		
	Non-responders (n = 3,020)	Responders (n = 3,485)	p-value	Non-responders (n = 4,237)	Responders (n = 3,082)	p-value	Non-responders (n = 6,267)	Responders (n = 4,259)	p-value
Age, mean (SD)	69 (11)	68 (10)	0.02	69 (11)	68 (10)	< 0.001	70 (11)	69 (10)	< 0.001
BMI, mean (SD)	26.8 (4.7)	27.0 (4.3)	0.03	27.3 (4.7)	27.2 (4.4)	0.5	27.3 (4.7)	27.4 (4.5)	0.5
Female sex[a]	1,986 (66)	2,225 (64)	0.1	2,787 (66)	1,922 (62)	0.003	4,148 (66)	2,698 (63)	0.003
Previous surgery on the joint	177 (5.9)	158 (4.5)	0.02	207 (4.9)	109 (3.5)	0.005	313 (5.0)	164 (3.9)	0.002
ASA classification[a]			< 0.001			< 0.001			< 0.001
I–II	2,190 (73)	2,772 (80)		3,107 (73)	2,515 (82)		4,647 (74)	3,327 (78)	
III–IV	829 (27)	712 (20)		1,127 (27)	565 (18)		1,620 (26)	932 (22)	
Type of hospital[a]			< 0.001			< 0.001			< 0.001
General	2,569 (85)	2,748 (79)		3,810 (90)	2,521 (82)		5,516 (88)	3,663 (86)	
Private	347 (11)	683 (20)		320 (7.6)	499 (16)		562 (9.0)	488 (11)	
Academic	104 (3.4)	54 (1.5)		107 (2.5)	62 (2.0)		189 (3.0)	108 (2.5)	
Indication[a]			< 0.001			< 0.001			< 0.001
Osteoarthritis	2,115 (70)	3,160 (91)		3,411 (81)	2,860 (93)		5,112 (82)	3,940 (93)	
Non-osteoarthritis	901 (30)	319 (9.2)		824 (19)	221 (7.2)		1,150 (18)	317 (7.4)	
Charnley classification[a]			< 0.001			< 0.001			< 0.001
A/B1/B2	2,438 (81)	3,304 (95)		3,660 (86)	2,921 (95)		5,589 (89)	4,051 (95)	
C	88 (2.9)	105 (3.0)		128 (3.0)	99 (3.2)		160 (2.6)	136 (3.2)	
N/A	479 (16)	58 (1.7)		429 (10)	50 (1.6)		495 (7.9)	62 (1.5)	
Socioeconomic status score[a]			0.2			0.9			0.1
3rd–5th quintile (highest)	1,648 (55)	1,954 (56)		2,353 (56)	1,707 (55)		3,371 (54)	2,407 (57)	
1st–2nd quintile (lowest)	1,347 (45)	1,497 (43)		1,850 (44)	1,349 (44)		2,723 (43)	1,830 (43)	

The PROMs responders were compared with the respective non-responders. Data is presented as n (%), unless reported otherwise.
[a] Numbers do not always add up due to missing or unknown values.

Table 4: Characteristics of primary THA patients who were operated in the same period based on the inclusion window in 2018 plus 2019 (control)

Factor	Preoperative			3-month postoperative			12-month postoperative		
	Non-responders (n = 12,110)	Responders (n = 11,522)	p-value	Non-responders (n = 11,543)	Responders (n = 9,387)	p-value	Non-responders (n = 11,640)	Responders (n = 7,752)	p-value
Age, mean (SD)	70 (10)	69 (10)	< 0.001	69 (11)	68 (9.9)	< 0.001	70 (11)	69 (10)	< 0.001
BMI, mean (SD)	27.1 (4.6)	27.4 (4.6)	< 0.001	27.2 (4.6)	27.3 (4.5)	0.03	27.2 (4.7)	27.4 (4.5)	0.005
Female sex[a]	7,698 (64)	7,415 (64)	0.02	7,606 (66)	5,873 (62)	< 0.001	7,781 (67)	4,992 (64)	< 0.001
Previous surgery on the joint	646 (5.3)	475 (4.1)	< 0.001	604 (5.2)	344 (3.7)	< 0.001	601 (5.2)	288 (3.7)	< 0.001
ASA classification[a]			0.03			< 0.001			< 0.001
I–II	9,333 (77)	9,018 (78)		8,861 (77)	7,558 (81)		9,013 (77)	6,394 (82)	
III–IV	2,775 (23)	2,504 (22)		2,680 (23)	1,829 (19)		2,625 (23)	1,358 (18)	
Type of hospital[a]			< 0.001			< 0.001			< 0.001
General	10,250 (85)	10,285 (89)		10,086 (87)	8,103 (86)		10,366 (89)	6,903 (89)	
Private	1,469 (12)	1,002 (8.7)		1,110 (9.6)	1,086 (12)		921 (7.9)	684 (8.8)	
Academic	391 (3.2)	235 (2.0)		347 (3.0)	198 (2.1)		353 (3.0)	165 (2.1)	
Indication[a]			< 0.001			< 0.001			< 0.001
Osteoarthritis	9,727 (80)	10,696 (93)		9,585 (83)	8,727 (93)		9,634 (83)	7,196 (93)	
Non-osteoarthritis	2,365 (20)	818 (7.1)		1,943 (17)	658 (7.0)		1,993 (17)	556 (7.2)	
Charnley classification[a]			< 0.001			< 0.001			< 0.001
A/B1/B2	10,678 (88)	10,945 (95)		10,410 (90)	8,956 (95)		10,504 (90)	7,379 (95)	
C	321 (2.7)	397 (3.5)		312 (2.7)	271 (2.9)		361 (3.1)	249 (3.2)	
N/A	1,056 (8.7)	146 (1.3)		773 (6.7)	127 (1.4)		724 (6.2)	95 (1.2)	
Socioeconomic status score[a]			0.3			< 0.001			< 0.001
3rd–5th quintile (highest)	6,538 (54)	6,357 (55)		6,144 (53)	5,404 (58)		6,252 (54)	4,382 (57)	
1st–2nd quintile (lowest)	5,333 (44)	5,058 (44)		5,236 (45)	3,853 (41)		5,282 (45)	3,302 (43)	

The PROMs responders were compared with the respective non-responders. Data is presented as n (%), unless reported otherwise.
[a] Numbers do not always add up due to missing or unknown values.

Table 5: Characteristics of primary TKA patients who were operated in the same period based on the inclusion window in 2020 (COVID-19)

	Preoperative			6-month postoperative			12-month postoperative		
Factor	**Non-responders (n = 1,857)**	**Responders (n = 2,144)**	**p-value**	**Non-responders (n = 5,079)**	**Responders (n = 3,745)**	**p-value**	**Non-responders (n = 4,608)**	**Responders (n = 3,094)**	**p-value**
Age, mean (SD)	69 (9.2)	69 (8.5)	0.2	69 (8.8)	69 (8.4)	< 0.001	70 (9.0)	69 (8.5)	< 0.001
BMI, mean (SD)	29.6 (5.1)	29.5 (5.0)	0.6	29.6 (5.0)	29.5 (4.8)	0.2	30.1 (5.1)	29.8 (4.9)	0.01
Female sex[a]	1,187 (64)	1,308 (61)	0.06	3,212 (63)	2,291 (61)	0.05	3,083 (67)	1,946 (63)	< 0.001
Previous surgery on the joint	446 (24)	527 (25)	0.7	1,074 (21)	948 (25)	0.001	941 (20)	787 (25)	< 0.001
ASA classification[a]			0.2			< 0.001			0.001
I–II	1,388 (75)	1,636 (76)		3,778 (74)	2,916 (78)		3,320 (72)	2,340 (76)	
III–IV	466 (25)	504 (24)		1,298 (26)	829 (22)		1,287 (28)	754 (24)	
Type of hospital[a]			< 0.001			< 0.001			< 0.001
General	1,470 (79)	1,715 (80)		4,301 (85)	3,064 (82)		3,974 (86)	2,563 (83)	
Private	342 (18)	412 (19)		677 (13)	617 (16)		533 (12)	470 (15)	
Academic	45 (2.4)	17 (0.8)		101 (2.0)	64 (1.7)		101 (2.2)	61 (2.0)	
Indication[a]			0.006			0.1			0.1
Osteoarthritis	1,750 (94)	2,056 (96)		4,879 (96)	3,626 (97)		4,418 (96)	2,990 (97)	
Non-osteoarthritis	102 (5.5)	79 (3.7)		189 (3.7)	114 (3.0)		175 (3.8)	98 (3.2)	
Charnley classification[a]			0.003			0.2			0.04
A/B1/B2	1,762 (95)	2,061 (96)		4,870 (96)	3,588 (96)		4,419 (96)	2,932 (95)	
C	72 (3.9)	59 (2.8)		174 (3.4)	135 (3.6)		150 (3.3)	135 (4.4)	
N/A	16 (0.9)	5 (0.2)		14 (0.3)	4 (0.1)		14 (0.3)	9 (0.3)	
Socioeconomic status score[a]			1			0.4			0.1
3rd–5th quintile (highest)	1,009 (54)	1,162 (54)		2,596 (51)	1,978 (53)		2,303 (50)	1,637 (53)	
1st–2nd quintile (lowest)	841 (45)	967 (45)		2,360 (46)	1,731 (46)		2,116 (46)	1,400 (45)	

The PROMs responders were compared with the respective non-responders. Data is presented as n (%), unless reported otherwise.
[a] Numbers do not always add up due to missing or unknown values.

Table 6: Characteristics of primary TKA patients who were operated in the same period based on the inclusion window in 2018 plus 2019 (control)

Factor	Preoperative			6-month postoperative			12-month postoperative		
	Non-responders (n = 8,653)	Responders (n = 8,398)	p-value	Non-responders (n = 9,786)	Responders (n = 7,389)	p-value	Non-responders (n = 9,130)	Responders (n = 6,120)	p-value
Age, mean (SD)	69 (9.2)	69 (8.9)	0.7	69 (9.2)	68 (8.4)	< 0.001	69 (9.6)	69 (8.7)	< 0.001
BMI, mean (SD)	29.9 (5.2)	29.7 (5.0)	0.01	29.7 (5.0)	29.4 (4.8)	< 0.001	30.0 (5.2)	29.7 (5.0)	0.001
Female sex[a]	5,521 (64)	5,390 (64)	0.6	6,188 (63)	4,458 (60)	< 0.001	6,009 (66)	3,819 (62)	< 0.001
Previous surgery on the joint	1,921 (22)	2,098 (25)	0.01	2,322 (24)	2,049 (28)	< 0.001	2,272 (25)	1,682 (27)	< 0.001
ASA classification[a]			0.1			0.01			< 0.001
I–II	6,604 (76)	6,322 (75)		7,695 (79)	5,930 (80)		6,912 (76)	4,912 (80)	
III–IV	2,047 (24)	2,075 (25)		2,089 (21)	1,459 (20)		2,215 (24)	1,208 (20)	
Type of hospital[a]			< 0.001			< 0.001			< 0.001
General	6,999 (81)	7,399 (88)		8,251 (84)	6,312 (85)		7,903 (87)	5,274 (86)	
Private	1,463 (17)	840 (10)		1,321 (13)	983 (13)		970 (11)	765 (13)	
Academic	191 (2.2)	159 (1.9)		214 (2.2)	94 (1.3)		257 (2.8)	81 (1.3)	
Indication[a]			0.1			< 0.001			< 0.001
Osteoarthritis	8,324 (96)	8,117 (97)		9,415 (96)	7,199 (97)		8,752 (96)	5,957 (97)	
Non-osteoarthritis	325 (3.8)	275 (3.3)		316 (3.2)	190 (2.6)		376 (4.1)	163 (2.7)	
Charnley classification[a]			< 0.001			< 0.001			0.04
A/B1/B2	8,332 (96)	7,984 (95)		9,394 (96)	7,053 (96)		8,750 (96)	5,843 (95)	
C	263 (3.0)	359 (4.3)		321 (3.3)	300 (4.1)		310 (3.4)	247 (4.0)	
N/A	32 (0.4)	20 (0.2)		39 (0.4)	7 (0.1)		31 (0.3)	13 (0.2)	
Socioeconomic status score[a]			0.6			< 0.001			< 0.001
3rd–5th quintile (highest)	4,388 (51)	4,404 (52)		5,002 (51)	4,009 (54)		4,541 (50)	3,273 (53)	
1st–2nd quintile (lowest)	3,984 (46)	3,926 (47)		4,609 (47)	3,304 (45)		4,409 (48)	2,807 (46)	

The PROMs responders were compared with the respective non-responders. Data is presented as n (%), unless reported otherwise.
[a] Numbers do not always add up due to missing or unknown values.

Table 7: Adjusted mean differences (AMD) with 95% confidence intervals (CIs) of EQ-5D index and OHS/OKS during the COVID-19 lockdown compared with the control periods.

	Preoperative		3-/6-month postoperative		12-month postoperative	
Factor	**AMD (CI)**	**p-value**	**AMD (CI)**	**p-value**	**AMD (CI)**	**p-value**
Total Hip Arthroplasty	n = 15,007		n = 12,469		n = 12,011	
Without interaction terms						
EQ-5D index[a]	−0.021 (−0.029 to −0.013)	< 0.001	0.008 (0.001 to 0.015)	0.03	0.010 (0.003 to 0.016)	0.003
With interaction terms	*no relevant interaction terms*					
BMI > 30 EQ-5D index[a]			−0.019 (−0.035 to – 0.002) 0.012 (0.005 to 0.020)	0.02 0.002		
Infections ≥ 50 per 100.000 EQ-5D index[a]					– 0.012 (−0.022 to −0.001) 0.014 (0.007 to 0.022)	0.03 < 0.001
Without interaction terms						
OHS[a]	−0.95 (−1.30 to −0.60)	< 0.001	0.44 (0.13 to 0.74)	0.005	0.71 (0.42 to 1.00)	< 0.001
With interaction terms	*no relevant interaction terms*		*no relevant interaction terms*			
Infections ≥ 50 per 100.000 OHS[a]					−0.58 (−1.03 to −0.13) 0.93 (0.60 to 1.27)	0.01 < 0.001
Total Knee Arthroplasty	n = 10,542		n = 11,134		n = 9,214	
Without interaction terms						
EQ-5D index[a]	−0.003 (−0.012 to 0.007)	0.6	0.000 (−0.007 to 0.007)	1.0	0.013 (0.006 to 0.021)	< 0.001
With interaction terms	–[b]		–[b]		*no relevant interaction terms*	
Without interaction terms						
OKS[a]	−0.03 (−0.36 to 0.41)	0.9	0.12 (−0.22 to 0.47)	0.5	0.45 (0.07 to 0.83)	0.02
With interaction terms	–[b]		–[b]		*no relevant interaction terms*	

Analyses were adjusted for age, sex, BMI, previous surgery on the operated joint, indication, ASA score, Charnley classification, type of hospital, and socio-economic status score.
[a] A higher EQ-5D index, OHS/OKS represents an improvement.
[b] Interaction terms were not analyzed in non-significant associations.

Table 8: Adjusted mean differences (AMD) with 95% confidence intervals (CIs) of EQ-VAS and HOOS-PS/KOOS-PS during the COVID-19 lockdown compared with the control periods.

	Preoperative		3-/6-month postoperative		12-month postoperative	
Factor	**AMD (CI)**	**p-value**	**AMD (CI)**	**p-value**	**AMD (CI)**	**p-value**
Total Hip Arthroplasty	n = 15,007		n = 12,469		n = 12,011	
Without interaction terms						
EQ-VAS[a]	0.11 (–0.71 to 0.93)	0.8	2.16 (1.40 to 2.92)	< 0.001	3.11 (2.38 to 3.85)	< 0.001
With interaction terms	–[c]		*no relevant interaction terms*		*no relevant interaction terms*	
Without interaction terms						
HOOS-PS[b]	2.15 (1.43 to 2.87)	< 0.001	–0.38 (–1.00 to 0.24)	0.2	–0.77 (–1.35 to –0.20)	0.009
With interaction terms			–[c]		*no relevant interaction terms*	
BMI >30 HOOS-PS[b]	–1.88 (–3.60 to –0.16) 2.60 (1.78 to 3.41)	0.03 < 0.001				
SES 1[st] –2[nd] quintile (lowest) HOOS-PS[b]	–1.55 (–3.00 to –0.09) 2.82 (1.88 to 3.76)	0.04 < 0.001				
Total Knee Arthroplasty	n = 10,542		n = 11,134		n = 9,214	
Without interaction terms:						
EQ-VAS[a]	0.76 (–0.20 to 1.73)	0.1	1.82 (1.08 to 2.57)	< 0.001	1.92 (1.06 to 2.79)	< 0.001
With interaction terms:	–[c]				*no relevant interaction terms*	
ASA ≥3 EQ-VAS[a]			2.74 (0.82 to 4.67) 1.23 (0.42 to 2.05)	0.005 0.003		
Age >70 EQ-VAS[a]			2.47 (0.98 to 3.96) 0.64 (–0.36 to 1.64)	0.001 0.2		
Without interaction terms:						
KOOS-PS[b]	0.28 (–0.49 to 1.04)	0.5	0.49 (–0.09 to 1.06)	0.1	–0.77 (–1.48 to –0.07)	0.03
With interaction terms:	–[c]		–[c]		*no relevant interaction terms*	

Analyses were adjusted for age, sex, BMI, previous surgery on the operated joint, indication, ASA score, Charnley classification, type of hospital, and socio-economic status score.
[a] A higher EQ-VAS represents an improvement.
[b] A lower HOOS-PS/KOOS-PS represents an improvement.
[c] Interaction terms were not analyzed in non-significant associations.

Chapter 7

A head-to-head comparison of the adult EQ-5D-5L and youth EQ-5D-Y-5L in adolescents with idiopathic scoliosis

Authors

Joshua M. Bonsel[1], MD

Charles M.M. Peeters[2,3], MD, PhD

Max Reijman[1], MSc, PhD

Tim Dings[1], MSc

Joost P.H.J. Rutges[1], MD, PhD

Diederik H.R. Kempen[4,5], MD, PhD

Jan A.N. Verhaar[1], MD, PhD, emeritus professor

Gouke J. Bonsel[6], MD, PhD, emeritus professor

Affiliations

[1] Department of Orthopaedics and Sports Medicine, Erasmus MC, University Medical Center, Rotterdam, The Netherlands

[2] Department of Orthopaedics, University Medical Center Groningen, Groningen, The Netherlands

[3] Department of Orthopaedics, Isala Hospital, Zwolle, The Netherlands

[4] Department of Orthopaedics, OLVG, Amsterdam, The Netherlands

[5] Department of Orthopaedics, Amsterdam University Medical Center, Amsterdam, The Netherlands

[6] EuroQol Research Foundation, Rotterdam, The Netherlands

Status

Published.

Citation: Bonsel JM, Peeters CMM, Reijman M, Dings T, Rutges J, Kempen DHR, et al. A head-to-head comparison of the adult EQ-5D-5L and youth EQ-5D-Y-5L in adolescents with idiopathic scoliosis. J Patient Rep Outcomes. 2025;9(1):13.

ABSTRACT

Background

Multiple diseases, such as Adolescent Idiopathic Scoliosis (AIS), present at adolescent age and the impact on quality of life (QoL) prolongs into adulthood. For the EQ-5D, a commonly used instrument to measure QoL, the current guideline is ambiguous whether the youth or adult version is to be preferred at adolescent age. To assess which is most suitable, this study tested for equivalence along predefined criteria of the youth (EQ-5D-5L) and adult (EQ-5D-Y-5L) version in an adolescent population receiving bracing therapy for AIS.

Methodology

107 adolescents were recruited from 4 scoliosis centers in the Netherlands between March 2022 and January 2023; they completed both EQ-5D's and the SRS-22r (scoliosis-specific questionnaire). The following criteria were evaluated using the individual and sum of domains (level-sum-score (LSS)). Our primary criterion for non-equivalence of the EQ-5D's was less than excellent (≤0.9) intra-individual agreement using Intraclass Correlation Coefficient (ICC) analysis for LSS and weighted (quadratic) kappa for domains. Secondary criteria were differences in ceiling using McNemar test; a different number of quantified hypotheses for construct validity achieved using the SRS-22r as comparator; differences in test-retest reliability by comparing ICC/kappa values using a Z-test.

Results

Adolescents had a mean age of 14 years (range 12-18), and 78% were female. Ceiling was mostly comparable between EQ-5D's, ranging from 78-81% for mobility and self-care, 52-54% for usual activities, and 31-36% for pain/discomfort. The EQ-5D-5L showed more ceiling (57%) compared to the EQ-5D-Y-5L (41%) on anxiety/depression (p=0.006). Agreement between the EQ-5D's did not meet our criterion for the LSS (ICC 0.79 (95% confidence interval 0.70, 0.85)), and decreased further at the domain-level. Both EQ-5D's achieved 5/7 validity hypotheses. Test-retest reliability was slightly better for EQ-5D-5L LSS (ICC 0.76 (0.64, 0.84)) compared to EQ-5D-Y-5L LSS (ICC 0.69 (0.55, 0.79)), although this was statistically insignificant (p=0.284). This pattern was similar for most domains.

Conclusions

The EQ-5D versions showed insufficient agreement, and cannot be considered fully equivalent. While they were similar in terms of validity and test-retest reliability, differences in score distribution were present. Taken together, we advise using the EQ-5D-5L to monitor the QoL in adolescent patients with AIS, as it avoids switching instruments and thus data discontinuities. Future studies should verify these findings in different patient groups and the general population.

BACKGROUND

Health-Related Quality of Life (HRQoL) in children and adults, preferably self-reported, is recognized as an essential outcome parameter in medical practice and research. The EQ-5D is a widely used instrument to measure HRQoL in adults[1], and 2 versions are available in terms of the number of response levels: the 3-level (EQ-5D-3L) and 5-level (EQ-5D-5L) version. A decade ago, a youth version was developed aimed at children from 8-11 years of age[2, 3]. The intended concept and general structure were the same as the adult version, while the wording and content were tailored towards children. Currently, the youth version of the EQ-5D is also available as 3-level (EQ-5D-Y-3L) and 5-level (EQ-5D-Y-5L) version. Contemporary evidence has shown that the adult EQ-5D-5L (adult) has superior discriminatory power with less ceiling and a similar psychometric pattern as the EQ-5D-Y-5L (youth) [4-7]. Therefore, our study uses the 5-level versions.

Our research focused on the age-specificity of both versions. Specifically, our study tests the equivalence of the EQ-5D-5L and EQ-5D-Y-5L with data from Adolescent Idiopathic Scoliosis (AIS) patients who receive bracing treatment. Current guidelines from the EuroQol Research Foundation suggest the EQ-5D-Y self-report to be used in the younger age range (8-11 years) for its better comprehensibility[8]. In adolescents (12-18 years) neither version is preferred. Indirect evidence suggests that the EQ-5D-5L and EQ-5D-Y-5L perform equally well regarding validity, reliability, and responsiveness in this adolescent population[4, 9, 10]. Yet, head-to-head comparative evidence is absent. If the EQ-5D-5L and EQ-5D-Y-5L indeed are psychometrically similar ('equivalent') in adolescents, and otherwise comparable in practical application, this would imply the versions can be used interchangeably. If true, this would signify a preference for the EQ-5D-5L as it avoids the switching of versions at an age threshold in longitudinal applications. If the versions are not equivalent and the EQ-5D-Y-5L performs better in terms of alignment with the experience, language, and reflective abilities of adolescents, then this version should be preferred up to the age of 17.

AIS is the most common type of scoliosis; about 3 to 5 per 1000 children are estimated to develop AIS requiring treatment[11]. Although AIS patients are generally healthy apart from the deformity, the disease often decreases the quality of life through the experienced pain and social impact. Moreover, due to various treatment modalities such as bracing or surgery, AIS patients also face problems with self-image and mental health[12-14]. As the disease impact, the associated burden, and the side-effects of treatment inevitably prolong into adulthood, this population is a prime example to study the continuity of HRQoL instruments longitudinally.

In this study, we hypothesize that the EQ-5D versions are equivalent in this adolescent population regarding (1) intra-individual agreement, (2) distributional properties, in particular ceiling, (3) performance in validity tests, and (4) test-retest reliability. The criteria norms are discussed in the methods section.

METHODOLOGY

Study design

Questionnaires and other data were collected prospectively. This study was approved by the Medical Ethical Review Board from University Medical Center Groningen (reference 202100536); study-site specific ethical approval of each participating center was also obtained. Although this study was not pre-registered, we developed a statistical plan before data collection was complete. This manuscript is written according to the Guidelines for Reporting Reliability and Agreement Studies and COSMIN reporting guideline for studies on measurement properties of Patient-Reported Outcome Measures (PROMs) [15, 16]. We aimed for at least 100 participants advised by the COSMIN guidelines.

Participants

Consecutive patients from 4 scoliosis centers were included at the outpatient clinics between March 2022 and January 2023 if they met the following inclusion criteria: diagnosis of AIS, under active treatment with bracing, and age between 12 and 18 years. The diagnosis of AIS is made after other causes for (secondary) scoliosis have been excluded or are deemed unlikely. The disease severity is typically measured using the Cobb angle on spine radiographs. Patients receive bracing therapy generally for moderate curvatures and upwards, i.e., a Cobb angle >20°, with the aim to prevent further curve progression and the need for spinal surgery[11, 17]. Patients were excluded who underwent surgery or inability to complete study questionnaires due to cognitive impairment or insufficient understanding of the Dutch language.

Procedures

Eligible patients (and their parent/guardian) received oral and standardized written information on the study, and participants were required to provide consent conform Dutch law. Adolescents aged 12 to 16 give are required to provide consent independently in addition to their parents or guardian. From 17 and older, adolescents sign themselves. After obtaining signed informed consent, patients were sent a first link to a set of questionnaires in an electronic data-capture system (Castor). The first set of questionnaires included (1) various demographics, (2) the EQ-5D-5L (and EQ Visual Analogue Scale (VAS)), (3) the SRS-22r which has no defined age-limits, and (4) the EQ-5D-Y-5L (and EQ VAS). No missing data were allowed; however, one patient aborted the survey too early resulting in one missing value for the EQ VAS. The order of the EQ-5D versions was individually randomized. On top of these questionnaires, 75% of patients also filled out a novel Brace Questionnaire (BrQ) to assess its validity; the results have been recently published and are not discussed or used in this study[18]. To assess test-retest reliability, patients were sent a second link 7-14 days after completion of the first set of questionnaires.

Questionnaires

Demographics

Obtained demographics included age, sex, education level, body mass index (BMI), menarche (if female) and Cobb angle at inclusion. In the Netherlands, education can be trichotomized into primary education (i.e., primary school), secondary education (i.e., preparatory vocational, secondary vocational education, preparatory general education, or preparatory university education), and tertiary education (i.e., higher professional education or university education) [19]. Secondary education is generally known as high school. We collapsed secondary and tertiary education in two groups: *practical education* which included preparatory vocational or secondary vocation education and *theoretical education* which included preparatory general and preparatory university education, and also higher professional and university education.

EQ-5D-5L and EQ-5D-Y-5L

The official Dutch translation of the five-level versions of the EQ-5D-5L and EQ-5D-Y-5L was used[20]. Both versions cover 5 domains (Mobility, Self-care, Usual activities, Pain/Discomfort, and Anxiety/Depression), and both have 5 response levels resulting in 3125 possible health states.

The EQ-5D-Y-5L differs from the EQ-5D-5L in the following: (1) 'walking about' is added as explanation to the domain header 'Mobility'; (2) the domain header 'Self-care' is changed into 'Looking after myself'; (3) child-relevant examples are listed after the domain header 'Usual activities' ('going to school, hobbies, sports, playing, doing things with family or friends'); (4) the domain header 'Pain/Discomfort' is changed into 'Pain or other complaints'; (5) the domain header 'Anxiety/Depression' is changed into 'Feeling worried, Sad or Unhappy'. The most obvious difference concerns (6) the response levels: supposedly more child-friendly terms for level 3 and 4 are used in the EQ-5D-Y-5L. (7) Also, the most extreme level 5 is formulated slightly different for the domains 'Mobility', 'Self-care' and 'Daily activities': the phrase 'I am not able to' is replaced with 'I cannot'. The changes of the Y-version were the result of extensive qualitative and quantitative testing[2, 3]. The question texts (in Dutch) are included in Supplemental File 1; the full versions can be requested from the EuroQol Research Foundation.

The EQ-5D-5L has country-specific preference-based value sets available (for both 3L and 5L), that transforms each health state into an aggregate score, including the Netherlands[21]. For the EQ-5D-Y-5L currently only 3L value sets are available, and 5L sets are on their way[22]. As the primary goal of our research is descriptive equivalence, and in view of the absence of valuation sets for the currently used EQ-5D-Y-5L version, we use the level sum score (LSS) to compare aggregate scores between the instrument versions. Using the LSS, the best possible score is 1+1+1+1+1=5, and the worst possible score is 5+5+5+5+5=25. This conforms to current practice in non-economic papers, including research into descriptive performance[23].

EQ VAS

The EQ VAS aims to measure overall quality of life, and is a combination between a traditional Numerical Rating Scale and a Visual Analogue Scale. It is presented vertically. At the top a label states 'the best imaginable health'. The scale ranges from 0 (worst) to 100 (best), with ticks on the scale at each increment of 10. The youth version of the EQ VAS differs from the adult version in the following: (1) an informal version of the Dutch pronoun 'you' is used, and (2) the term 'measuring scale' is replaced by 'line'.

SRS-22r

The SRS-22r is a commonly used AIS-specific questionnaire developed and validated for adolescents, which we used as the comparator/reference for validity analysis[12, 24, 25]. It covers the domains function, pain, self-image, mental health, and satisfaction/dissatisfaction with management. Each domain consists of 5 items except for satisfaction/dissatisfaction, which consists of 2 items. Domain and aggregate scores are calculated by averaging the item-scores for each domain, and all items, respectively; scores range from 1 to 5, where a higher score indicates a better outcome.

Statistical analysis

General

In view of our research goal, the null hypothesis (to be rejected) is that the two EQ-5D versions are not equivalent, while the alternative hypothesis claims equivalence. Hence, equivalence is to be proven. To test for the equivalence of a new version or collection modality of HRQoL instruments in comparison to a default version several recommendations are available[26, 27]. This entails non-inferiority testing of the new version, which evaluates whether the new version is not worse than the default version. In our study, we test for true equivalence (rather than non-inferiority) as there is no default; in other words, either version may be better than the other. We derived our set of criteria from the above recommendations, taking the absence of a default into consideration.

The primary criterion is head-to-head (intra-individual) agreement of ≥0.91 expressed by Intraclass Correlation Coefficients (ICC) for aggregate scores and kappa values for domains, conform the recommendations for application of PROMs at the individual level. Of note, for application at the group level, recommendations are more lenient and ICC and kappa values of ≥0.7 and ≥0.8 are considered acceptable, respectively. Three secondary psychometric criteria were: distributional properties (lack of ceiling in particular), validity, and test-retest reliability. In the context of longitudinal use of EQ-5D in registries covering adolescent and adult age, test-retest reliability has specific relevance. If the versions are equivalent based on the primary criterion, and are similar in practical features, we conclude that they are interchangeable. If the EQ-5D versions are not equivalent, we will prefer the version with the best psychometric performance on secondary criteria where test-retest reliability has extra weight.

For further statistical testing of strength of association, ICC, kappa and Spearman rank correlation analysis were used. ICC and kappa coefficients were interpreted as follows: poor (≤0.39), fair (0.40-0.59), good (0.60-0.74), and excellent (0.75-1.00) reliability[28]. Spearman rank coefficients (rho) were interpreted as: negligible (≤0.10), weak (0.11-0.39), moderate (0.40-0.69), strong (0.70-0.89), and very strong (≥0.90) correlation[29].

Below we provide details on the statistical analysis. All analyses were performed in R version 4.3.1[30]. Where appropriate 95% confidence intervals (95% CIs) were reported, and a p-value <0.05 was considered significant. R packages used are included in Supplemental File 2.

Sample description

Sample characteristics were summarized, and conventional descriptive statistics for the EQ-5D-5L, EQ-5D-Y-5L, and SRS-22r responses were calculated. Aggregate scores between EQ-5D versions were compared using the Wilcoxon signed-rank test, while domains were compared using the Bowker's test for symmetry.

Distributional characteristics: ceiling and floor effects

The proportion of patients reporting 'no problems' (ceiling) and 'extreme problems' (floor) for the LSS and each domain, were compared between the EQ-5D versions using the McNemar test. For reference, these procedures were also conducted for the EQ-VAS and the SRS-22r. Overall, we expected relatively high ceiling and any significant difference between EQ-5D versions was considered potentially relevant.

Intra-individual agreement

ICCs based on single measurement, absolute-agreement, two-way random effects model were calculated for the LSS of the EQ-5D versions[31]. An ICC absolute-agreement was selected for all comparisons, as systematic differences are also relevant in the overall appraisal of QoL. ICC absolute-agreement typically results in lower ICC estimates compared to ICC consistency, which excludes systematic differences. Weighted (quadratic) kappa values were calculated for domains. A relevant disagreement was defined as an ICC or kappa ≤0.90, as described above. If indeed intra-individual agreement was less than hypothesized, we explored the observed disagreement with Bland-Altman plots[32]. ICC's and kappa are reliability parameters which relate the measurement error to the variation in the studied population, while Bland-Altman plots provide specific insights into the measurement error component. The Limits of Agreement (LOA), which were set at 95%, describe the size of measurement error between EQ-5D versions[33]. The dispersion of datapoints illustrate whether measurement error is random or systematic in nature. In case of the latter, future work may investigate the adjustability of this variation. Difference scores were assessed graphically and found to be roughly normaliy distributed, hence no data transformation was applied. Similar procedures were applied to the EQ VAS as reference.

Convergent and divergent validity

The strength of association using Spearman rank correlation was established between the EQ-5D-5L and the SRS-22, and the EQ-5D-Y-5L and the SRS-22r, respectively. The COSMIN guidelines states that 75% of hypotheses should be met to assume validity. Associations were established between total scores, between similar domains (convergent validity, rho≤-0.40) and between conceptually unrelated domains (divergent validity, rho≥-0.39), based on previous literature[4, 9, 10]. We expected only negative associations given the EQ-5D is the only questionnaire for which lower scores reflect better health. For convergent validity, we compared EQ-5D self-care to SRS-22r function, EQ-5D pain to SRS-22r pain, EQ-5D anxiety/depression to SRS-22r self-image and EQ-5D anxiety/depression to SRS-22r mental health. For divergent validity, we compared EQ-5D mobility to the SRS-22r function and EQ-5D usual activities to the SRS-22r function. Finally, we inspected whether either questionnaire in general outperformed the other in terms of validity, considering a difference in number of thresholds achieved of 1 or more to be relevant.

Test-retest reliability

Using the same approach as under intra-individual agreement, ICCs and kappa values were calculated for the LSS and domains between the first and second measurements, for the EQ-5D-5L and EQ-5D-Y-5L separately. We applied the same thresholds for and expected test-retest reliability to exceed ≥0.91 for both EQ-5D versions. To evaluate differences in test-retest reliability among EQ-5D versions, we applied Fisher's r-to-Z transformation to the coefficients and used a Z-test (Steiger's) for dependent groups to determine statistical significance[34, 35]. Similarly, Bland-Altman plots were used to illustrate the measurement error from first to second measurement.

Sensitivity analysis

To check the robustness of the findings regarding intra-individual agreement and test-retest reliability in particular, we re-ran these analyses within known subgroups which reflect more vs. less severe disease based on previous literature[4, 9, 10]. ICCs and kappa values were recalculated in the following subgroups: a Cobb angle ≥30 vs. <30; SRS-22r sum-score best 50% vs. worst 50%; practical vs. theoretical education; age oldest 50% vs. youngest 50%.

Due to the small number of children who were still in primary school (n=8), these were not used in the comparison according to education.

RESULTS

Out of 175 eligible patients with AIS undergoing brace treatment, 107 provided informed consent and completed the first survey. Seventy-eight (75%) responded to the second survey at an average follow-up of 27 days (Standard Deviation (SD) 16, range 9-73). Patients were included at a mean age of 14 years (SD 1.4, range 12–18), and 83 (78%) were female (Table 1).

Table 1: Characteristics of study population

Total sample, n=107	
Age in years, mean (SD)	14.3 (1.4)
Female, n (%)	83 (78)
Highest completed education, n (%)	
Primary education	8 (8)
Practical education	42 (40)
Theoretical education	57 (52)
Body mass index (kg/m2), mean (SD)	18.0 (2.6)
Menarche (if female, n=83), n (%)	62 (75)
Cobb angle at inclusion*, n (%)	
≤30	46 (43)
>30	60 (57)

A higher Cobb angle indicates more severe scoliosis.
* Data is missing from 1 patient.

The sample was relatively healthy, with high (low for LSS) average scores on all questionnaires (Table 2A, Figure 1). The EQ-5D's were similar with regard to aggregate scores: the median LSS was 7 (Interquartile Range (IQR) 6–9) for both the EQ-5D-5L and EQ-5D-Y-5L (p=0.243). At the domain level on both EQ-5D's, mobility and self-care were rated slightly better compared to usual activities, pain, and anxiety/depression. Median values of domain scores were also similar between EQ-5D's. The median value for the aggregate SRS-22r score was 4.0 (IQR 3.5–4.4). Corresponding domains in SRS-22r and EQ-5D tended to produce a similar distributional pattern (Table 2B).

Table 2A: Descriptive statistics of EQ-5D versions

	EQ-5D-5L			EQ-5D-Y-5L				
	Median (IQR)	Range	Ceiling, n (%)	Median (IQR)	Range	Ceiling, n (%)	p-value (diff. in median)**	p-value (diff. in ceiling)***
Aggregate								
LSS	7 (6 – 9)	5 – 18	19 (18)	7 (6 – 9)	5 – 17	14 (13)	0.243	0.359
VAS*	87 (70 – 95)	42 – 100	15 (14)	85 (73 – 94)	45 – 100	13 (13)	0.785	1.000
Domain								
Mobility	1 (1 – 1)	1 – 5	83 (78)	1 (1 – 1)	1 – 4	84 (79)	0.795	1.000
Self-care	1 (1 – 1)	1 – 3	87 (81)	1 (1 – 1)	1 – 3	87 (81)	0.753	1.000
Usual act.	1 (1 – 2)	1 – 5	58 (54)	1 (1 – 2)	1 – 4	56 (52)	0.830	0.864
Pain/disc.	2 (1 – 2)	1 – 4	33 (31)	2 (1 – 2)	1 – 4	39 (36)	0.624	0.327
Anx./depr.	1 (1 – 2)	1 – 5	61 (57)	2 (1 – 2)	1 – 5	44 (41)	0.267	0.006

Ceiling effects were defined as the best score attainable. For the LSS and domain scores a lower score indicates better health, while for the SRS-22r and VAS a higher score indicates better health.
*Data of the VAS (EQ-5D-Y-5L) is missing in 1 patient.
**For aggregate scores the Wilcoxon signed-rank test was used, while for domain scores the Bowker test was used.
***For all comparisons the McNemar test was used.
Abbreviations: LSS = level-sum-score; VAS = Visual Analogue Scale; diff. = difference; disc.=discomfort; anx.=anxiety; depr.=depression; IQR = Interquartile Range

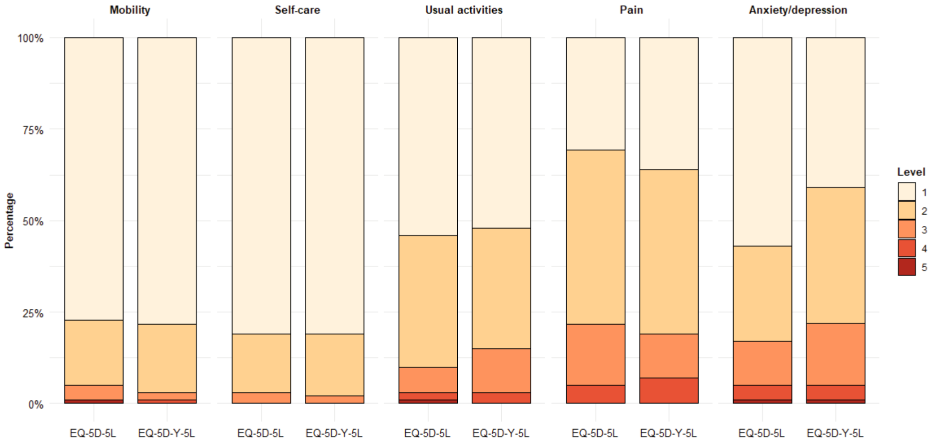

Figure 1: Distribution of the domain responses of the EQ-5D versions

Table 2B: Descriptive statistics of SRS-22r

	Median (IQR)	Range	Ceiling, n (%)
Aggregate			
Sum-score	4.0 (3.5 – 4.4)	2.2 – 4.8	0
Domain			
Function	4.4 (4.0 – 4.8)	2.8 – 5.0	16 (15)
Pain	4.2 (3.8 – 4.5)	1.4 – 5.0	9 (8)
Self-image	3.6 (3.0 – 4.1)	1.6 – 5.0	2 (2)
Mental health	3.8 (3.1 – 4.2)	1.0 – 5.0	3 (3)
Satisfaction with treatment	4.0 (3.5 – 4.5)	2.0 – 5.0	15 (14)

Ceiling and floor

Both EQ-5D versions produced no floor regarding aggregate scores and max. 1% for domains. Ceiling was prominent: with regard to the LSS, the ceiling was slightly larger for the EQ-5D-5L (18%) compared to the EQ-5D-Y-5L (13%), although this did not differ significantly (p=0.359). Ceiling was about similar for most domains of EQ-5D versions, and did not differ signficantly. The highest ceiling was observed for mobility (78% and 79%, for EQ-5D-5L and EQ-5D-Y-5L, respectively) and self-care (81% and 81%), and the lowest for pain (31% and 36%); usual activities was in-between (54% and 52%). The ceiling of the anxiety/depression domain was significantly higher for EQ-5D-5L (57%) compared to EQ-5D-Y-5L (41%) (p=0.006).

Intra-individual agreement

The agreement (ICC) between EQ-5D's was 0.79 (95% CI 0.70, 0.85) for LSS and 0.80 (95% CI 0.72, 0.86) for VAS (Table 3). At the domain level, kappa values were smaller; they were highest for self-care and pain/discomfort, and lowest for usual activities and anxiety/depression. All ICC/kappa values were lower than our predefined threshold of ≥0.91.

Table 3: Agreement between EQ-5D versions

	Predefined hypothesis	ICC (95% CI)
Aggregate		
VAS	N/A	0.80 (0.72, 0.86)
LSS	≥0.91	0.79 (0.70, 0.85)
		Kappa (95% CI)
Domain		
Mobility	≥0.91	0.62 (0.38, 0.86)
Self-care	≥0.91	0.76 (0.58, 0.94)
Usual act.	≥0.91	0.48 (0.31, 0.65)
Pain	≥0.91	0.69 (0.56, 0.81)
Anx./depr.	≥0.91	0.60 (0.44, 0.76)

ICC's were calculated for the aggregrate scores, between the EQ-5D-A and the EQ-5D-Y. Kappa analysis was used to assess agreement for domains.
*Indicates if the predefined hypotheses was met (not the case for any comparison).
Abbreviations: LSS = level-sum-score; VAS = Visual Analogue Scale; N/A= not applicable; ICC = Intraclass Correlation Coefficient; 95% CI = 95% confidence interval

Bland-Altman plots were created to gain insights into the measurement error between the EQ-5D versions (Figure 2 and 3). For the LSS, the mean difference was -0.15 (95% CI -0.46, 0.16). The upper LOA was 3.00 (95% CI 2.47, 3.53) and the lower LOA was -3.30 (95% CI -3.82, -2.77). In other words, 95% of differences between the LSS of EQ-5D's fall between approximately -3 and +3. For the VAS, the mean difference was 0.29 (95% CI -1.99, 1.40), upper LOA 16.94 (95% CI 14.00, 18.87), lower LOA -17.52 (95% CI -20.45, -14.59). Overall, the plots suggested that disagreement was largely due to random variation, for both the LSS and VAS scores.

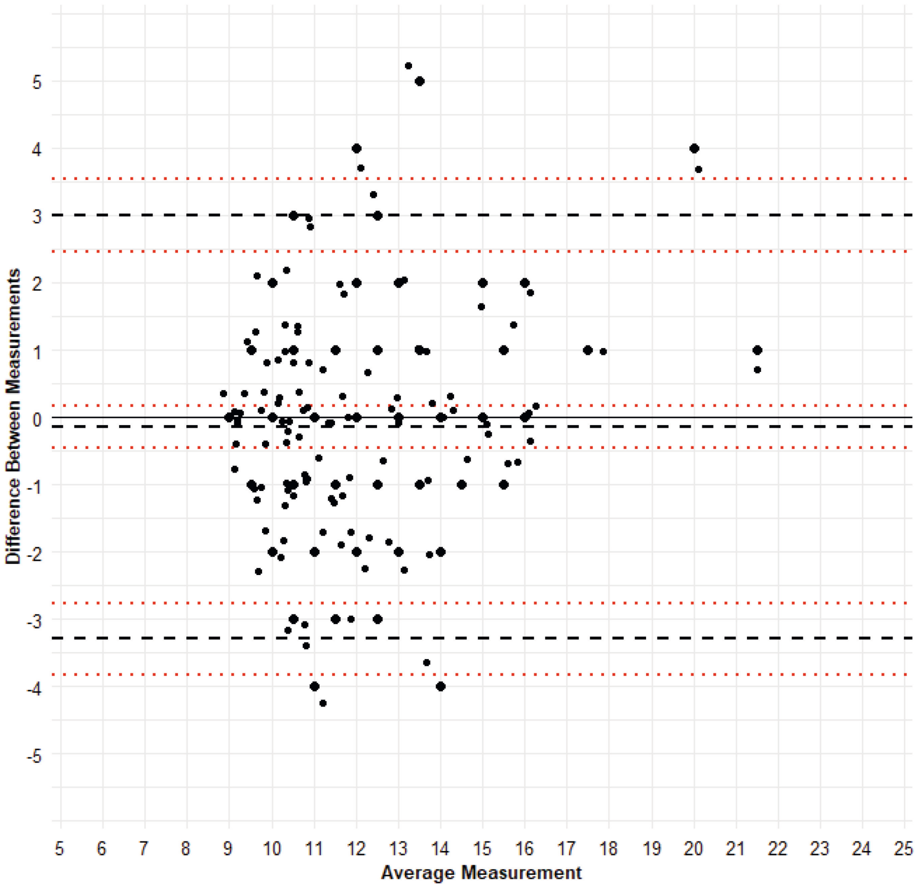

Figure 2: Bland-Altman plot for the LSS of the EQ-5D versions

The y-axis depicts the difference between the intra-individual measurement of the EQ-5D-5L and EQ-5D-Y-5L. The x-axis depicts the average of these two measurements. The dashed lines indicate the mean difference between EQ-5D versions and 95% limits of agreement. The red dotted lines represent the 95% confidence intervals for these estimates.

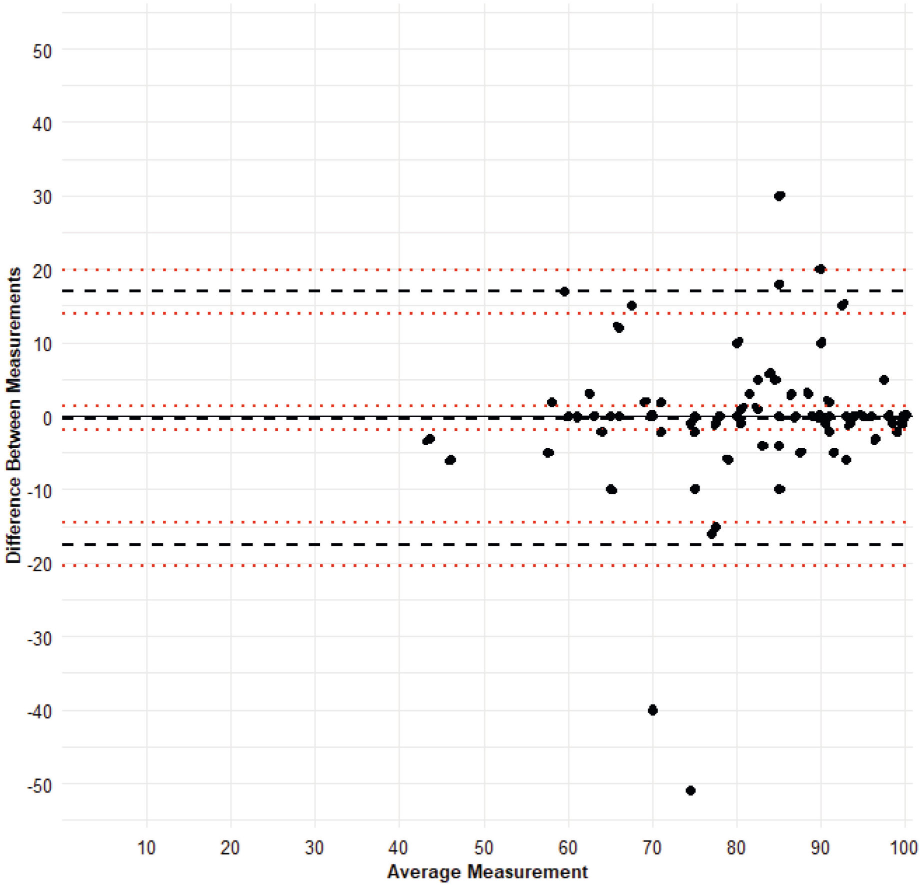

Figure 3: Bland-Altman plot for the VAS of the EQ-5D versions

The y-axis depicts the difference between the intra-individual measurement of the VAS obtained with the EQ-5D-5L and the VAS obtained with the EQ-5D-Y-5L. The x-axis depicts the average of these two measurements. The dashed lines indicate the mean difference between VAS versions and 95% limits of agreement. The red dotted lines represent the 95% confidence intervals for these estimates.

Convergent and divergent validity

The pre-defined hypotheses with regard to validity were met for 5 out of 7 hypotheses pertaining to the LSS or domains, for both the EQ-5D-5L and EQ-5D-Y-5L (Table 4).

Table 4: Convergent and divergent validity of EQ-5D versions

		EQ-5D-5L	EQ-5D-Y-5L
	Predefined hypothesis	Rho (95% CI)	Rho (95% CI)
Aggregrate			
EQ VAS vs. SRS sum-score	N/A	0.57 (0.40, 0.68)	0.52 (0.35, 0.65
EQ-5D LSS vs. SRS sum-score	≤-0.40	-0.71* (-0.58, -0.80)	-0.68* (-0.54, -0.78)
EQ-5D LSS vs. EQ VAS	≤-0.40	-0.57* (-0.40, -0.68)	-0.48* (-0.30, -0.62)
Domain			
EQ-5D mobility vs. SRS function	≥-0.39	-0.36* (-0.18, -0.52)	-0.25* (-0.07, -0.43)
EQ-5D self-care vs. SRS function	≤-0.40	-0.16 (0.04, -0.34)	-0.08 (0.12, -0.27)
EQ-5D usual act. vs. SRS function	≥-0.39	-0.61 (-0.46, -0.73)	-0.44 (-0.27, -0.59)
EQ-5D pain vs. SRS pain	≤-0.40	-0.64* (-0.50, -0.75)	-0.61* (-0.46, -0.73)
EQ-5D anx./depr vs. SRS self-image	≤-0.40	-0.49* (-0.32, -0.63)	-0.54* (-0.39, -0.67)
EQ-5D anx./depr vs. SRS mental health	≤-0.40	-0.63* (-0.48, -0.74)	-0.65* (-0.51, -0.76)

Spearman rank correlations were calculated between the aggregate and domain scores. A higher EQ-5D domain/aggregate score indicates worse health, while a higher EQ VAS and SRS-22r domain/aggregate score indicates better health.
*indicates if the predefined hypotheses was met.
Abbreviations: LSS = level-sum-score; VAS = Visual Analogue Scale; N/A = not applicable; 95% CI = 95% confidence interval

Test-retest reliability

ICCs were 0.76 (95% CI 0.64, 0.84) for the EQ-5D-5L LSS and 0.69 (95% CI 0.55, 0.79) for the EQ-5D-Y-5L; see Table 5. Test-retest reliability was lower at the domain-level, with the lowest kappa value observed for the self-care domain (EQ-5D-5L: 0.29 (95% CI 0.03, 0.56), EQ-5D-Y-5L: 0.19 (95% CI -0.06, 0.43)) and the highest for the anxiety/depression domain (EQ-5D-5L: 0.67 (95% CI 0.48, 0.85), EQ-5D-Y-5L: 0.69 (95% CI 0.56, 0.82)). Slightly higher point-estimates were generally observed for aggregate and domain scores of the EQ-5D-5L as compared to EQ-5D-Y-5L, however, these were not statistically significantly different. The Bland-Altman plots suggested that the difference between baseline and second measurement were mainly attributable to random variation rather than due to true change (Supplemental File 3, Figure 1–4).

Table 5A: Test-retest reliability of EQ-5D versions

		EQ-5D-5L	EQ-5D-Y-5L	
	Predefined hypothesis	**ICC (95% CI)**	**ICC (95% CI)**	**p-value (diff. in ICC/kappa)****
Aggregrate				
VAS	N/A	0.45 (0.26, 0.61)	0.50 (0.32, 0.65)	0.621
LSS	≥0.91	0.76 (0.64, 0.84)	0.69 (0.55, 0.79)	0.284
		Kappa (95% CI)	**Kappa (95% CI)**	
Domain				
Mobility	≥0.91	0.40 (0.19, 0.60)	0.50 (0.31, 0.68)	0.376
Self-care	≥0.91	0.29 (0.03, 0.56)	0.19 (-0.06, 0.43)	0.442
Usual act.	≥0.91	0.64 (0.46, 0.81)	0.51 (0.32, 0.70)	0.156
Pain	≥0.91	0.66 (0.53, 0.79)	0.58 (0.41, 0.75)	0.360
Anx./depr.	≥0.91	0.67 (0.48, 0.85)	0.69 (0.56, 0.82)	0.732

ICC's and kappa values were calculated for the aggregate and domain scores, between the first and second measurement at least 7 days later (average 27 days later).
*indicates if the predefined hypotheses was met (not the case for any comparison).
**To compare ICC and kappa values, a Fisher's r-to-Z transformation was applied and a Z-test (Steiger) was used to determine statistical significance.
Abbreviations: LSS = level-sum-score; VAS = Visual Analogue Scale; ICC = Intraclass Correlation Coefficient; diff. = difference; N/A = not applicable; 95% CI = 95% confidence interval

Table 5B: Test-retest reliability of SRS-22r

	ICC (95% CI)
Aggregrate	
Sum-score	0.87 (0.80, 0.92)
Domain	
Function	0.70 (0.61, 0.83)
Pain	0.76 (0.65, 0.84)
Self-image	0.84 (0.76, 0.90)
Mental health	0.79 (0.69, 0.86)
Satisfaction with treatment	0.67 (0.53, 0.78)

Sensitivity analysis

The intra-individual agreement was relatively higher in subgroups with more severe scoliosis as defined by the SRS-22r or Cobb angle for both versions (Supplemental File 4, Tables 1–8). In contrast, agreement was lower in patients less affected by scoliosis. The subgroups education and age appeared to not affect the agreement. Test-retest reliability was similar according to Cobb angle, education and age, while better reliability was observed in patients with worse SRS-22r scores. The differences in points-estimates between the EQ-5D-5L and EQ-5D-Y-5L generally persisted (Supplemental File 4, Tables 9–16).

DISCUSSION

Main findings

In this study, we compared the EQ-5D-5L and EQ-5D-Y-5L in a sample of AIS patients treated with a brace. Intra-individual agreement across versions was found to be excellent for the LSS (ICC 0.79 (95% CI 0.70, 0.85)), however, did not meet our primary criterion for equivalence. Agreement further dropped at the domain level, in particular for *mobility*, *usual activities*, and *anxiety/depression*. Regarding psychometric properties, ceiling was comparable for most domains and the LSS, except for the *anxiety/depression* domain which showed sigifiicantly more ceiling for the EQ-5D-5L (57%) compared to the EQ-5D-Y-5L (41%). This may be attributed to the different wording of both question and response. Both the EQ-5D-5L and EQ-5D-Y-5L demonstrated comparable validity, achieving 5 out of 7 hypotheses (close to the commonly used 75% threshold). With regard to test-retest reliability, point-estimates were slightly higher for the EQ-5D-5L (LSS 0.76 (95% CI 0.64, 0.84)) as compared to the EQ-5D-Y-5L (LSS 0.69 (0.55, 0.79)), although these differences did not reach significance. As secondary psychometric criteria overall were roughly similar between EQ-5D versions, we think that in the context of patient monitoring from adolescence to adulthood the EQ-5D-5L is the preferred instrument. This avoids potential data discontinuities resulting from switching between versions and hence facilitates longitudinal follow-up from adolescence into adulthood.

Comparison with other literature

This study is based on adopted criteria, which can greatly influence the judgement of determining (non-)equivalence. We chose to require intra-individual agreement (and test-retest reliability) to achieve strict thresholds, as we believe using EQ-5D versions interchangeably requires the instruments to align very strongly. However, for the purpose of larger group comparisons, more lenient thresholds may be used, as described in the methods section. Both EQ-5D versions showed acceptable intra-individual agreement and test-retest reliability for the LSS using these thresholds, but not at the domain level. Although no studies are available to compare the level of intra-individual agreement, test-retest reliability findings of both EQ-5Ds were in line with previous studies[9, 10]. In retrospect, it was unlikely for the reliability of EQ-5Ds to achieve the strict threshold we applied.

Lack of reliability of the EQ-5Ds was mostly attributable to random error, presumably because each domain includes only one question[36]. For longitudinal follow-up of patients, higher test-retest reliability translates into being able to more precisely capture a given health state. Treatment decisions may be contingent on the measured health state, and inaccuracies may have important implications. Hence, it is imaginable that the version with a trend of higher estimates may be the preferred option in this adolescent AIS population, i.e., the EQ-5D-5L.

As both EQ-5D versions have the same number of response levels, three underlying mechanisms may explain the disagreement between instrument versions for the domains *mobility*, *usual activities*, and *anxiety/depression*. Firstly, due to different wording of the question these domains cover a different underlying idea/concept. Secondly, they cover the same idea/concept, but the average distribution of scores is shifted lower or higher in general. Thirdly, due to different wording of the five severity labels, the distribution of the numbers (response) is different. In the first case one expects, if tested against an external anchor such as the SRS-22r, that the ranking of the responses of both versions is different. As this was not the case, the first explanation seems unlikely. In the second and third mechanism, one would expect the ranking to be similar despite a different use of the scale (distribution). In view of the fairly limited textual adaptations of the youth version, the results seem to match these explanations. The second mechanism is exemplified by the higher ceiling for *anxiety/depression* for the EQ-5D-5L compared to the EQ-5D-Y-5L. The EQ-5D-5L describes this domain as "fear/sadness", while the EQ-5D-Y-5L describes it as "worrying, sadness or unhappiness". In this situation, the underlying response scale may be shifted upwards in a constant fashion, hence patients use extreme values (ceiling) more often while correlation between measures remains relatively preserved. The third mechanism is expected to apply to the *mobility* and *usual activities* domains.

The EQ-5D-5L and EQ-5D-Y-5L demonstrated comparable validity. The validity findings were generally compatible with previous studies, and were close to the currently accepted 75% guideline for demonstrating validity[9, 10, 37, 38]. The LSS and SRS-22r sum scores were strongly correlated, suggesting that the EQ-5D is able to capture the relevant disease burden and HRQoL of AIS patients treated with a brace. We found insufficient association between the EQ-5D domain self-care and SRS-22r function domain (rho -0.16 (EQ-5D-5L) and -0.08 (EQ-5D-Y-5L) instead of ≥-0.40). A higher than expected association was found between the EQ-5D domain usual activities and the SRS-22r function domain (rho -0.61 (EQ-5D-5L) and -0.44 (EQ-5D-Y-5L) instead of ≤-0.39) [9]. The SRS-22r function domain focuses on the level of activity, on limitations in doing things around the house, financial difficulties due to AIS, and limits in going out with friends[12, 24]. These (mild) differences between our study and previous papers may be attributable to differences between samples: only 11% of the sample in the study by Adobor et al. was undergoing brace treatment at the time of filling out the questionnaire, and a larger percentage had surgery (39%) or were scheduled for surgery (30%), hence representing a population with more severe scoliosis. It is imaginable that a patient with more severe scoliosis have increased problems with *self-care* thus correlating more strongly with the SRS-22r function domain.

Strengths and limitations

The present study had some limitations. Firstly, a sample size of 107 can be considered small, however, it does meet the current COSMIN criteria and the homogeneity of the sample permits careful testing[15]. Secondly, we did not include a question on experienced health change at the second measurement. Generally, excluding patients who report a change in health may benefit test-retest reliability. However, this would have added to the questionnaire burden already consisting of two close to identical questionnaires and a comparator. Also, we think a health change is unlikely in these rather healthy persons, as they were approached after they had already initiated bracing therapy and were still required to wear their brace until at least the subsequent visit which in general is 6 months later. Thirdly, as the study population was rather healthy, data was skewed. This affected the size of the kappa, resulting in lower values than would be expected for the observed absolute agreement. Finally, the current study is performed in a selected AIS population undergoing bracing treatment, and is inevitably not generalizable to all AIS patients. While AIS patients show a wide range of symptoms, specific patient groups may exist where the instrument versions show larger differences, or no difference at all.

CONCLUSION

This is the first head-to-head comparison of the EQ-5D-5L and EQ-5D-Y-5L in an adolescent AIS population treated with a brace, using a strict testing format to reject or establish equivalence. The EQ-5D versions show insufficient intra-individual agreement and cannot be considered fully equivalent, and thus and cannot be used interchangeably. Although they were roughly similar in terms of validity and test-retest reliability, specific differences in score distribution were present. If longitudinal measurement of HRQoL from adolescence into adulthood is foreseen, and we think the EQ-5D-5L is the preferred choice with the added benefit that potential data discontinuities are avoided. Future studies should verify if this finding holds in different patient groups and the general population.

REFERENCES

1. Brooks R. EuroQol: the current state of play. Health Policy 1996; 37: 53-72.
2. Wille N, Badia X, Bonsel G, Burström K, Cavrini G, Devlin N, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. Qual Life Res 2010; 19: 875-886.
3. Ravens-Sieberer U, Wille N, Badia X, Bonsel G, Burström K, Cavrini G, et al. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. Qual Life Res 2010; 19: 887-897.
4. Wong CKH, Cheung PWH, Luo N, Cheung JPY. A head-to-head comparison of five-level (EQ-5D-5L-Y) and three-level EQ-5D-Y questionnaires in paediatric patients. Eur J Health Econ 2019; 20: 647-656.
5. Kreimeier S, Åström M, Burström K, Egmar AC, Gusi N, Herdman M, et al. EQ-5D-Y-5L: developing a revised EQ-5D-Y with increased response categories. Qual Life Res 2019; 28: 1951-1961.
6. Verstraete J, Scott D. Comparison of the EQ-5D-Y-5L, EQ-5D-Y-3L and PedsQL in children and adolescents. Journal of Patient-Reported Outcomes 2022; 6: 67.
7. Buchholz I, Janssen MF, Kohlmann T, Feng Y-S. A Systematic Review of Studies Comparing the Measurement Properties of the Three-Level and Five-Level Versions of the EQ-5D. PharmacoEconomics 2018; 36: 645-661.
8. EuroQol Research Foundation. https://euroqol.org/publications/user-guides/, accessed November 2023.
9. Adobor RD, Rimeslåtten S, Keller A, Brox JI. Repeatability, Reliability, and Concurrent Validity of the Scoliosis Research Society-22 Questionnaire and EuroQol in Patients With Adolescent Idiopathic Scoliosis. Spine 2010; 35: 206-209.
10. Cheung PWH, Wong CKH, Samartzis D, Luk KDK, Lam CLK, Cheung KMC, Cheung JPY. Psychometric validation of the EuroQoL 5-Dimension 5-Level (EQ-5D-5L) in Chinese patients with adolescent idiopathic scoliosis. Scoliosis Spinal Disord 2016; 11: 19.
11. JEH P. School screening for scoliosisproefschrift. Utrecht: Universiteit Utrecht. 1996.
12. Schlösser TP, Stadhouder A, Schimmel JJ, Lehr AM, van der Heijden GJ, Castelein RM. Reliability and validity of the adapted Dutch version of the revised Scoliosis Research Society 22-item questionnaire. Spine J 2014; 14: 1663-1672.
13. Tones M, Moss N, Polly DW, Jr. A review of quality of life and psychosocial issues in scoliosis. Spine (Phila Pa 1976) 2006; 31: 3027-3038.
14. Zhang J, He D, Gao J, Yu X, Sun H, Chen Z, Li M. Changes in life satisfaction and self-esteem in patients with adolescent idiopathic scoliosis with and without surgical intervention. Spine (Phila Pa 1976) 2011; 36: 741-745.
15. Gagnier JJ, Lai J, Mokkink LB, Terwee CB. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. Qual Life Res 2021; 30: 2197-2218.
16. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. Journal of Clinical Epidemiology 2011; 64: 96-106.
17. Konieczny MR, Hieronymus P, Krauspe R. Time in brace: where are the limits and how can we improve compliance and reduce negative psychosocial impact in patients with scoliosis? A retrospective analysis. Spine J 2017; 17: 1658-1664.

18. Peeters CMM, Bonsel JM, Munnik-Hagewoud R, Mostert AK, Van Solinge GB, Rutges J, et al. Validity and reliability of the adapted Dutch version of the Brace Questionnaire (BrQ). Acta Orthop 2023; 94: 460-465.
19. Dutch Government https://www.government.nl/topics/themes/education, accessed November 2023. 2023.
20. EuroQol Research Foundation. 2024.
21. M MV, K MV, S MAAE, de Wit GA, Prenger R, E AS. Dutch Tariff for the Five-Level Version of EQ-5D. Value Health 2016; 19: 343-352.
22. Roudijk B, Sajjad A, Essers B, Lipman S, Stalmeier P, Finch AP. A Value Set for the EQ-5D-Y-3L in the Netherlands. PharmacoEconomics 2022; 40: 193-203.
23. Verstraete J, Marthinus Z, Dix-Peek S, Scott D. Measurement properties and responsiveness of the EQ-5D-Y-5L compared to the EQ-5D-Y-3L in children and adolescents receiving acute orthopaedic care. Health and Quality of Life Outcomes 2022; 20: 28.
24. Asher M, Min Lai S, Burton D, Manna B. The reliability and concurrent validity of the scoliosis research society-22 patient questionnaire for idiopathic scoliosis. Spine (Phila Pa 1976) 2003; 28: 63-69.
25. Asher M, Min Lai S, Burton D, Manna B. Discrimination validity of the scoliosis research society-22 patient questionnaire: relationship to idiopathic scoliosis curve pattern and curve size. Spine (Phila Pa 1976) 2003; 28: 74-78.
26. Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, et al. Recommendations on Evidence Needed to Support Measurement Equivalence between Electronic and Paper-Based Patient-Reported Outcome (PRO) Measures: ISPOR ePRO Good Research Practices Task Force Report. Value in Health 2009; 12: 419-429.
27. O'Donohoe P, Reasner DS, Kovacs SM, Byrom B, Eremenco S, Barsdorf AI, et al. Updated Recommendations on Evidence Needed to Support Measurement Comparability Among Modes of Data Collection for Patient-Reported Outcome Measures: A Good Practices Report of an ISPOR Task Force. Value in Health 2023; 26: 623-633.
28. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment 1994; 6: 284-290.
29. Chan YH. Biostatistics 104: correlational analysis. Singapore Med J 2003; 44: 614-619.
30. R Core Team (2024). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
31. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 2016; 15: 155-163.
32. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999; 8: 135-160.
33. Gerke O. Reporting Standards for a Bland-Altman Agreement Analysis: A Review of Methodological Reviews. Diagnostics (Basel) 2020; 10.
34. Tests for comparing elements of a correlation matrix. vol. 87. US: American Psychological Association 1980:245-251.
35. Long D, Polinder S, Bonsel GJ, Haagsma JA. Test-retest reliability of the EQ-5D-5L and the reworded QOLIBRI-OS in the general population of Italy, the Netherlands, and the United Kingdom. Qual Life Res 2021; 30: 2961-2971.
36. Megan D, Jennifer K. Patient-reported outcome measures (PROMs): how should I interpret reports of measurement properties? A practical guide for clinicians and researchers who are not biostatisticians. British Journal of Sports Medicine 2014; 48: 792.

37. Lin J, Wong CKH, Cheung JPY, Cheung PWH, Luo N. Psychometric performance of proxy-reported EQ-5D youth version 5-level (EQ-5D-Y-5L) in comparison with three-level (EQ-5D-Y-3L) in children and adolescents with scoliosis. Eur J Health Econ 2022; 23: 1383-1395.

38. Wong CKH, Cheung PWH, Samartzis D, Luk KD, Cheung KMC, Lam CLK, Cheung JPY. Mapping the SRS-22r questionnaire onto the EQ-5D-5L utility score in patients with adolescent idiopathic scoliosis. PLoS One 2017; 12: e0175847.

DECLARATIONS

Ethics approval: This study was approved by the Medical Ethical Review Board from University Medical Center Groningen (RR-number: 202100536); study-site specific ethical approval of each participating center was also obtained.

Consent for publication: not applicable

Consent to participate: Eligible patients (and their parent/guardian) received oral and standardized written information on the study, and participants were required to provide consent conform Dutch law. Adolescents aged 12 to 16 give are required to provide consent independently in addition to their parents or guardian. From 17 and older, adolescents sign themselves (if deemed capable). Although not required by Dutch law, two versions of written information were provided. The first (standard) was tailored towards adults including adolescents aged >16, while the second (additional) was tailored towards children aged 12 to 16. The latter used child-friendly language and terminology. At each site, consent was obtained by researchers and orthopedic surgeons with knowledge of the patient population and the study.

Availability of data and material: the currently used dataset have been archived in a data repository (link: https://doi.org/10.34894/PDJZXH) and are available upon reasonable request, after approval by the author team. As the data are sensitive in nature, there are restrictions in place with regard to the availability of the data. Codes used to conduct the analyses are obtainable from the corresponding author.

Competing interests: the authors declare that they have no competing interests.

Funding: this work was funded by a PhD grant (PHD-287) provided by the EuroQol Research Foundation. The funder had no role in the design and conduct of the study, or approval of the manuscript, nor the decision to submit the manuscript for publication. The views expressed are those of the individual authors and do not necessarily reflect the views of the EuroQol Research Foundation.

Authors' contributions: all authors contributed to the study conception and design. Funding acquisition was performed by Joshua Bonsel, Max Reijman, Jan Verhaar and Gouke Bonsel. Data collection and analysis were performed by Joshua Bonsel and Charles Peeters. The first draft of the manuscript was written by Joshua Bonsel and Tim Dings, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Acknowledgements: not applicable.

Supplementary data: Supplementary data associated with this article can be found in the online version at doi: 10.1186/s41687-025-00842-z

Chapter 8

General discussion

INTRODUCTION

In this thesis, the use of PROMs data for improving the quality of care was explored in both routine applications and empirical research. The latter focused on specific factors that modify the quality of care, such as socioeconomic inequalities and the impact of the COVID-19 pandemic. Several methodological challenges that emerged in this context were also studied. For example, in adolescent patients with scoliosis, a comparison between a child-specific and an adult quality of life instrument was conducted, to identify the optimal instrument to monitor these patients into adulthood.

This chapter is organized as follows. Three topics of particular relevance to this thesis are discussed, namely 1) the use of PROMs in the context of quality of care according to the level of potential impact (micro-, meso-, and macro-level); 2) The measurement and interpretation of socioeconomic inequalities in health care, where we consider their absence a quality feature; and 3) The approach to addressing methodological challenges emerging in these contexts. We conclude this chapter with our perspectives for the future.

PROMS IN QUALITY OF CARE

Micro-level routine care applications of PROMs

In **Chapter 2**, our review showed that the use of PROMs at the micro-level (patient-level) is well established. We identified two promising types of PROMs applications in orthopedics, namely shared-decision making and post-operative monitoring tools.

Shared decision-making tools

Three studies utilized PROM-based tools to inform patients about expected benefits from arthroplasty and conservative treatment options for osteoarthritis; benefits were expressed in terms of disease burden and quality of life[1-3]. These tools also included an educational component on the disease, unlike the LROI tool 'Patients Like Me'. A significant percentage of patients, more in TKA than in THA, report no improvement or even deterioration when measured with PROMs, or express dissatisfaction with the final result[4]. One commonly cited reason for this is differing expectations of the procedure, whether realistic or not, which are not adequately met; these tools may help modify these expectations. The identified studies found evidence of improved decision quality, although it is uncertain to what extent decision quality is associated with better overall patient experienced or clinical outcomes[5].

Short-term post-operative monitoring tools

We identified one study that used an eHealth app in which patients reported daily symptoms and uploaded images of their wound, combined with educational material regarding the procedure, following oncological colorectal surgery[6]. The intervention moderately improved PROMs, but it did not reduce other important clinical outcomes such as emergency department visits or readmissions. While appealing, PROMs are designed to capture outcomes related to the disease under study and may miss important symptoms related to surgical recovery. The questionnaire used in the identified study, for example, did not include essential symptoms such as fever or wound leakage. Tailored symptom inventory lists have been suggested as more suitable for this purpose[7]. A similar eHealth app is under development in orthopedics which uses a tailored symptom list to identify early postoperative problems such as (wound) infection[8]. Although this tool has the potential to replace part of the short-term outpatient visits or phone calls, thus decreasing the burden on clinical practitioners, it comes with certain challenges. First, determining responsibility for checking app reports and subsequently establishing clear action plans for abnormal findings requires careful consideration. Second, the threshold of the PROM or symptom inventory requiring action should be low and prevent false-negative conclusions of infection with potentially severe consequences. Rigorous testing is necessary to determine the limits of this type of monitoring, with due consideration for patient reassurance. Clinical assessment and patient-doctor interactions are essential for determining whether treatment is necessary, as self-reported results alone are insufficient for making such decisions. The likelihood of a postoperative infection is determined through a comprehensive assessment of symptoms, physical examination, and blood work. Clinical expertise is needed to interpret these results, and to our knowledge, there is currently no prediction model that can precisely identify which patients require treatment, although Artificial Intelligence may change the state of play. This is particularly relevant for patients in the ‘grey area’, i.e., without clear infection symptoms such as persistent wound leakage.

Long-term post-operative monitoring tools

Another study used long-term PROM data to provide insights into progress or deterioration for physiotherapists in an exercise program in patients with osteoarthritis (OA)[9]. Overall, moderately improved outcomes were found. This study underlines an important consideration applicable to post-operative monitoring tools, and the use of PROMs in general. Their use typically forms part of a multifaceted intervention, and it is impossible to determine the precise effect attributable to the collection and use of PROMs. This also applies to the previously described applications. In this study, it was not explained how and to what extent PROM data were used to support the exercise program.

Common considerations for the incorporation of these tools in daily practice

All identified tools rely on a digital environment. As such, the inclusivity of care may be negatively impacted because not all patients have sufficient (digital) knowledge to use these tools or even have a cellphone[10]. Alternative information provision (e.g., printed materials) and classical follow-up methods should remain available. Secondly, these tools do not include non-medical considerations and do not easily capture the varying importance that patients place on different treatment aspects and outcomes. These assessments typically are not integrated into the tool itself (ref [11] is an exception), and still, there remains room for the patient-doctor consultation to discuss the results from the tools. Thirdly, to realize effective use of PROMs and symptom inventories, training for both patients and clinicians is essential[12]. Clinical users, for example, have limited experience with PROM data and their interpretation and presentation in these tools; they may need guidance on how to deal with unfavorable PROM outcomes, similar to how they are trained to deal with deviant radiographical outcomes. This list of considerations can be extended.

Meso-/macro-level routine care applications of PROMs

At the meso-/macro-level, evidence of the use of PROMs was scarce, though actual use is presumably more widespread than what is published in medical journals. Two applications were identified, namely the use of PROMs for benchmarking purposes and as input for quality improvement initiatives through in-depth data analysis or plan-do-check-act (PDCA) cycles. The evidence did not suggest a benefit from benchmarking alone[13-16]. Effective applications were found for the latter, including using PROM data to guide the implementation of a hospital-wide pain protocol and the introduction of a new type of knee prosthesis in a national healthcare system[17-20].

A key observation may substantiate the difference in effectiveness between benchmarking and using data as input for quality improvement initiatives. In benchmarking, data applications 'simply' provide outcome data to providers or other relevant stakeholders. More advanced, data-guided specific initiatives consist of well-thought-out plans and different feedback loops. A report issued in 2022 by Zorginstituut Netherlands on how stakeholders use quality indicators reflects the current state of play[21]. Effective applications seem to require support from quality committees, preferably with input from physicians regarding what factors within reach could be responsible for suboptimal (or superior) performance, e.g., across patient groups.

Specific challenges apply. As already observed at the micro-level, clinical users often have limited experience with PROMs and may require guidance on how to respond to negative PROM outcomes[22]. In specialized care, particularly when average performance is excellent, the multifactorial background and often small patient numbers per professional make it difficult to assign suboptimal performance to a specific factor, whether provided-related or not. Even if a measured quality difference can be transformed into an opportunity to improve care, changing the clinical pathway or staff competence is a costly and time-consuming process, even with optimal support available. The review did not identify papers or reports on the use and impact of clinical dashboards at the meso-level, despite the increasing use of these tools[23]. We presume this underrepresentation to be the consequence of an ambiguous status on whether reporting experiences represents 'science' or just practice. Arguably, the line between PROMs used in routine clinical care and empirical research applications at the meso-/macro-level is somewhat arbitrary. This is less the case for the research we presented on socioeconomic inequalities and the impact of the COVID-19 lockdown. **Chapters 3**, **4**, and **6** provided insights that could inform guidelines and policy decisions, potentially improving quality of care. The impact of such studies on guidelines or practice is yet to be established.

SOCIOECONOMIC INEQUALITIES IN HEALTHCARE

Findings from the Dutch orthopedic data and the panel discussion

In **Chapters 3**, **4**, and **6** we conducted multiple empirical studies using LROI PROMs data of THA and TKA patients to identify opportunities for improved clinical care. **Chapter 3** and **4** are highlighted here, where we studied socioeconomic inequalities in PROMs of these patients. Two separate pathways to inequalities were discovered. Firstly, low SES was associated with worse PROMs before surgery, which could be interpreted as a selection effect. Secondly, low SES was associated with worse PROMs after surgery, in addition to the selection effect, which might indicate suboptimal recovery, i.e., a prognostic effect. These subtle findings serve as a prime example of how PROMs can provide in-depth understanding of inequality mechanisms.

We invited a panel of Dutch orthopedic experts to discuss the paper and subsequently determine whether the differences were relevant and what potential next steps could be. The observed differences were unanimously accepted as being of relevance. Yet, the second question on what to do led to a mixed response. Most experts preferred to postpone action until more was known about underlying pathways. Although pathways considered were mainly patient-related, some provider-related pathways were suggested. The experts had difficulty designing an intervention at this stage that would reduce differences. The group agreed on how to proceed: first, a subsequent project could be initiated using LROI data to unravel underlying drivers for inequalities. This could be done by linking various data sources (e.g., insurance, general practitioner (GP), hospital data) to the LROI or by enhancing the registry using local hospital data. However, this is a complex task, as for example insurers do not easily share their data, and vice versa. Another approach would be to collect new (survey) data.

Targets for reduction of socioeconomic inequalities

Based on the literature review and the panel discussion, we identified potential intermediary factors contributing to inequalities, which may serve as targets for improvement strategies (Table 1). The inventory is categorized into patient, contextual, provider, and insurer factors. Potential strategies may include (a) primary prevention, such as education and policy changes aimed at reducing exposure to these factors, (b) secondary prevention, such as lifestyle programs designed to modify patient characteristics, (c) professional performance programs focusing on improving physician performance, and (d) insurance-targeted programs to address disparities related to coverage and cost.

Table 1: potential determinants for socioeconomic inequalities

Selection effect
Patient-level factors
More severely experienced pain and poorer joint function (more severe disease)
Other joints affected
More comorbidities (e.g., ASA, BMI, diabetes)
Worse mental health status
Unhealthy behaviors (e.g., smoking, diet, less movement/sports)
Lack of knowledge (e.g., unclear expectations, health literacy, delayed presentation at physician, long travel distance/time, uncertainty about treatment preferences and leaving the decision up to the surgeon)
Different preferences (e.g., regarding timing of presentation at physician)
Contextual Factors
Poorer living situation
Job-related challenges: more physical labor, required to return to work earlier after surgery, more frequent unemployment
More debts
Higher levels of stress
Minority ethnicity
Lower income
Lack of a social safety network
Provider-level factors
Later referral by GP
Differential indication by surgeon
Variation in (quality of) healthcare delivery
Variation in case-mix or number of deprived patients presenting at hospital
Insurer-level factors
Co-payments (e.g., deductible)
Coverage and contract differences (e.g. physiotherapy not covered in basic insurance package)
Prognostic: Recovery effect
All the above factors
Less use of physical rehabilitation
Lower bone quality / cementation techniques
Lower compliance with post-operative instructions
Different expectations of the procedure

Patient factors

Patient factors may influence the selection and prognostic effect, or both, thereby further complicating how inequalities are driven. Orthopedic surgeons arguably have the most experience with factors related to a patient's physical health status. These factors include **unhealthy behaviors,** often leading to an increased body mass index (BMI) and the presence of **comorbidities** such as diabetes. These factors are known to be associated with poorer outcomes and higher complication rates following THA and TKA[24, 25] and tend to be more prevalent among the deprived[26]. One study observed a higher prevalence of these factors among minorities undergoing THA or TKA, which is linked to SES. Furthermore, these patients experienced more complications as a result of these factors overall[27].

The increased prevalence of the physical health factors at least partly explains the differential use of THA and TKA in deprived patients, i.e., the selection pathway[28]. Surgeons are typically less inclined to prescribe arthroplasty if they think the expected benefits do not outweigh the risks. Similarly, patients may also be less inclined to undergo arthroplasty for the same reasons. We are unaware of studies eliciting these provider/patient **preferences**. The previously described shared-decision-making tools could be useful if preferences are insufficiently aligned between the patient and the provider. This seems particularly useful for elective arthroplasty, where the benefit-risk judgment is multifactorial.

The LROI cannot provide detailed information on all comorbidity factors but did allow us to shed light on the effect of BMI and a proxy for overall physical health status measured using the American Society of Anesthesiology (ASA) score[29]. These factors explained a relatively small part of the observed inequality (up to 9% for BMI and 16% for ASA), both pre- and postoperatively. Risk-optimizing interventions such as BMI reduction have been found to effectively reduce postoperative complication rates[30], and could modify the inequality observed in PROMs as well. Additionally, enhanced guidance from the surgeon or a related professional (e.g., a nurse) may be valuable. Preoperatively, the lower PROM scores in patients with higher BMI may reflect the more stringent criteria applied (by the patient or provider) in these patients.

Potential factors specifically underlying the selection effect are that low SES patients may lack **communication skills,** or express themselves differently. For example, high SES patients may be able to make their case more pertinent, i.e., be more persuasive[31]. Conversely, physicians may also lack communication skills and physicians may underestimate the true disease burden of low SES patients or they may prioritize lifestyle interventions (e.g., weight loss). Moreover, low vs. high SES patients may have different **expectations** regarding the procedure, which the previously shared decision-making tools may modify. Additionally, low SES patients may more often be unable to continue working in the same capacity (especially in physical jobs), or face **financial barriers**, such as the deductible[32]. On the other hand, employers may be insufficiently accommodating to the patient's condition, and current insurance terms may not adequately account for these barriers[33]. Overall, these factors may result in low SES patients experiencing longer waiting times before requesting surgical relief for OA[34].

Another important determinant is **health literacy**, or lack thereof. Limited health illiteracy may contribute to selection and prognostic disadvantage through various mechanisms. For instance it is likely that patients with low SES struggle with topics such as self-management of OA, rehabilitation after arthroplasty, and more generally with how to navigate the health system. The previously mentioned communication barriers may lead to suboptimal treatment decisions, particularly in terms of timing. Various effective options exist to improve health literacy or to provide tailored information for different levels of health literacy[35]. The ongoing project at Maastricht University Medical Center does precisely that: it targets hip and knee OA patients with limited health literacy, creates more time during consultations to improve decision-making, and provides targeted guidance if deemed necessary[36].

Provider and insurer factors

Factors related to the broader performance of healthcare providers and hospitals are acknowledged as important factors affecting the quality of care. Their impact may also contribute to SES inequalities in PROMs, particularly via the process of care delivery (e.g., through access mechanisms, specific medical decisions and actions, and follow-up monitoring and support provided). These factors have not, to our knowledge, been investigated using registry data; currently, the extent of SES inequalities per provider is not interpreted as a quality indicator. Hospital volume is one such factor: a study in the US used national insurance data and found that minority patients were more likely to receive arthroplasty in low-volume centers and/or by providers performing fewer procedures per year, leading to higher complication rates in such centers[37]. Waiting times (from the time of indication) for arthroplasty have also been found to be longer in deprived patients in the United Kingdom[38]. These examples also demonstrate the difficulty in reducing inequalities. Centralizing procedures like THA and TKA to achieve high-volume centers everywhere can have a twofold effect: increased hospital volumes may improve outcomes, but increased travel time and transport barriers may have the opposite effect in low SES groups[39]. The Netherlands has shorter waiting times in general, and we do not know whether waiting times are longer in the deprived population.

Insurance factors can play a role as well. Reimbursements/payments typically cover the treatment and postoperative care, including treatment of potential complications. Current reimbursement models only to a limited extent compensate for patient characteristics that are associated with increased risks for complications. In the United States, hospitals have heavily invested in prediction models to identify optimal candidates for surgery. This is typically referred to as 'cherry picking' or 'lemon dropping,' where patients with a low risk of complications are preferentially selected for surgery, while patients with high risk of complications are refused care or exposed to barriers, respectively. This phenomenon may contribute to provider-induced inequalities according to SES, as the deprived typically experience these risk factors more frequently (see earlier)[40]. We are unaware of such prediction models using Dutch data. However, based on anecdotal evidence, certain surgeons and/or hospitals may also conduct such selection procedures.

The direct performance of healthcare providers with regard to SES may also be relevant. One example is the differential uptake of guidelines, presumably unintentionally. In **Chapter 5**, we examined SES and racial disparities in the use of peripheral nerve blocks (PNBs) for perioperative pain management in THA and TKA in the United States (US), which reflects access to care to some extent. While we found marked variation in PNB use across US hospitals, this did not translate into SES-inequalities. This does reflect individual inequalities to some extent, as patients may receive different treatments, such as PNBs, based on which healthcare provider they visit. Various reasons may substantiate the differential use of PNB, including different training received or resources available to the provider. This phenomenon requires careful investigation, as providers may deviate from guidelines for valid reasons, and the simple assumption of guideline non-adherence as the driver of inequalities may be overly simplistic. Evaluating performance differences in general is complex when using routinely collected data only. The interrelatedness of factors requires a prudent case-mix adjustment approach, as over-adjustment can take away avoidable differences. On the other hand, lack of case-mix adjustment may result in a situation where cases are insufficiently comparable, hence resulting in an unfair comparison. A case-by-case or audit-like investigation of a subset of patients (e.g., those who underwent reinterventions) and/or providers may provide insight into this phenomenon.

Other factors may contribute to inequalities. There is no single definitive answer or remedy, but gradual change can begin now.

METHODOLOGICAL CHALLENGES

General

The application of PROMs in the context of improving the quality of care and inequality research is not without difficulty, particularly regarding the methods used. Several methods were explored in this thesis.

Lifespan consistent measurement of quality of life

The first issue pertains to the consistent measurement of quality of life across the transition from adolescence into adulthood. In **Chapter 7**, we studied this issue in adolescents with

Adolescent Idiopathic Scoliosis (AIS), and showed that the child-specific and adult version of the EQ-5D had similar validity and test-retest reliability. However, they did not correspond perfectly. We conducted a sensitivity analysis to determine whether these findings differed between *younger* (12-14) and *older* adolescents (15-17), but this was not the case.

Taken together, we concluded that the adult EQ-5D was slightly more suitable for adolescents as it prevents data discontinuities from adolescence into adulthood. This advice applies to both routine clinical care applications and empirical research in patients with AIS. A main limitation of our study is that we used data from a specific orthopedic and diseased population. As a result, these findings may not be generalizable to other common orthopedic conditions in adolescents or the general population.

If verified in other populations, the recommendation to use the adult version in adolescents could be applied more broadly. Moreover, it would facilitate cost-effectiveness evaluations using available EQ-5D value sets. However, the appropriateness of using an adult value set in adolescents remains a subject of ongoing debate[41].

What constitutes a relevant health change, individually and at the aggregate level

The second challenge is to agree on one (or more) definitions of what constitutes a relevant health change, which can also be applied to a PROM score difference. The concept of relevance concerns the interpretation of a difference, differentiating between the individual and aggregate levels, and extends beyond merely determining whether a change occurred and its associated probability (e.g., using confidence intervals). Defining relevance was a recurrent challenge throughout **Chapters 2**, **3**, **4**, and **6**.

In terms of routine clinical care applications, at the micro level, monitoring efforts in orthopedic patients lacked a formal definition or threshold for what constitutes ‘abnormal recovery’ (**Chapter 2**). In other words, at what point does a worsening symptom score necessitate evaluation and/or treatment? This differs from the use of PROMs or symptom inventories to identify infections in the previously mentioned eHealth app[8]. In the former case, the PROM itself is the outcome requiring assessment of relevance, while in the latter case, the PROM or symptom inventory is used to predict or identify a potentially harmful complication. In **Chapters 3**, **4** and **6**, we encountered this issue at the meso-/macro-levels, facing the question of whether the observed PROM differences related to SES and the COVID-19 lockdown were relevant.

It is important to quantitatively express the relevance of the impact on PROM scores. More specifically, when is a health outcome difference large enough to indicate a need for a clinical change, such as re-intervention at the individual level, or policy adjustment at the group level[42, 43]? The concept of Minimal Important Differences (MID) or Minimal Clinically Important Differences (MCIDs) has been proposed to define a relevant health change[42]. This parameter aims to capture the smallest worthwhile longitudinal change in PROM perceived by the patient, and is often applied at the group level. While straightforward in theory, applying it in practice is challenging. A key consideration is that all MIDs apply to a pre-post changes in PROM datasets, meaning one cannot use MIDs to compare differences between groups (and not to cross-sectional differences).

There are three global methods to calculate the MID: distribution-based, individual or within-patient anchor-based, and consensus-based[44]. Distribution-based methods are data-driven and assert that a convenient statistical metric reflects the required patient relevance; mostly, the standard deviation is used. An important caveat is that in a homogeneous patient sample, MIDs automatically become smaller because the standard deviation is smaller. As a result, distribution-based methods are relatively infrequently used.

Anchor-based approaches are most popular, particularly in orthopedics. Typically, Global Rating of Change (GROC) questions are used, which assess whether patients improved, deteriorated, or stayed the same in terms of their health. For example, the within-patient anchor-based approach often defines the difference between patients who experienced ‘no improvement’ and ‘slight improvement’ on the GROC question as the MID. Anchor-based approaches are less arbitrary, as they relate a measurable quantity of change (in PROMs terms) to a consequential observation or the patient’s elicited judgment of the experienced change.

However, there are several issues with this parameter. First, these approaches are influenced by the choice of anchor and the threshold values selected by the user. Second, estimates of MIDs vary depending on the initial condition (pre-score of the respective PROM), and patient characteristics (e.g., SES), as patients may assess the MID from different perspectives[45]. Third, MIDs are population-specific. Fourth, experienced change is asymmetrical: an improvement from 80 to 90 could be perceived as smaller than the deterioration from 90 to 80. This asymmetry is also likely to depend on the pre-score. Fifth, commonly used health change questions are retrospective and prone to external biases, such as recall bias and generally have worse measurement characteristics compared to the PROM scores they are meant to complement[46]. Finally, while anchor-based MIDs provide an empirical basis for defining relevant health changes, they do not capture societal preferences. The distinction between ‘no improvement’ and ‘slight improvement’ alone does not justify policy change. To what extent would a patient be willing to sacrifice time, money, or accept health risks to move from no improvement to slight improvement?

Consensus-based approaches have also been proposed as methods to arrive at the MID. These methods may involve Delphi studies or panel discussions to gather expert opinions and determine or triangulate the MID. However, they are largely affected by the selected panel members. The earlier-referred panel discussion on SES inequalities that we organized, shared some similarities with this approach. Although average differences between SES groups did not always reach currently accepted MIDs/MCIDs, in particular postoperatively, they were still considered important by panel members.

The use of a living-area-based proxy for individual socioeconomic status

Another challenge is the measurement of individual SES. An area-based proxy, as used in **Chapters 3**, **4,** and **5**, approximates individual SES measures such as household income, education level and employment; however, it does not correspond 1-on-1[47-49]. Nevertheless, an area-based measure offers advantages, particularly for small areas, as it captures relevant environmental SES information while also serving as an indicator of individual SES. In **Chapters 3** and **4**, we used an area-based SES indicator linked at the 4-digit postal code level. This indicator, for example, is known to strongly correlate with measures of local living quality. Moreover, its statistical explanatory power in socioeconomic studies is considerable (independent of earlier described factors), which adds credibility to its use as a valid proxy for individual SES. Finally, an area-based measure of SES avoids additional registration burden of individual SES measures for the clinician or patient. In our case, the use of an area-based measure was pragmatic, as individual SES measures are not collected in the LROI.

Inequalities beyond socioeconomic disparity

We were aware of our omission of other widely accepted health inequality factors, represented by the PROGRESS-plus acronym. The PROGRESS core set includes place of residence, race/ethnicity, occupation, gender/sex, religion, education, socioeconomic status, and social capital. While some of these factors may be indirectly captured through our area-based SES measure, we acknowledge that additional inequalities and underlying mechanisms linked to these variables likely exist, either coinciding or interacting alongside SES.

Measurement error of PROMs

In **Chapter 4**, SES inequalities in a range of PROMs in THA/TKA patients were explored. The largest negative difference by SES was observed in measures of joint functioning, pain, and quality of life. However, this did not translate to worse overall general health as measured by the EQ Visual Analogue Scale (VAS) or to greater dissatisfaction among patients with low SES. One potential explanation is the phenomenon of reporting heterogeneity (RH), which refers to the tendency of individuals to interpret questions or response scales of PROMs differently. The EQ VAS and satisfaction questionnaires may be disproportionately impacted by RH. Current LROI data do not allow for a definitive conclusion on the presence of RH.

The findings underline that, in the evaluation of inequalities, the choice of instrument can significantly affect the degree of the observed inequalities. One method to assess RH is by collecting external anchoring vignettes, for example, from a random sub-sample of patients captured in QRs[50]. A vignette briefly describes a hypothetical situation regarding a patient's life and disease, with particular attention to the measured health domains (e.g., OHS/OKS or EQ-5D). By asking patients to rate the same hypothetical situation, one can estimate for each patient how much their report of their own health status on the domains of interest deviates from what respondents normally would rate. This deviance in reporting can then be adjusted for, providing a better estimate of the objective change in self-reported health.

FUTURE PERSPECTIVES

PROMs in quality of care

Looking ahead, the routine application of PROMs holds promising potential to transform orthopedic care. However, merely collecting of PROMs data does not improve the quality of care. It requires a coordinated effort from researchers, clinicians, and other stakeholders to translate these data into quality improvements. Their use in routine clinical care should be rigorously tested, and areas or patient populations where PROMs are effective (or ineffective) in improving patients outcomes should be identified.

Attention should also be directed toward the cost-effectiveness of these applications, as collecting PROMs data under secure conditions is labor-intensive and costly. This aspect was not addressed in our review, as it was not included in our search criteria[51]. The selection and procurement of PROMs and other potential explanatory data for various purposes is a complex endeavor. Effective reuse of data, combined with a stepwise implementation in successful areas, appears to be the best approach, as it helps reduce relative costs and serves the best interest of patients. Overall, insurers could play a significant role in stimulating and funding these initiatives, as they may result to better outcomes for all.

Successful applications include innovative shared-decision making tools such as the 'Patients Like Me' tool (LROI) [52]. The addition of an educational component to this tool presents an opportunity for improvement. Overall, studies employing shared-decision making tools have found enhanced decision quality; future research should clarify whether this translates into less variation in PROMs after surgery. Additionally, the impact of these tools on patient-reported experience measures (PREMs) should also be investigated.

Another promising area is the integration of PROMs into eHealth applications that combine educational material with wound imaging features to improve the monitoring of short-term surgical recovery. Future studies should focus on defining clear strategies for when a symptom or combination of symptoms is considered 'abnormal' and requires clinical attention. While the integration of PROMs with other data has the potential to improve quality of care at the meso-/macro-level, evidence supporting this is currently lacking. There is an urgent need for practical applications at this level, such as their incorporation into clinical dashboards alongside other clinical outcome data and their use in quality improvement cycles (e.g., PDCA).

Integration with existing processes is essential. Any newly added questionnaires should be well justified, particularly those intended for micro-level use. Moreover, their collection must be streamlined to minimize the burden on both patients and providers, especially since patients already complete pre- and post-operative PROMs for many orthopedic procedures. 'Patient journey' apps have the potential to efficiently incorporate these questionnaires, centralizing all assessments and enhancing the user experience. Such apps also enable both patients and providers to view their PROM results, which can lead to increased engagement and reduce the perceived questionnaire burden.

Socioeconomic inequalities in healthcare

PROMs collected in the Dutch implants registry LROI have proven valuable for studying socioeconomic inequalities in orthopedic outcomes. They offer opportunities to unravel the sources of observed inequalities in order to determine ways to reduce them. One approach may be to gather information on potential drivers and assess their role in inequality pathways. Ideally, LROI data should be enhanced to achieve this goal. The LROI will launch a modular registry in the near future, which could facilitate the collection of additional information on inequality measures and intermediary factors in a subset of Dutch orthopedic hospitals. This initiative might provide strong evidence for identifying which targets should be prioritized to reduce inequalities. Traditionally, targets have been approached from the patient perspective (e.g., the patient lacks understanding); however, attention should also be directed toward the provider perspective (e.g., the doctor insufficiently explains information). Furthermore, pursuing data linkage as described earlier, while maintaining the anonymity and privacy of individual patients, could aid in identifying targets and improving overall quality of care.

At the same time, we should consider investing in the development of an 'intervention' aimed at reducing inequalities. One potential approach is an integrative approach that acknowledges the complexities of determinants of inequalities, and their interactions. This recognition has been identified as one of the most significant barriers to successful inequality-reducing interventions in general[53]. A centralized platform, potentially hosted under the Dutch Orthopedic Society (Dutch translation: NOV) could manage, facilitate, and coordinate different stakeholders across various levels to target multiple potential drivers of inequalities. The orthopedic patient's journey typically begins with the onset of osteoarthritis and may culminate in arthroplasty. However, it is essential to recognize that many osteoarthritis patients never require such procedures. Throughout this journey, patients interact with various healthcare providers, including general practitioners, physiotherapists, and orthopedic surgeons. A coordinating platform could enable these stakeholders to collaborate in addressing inequalities, potentially yielding synergistic effects.

In this integrative network, the orthopedic surgeons could take on a leadership role, which aligns with the growing trend of orthopedic specialists becoming more actively involved in preventative care, promoting mobility, and encouraging physical activity. Developing and implementing new interventions to reduce inequalities can be costly and often necessitates layering these interventions on top of existing ones. An integrative platform would modify the existing healthcare system, empowering relevant care providers to take a leadership role in reducing inequalities, facilitates mutual learning, and specifically evaluating and implementing multiple (existing) interventions for inequality reduction simultaneously. Ultimately, this approach may prove to be relatively more cost-effective.

FINAL REMARKS

This thesis has illustrated the potential use of PROMs to improve quality of care in orthopedics. Our research, along with that of colleagues, highlights the potential of novel PROM applications, such as shared-decision making and monitoring tools. In these applications, PROMs data serve as valuable complement to, rather than replacement for, other essential data such as clinical parameters and symptom inventory lists. Importantly, the findings demonstrate the value of PROMs in elucidating socioeconomic inequalities. The findings underscore the need for targeted interventions to reduce these inequalities. Moreover, PROMs were effective to gain insight into the impact of a national event, namely the COVID-19 pandemic, on orthopedic outcomes.

However, methodological challenges remain, including the interpretation of clinical relevance, and identifying and accounting for potential sources of bias, such as reporting heterogeneity. Continued effort should be put toward providing examples of the effective use of PROMs data and methods to address such methodological challenges, paving the way for PROMs to have an even greater impact on improving patient-centered care in the future.

REFERENCES

1. Jayakumar P, Moore MG, Furlough KA, Uhler LM, Andrawis JP, Koenig KM, et al. Comparison of an Artificial Intelligence-Enabled Patient Decision Aid vs Educational Material on Decision Quality, Shared Decision-Making, Patient Experience, and Functional Outcomes in Adults With Knee Osteoarthritis: A Randomized Clinical Trial. JAMA Netw Open 2021; 4: e2037107.
2. Bansback N, Trenaman L, MacDonald KV, Durand D, Hawker G, Johnson JA, et al. An online individualised patient decision aid improves the quality of decisions in patients considering total knee arthroplasty in routine care: A randomized controlled trial. Osteoarthr Cartil Open 2022; 4: 100286.
3. Volkmann ER, FitzGerald JD. Reducing gender disparities in post-total knee arthroplasty expectations through a decision aid. BMC Musculoskelet Disord 2015; 16: 16.
4. Hafkamp FJ, Gosens T, de Vries J, den Oudsten BL. Do dissatisfied patients have unrealistic expectations? A systematic review and best-evidence synthesis in knee and hip arthroplasty patients. EFORT Open Rev 2020; 5: 226-240.
5. Shay LA, Lafata JE. Where is the evidence? A systematic review of shared decision making and patient outcomes. Med Decis Making 2015; 35: 114-131.
6. Pooni A, Brar MS, Anpalagan T, Schmocker S, Rashid S, Goldstein R, et al. Home to Stay: A Randomized Controlled Trial Evaluating the Effect of a Postdischarge Mobile App to Reduce 30-Day Readmission Following Elective Colorectal Surgery. Ann Surg 2023; 277: e1056-e1062.
7. Gawande, A. (2011). The checklist manifesto. Profile Books.
8. Scheper H, Derogee R, Mahdad R, van der Wal RJP, Nelissen R, Visser LG, de Boer MGJ. A mobile app for postoperative wound care after arthroplasty: Ease of use and perceived usefulness. Int J Med Inform 2019; 129: 75-80.
9. Holm I, Pripp AH, Risberg MA. The Active with OsteoArthritis (AktivA) Physiotherapy Implementation Model: A Patient Education, Supervised Exercise and Self-Management Program for Patients with Mild to Moderate Osteoarthritis of the Knee or Hip Joint. A National Register Study with a Two-Year Follow-Up. J Clin Med 2020; 9.
10. Kontos E, Blake KD, Chou WY, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014; 16: e172.
11. Marshall DA, Trenaman L, MacDonald KV, Johnson JA, Stacey D, Hawker G, et al. Impact of an online, individualised, patient reported outcome measures based patient decision aid on patient expectations, decisional regret, satisfaction, and health-related quality-of-life for patients considering total knee arthroplasty: Results from a randomised controlled trial. J Eval Clin Pract 2023; 29: 513-524.
12. Al Sayah F, Lahtinen M, Bonsel GJ, Ohinmaa A, Johnson JA. A multi-level approach for the use of routinely collected patient-reported outcome measures (PROMs) data in healthcare systems. J Patient Rep Outcomes 2021; 5: 98.
13. Varagunam M, Hutchings A, Neuburger J, Black N. Impact on hospital performance of introducing routine patient reported outcome measures in surgery. J Health Serv Res Policy 2014; 19: 77-84.
14. Kumar RM, Fergusson DA, Lavallée LT, Cagiannos I, Morash C, Horrigan M, et al. Performance Feedback May Not Improve Radical Prostatectomy Outcomes: The Surgical Report Card (SuRep) Study. J Urol 2021; 206: 346-353.
15. Weingarten SR, Kim CS, Stone EG, Kristopaitis RJ, Pelter M, Sandhu M. Can peer-comparison feedback improve patient functional status? Am J Manag Care 2000; 6: 35-39.

16. Boyce MB, Browne JP. The effectiveness of providing peer benchmarked feedback to hip replacement surgeons based on patient-reported outcome measures--results from the PROFILE (Patient-Reported Outcomes: Feedback Interpretation and Learning Experiment) trial: a cluster randomised controlled study. BMJ Open 2015; 5: e008325.
17. Garduño-López AL, Acosta Nava VM, Castro Garcés L, Rascón-Martínez DM, Cuellar-Guzmán LF, Flores-Villanueva ME, et al. Towards Better Perioperative Pain Management in Mexico: A Study in a Network of Hospitals Using Quality Improvement Methods from PAIN OUT. J Pain Res 2021; 14: 415-430.
18. Zaslansky R, Chapman CR, Baumbach P, Bytyqi A, Castro Lopes JM, Chetty S, et al. Improving perioperative pain management: a preintervention and postintervention study in 7 developing countries. Pain Rep 2019; 4: e705.
19. Haller G, Agoritsas T, Luthy C, Piguet V, Griesser AC, Perneger T. Collaborative quality improvement to manage pain in acute care hospitals. Pain Med 2011; 12: 138-147.
20. Partridge T, Carluke I, Emmerson K, Partington P, Reed M. Improving patient reported outcome measures (PROMs) in total knee replacement by changing implant and preserving the infrapatella fatpad: a quality improvement project. BMJ Qual Improv Rep 2016; 5.
21. Sira Consulting B.V., Report on the use of quality information in the context of quality of care [In gesprek over kwaliteit van zorg - Het gebruik van kwaliteitsinformatie in zes zorgsectoren], available at: https://www.zorginstituutnederland.nl/publicaties/rapport/2022/07/14/onderzoek-naar-het-verzamelen-en-gebruik-van-kwaliteitsinformatie-in-de-zorg, accessed on: 05-08-2024.
22. OECD. Health at a Glance2023.
23. "Dashboard PROMs LUMC stimuleert Samen beslissen - Monique Baas, LUMC", Linnean initiative, available at: https://www.linnean.nl/nieuws+linnean/2028853.aspx, accesed on: 15-08-2024.
24. Qin W, Huang X, Yang H, Shen M. The Influence of Diabetes Mellitus on Patients Undergoing Primary Total Lower Extremity Arthroplasty: A Systematic Review and Meta-Analysis. BioMed Research International 2020; 2020: 6661691.
25. Zusmanovich M, Kester BS, Schwarzkopf R. Postoperative Complications of Total Joint Arthroplasty in Obese Patients Stratified by BMI. The Journal of Arthroplasty 2018; 33: 856-864.
26. Algren MH, Bak CK, Berg-Beckhoff G, Andersen PT. Health-Risk Behaviour in Deprived Neighbourhoods Compared with Non-Deprived Neighbourhoods: A Systematic Literature Review of Quantitative Observational Studies. PLoS One 2015; 10: e0139297.
27. Liimakka AP, Amen TB, Weaver MJ, Shah VM, Lange JK, Chen AF. Racial and Ethnic Minority Patients Have Increased Complication Risks When Undergoing Surgery While Not Meeting Clinical Guidelines. J Bone Joint Surg Am 2024; 106: 976-983.
28. McLaughlin J, Elsey J, Kipping R, Owen-Smith A, Judge A, McLeod H. Access to hip and knee arthroplasty in England: commissioners' policies for body mass index and smoking status and implications for integrated care systems. BMC Health Services Research 2023; 23: 77.
29. Dripps RD. New classification of physical status. Anesthesiology 1963; 24: 111.
30. Lau LCM, Chan PK, Lui TWD, Choi SW, Au E, Leung T, et al. Preoperative weight loss interventions before total hip and knee arthroplasty: a systematic review of randomized controlled trials. Arthroplasty 2024; 6: 30.
31. Willems S, De Maesschalck S, Deveugele M, Derese A, De Maeseneer J. Socio-economic status of the patient and doctor-patient communication: does it make a difference? Patient Educ Couns 2005; 56: 139-146.

32. Adu Y, Hurley J, Ring D. Are There Racial and Ethnic Variations in Patient Attitudes Toward Hip and Knee Arthroplasty for Osteoarthritis? A Systematic Review. Clin Orthop Relat Res 2024; 482: 1417-1424.
33. Lopes FV, Riumallo Herl CJ, Mackenbach JP, Van Ourti T. Patient cost-sharing, mental health care and inequalities: A population-based natural experiment at the transition to adulthood. Soc Sci Med 2022; 296: 114741.
34. Neuburger J, Hutchings A, Allwood D, Black N, van der Meulen JH. Sociodemographic differences in the severity and duration of disease amongst patients undergoing hip or knee replacement surgery. J Public Health (Oxf) 2012; 34: 421-429.
35. Visscher BB, Steunenberg B, Heijmans M, Hofstede JM, Devillé W, van der Heide I, Rademakers J. Evidence on the effectiveness of health literacy interventions in the EU: a systematic review. BMC Public Health 2018; 18: 1414.
36. "Samen Beslissen bij artrose: Het Beweeghuis voor iedereen", available at: https://www.uitkomstgerichtezorg.nl/publicaties/artikelen/2023/08/31/samen-beslissen-bij-artrose-het-beweeghuis-voor-iedereen, accessed 12-07-2024.
37. Cohen-Rosenblum A, Richardson MK, Liu KC, Wang JC, Piple AS, Hansen C, et al. Medicaid Patients Undergo Total Joint Arthroplasty at Lower-Volume Hospitals by Lower-Volume Surgeons and Have Poorer Outcomes. J Bone Joint Surg Am 2023; 105: 979-989.
38. Laudicella M, Siciliani L, Cookson R. Waiting times and socioeconomic status: Evidence from England. Social Science & Medicine 2012; 74: 1331-1341.
39. Orringer M, Roberts H, Ward D. Racial and Socioeconomic Differences in Distance Traveled for Elective Hip Arthroplasty. J Am Acad Orthop Surg Glob Res Rev 2022; 6.
40. Bernstein DN, Reitblat C, van de Graaf VA, O'Donnell E, Philpotts LL, Terwee CB, Poolman RW. Is There An Association Between Bundled Payments and "Cherry Picking" and "Lemon Dropping" in Orthopaedic Surgery? A Systematic Review. Clin Orthop Relat Res 2021; 479: 2430-2443.
41. Devlin N, Pan T, Kreimeier S, Verstraete J, Stolk E, Rand K, Herdman M. Valuing EQ-5D-Y: the current state of play. Health and Quality of Life Outcomes 2022; 20: 105.
42. Dekker J, de Boer M, Ostelo R. Minimal important change and difference in health outcome: An overview of approaches, concepts, and methods. Osteoarthritis and Cartilage 2024; 32: 8-17.
43. Maredupaka S, Meshram P, Chatte M, Kim WH, Kim TK. Minimal clinically important difference of commonly used patient-reported outcome measures in total knee arthroplasty: review of terminologies, methods and proposed values. Knee Surg Relat Res 2020; 32: 19.
44. Kallogjeri D, Spitznagel EL, Jr., Piccirillo JF. Importance of Defining and Interpreting a Clinically Meaningful Difference in Clinical Research. JAMA Otolaryngology–Head & Neck Surgery 2020; 146: 101-102.
45. Sheean AJ, Tenan MS, DeFoor MT, Cognetti DJ, Bedi A, Lin A, et al. Minimal important clinical difference values are not uniformly valid in the active duty military population recovering from shoulder surgery. J Shoulder Elbow Surg 2024.
46. Schmitt J, Di Fabio RP. The Validity of Prospective and Retrospective Global Change Criterion Measures. Archives of Physical Medicine and Rehabilitation 2005; 86: 2270-2276.
47. Subramanian SV, Chen JT, Rehkopf DH, Waterman PD, Krieger N. Comparing individual- and area-based socioeconomic measures for the surveillance of health disparities: A multilevel analysis of Massachusetts births, 1989-1991. Am J Epidemiol 2006; 164: 823-834.

48. Mesterton J, Willers C, Dahlström T, Rolfson O. Comparison of individual and neighbourhood socioeconomic status in case mix adjustment of hospital performance in primary total hip replacement in Sweden: a register-based study. BMC Health Serv Res 2020; 20: 645.
49. Moss JL, Johnson NJ, Yu M, Altekruse SF, Cronin KA. Comparisons of individual- and area-level socioeconomic status as proxies for individual-level measures: evidence from the Mortality Disparities in American Communities study. Population Health Metrics 2021; 19: 1.
50. d'Uva TB, O'Donnell O, van Doorslaer E. Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans. Int J Epidemiol 2008; 37: 1375-1383.
51. Pronk Y, Pilot P, Brinkman JM, van Heerwaarden RJ, van der Weegen W. Response rate and costs for automated patient-reported outcomes collection alone compared to combined automated and manual collection. J Patient Rep Outcomes 2019; 3: 31.
52. Patients like me [Patienten zoals ik], Dutch Orthopedic Society (NOV) and Dutch Arthroplasty Registry (LROI), available at: https://www.zorgvoorbeweging.nl/patienten-zoals-ik, accessed 10-07-2024.
53. Wilderink L, Bakker I, Schuit AJ, Seidell JC, Pop IA, Renders CM. A Theoretical Perspective on Why Socioeconomic Health Inequalities Are Persistent: Building the Case for an Effective Approach. Int J Environ Res Public Health 2022; 19.

Appendix

English summary

Dutch summary

List of publications

Acknowledgements (‘dankwoord’)

PhD portfolio

List of abbreviations

ENGLISH SUMMARY

Patient-Reported Outcome Measures (PROMs) are self- or proxy-reported questionnaires that assess the experienced burden of disease and/or quality of life. They play a major role in evaluating the effectiveness of interventions. Currently, the World Health Organization and other global bodies advise incorporating PROMs throughout health systems, as they offer a deeper understanding of what truly matters to the patient and are thus important reflectors of quality of care. The field of orthopedics has also responded to this call, and PROMs now form an important outcome in the evaluation of quality of care alongside other clinical metrics such as revision rates. Over time, PROMs have become an essential component of quality registries (QRs) alongside other outcome data, and patient and procedure characteristics. In the Netherlands, the Dutch Arthroplasty Registry (LROI) was founded in 2007 and has collected PROMs since 2014. Extensive efforts from all relevant personnel and stakeholders have resulted in high-quality registry data. This thesis illustrates how orthopedic QR PROM data can be used in routine clinical care applications and empirical research to identify opportunities for enhanced quality of care. Moreover, it addressed specific methodological challenges in this context.

In **Chapter 2**, we conducted a systematic review on the impact of PROMs in routine clinical care applications, with our scope not restricted to orthopedics alone. At the micro- (patient) level, routine PROM application has matured notably, and various patient communication, symptom screening, and monitoring initiatives have been found to positively affect patient outcomes. Applications at the meso- (organizational) and macro-level (national), such as comparing the performance of providers or insurance parties using PROMs, are still in their infancy. Specific effective examples in orthopedics were highlighted. First, we found several studies effectively using PROMs in shared decision-making tools to inform patients undergoing knee arthroplasty, thereby improving their decision quality regarding whether to undergo the procedure. One study successfully employed a plan-do-check-act cycle to evaluate the implementation of a novel knee prosthesis in a national healthcare system. Determinants of PROM applications that were associated with an increased chance of success were also identified, such as a clear link between the PROM application and clinical actions and/or care pathways.

In the subsequent chapters, we shifted towards empirical research applications to illustrate the potential of PROMs and to identify opportunities for improved orthopedic care. We conducted several studies on the presence of socioeconomic inequalities and on the impact of the COVID-19 pandemic on orthopedic outcomes.

In **Chapter 3**, we conducted a national (Dutch) study using LROI data on socioeconomic inequalities in PROMs of Total Hip and Knee Arthroplasty (THA, TKA) patients with the

indication of osteoarthritis. An area-based measure of socioeconomic status (SES) indicator based on neighborhood income, unemployment rate, and education level was used as the exposure variable. The main outcomes were the EQ-5D reflecting the general quality of life, and the Oxford Hip/Knee Score (OHS/OKS) reflecting joint pain and functioning. The PROMs obtained were measured preoperatively and at 12-month follow-up. We discerned two pathways to inequalities. Firstly, deprived patients had lower quality of life and joint functioning before surgery, indicating a selection effect either by patient or provider. These differences decreased but persisted when measured using 12-month follow-up PROMs. These findings indicated that deprived patients experience a worse recovery trajectory. This result enables a more precise definition of targets for decreasing socioeconomic inequalities in the future.

In **Chapter 4**, we used the same dataset as in **Chapter 3** to study how SES inequalities manifest across various PROMs and which domains were most affected in THA and TKA patients. We additionally obtained, in addition to the abovementioned PROMs, the EQ Visual Analogue Scale (VAS) which represents overall experienced health and Numerical Rating Scales (NRS) for pain in the joint. Moreover, an NRS satisfaction with the undergone procedure was obtained, available for TKA patients only at 12-month follow-up. The findings revealed that SES inequalities varied across different PROM instruments. Deprived patients experienced similar negative differences on the EQ-5D and OHS/OKS, although the difference on the OKS in TKA patients was larger compared to the EQ-5D preoperatively. NRS for pain demonstrated differences comparable to those of the EQ-5D and OKS The EQ VAS showed smaller differences by SES, compared to EQ-5D and OHS/OKS, NRS satisfaction showed no inequality at 12 months for TKA patients. Domain analysis of the EQ-5D revealed that usual activities explained up to 46% of differences in both THA and TKA patients, while anxiety/depression had a limited role (up to 17%). For OHS/OKS, functioning was more impacted in THA (up to 61%) and pain in TKA (up to 68%), especially at 12 months. The findings illustrate that deprived patients experience worse joint functioning, pain, and quality of life. Yet, this did not translate into deprived patients rating their overall general health as worse or experiencing dissatisfaction with the arthroplasty received. Future research should explore whether PROMs like EQ VAS and satisfaction scales accurately reflect true inequality-related effects or are disproportionally affected by measurement error such as differential interpretation of wording.

In **Chapter 5**, we broadened our view of inequalities by studying the role of socioeconomic, patient, and hospital characteristics in the perioperative utilization of peripheral nerve blocks (PNBs) in THA and TKA patients from the United States. We found that, contrary to our hypotheses, socioeconomic variables including SES and race/ethnicity had limited to no effect. Rather, practice variation was the result of clinical (e.g., less use in semi- and non-elective arthroplasty) and hospital characteristics. These findings emphasized individual and practice variations in PNB use. Subsequently, the association between PNB utilization and complication rates, readmission rates, and length of stay was analyzed, controlling for potential confounders. In TKA, PNB use was associated with reduced complications and length of stay; this association was not found in THA. The findings emphasize the need for more standardized provision of PNBs. Moreover, more detailed information is needed to conclude what the drivers are (e.g., financial and time incentives or the role of specialist training).

In **Chapter 6**, we conducted a national (Dutch) study using LROI data on the impact of the COVID-19 pandemic and lockdown on PROMs in THA and TKA patients, for any indication. We again obtained data on the EQ-5D and OHS/OKS. We found that during the COVID-19 lockdown, THA patients had slightly worse preoperative quality of life and joint functioning, suggesting the selection of patients with greater urgency. This effect was not present in TKA patients, for which we did not have an explanation. Postoperative outcomes hardly differed, indicating that the recovery trajectory after arthroplasty was probably not affected by the COVID-19 pandemic or lockdown.

In **Chapter 7**, we studied a methodological issue pertaining to the longitudinal follow-up of quality of life from adolescence into adulthood. We collected and compared a child-specific (EQ-5D-Y-5L) and an adult (EQ-5D-5L) generic quality of life instrument in adolescents with Adolescent Idiopathic Scoliosis (AIS) receiving bracing therapy. We specifically tested whether they were equivalent, and whether they performed similarly regarding important instrument characteristics. While they were similar in terms of validity and test-retest reliability, differences in score distribution were present. Taken together, we advise using the EQ-5D-5L to monitor the quality of life in adolescent patients with AIS, as this avoids switching instruments and which would create data discontinuities when patients transition from adolescence into adulthood (>17 year of age).

In **Chapter 8**, we provide a general discussion of the above-mentioned chapters, specifically on three topics, namely quality of care, socioeconomic inequalities, and methodological challenges in these contexts. Some highlights include the following.

Quality of care: In terms of routine clinical care applications of PROMs, the identified shared decision-making tools and post-operative monitoring apps (micro-level) show promise but require further evaluation. In particular, it is unclear whether improved decision quality will also improve patient outcomes such as quality of life. At the meso-/macro-level, evidence is scarce, and we are in need of examples of how to employ PROMs at this level, supporting or refuting their effectiveness.

Socioeconomic inequalities: PROMs provided insights into mechanisms underlying inequalities. However, we lack knowledge about intermediary determinants driving inequalities. An example may be that patients with low SES have worse health literacy, which reflects an individual's knowledge and capability to navigate a healthcare system. Moreover, provider-related factors may play a role, such as the (unconscious) selection of optimal candidates for arthroplasty by the provider. We provided a potential long-term strategy for reducing inequalities in orthopedics.

Methodological challenges: One issue we encountered across most of our studies was determining the clinical relevance of differences in PROMs and PROM changes. Advice is provided on how to approach this issue, particularly in in the context of measurement of inequalities.

Final remarks

This thesis used PROM data to identify opportunities for enhanced clinical practice in orthopedic surgery while deepening our understanding of their application in this field. Novel PROM applications were highlighted, such as shared decision-making tools and monitoring of recovery after surgery using eHealth apps. Moreover, we showed how PROMs can be used to elucidate socioeconomic inequalities in orthopedic outcomes. The findings underscored the need for more research into underlying determinants of these inequalities and the need for targeted interventions to reduce these differences. Methodological challenges remain, including the interpretation of clinical relevance and accounting for potential bias due to measurement error. Continued effort should be put toward providing examples of effective use of PROMs data and methods to address such methodological challenges. This will aid in further paving the way for PROMs to have an even greater impact on patient-centered care.

DUTCH SUMMARY

Patiënt Gerapporteerde Uitkomstmaten (Engelse afkorting: PROMs) zijn door de patiënt zelf of als dat niet kan ouders of begeleiders (‘proxy’)-gerapporteerde vragenlijsten die de ervaren ziektelast en/of levenskwaliteit beoordelen. Ze spelen een belangrijke rol in het evalueren van de effectiviteit van behandeling. De Wereldgezondheidsorganisatie en andere instanties adviseren om PROMs in gezondheidssystemen op te nemen als indicator voor kwaliteit van zorg, omdat ze een beeld geven wat echt belangrijk is voor de patiënt. Ook in de orthopedie wordt veel gebruik gemaakt van PROMs. Naast andere klinische uitkomsten zoals revisie risico vormen PROMs inmiddels een belangrijke uitkomst in de evaluatie van de kwaliteit van zorg. PROMs zijn ook een essentieel onderdeel geworden van kwaliteitsregistraties (Engelse afkorting: QRs). De Landelijke Registratie Orthopedische Implantaten (LROI) is in 2007 opgericht. Sinds 2014 verzamelt de LROI ook PROMs. Dagelijkse inspanningen van alle relevante medewerkers en belanghebbenden hebben geresulteerd in hoge kwaliteit registergegevens. Dit proefschrift toont aan hoe PROM-gegevens in orthopedische QRs gebruikt kunnen worden in routinematige klinische zorg maar ook in empirisch onderzoek om de kwaliteit van de zorg te verbeteren. Daarnaast worden enkele methodologische uitdagingen in deze context bestudeerd.

In **hoofdstuk 2** wordt een systematische review besproken, uitgevoerd naar de impact van PROMs in routinematige klinische zorgtoepassingen, waarin we naast de orthopedie ook keken in andere specialismen. Op micro- (patiënt) niveau is de routinematige toepassing van PROMs aanzienlijk ontwikkeld. Van verschillende initiatieven op het gebied van patiëntcommunicatie, symptoom screening en monitoring is een positief effect op patiëntuitkomsten vastgesteld. Toepassingen op meso- (organisatorisch) en macro-niveau (nationaal), zoals het vergelijken van uitkomsten van zorgverleners of verzekeringspartijen met behulp van PROMs, staan nog aan het begin van verdere ontwikkeling. Specifieke effectieve voorbeelden in de orthopedie worden uitgelicht. Ten eerste vonden we verschillende onderzoeken waarin PROMs effectief werden gebruikt in tools om mogelijke kandidaten voor een knie prothese te informeren en gezamenlijke besluitvorming te faciliteren. De tools leken de kwaliteit van besluitvorming over het doel van de voorgestelde ingreep te verbeteren. Eén onderzoek paste met succes PROMs toe in een plan-do-check-act cyclus toe om de implementatie van een nieuwe knieprothese in een nationaal gezondheidszorgsysteem te evalueren. Determinanten van PROM-toepassingen die leiden tot een verhoogde kans op effectiviteit werden ook geïdentificeerd, zoals een duidelijke link tussen de PROM toepassing en klinische acties en/of zorgpaden.

In de overige hoofdstukken komen empirische toepassingen van PROMs aan de orde om het potentieel daarvan te illustreren en om mogelijkheden voor verbeterde orthopedische zorg te identificeren. We hebben verschillende onderzoeken uitgevoerd naar de aanwezigheid

van sociaal-economische ongelijkheden en naar de impact van de COVID-19 pandemie op orthopedische uitkomsten.

In **hoofdstuk 3** presenteerden we een landelijke (Nederlandse) studie uit met LROI gegevens naar socio-economische ongelijkheden in PROMs van patiënten die een Totale Heup of Knie Prothese (THP, TKP) kregen voor artrose. Een wijk-indicator voor sociaal-economische status (SES), gebaseerd op buurtinkomen, werkloosheidspercentage en opleidingsniveau, werd gebruikt als blootstellingsvariabele. De hoofduitkomsten waren de EQ-5D en de Oxford Hip/Knee Score (OHS/OKS), die respectievelijk de algemene kwaliteit van leven en pijn/functioneren van het gewricht weerspiegelen. We ontdekten twee onderliggende mechanismen die ten grondslag liggen aan ongelijkheden. Ten eerste hadden patiënten met een lagere (SES) een lagere kwaliteit van leven en een lagere score ten aanzien van het functioneren van het gewricht vóór de operatie in vergelijking met patiënten met een hogere SES. Dit zou kunnen passen bij een mogelijk selectie-effect door de patiënt of de zorgverlener. Daarnaast hadden patiënten met een lagere SES, lagere scores na de operatie. Dit kan passen bij een slechter hersteltraject, alhoewel dit verschil kleiner was dan het preoperatieve verschil. Deze bevindingen maakt het mogelijk om nadere doelen te stellen om SES ongelijkheden in de toekomst te verminderen.

In **hoofdstuk 4** gebruikten we dezelfde dataset als in **hoofdstuk 3** om te onderzoeken hoe SES-ongelijkheden zich manifesteerden in verschillende PROMs en welke domeinen het meest werden beïnvloed bij THP en TKP patiënten. Naast de eerder genoemde PROMs hebben we ook de EQ Visual Analogue Scale (VAS) voor algemene ervaren gezondheid en Numerieke Beoordelingsschalen (NRS) voor gewrichtspijn gebruikt. Bovendien werd een NRS voor tevredenheid met de uitgevoerde ingreep verkregen. Die was alleen beschikbaar voor TKP patiënten bij de 12-maand follow-up. Onze bevindingen toonden aan dat SES-ongelijkheden verschilden tussen PROM-instrumenten. Patiënten met een lagere SES vertoonden vergelijkbare negatieve verschillen op de EQ-5D en OHS/OKS, hoewel het verschil op OKS bij TKP patiënten preoperatief groter was dan bij EQ-5D. Patiënten met een lagere SES vertoonden kleinere verschillen op de EQ VAS in vergelijking met de EQ-5D en OHS/OKS. NRS voor pijn toonde verschillen die vergelijkbaar waren met EQ-5D en OKS, terwijl NRS voor tevredenheid geen ongelijkheid liet zien na 12 maanden follow-up bij TKP patiënten. Domeinanalyse van de EQ-5D toonde dat dagelijkse activiteiten tot 46% van de verschillen verklaarden bij zowel THP als TKP patiënten, terwijl angst/depressie een beperkte rol speelde (tot 17%). Bij OHS/OKS werd het functioneren meer beïnvloed bij THP (tot 61%) en pijn bij TKP (tot 68%), vooral na 12 maanden. De bevindingen illustreren dat patiënten met een lagere SES slechter functioneren van het gewricht, meer pijn en een lagere kwaliteit van leven ervaren. Dit vertaalde zich echter niet in een slechtere beoordeling van hun algemene gezondheid of ontevredenheid over de ontvangen THP/TKP. Toekomstig onderzoek moet onderzoeken of PROMs zoals EQ VAS en tevredenheidsschalen nauwkeurig de werkelijke ongelijkheid-gerelateerde effecten weergeven of onevenredig worden beïnvloed door meetfouten, zoals verschillen in interpretatie van de vraagstelling.

In **hoofdstuk 5** hebben we onze kijk op ongelijkheden verbreed door de rol van sociaal-economische, patiënt- en ziekenhuiskenmerken te bestuderen in het perioperatieve gebruik van perifere zenuwblokkades (Engelse afkorting: PNBs) in THP en TKP patiënten in de Verenigde Staten. We ontdekten dat, in tegenstelling tot onze hypothesen, sociaal-economische variabelen waaronder SES en ras/etniciteit weinig tot geen effect hadden. Praktijkvariatie was eerder het gevolg van klinische (bijv. minder gebruik voor semi- en niet-electieve indicaties) en ziekenhuiskenmerken. Deze bevindingen benadrukten individuele en praktijkvariaties in het gebruik van PNB. De impact van PNBs op complicatie aantallen, heropnames en duur van opname werd geanalyseerd, rekening houdend met mogelijke confounders. In TKP bleek PNB gebruik te associëren met verminderd aantal complicaties en kortere duur van opname; deze associatie werd niet gezien in THP patiënten. De bevindingen benadrukken de noodzaak voor meer gestandaardiseerde toepassing van PNBs. Er is meer gedetailleerde informatie nodig om conclusies te kunnen trekken over onderliggende oorzaken (bijv. financiële en tijdsprikkels of de rol van training van de specialist).

In **hoofdstuk 6** bespreken we een nationaal (Nederlandse) onderzoek met LROI gegevens naar de impact van de COVID-19 pandemie en lockdown op PROMs in THP en TKP patiënten, voor elke indicatie. Opnieuw gebruikten we de EQ-5D en OHS/OKS. We ontdekten dat THP patiënten tijdens de COVID-19 lockdown een iets slechtere preoperatieve kwaliteit van leven en gewricht functioneren hadden wat zou kunnen passen bij selectie van patiënten die een grotere urgentie voor de ingreep hadden. Dit effect was niet aanwezig bij TKP patiënten waarvoor we geen verklaring hebben. Postoperatieve uitkomsten verschilden nauwelijks. Dit geeft aan dat het hersteltraject na de operatie waarschijnlijk niet of nauwelijks beïnvloed werd door de COVID-19 pandemie of de lockdown.

In **hoofdstuk 7** hebben we een methodologisch probleem met betrekking tot de longitudinale follow-up van kwaliteit van leven van adolescentie naar volwassenheid bestudeerd. We bestudeerden dit probleem in adolescenten met Adolescente Idiopathische Scoliose (AIS) die hiervoor een brace droegen. We verzamelden twee varianten van een generiek kwaliteit van leven instrument, één gericht op kinderen (EQ-5D-Y-5L) en de ander op volwassen (EQ-5D-5L). Vervolgens vergeleken we de varianten onderling en hebben we specifiek getest of ze equivalent waren, en of de één het beter deed dan de ander op instrument karakteristieken. Hoewel ze vergelijkbaar waren wat betreft validiteit en test-hertest betrouwbaarheid, waren er verschillen in de verdeling van de scores aanwezig. Alles bij elkaar genomen, kwamen we tot het advies de EQ-5D-5L te gebruiken om de kwaliteit van leven te monitoren in adolescente patiënten met AIS. Dit voorkomt de noodzaak om te wisselen van instrument, en daarmee discontinuïteit in data wanneer patiënten overgaan van adolescentie naar volwassenheid (>17 jaar oud).

In **hoofdstuk 8** volgt een algemene bespreking van de bovengenoemde hoofdstukken, specifiek gericht op drie onderwerpen: kwaliteit van zorg, sociaal-economische ongelijkheden en methodologische uitdagingen in deze contexten. Kort weergegeven:

Kwaliteit van zorg: De geïdentificeerde tools voor gedeelde besluitvorming en apps voor postoperatieve monitoring (micro-niveau) lijken veelbelovend, maar vereisen verdere evaluatie. Het is vooral onduidelijk of verbeterde besluitvorming ook patiëntuitkomsten zoals kwaliteit van leven zal verbeteren. Op meso-/macro-niveau is het bewijs schaars en zijn voorbeelden nodig van hoe PROMs effectief kunnen worden ingezet.

Sociaal-economische ongelijkheden: PROMs gaven inzicht in onderliggende mechanismen van ongelijkheden. We missen echter kennis over de tussenliggende determinanten. Een voorbeeld is dat patiënten met lage SES mogelijk verminderde gezondheidsvaardigheden hebben. Ook kunnen zorgverlener-gerelateerde factoren een rol spelen, zoals (onbewuste) selectie van optimale kandidaten voor THP/TKP. Een langetermijnstrategie voor het verminderen van ongelijkheden in de orthopedie wordt in dit hoofdstuk besproken.

Methodologische uitdagingen: één van de terugkerende uitdagingen in onze studies was het bepalen van de klinische relevantie van verschillen in PROMs en PROM-veranderingen.

Slotopmerkingen

Dit proefschrift heeft PROMs gebruikt om kansen voor verbeterde kwaliteit van orthopedische zorg te identificeren, alsmede onze kennis van het gebruik van PROMs in deze context te vergroten. Nieuwe PROM toepassingen werden belicht, zoals tools voor gedeelde besluitvorming en het monitoren van het herstel na een operatie door middel van eHealth-apps. Daarnaast hebben we besproken hoe PROMs gebruikt kunnen worden om sociaal-economische ongelijkheden in orthopedische uitkomsten uit te diepen. De bevindingen onderstrepen de noodzaak voor onderzoek naar onderliggende determinanten van deze ongelijkheden, en noodzaak voor gerichte interventies om deze verschillen te verkleinen. Er blijven methodologische uitdagingen bestaan, waaronder de interpretatie van klinische relevantie en mogelijke bias door meetfouten. Er zijn meer voorbeelden en analyses nodig over het effectief gebruik van PROM gegevens en methoden om bovengenoemde methodologische uitdagingen aan te pakken. Dit draagt er aan bij dat PROMs in de toekomst een nog grotere bijdrage kunnen hebben aan patiëntgerichte zorg.

LIST OF PUBLICATIONS

In this thesis:

Bonsel JM, Itiola AJ, Huberts AS, Bonsel GJ, Penton H. The use of patient-reported outcome measures to improve patient-related outcomes - a systematic review. Health Qual Life Outcomes. 2024;22(1):101.

Bonsel JM, Reijman M, Verhaar JAN, van Steenbergen LN, Janssen MF, Bonsel GJ. Socioeconomic inequalities in patient-reported outcome measures of Dutch primary hip and knee arthroplasty patients for osteoarthritis. Osteoarthritis Cartilage. 2024;32(2):200-9.

Bonsel JM, Reijman M, Macri EM, van Steenbergen LN, Verhaar JAN, Bonsel GJ. Socioeconomic Inequalities in Patient-Reported Outcome Measures among Total Hip and Knee Arthroplasty Patients: A Comprehensive Analysis of Instruments and Domains. Submitted.

Bonsel JM, Kodali H, Poeran J, et al. Socioeconomic, Patient, and Hospital Determinants for the Utilization of Peripheral Nerve Blocks in Total Joint Arthroplasty. Anesth Analg. 2025; 140: 675-86.

Bonsel JM, Groot L, Cohen A, Verhaar JAN, Gademan MGJ, Spekenbrink-Spooren A, et al. Impact of the COVID-19 lockdown on patient-reported outcome measures in Dutch hip and knee arthroplasty patients. Acta Orthop. 2022;93:808-18.

Bonsel JM, Peeters CMM, Reijman M, Dings T, Rutges J, Kempen DHR, et al. A head-to-head comparison of the adult EQ-5D-5L and youth EQ-5D-Y-5L in adolescents with idiopathic scoliosis. J Patient Rep Outcomes. 2025;9(1):13.

Not in this thesis:

Pollet V, **Bonsel J**, Ganzeboom B, Sakkers R, Waarsing E. Morphological variants to predict outcome of avascular necrosis in developmental dysplasia of the hip. Bone Joint J. 2021;103-B(5):999-1004.

Bonsel JM, Gielis WP, Pollet V, Weinans HH, Sakkers RJB. Statistical Shape Modeling of US Images to Predict Hip Dysplasia Development in Infants. Radiology. 2022;303(2):425-32.

Peeters CMM, **Bonsel JM**, Munnik-Hagewoud R, Mostert AK, Van Solinge GB, Rutges J, et al. Validity and reliability of the adapted Dutch version of the Brace Questionnaire (BrQ). Acta Orthop. 2023;94:460-5.

Visser TS, Brul S, Deng J, **Bonsel J**, van Es E, Eygendaal D, de Vos RJ. Low socioeconomic status is associated with worse treatment outcomes in patients with Achilles tendinopathy. Br J Sports Med. 2024;58(11):579-85.

van Bussel E, van Marle L, **Bonsel JM**, de Vrij D, Weinans H, Sakkers RJB. Statistical shape modeling on ultrasound imaging: a promising tool for prognosis in Graf type D, III and IV developmental dysplasia of the hip in infants. Accepted at The Ultrasound Journal.

Eijkens J, **Bonsel JM**, Feng YS, Janssen MF, Lubetkin E, Haagsma J. Health inequalities in Health-Related Quality of Life during the COVID-19 pandemic in 6 countries. Submitted.

ACKNOWLEDGEMENTS ('DANKWOORD')

Dit proefschrift heeft mij de afgelopen jaren beziggehouden en ik ben ontzettend trots op het resultaat. Het realiseren ervan was niet mogelijk zonder de hulp van verschillende personen, die ieder op hun eigen manier een belangrijke bijdrage hebben geleverd. Niet alleen aan het proefschrift zelf, maar ook op aangrenzende vlakken. Een aantal personen wil ik in het bijzonder bedanken.

Beste **dr. Reijman**, beste Max, dankjewel voor je fijne en betrokken begeleiding in dit promotietraject. Ik heb ontzettend veel van jou geleerd om als onafhankelijk onderzoeker aan de slag te kunnen gaan. Elke voortgangsmeeting opende met even bijpraten over zaken als Italiaanse koffie en wijn, gevolgd door duidelijke adviezen waarmee ik verder kon. Vooral erg prettig vond ik het feit dat je deur altijd openstond voor vragen of een potje tafelvoetbal. Ik kijk uit naar het vervolg dat ons werk heeft!

Beste **prof.dr. Verhaar**, beste Jan, ik wil je bedanken voor je waardevolle bijdragen en begeleiding. Elk manuscript werd nauwkeurig ontleed, met vrijwel altijd weer een duidelijkere versie als gevolg; hiervan heb ik ontzettend veel geleerd. Je passie voor de orthopedische chirurgie vind ik bewonderingswaardig en heeft mijn passie ervoor vergroot.

Beste **prof.dr. Bonsel**, beste Gouke, je was een bijzonder betrokken begeleider, ik noem je dan ook mijn schaduw-promotor. Je hebt mij geïntroduceerd bij de EuroQol groep, een team gepassioneerde onderzoekers dat zich ontfermt over de EQ-5D. We hebben samen met Max en Jan dit promotie-traject op touw gezet. Je durft op elk front mee te denken, zowel op inhoudelijk als carrière-technisch gebied, en neemt hier de tijd voor; dit in het bijzonder heb ik gewaardeerd. Het is inspirerend om te zien hoe je op de hoogte bent van wat er speelt en welke aanspreekpunten voor bepaalde zaken belangrijk zijn, niet alleen in ons vakgebied, maar ook daarbuiten. Alhoewel het soms moeilijk was, hebben we - denk ik - succesvol 'familie' en 'werk' gecombineerd. Dankjewel voor alle inzet!

Beste **prof.dr. Eygendaal**, beste Denise, ik wil je bedanken voor je aandacht en tijd als lid van de beoordelingscommissie en voor de kansen die je mij hebt aangereikt. Zodoende kon ik gedurende mijn promotietraject klinisch aan de slag als 'Phanios'. Hierdoor heb ik mijn passie voor de orthopedie verder kunnen ontwikkelen, daarnaast heb ik simpelweg een hele leuke tijd gehad. Ik kijk ernaar uit om nog veel van je te leren, zowel op klinisch als onderzoeksgebied.

Daarnaast bedank ik graag **prof.dr. Nils Gutacker** en **dr. Juanita Haagsma** voor hun tijd en aandacht als lid van de beoordelingscommissie. Juanita, dankjewel voor je terloopse adviezen en begeleiding, ik zie uit naar onze samenwerking in de toekomst. Dank ook aan

prof.dr. Taco Gosens, **prof.dr. Radboud Dolhain** en **dr. Jashvant Poeran** voor hun rol in de promotiecommissie. Jashvant, ik wil je in het bijzonder bedanken voor je betrokkenheid bij de totstandkoming van de exchange naar New York en mijn warme ontvangst aldaar.

Prof.dr. Elly Stolk, dankjewel voor je adviezen en (mede)introductie binnen de EuroQol groep, vooral bij het op touw zetten van onze projecten. **Dr. Bas Janssen**, dankjewel voor je hulp bij de analyses in het begin van mijn promotietraject, zonder was het een stuk moeilijker geweest!

Dr. Hanish Kodali and **dr. Brocha Stern**, thank you both for your help, guidance, and warm welcome in New York. The exchange would not have been possible without both of you! Hanish, some days we were the only ones in office. I enjoyed our coffee walks through Central Park very much, thanks for that.

Dr. Liza van Steenbergen en **Anneke Spekenbrink-Spooren**, dank jullie wel voor jullie bijdrage aan dit proefschrift, vooral in jullie onmisbare rol namens de Landelijke Registratie Orthopedische Implantaten (LROI). Dank ook aan alle orthopedisch chirurgen die dagelijkse de moeite nemen om de ingrepen te registreren, met als gevolg zeer hoge kwaliteit nationale data.

Mijn **coauteurs** wil ik bedanken voor hun bijdragen aan dit proefschrift. In het bijzonder wil ik **Ademola Itiola** en **Anouk Huberts** benoemen voor hun bijdrage aan de systematic review die toch soms ons doorzettingsvermogen op proef stelde. **Dr. Hannah Penton**, dankjewel voor je geduld en begeleiding bij deze review. **Dr. Carlo Peeters**, dankjewel dat wij mochten aanhaken bij jouw studie, en voor je harde werk om de inclusies rond te krijgen. Ik wil dan ook de afdelingen Orthopedie van het UMCG, OLVG, Isala en EMC bedanken voor hun bijdrage aan de inclusies. Maar bovenal, dank aan de **patiënten** voor hun deelname aan de studie.

AIOS, **inmiddels orthopedisch chirurgen** en **stafleden** van het Erasmus MC, dank jullie wel voor de leerzame en leuke ANIOS-tijd. In het bijzonder wil ik team Arte (**dr. Koen Bos**, **dr. Wout Veltman**, **dr. Jakob van Oldenrijk**, **Coen Otterspeer**) bedanken voor de kans om mij verder te ontwikkelen binnen de orthopedie, in de directe patiëntenzorg én op andere fronten.

Daarnaast bedank ik graag ook de andere opleiders (**dr. Gerald Kraan** en **dr. Olav van der Jagt**) voor het vertrouwen. Ik zie enorm uit naar de aankomende jaren, om de nodige kennis en kunde binnen de orthopedie op te doen.

Dr. Ralph Sakkers, Ralph, wat begon als een master-student project liep toch wat uit de hand, met meerdere leuke projecten buiten dit proefschrift als gevolg; gedurende deze tijd heb ik veel van je geleerd over de kinderorthopedie en het goed uitvoeren van onderzoek. Ik wil ook **dr. Willem-Paul Gielis**, **Erik van Bussel**, en **prof.dr. Harrie Weinans** bedanken voor hun begeleiding in deze projecten.

Dr. Duncan Meuffels, dankjewel voor de kennismaking met de onderzoekswereld in de orthopedie. Samen met Max begeleidde je mij in mijn allereerste onderzoeksproject. Ik vind het inspirerend hoe je met veel enthousiasme en trots het vak orthopedie uitoefent; dit heeft mijn passie in het begin van mijn carrière zeker mede aangestoken.

Dr. Benjamin Craig, it was great to set up the EuroQol PhD Network together. I admire how much energy and effort you have put into the network, allowing PhD students across the world to have opportunities to share their work and experiences. **Jia Jia Lee**, you joined our team at a later stage and have now taken over the leadership of the Network. Thank you for your valuable contributions.

Annet en **Simone**, bedankt voor al jullie hulp met het oplossen van de alledaagse problemen waar je tegen aan loopt gedurende een promotietraject. Jullie zijn voor mij en de andere PhD studenten een ontzettend belangrijke steun.

Eline, **Erin**, **Tjerk**, **Floris**, **Mark**, **Noor**, **Fleur**, **Myrthe**, **Merel**, **Britt**, **Derek**, **Niels**, **Michiel**, **Irene**, **Brechtje**, **Rintje**, **Maud**, **Mariska**, **Delong**, **Jinchi** en **Zhongcan**, ofwel mijn **collega's van Nc-424**. Dank jullie wel voor de ontzettend gezellige tijd, hulp en natuurlijk het dagelijkse potje tafelvoetbal!

Thomas, **Frank**, **Sebas**, **Gijs**, **Age**, **Robin** en **Rowdy**, dank voor jullie vriendschap, al sinds de middelbare school. We hebben hele leuke tijden en vakanties meegemaakt. Wat ik heel fijn vind is dat jullie (bijna) altijd wel te porren zijn voor een biertje, met de soms nodige afleiding van dit promotietraject. **Frank**, in het bijzonder, leuk dat je er vandaag bij bent als paranimf.

Laurens, **Steven**, **Jaap**, **Kristian**, **Andries** en **Flores**, ik heb jullie leren kennen tijdens de studie en bij SKADI Rotterdam. Ik heb altijd genoten van jullie vriendschap en zie uit naar de aankomende bijzondere activiteiten die jullie altijd weten te verzinnen.

Edwin en **Sacha**, ik had me geen betere schoonouders kunnen wensen. Jullie waren een ontzettende steun in soms toch uitdagende tijden om dit proefschrift tot een mooi einde te brengen, waarvoor dank.

Papa en **mama**, jullie onvoorwaardelijke steun krijg ik inmiddels al 30 jaar. De drijfveer om iets moois te maken van het leven en hard te werken om je doelen te halen, heb ik zonder twijfel van jullie geleerd. Dit promotietraject, en alles daarvoor en daarna, was niet mogelijk geweest zonder jullie, en daarvoor ben ik jullie ontzettend dankbaar. **Rachel**, **Duco** en sinds enige tijd ook **Oscar**, dit geldt natuurlijk ook voor jullie! **Grandpa** en **granny, opa** en **oma, Karin, Catharina, Terry** en **Nick**, dank jullie wel voor jullie oprechte interesse gedurende dit traject en jullie steun op meerdere fronten.

Lieve **Ellis**, mijn vriendin en aanstaande vrouw, en onze twee boefjes **Claire** en **Carter**. We leerden elkaar kennen tijdens de studie en sindsdien zijn we door dik en dun samen. We hebben een hoop mooie dingen samen meegemaakt. Je was altijd al, maar ook gedurende dit promotietraject, mijn steun en toeverlaat. Aangezien je je in eenzelfde traject bevond, konden we heerlijk stoom afblazen als het even nodig was. Ik kan niet wachten om nog meer ervaringen met je te delen en ons leven samen verder op te bouwen.

PHD PORTFOLIO

Personal details	
Name PhD student	Joshua Michael Bonsel
Department	Orthopedics and Sports Medicine, Erasmus MC University Medical Center, Rotterdam, The Netherlands
Research School	Erasmus MC Graduate School
PhD period	01-10-2021 – 30-09-2024
Promotor	prof.dr. J.A.N. Verhaar
Supervisors	dr. M. Reijman
Other advisors	Not applicable

Month-Year	General courses	ECTS
02-2022	Biomedical English Writing and Communication	1.5
02-2022	Research Integrity	0.3
08-2023	eBROK (‘Basiscursus Regelgeving Klinisch Onderzoek’)	1.5
03-2023	In NYC: 1) Rigor, Reproducibility, and Ethical Behavior in Biomedical Research, 2) Basic Good Clinical Practice course.	0.5
	Specific courses	
11-2021	Castor online certificates + basic training	0.5
11-2021	Systematic searching in Pubmed/Embase	0.4
01-2022	Personal Leadership & Communication Workshop	1.0
10-2022	Biostatistics I [CK020]	4.5
01-2023	Presentation training	0.2
11-2023	Biostatistics II [CK030]	4.5
08-2024	Causal Mediation Analysis [ESP69]	1.4
	Seminars and workshops	
11-2021	Visualization workshop	0.1
Since 2021	Zuid-West Nederland Overleg Traumachirurgie (ZWOT) trauma meeting, two meetings per year	0.5
	Oral presentations	
10-2018	‘De uitkomst in Avasculaire Kopnecrose voorspellen met de heupvorm’ *Netherlands Orthopedic Association (NOV) annual conference, ’s-Hertogenbosch, The Netherlands*	1
10-2021	‘Impact of the intelligent COVID-19 lockdown on the quality of life in hip and knee arthroplasty patients in the Netherlands’ *NOV annual conference, ’s-Hertogenbosch, The Netherlands*	1

11-2021	'Impact of the intelligent COVID-19 lockdown on the quality of life in hip and knee arthroplasty patients in the Netherlands' *International Society of Arthroplasty Registries (ISAR) annual conference, virtual*	1
10-2022	'Socio-economische ongelijkheden in zelf gerapporteerde uitkomsten van patiënten met een primaire heup of knie prothese voor artrose – een landelijke registratie studie' *NOV annual conference, Leeuwarden, The Netherlands*	1
10-2023	'Socioeconomic inequality in self-reported outcomes of Dutch primary hip and knee arthroplasty patients for osteoarthritis' *Osteoarthritis Research Society International annual conference, Denver, USA*	1
05-2023	'Socioeconomic inequality in self-reported outcomes of Dutch primary hip and knee arthroplasty patients for osteoarthritis' *ISAR annual conference, Montreal, Canada*	1
06-2024	'Socio-economic, patient, and hospital determinants for the utilization of peripheral nerve blocks in total joint arthroplasty – a United States experience' *Combined 61st Nordic Orthopaedic Federation (NOF) Congress and NOV annual conference, Rotterdam, The Netherlands*	1
06-2024	'A comparison of the adult and youth version of the EQ-5D-5L in adolescents with idiopathic scoliosis' *Combined 61st NOF Congress and NOV annual conference, Rotterdam, The Netherlands*	1
03-2023	Discussant for 'Does perspective really matter in proxy-reported outcomes assessment? A head-to-head comparison of two proxy versions for EQ-5D-Y-3L and EQ-5D-Y-5L in paediatric Asthma patients' *EuroQol Academy conference, Milan, Italy*	0.3
	Poster presentations	
03-2023	'Analysing socio-economic health inequalities using self-reported outcomes - an example using Dutch primary hip and knee' *EuroQol Academy conference, Milan, Italy*	0.5
03-2024	'The use of patient-reported outcome measures to improve quality of care – a systematic review' *EuroQol Academy conference, Copenhagen, Denmark*	0.5
04-2024	'A comparison of the adult and youth version of the EQ-5D-5L in adolescents with idiopathic scoliosis' *Sophia Research Day, Rotterdam, The Netherlands*	0.5

	(Inter)national conferences	
	All conferences/meetings described above were attended.	
10-2023	EuroQol Plenary Meeting, Rome, Italy	0.5
11-2024	EuroQol Plenary Meeting, Noordwijk, The Netherlands	0.5
	Department meetings	
2022-2024	Bi-weekly evidence-based espresso (journal club, discussing a scientific paper	1.5
2022-2024	Weekly clinical education on various topics	1.5
	Miscellaneous	
2021-2023	Promeras Treasurer at Promeras, a PhD student representative body at Erasmus MC. Within Promeras, I co-initiated the Career Development Working Group, organizing 1-2 events focussed on career development per year.	2
2022-2024	EuroQol PhD Network Cofounded the EuroQol PhD Network in 2022. As of now we have ~40 PhD-student members working on EuroQol projects, and ~100 additional participants registered for our newsletter. We have approximately 10 educational webinars a year, and various other activities, aimed at teaching peers, facilitating networking among researchers, and provide PhD support in a broad sense if needed.	3
2023	Organization of ski-trip Orthopedics and Sports Medicine at EMC	0.5
2024	Organization of ski-trip ROGO Rotterdam (EMC, ETZ, RHOC)	0.5
	Teaching	
2022	Supervising bachelor students performing a systematic review	0.5
2023	Minor ‘Orthopedie & Sportgeneeskunde’, tutoring writing assignment	0.8
2021 - 2024	Teaching master students doing their mandatory research program for their master (e.g., medicine), specifically statistics and research methodology	0.5
	Supervising Master’s theses and students	
2023	Supervision of Tim Dings Health, Economics, Policy and Law master-thesis (head-to-head comparison of the EQ-5D adult and youth in an adolescent scoliosis population treated with a brace).	1.5
2021-2022	Supervision of master student (medicine), mostly on methodology and analyses (N. Harlianto, Statistical Shape Modelling (SSM) on hip ultrasound, collaboration with dr. R. Sakkers (UMCU))	1.5
2022-2023	Supervision of master student (medicine), mostly on methodology/ analyses (L. Marle, SSM on hip ultrasound, collaboration with dr. R. Sakkers (UMCU))	1
	Total	43.5

European Credit Transfer and Accumulation System, ECTS

LIST OF ABBREVIATIONS

AIS = Adolescent Idiopathic Scoliosis
AMD = Adjusted Mean Difference
APERSU = Alberta PROMs and EQ-5D Research and Support Unit
ASA = American Society for Anesthesiologists
BMI = body mass index
BrQ = Brace Questionnaire
CBS = 'Centraal Bureau voor de Statistiek' (Statistics Netherlands)
CDSH = Commission on Social Determinants of Health
CI = confidence interval
CMS = Centers for Medicare and Medicaid Services
COSMIN = COnsensus-based Standards for the selection of health Measurement Instruments
CPT = Current Procedural Terminology
EQ-5D = a concise, generic measure of self-reported health
GDPR = General Data Protection Regulation
GP = General Practitioner
GROC = Global Rating of Change
HOOS-PS = Hip dysfunction and Osteoarthritis Outcome Score Physical Short-Form
HRQoL = Health-Related Quality of Life
ICC = Intraclass Correlation Coefficient
ICD = International Classification of Diseases
ICMJE = International Committee of Medical Journal Editors
IPPS = Inpatient Prospective Payment System
IQR = Interquartile Ranges
KOOS-PS = Knee Injury and Osteoarthritis Outcome Score Physical Short-Form
LROI = Dutch Arthroplasty Register
LSS = Level Sum Score
MCID = Minimal Clinically Important Difference
MID = Minimal Important Difference
NHS = National Health Service
NOV = Dutch Orthopedic Association
NRS = Numerical Rating Scale
OA = osteoarthritis
OECD = Organization for Economic Cooperation and Development
OHS = Oxford Hip Score
OKS = Oxford Knee Score
OR = odds ratio
PAR = Population Attributable Risk
PDCA = plan-do-check-act (cycle)

PNB = Peripheral Nerve Block
PREM = Patient-Reported Experience Measures
PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROM = Patient-Reported Outcome Measure
QALY = Quality adjusted life year
QR = Quality Registries
RH = Reporting Heterogeneity
SCP = 'Sociaal Cultureel Planbureau' (Netherlands Institute for Social Research)
SD = Standard Deviation
SDI = Social Deprivation Index
SES = Socioeconomic Status (or Socio-Economic Status)
STROBE = Strengthening the Reporting of Observational Studies in Epidemiology
THA = Total Hip Arthroplasty
TKA = Total Knee Arthroplasty
UK = United Kingdom
US = United States
VAS = Visual Analogue Scale
WHO = World Health Organization

Curriculum Vitae

Joshua Michael Bonsel werd geboren op 10 januari 1995 te Rotterdam. Na zijn middelbare schooltijd startte hij in 2013 met de studie Geneeskunde aan de Erasmus Universiteit Rotterdam. Al vroeg in de bachelor ontstond zijn interesse in de orthopedie, die hij verkende door te assisteren bij verschillende onderzoeksprojecten. In 2020 behaalde hij zijn artsendiploma, alsmede de master 'Health Economics, Policy & Law' aan de Erasmus Universiteit Rotterdam. Nadien werkte Joshua onder andere bij de Huisartsenpost Spijkenisse en de COVID-19 afdeling van het Franciscus Gasthuis te Rotterdam.

Aansluitend op zijn interesses initieerde Joshua mede een promotietraject, dat in 2021 werd gefinancierd door een PhD grant van de EuroQol Research Foundation, met ondersteuning van de afdeling Orthopedie & Sportgeneeskunde van het Erasmus MC. Gedurende deze tijd was hij ook medeoprichter van het internationale EuroQol PhD Network en volbracht hij een onderzoeksuitwisseling bij de afdelingen 'Population Health Science & Policy' en 'Orthopedics' van het Mount Sinaï Hospital in New York. Joshua combineerde het promotietraject met klinische werkzaamheden als ANIOS orthopedie.

In januari 2025 startte hij als AIOS orthopedie in het Ikazia Ziekenhuis te Rotterdam, het Erasmus MC, en het Reinier Haga Orthopedisch Centrum te Zoetermeer. In de toekomst hoopt Joshua de opleiding te kunnen combineren met zijn onderzoeksinteresse. In zijn vrije tijd geniet Joshua van het luisteren naar muziek ((soft)rock, elektronisch), films kijken en sporten (krachttraining, fietsen).