

## RESEARCH ARTICLE

# Availability of specific tools to assess patient reported outcomes in hip arthroplasty in Spain. Identifying the best candidates to incorporate in an arthroplasty register. A systematic review and standardized assessment

Jorge Arias-de la Torre<sup>1,2,3\*</sup>, Elisa Puigdomenech<sup>1,4</sup>, Jose M. Valderas<sup>5</sup>, Jonathan P. Evans<sup>5,6</sup>, Vicente Martín<sup>2,3</sup>, Antonio J. Molina<sup>3</sup>, Nuria Rodríguez<sup>1</sup>, Mireia Espallargues<sup>1,7</sup>

**1** Agency for Health Quality and Assessment of Catalonia (AQuAS), Barcelona, Spain, **2** CIBER Epidemiology and Public Health (CIBERESP), Madrid, Spain, **3** Institute of Biomedicine (IBIOMED), University of León, León, Spain, **4** eHealth Lab Research Group, School of Health Sciences, Universitat Oberta de Catalunya, Catalonia, Spain, **5** Health Services and Policy Research Group, University of Exeter Medical School, Exeter, United Kingdom, **6** Royal Devon and Exeter NHS Foundation Trust, Exeter, United Kingdom, **7** Health Services Research on Chronic Patients Network (REDISSEC), Madrid, Spain

\* These authors contributed equally to this work.

\* [jariasdelatorre@gencat.cat](mailto:jariasdelatorre@gencat.cat)



## OPEN ACCESS

**Citation:** Arias-de la Torre J, Puigdomenech E, Valderas JM, Evans JP, Martín V, Molina AJ, et al. (2019) Availability of specific tools to assess patient reported outcomes in hip arthroplasty in Spain. Identifying the best candidates to incorporate in an arthroplasty register. A systematic review and standardized assessment. PLoS ONE 14(4): e0214746. <https://doi.org/10.1371/journal.pone.0214746>

**Editor:** Daniel Pérez-Prieto, Consorci Parc de Salut MAR de Barcelona, SPAIN

**Received:** December 19, 2018

**Accepted:** March 19, 2019

**Published:** April 1, 2019

**Copyright:** © 2019 Arias-de la Torre et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its Supporting Information files.

**Funding:** The present study was funded by CIBERESP as part of the aid granted to Jorge Arias-de la Torre in 2017 and 2018 for short internships.

**Competing interests:** The authors have declared that no competing interests exist.

## Abstract

### Purpose

1) To systematically review the available scientific literature regarding specific instruments developed and/or tested in a Spanish population, to assess these PROMs in hip arthroplasty; 2) to carry out a standardized assessment of their measurement properties; and 3) to identify the best tools for use in Spain in an arthroplasty registry context.

### Methods

A systematic review of PubMed/MEDLINE and EMBASE and CINAHL was done. Furthermore, a standardized assessment of the questionnaires identified using the Evaluating the Measurement of Patient-Reported Outcomes (EMPRO) tool was performed. All developments, validation and studies aiming to assess the measurement properties of PROMs in hip arthroplasty in the Spanish population were included. Data from the questionnaires on metric properties was taken into account to identify the best candidates for inclusion in a register.

### Results

A total of 853 documents were found. After screening title and abstract, 13 full text documents were reviewed and 8 questionnaires adapted and validated to assess some of the

aspects of hip arthroplasty in the Spanish population were identified. After the EMPRO assessment, 4 questionnaires showed suitable properties (WOMAC, OAKHQOL, mini-OAKHQOL and PFH).

## Conclusions

In Spain, there are a few suitable hip-specific questionnaires currently available to assess PROMs in hip arthroplasty surgery. Some of the more widely used questionnaires, like the OHS and HOOS, have not been validated in the Spanish population until now. Identified tools are suitable for use in a clinical context, however their use in an arthroplasty register is more questionable due to the lack of validation studies of the widely used tools in other registers.

## Background

Hip arthroplasties are presently one of the most frequent elective surgeries worldwide. The evidence related to these surgeries points to a significant improvement in the physical function and health-related quality of life of the individuals, as well as the cost-effectiveness and long-term results of these procedures [1,2].

Traditionally, studies on hip arthroplasty surgery have focused on different outcomes, commonly related to patient survival and type of prosthesis used for the arthroplasty [3,4]. These studies usually aimed to assess the mortality rate of patients with a specific type of prosthesis or the revision risk of their prosthesis. Despite this, when assessing the results of a hip arthroplasty, evaluating Patient-Reported Outcome Measures (PROMs) such as quality of life, pain or physical function, among others [3,5–7] has gained importance in recent years, since they can provide different aspects not covered by traditional outcomes like patient-centred care, clinical decision-making or their possible implications in health policy.

There is currently a consensus, both in clinical and in research communities regarding the importance of considering PROMs in hip arthroplasties before and after surgery [2,6,8,9]. In the context of arthroplasty registers, this consensus is made explicit with the continuous incorporation and development of PROMs programs in registers around the world [10,11]. However, this consensus is not such for the most suitable tools to assess these PROMs [12]. As seen from the conclusions of several systematic reviews, until now there have been differences regarding the questionnaires recommended and used, in terms of the quality of different instruments [13–17] and the tools used by arthroplasty registries and clinical trials [10,11].

The instruments used to assess PROMs in hip arthroplasties could be classified in terms of the population they focus on. Specific tools designed to be used solely in populations with hip pathology, e.g. Oxford hip Score (OHS), Hip disability and Osteoarthritis Outcome Score (HOOS), or Harris Hip Score (HHS); and general tools designed for use by the population as a whole, e.g. Short Form-12 or 36 (SF-12 or SF-36), EuroQol-5D (EQ-5D), or World Health Organization Quality of Life Instrument (WHO-QoL) [6,13,14,18–20]. Some of these tools have been deemed suitable measurement properties for use in hip arthroplasty populations worldwide [10,13,14]. In spite of this evidence, and bearing in mind the differences in results among high quality reviews, particularly with specific tools [13–15,17,21], it might be useful to carry out a standardized assessment and systematic comparison when choosing the best tool possible, depending on its context of use. On one hand, a standardized assessment and

comparison could be useful to recommend a concrete tool that assesses PROMs in a clinical context, while on the other hand, it might be useful to select the most appropriate tool to be included in an arthroplasty register [22,23].

Focusing on hip-specific PROMs, their selection is generally context-dependent and guided by different criteria of which could be remarkable, among others, the tradition of use in a specific context or country or the availability of a tool [11]. Despite the possible high quality of the selected tools, these criteria are not generally based on a systematic review and comparison of the properties of the tools. Thus, to perform systematic reviews and standardised comparisons of these tools for specific contexts, it could be highly valuable to the decision making process when selecting a questionnaire to assess PROMs in hip arthroplasty.

Therefore, the aims of this study are: 1) to systematically review the scientific literature available about specific instruments developed and/or tested in a Spanish population and Spanish language, to assess PROMs in hip arthroplasty; 2) to evaluate and carry out a standardized assessment of their measurement properties; and 3) to identify the best candidates to be used in an arthroplasty registry context in Spain.

## Methods

To find relevant studies on available hip-specific tools in the Spanish population, a systematic review was performed in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (S1 Checklist) guidelines [24] and registered with PROSPERO under registry number CRD42018083626 [25].

The search was done up to 01/01/2018 and there was no restriction on the start date. Searches were conducted in the following databases: PubMed/MEDLINE, EMBASE and CINAHL. They were done using a filter developed specifically for PubMed/MEDLINE, which was then adapted for the other databases (S1 Filter). In addition, a manual search was conducted to retrieve any studies that might not have been included in the review. The search strategy and filter development was guided by previous systematic review filters used to select PROMs tools and the Spanish population [13,14,26,27]. Furthermore, references of the retrieved articles were screened for relevant studies, and relevant authors identified in the developmental studies of the different tools were contacted. After the systematic review, the tools assessed in the studies were identified and a standardized assessment of the adequacy of their measurement properties was done using the Evaluating the Measurement of Patient-Reported Outcomes (EMPRO) tool.

## Study inclusion and exclusion criteria and the review process

Population, Intervention, Comparison, Outcome (PICO) criteria were used. All development, adaptation, validation and studies aimed to assess the metric properties of procedure- or condition-specific tools used to assess PROMs in hip arthroplasty in the Spanish population were included. The following exclusion criteria were used: studies published in a language other than English or Spanish, studies focused on general tools, studies not focused on patients undergoing a hip arthroplasty or on the waiting list for one (e.g. studies focused only on osteoarthritis of the hip and not on hip arthroplasty), studies where it was impossible to determine the joint operated on, surgical technique papers and case studies or studies involving fewer than 10 patients. Due to the cross-cultural adaptation and the possible bias related to the comparability of metric properties between populations, only full-texts of tools developed and/or tested in the Spanish population (from Spain) were included [28–30].

After identifying the studies, a screening was done based on the outlined inclusion/exclusion criteria, first by title and abstract and then by full text. All papers that did not meet the

inclusion criteria were excluded from the review. All documents identified were revised independently by 2 expert reviewers (JA and JE). If the reviewers did not reach an agreement regarding one or more papers or attributes, a third reviewer (JMV) revised the documents and assessed the questionnaires to arrive at a consensus. After the revision, a standardized assessment of metric properties of the identified tools was done by 2 reviewers (JA and EP) using the EMPRO tool. Before the review, the comprehension of the research aims was assessed using a 10-study pilot and measured using Cohen's Kappa statistic. None of the reviewers was involved in developing the measures assessed.

### Data extraction and standardized evaluation of proms questionnaires

Data about the instruments identified in the studies included after the screening was retrieved. Information regarding their measurement properties was considered, following the Medical Outcomes Trust classification [31]. In addition, a narrative synthesis of the evidence found in each questionnaire was performed. The quality of the identified questionnaires was assessed and compared based on the evidence found in the studies included.

For the standardized assessment of the questionnaires, the EMPRO tool was used [22]. EMPRO is a tool designed for the standardized evaluation of the quality of instruments used to assess PROMs, based on the Medical Outcomes Trust criteria [31,32]. The questionnaire consists of 39 Likert-type items with a response scale from 1 ("strongly disagree") to 4 ("strongly agree") that are distributed among 8 attributes described in Table 1: Conceptual and measurement model (7 items); Reliability (8 items); Validity (6 items); Responsiveness (3 items); Interpretability (3 items); Administration burden (7 items); Alternative modes of administration (2 items); and Cross-cultural and linguistic adaptations (3 items). Additionally, the questionnaire provides a space for comments and references in each item and some items have the response option "no information" or "not applicable" when the information is insufficient or not suitable. To conclude the questionnaire, an overall recommendation is provided on a scale with the following response categories: "Strongly recommended", "Recommended with provisos or alterations", "Would not recommend" and "Unsure" with a rationale for the recommendation. A score for each attribute was calculated using the mean of the responses to all items composing that attribute and a linear transformation was done to obtain a score from 0 (the worst possible score) to 100 (the best possible score). Finally, an overall score for the questionnaire based on the mean score of the metric-related attributes (conceptual and measurement model, reliability, validity, responsiveness and interpretability) was obtained.

## Results

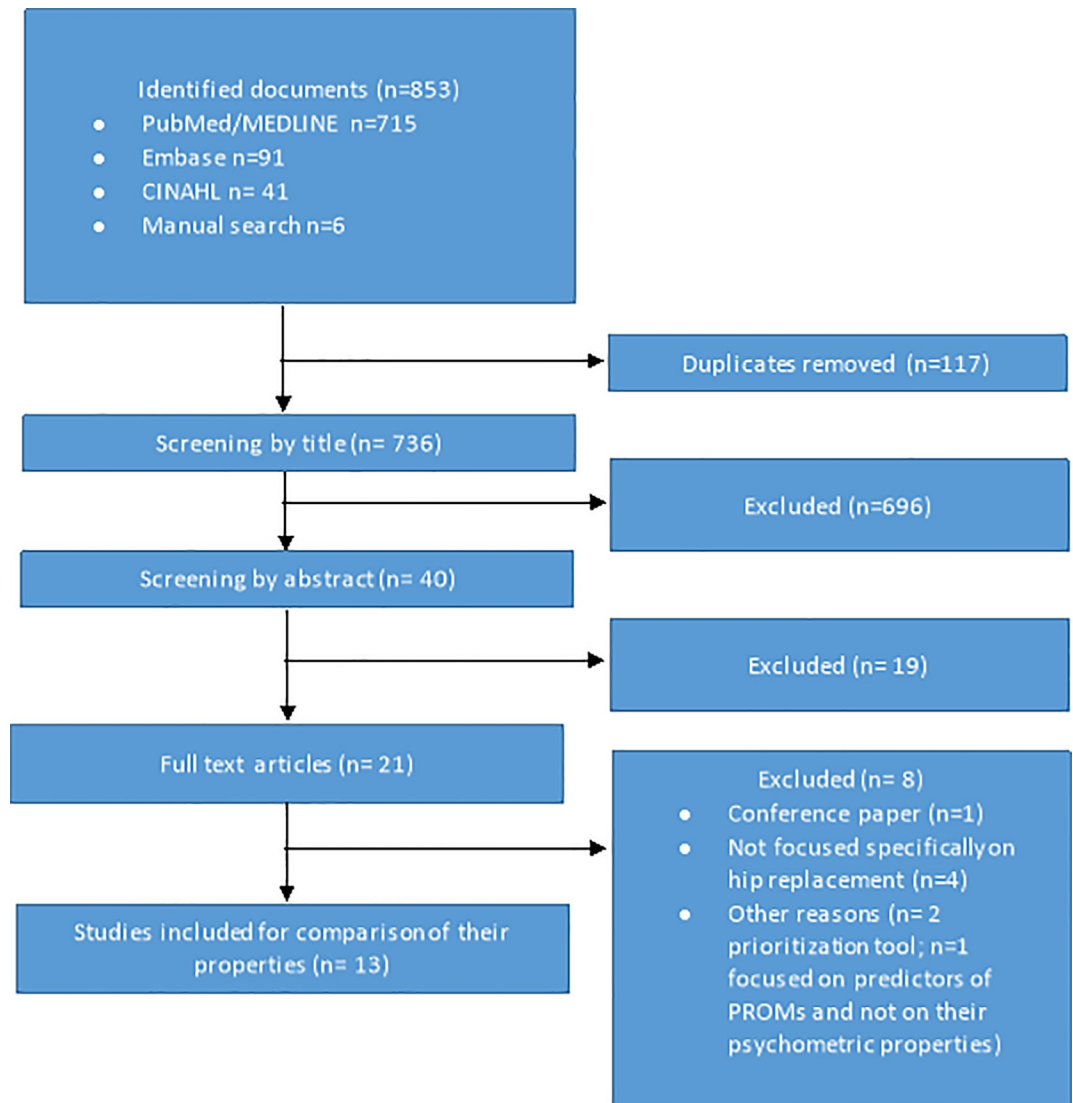
Fig 1 shows a flow diagram of the review process. Using the filter developed for systematic review, a total of 853 documents were identified. Of these, 715 (83.8%) were identified in PubMed/MEDLINE, 91 (10.7%) in EMBASE, 41 (4.8%) in CINAHL and 6 (0.7%) through a manual search. After checking for duplicates, 117 (13.7%) were removed. Of the remaining 736 titles, 696 were excluded, leaving a total of 40 abstracts to screen. After screening the abstracts, 21 full text articles were considered. Of these documents, 8 were excluded: 1 was a conference abstract, 4 were focused on hip osteoarthritis without considering hip replacement or it was impossible to determine the operated joint, 2 aimed to establish prioritisation systems rather than assess PROMs in hip replacement, and 1 was focused on variables that might act as predictors of PROMs and not on their specific measurement properties. Finally, a total of 13 full text articles published between 1997 and 2017 were included in the data extraction [33–45].

**Table 1. Attributes assessed using the Evaluating the Measurement of Patient-Reported Outcomes (EMPRO) tool.**

Attribute	Definition	Items included
<b>Conceptual and measurement model</b>	The rationale for and description of the concept and the populations that a measure is intended to assess and the relationship between these concepts.	1. Concept of measurement stated 2. Obtaining and combining items described 3. Rationality for dimensionality and scales 4. Involvement of target population 5. Scale variability described and adequate 6. Level of measurement described 7. Procedures for deriving scores
<b>Cultural adaptation</b>	Cultural and linguistic adaptation of the instrument.	8. Linguistic equivalence 9. Conceptual equivalence 10. Differences between the original and the adapted versions
<b>Reliability</b>	The degree to which an instrument is free from random error.	Internal consistency: 11. Data collection methods described 12. Cronbach's alpha adequate 13. IRT estimates provided 14. Testing in different populations Reproducibility: 15. Data collection methods described 16. Test-retest and time interval adequate 17. Reproducibility coefficients adequate 18. IRT estimates provided
<b>Validity</b>	The degree to which the instrument measures what it purports to measure.	19. Content validity adequate 20. Construct/criterion validity adequate 21. Sample composition described 22. Prior hypothesis stated 23. Rational for criterion validity 24. Tested in different populations
<b>Responsiveness</b>	An instrument's ability to detect change over time.	25. Adequacy of methods 26. Description of estimated magnitude of change 27. Comparison of stable and unstable groups
<b>Interpretability</b>	The degree to which one can assign easily understood meaning to an instrument's quantitative scores.	28. Rational of external criteria 29. Description of interpretation strategies 30. How data should be reported stated
<b>Burden</b>	The time, effort, and other demands placed on those to whom the instrument is administered (respondent burden) or on those who administer the instrument (administrative burden).	Respondent: 31. Skills and time needed 32. Impact on respondents 33. Not suitable circumstances Administrative: 34. Resources required 35. Time required 36. Training and expertise needed 37. Burden of score calculation
<b>Alternative modes of administration</b>	Alternative modes of administration used for the administration of the instrument	38. The metric characteristics and use of each alternative mode of administration 39. Comparability of alternative modes of administration

<https://doi.org/10.1371/journal.pone.0214746.t001>

Taking into account the instruments included in the studies (Table 2), a total of 8 tools were identified, four of which were variants or sub-scales of other main tools. In terms of specific questionnaires, we found: The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) [33–36,44], composed by 24 Likert-type items designed to assess pain, stiffness and physical function; two different versions of the WOMAC Short Form, each composed of 11 items with 5 Likert-type items [35,37]; The WOMAC Short Form (Function dimension) composed of 7 items validated to assess functionality in joint replacement [38]; The Pain and Function of the Hip scale (PFH) composed of 9 Likert-type items to assess pain,



**Fig 1. Review process.** Flow diagram.

<https://doi.org/10.1371/journal.pone.0214746.g001>

function and mobility/strength [41,42,45]; The Osteoarthritis Knee and Hip Quality of Life (OAKHQOL) composed of 43 Likert-type items, 40 of which are used to assess physical activity, mental health, pain, social support and social functioning [39]; The Mini-OAKHQOL composed of 20 Likert-type items to assess the same dimensions as the long version (OAKHQOL); and the Harris Hip Score (HHS), with 10 Likert-type items that assess pain, function, amplitude of movement and absence of deformity [43].

Considering the properties of the identified tools, Table 3 shows their adaptation according to the EMPRO standardized assessment. Based on the overall score, it was observed that only OAKHQOL, mini-OAKHQOL, WOMAC and PFH could be considered as reasonably acceptable (EMPRO overall score  $\geq 50$ ). The highest overall score corresponded to WOMAC with a score of 65.47, followed by OAKHQOL with a score of 58.87, mini-OAKHQOL with 58.57 points and PFH with 50.71 points. In addition, focusing on the specific metric attributes it was observed that all the assessed tools have EMPRO scores equal to or higher than 50 in reliability,

**Table 2. General characteristics of the instruments identified from validation studies in hip arthroplasty population.**

Questionnaire	Author (year)	Number of items and type	Assessed dimensions (number of items)	Punctuation
Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)	Escobar et al. (2002) Quintana et al. (2005) Lopez Alonso et al. (2009) Escobar et al. (2012) Quintana et al. (2012)	24 Likert-type items with a 5-point response scale	Pain (5 items), stiffness (2 items) and physical function (17 items)	Standardized from 0 (best health status) to 100 (worst health status)
WOMAC Short Form (v1)	Bilbao et al. (2011)	11 Likert-type items with a 5-point response scale	Pain (3 items) and physical function (8 items)	Standardized from 0 (best health status) to 100 (worst health status)
WOMAC Short Form (v2)	Lopez Alonso et al. (2009)	11 Likert-type items with a 5-point response scale	Symptomatology and physical disability in: repose (4 items), movement on stairs (3 items), put on or take off shocks (2 items), and Stiffness (2)	NS
WOMAC (Short Form. Function dimension)	Escobar et al. (2011)	7 Likert-type items with a 5-point response scale	Functional capacity (capacity to do activities)	Standardized from 0 (best health status) to 100 (worst health status)
Osteoarthritis Knee and Hip Quality of Life (OAKHQOL)	Gonzalez Saenz de Tejada et al. (2011)	43 Likert-type items with a 10-point response scale	Physical activity (16 items), mental health (13 items), pain (4 items), social support (4 items), and social functioning (3 items) and three independent items addressing sex life, professional life and fear of being dependent	Standardized from 0 (worst quality of life) to 100 (best quality of life)
Mini-OAKHQOL	Gonzalez Saenz de Tejada et al. (2017)	20 Likert-type items with a 10-point response scale	Physical activity (7 items), mental health (3 items), pain (3 items), social support (2 items), and social functioning (2 items) and three independent items addressing sex life, professional life and fear of being dependent	Standardized from 0 (worst quality of life) to 100 (best quality of life)
Pain and Function of the Hip scale (PFH)	Valls et al. (1997) Alonso et al. (2000) Marti-Valls et al. (2000)	9 Likert-type items with different point response scales depending on the specific dimension.	Pain (2 items), function (3 items) and mobility/ Strength (4 items)	From 0 (total functional limitation) to 85 (absence of functional limitation)
Harris Hip Score (HHS)	Navarro Collado et al. (2005)	10 Likert-type items (different points for each response scale)	Pain (1 item), function (7 items), amplitude of movement (1 item) and absence of deformity (2 items)	From 0 (worst possible functional capacity) to 100 (best possible functional capacity)

v1: version 1; v2: version 2; NS: Not explicitly specified

<https://doi.org/10.1371/journal.pone.0214746.t002>

validity and responsiveness, except PFH in validity. Only WOMAC and PFH had a score equal to or higher than 50 points in conceptual and measurement model and in interpretability. Regarding non-metric attributes (cultural adaptation and burden), only OAKHQOL had a score equal to or higher than 50 points.

## Discussion

Nowadays in Spain, the availability of specific tools to assess PROMs in hip arthroplasty is limited. After the systematic review of the available literature, only eight instruments (including short forms) were identified. Of these instruments only WOMAC, OAKHQOL, mini-OAKHQOL, and PFH, have shown suitable measurement properties to be used in the Spanish population. In addition, none of the most widely used hip-specific questionnaires was identified in arthroplasty registers around the world [10,11,21]. This evidence suggests that there are some

**Table 3. Attributes of each of the questionnaires identified from the studies included in the systematic review. EMPRO.**

Questionnaire	Conceptual and measurement model	Cultural adaptation	Reliability	Validity	Responsiveness	Interpretability	Burden	Alternative modes of administration	Overall score
WOMAC	++	+	++	++	+++	++	+	+	65.47
WOMAC SF (v1)	+	+	++	++	++	+	+	-	48.99
WOMAC SF (v2)	+	+	++	+	+	+	+	-	-
WOMAC SF (FD)	+	+	++	+	+++	+	+	-	-
OAKHQOL	+	+++	++	++	+++	+	++	-	58.87
Mini-OAKHQOL	+	+	+++	++	+++	+	+	-	58.57
PFH	++	+	++	+	++	++	+	-	50.71
HHS	+	+	+	+	++	+	+	-	-

WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index; SF: Short form; v1: version 1; v2: version 2; FD: Function dimension; OAKHQOL: Osteoarthritis Knee and Hip Quality of Life; PFH: Pain and Function of the Hip Scale; HHS: Harris Hip Score

+: EMPRO score 25–49; ++: EMPRO score 50–74; +++: EMPRO score 75–100; -: EMPRO score not applicable or not calculable

<https://doi.org/10.1371/journal.pone.0214746.t003>

suitable tools to be included in an arthroplasty register currently in Spain but, before deciding to do so, it might be valuable to carry out validation studies of widely used hip-specific tools in other registers worldwide as OHS and HOOS.

Taking into account the specific tools assessed in our study, only 4 were identified as suitable candidates and 1 of them, mini-OAKHQOL, was a shorter version of one of the main tools. Of these tools, the most acceptable, according to EMPRO guidelines, was WOMAC, which was the only tool used by 4 other arthroplasty registers worldwide [11]. In addition, the similarity between the EMPRO overall scores from OAKHQOL and the shorter version should be highlighted. Given the similarity in metric properties, we could argue that it would be better to use the mini-OAKHQOL tool over the full version since it has a lower burden, but this hypothesis requires more evidence to support it.

Several reviews assessing the acceptability of instruments to evaluate PROMs in hip arthroplasty were previously done in other countries [13–15,17]. From these reviews, specific instruments used with better properties were identified, which in most cases were the OHS and the HOOS. In addition, the most common instruments used in other registries were also the OHS and HOOS [10,21].

Regarding OHS, this questionnaire is one of the most widely used for different reasons, among which could be highlighted its length, including only 12 items, and its acceptable metric properties to be used in a wide range of languages and countries. Despite this, for the Spanish population, only one validation study was found [46]. This study was condition-focused and not procedure-focused, which was why we decided to exclude it. Patient characteristics of the population included in this study might be different from arthroplasty patients. For this reason, and bearing in mind one of the objectives was to select the best candidate tool to be included in an arthroplasty register, we decided to discard this study, again showing the need for procedure-focused validation studies of this questionnaire. The HOOS questionnaire is also one of the most widely used questionnaires in hip arthroplasties, used by the Swedish Arthroplasty Register among others [47], and has shown its acceptability in metric terms in other contexts [13,14,18,20,48]. With these reasons in mind, we propose that the adaptation and validation of OHS and HOOS in the hip arthroplasty population could be valuable in our



context, both for clinical use and to know what could be the most acceptable tool for inclusion in an arthroplasty register.

Finally, and despite the suitability of the metric properties of all tools in which was possible to calculate an EMPRO overall score, except WOMAC Short Form whose score was on the border of acceptability, we should highlight that they might only be considered as partially suitable candidates for inclusion in an arthroplasty register. As these tools are not among those most used by other registries, it is difficult to pool data and compare them. Additionally, when the properties of PROMs tools were compared in other contexts and countries [13,14,17], these tools were not present in some cases, OAKHQOL mini-OAKQOL and PFH, for example, or they were not the best candidates for adequacy or specificity. On the other hand, focusing on WOMAC, it should also be noted that the full questionnaire is included in HOOS, which is one of the more widely used questionnaires to assess PROMS in hip arthroplasties [14,20,48,49].

Some limitations of this review should be discussed. Firstly, we should mention that we only examined tools developed and/or tested in the population from Spain. This inclusion criteria, seems to limit the scope of the study. In spite of this, we think that with this approach, our results are much more accurate. Thus, we deem that they could be much more useful for the clinical community of Spain and, furthermore, they may serve as baseline for future comparative studies between countries. Besides, we have to remark the possible publication bias that could affect the results shown. Studies about PROMs could be unpublished or published as grey literature. Despite this, we have tried to be as exhaustive as possible in the literature search and have also tried to contact the most influential authors. As such, we feel that we have identified at least the main validation literature focused on hip arthroplasty specific tools. Another limitation is that we have excluded studies focused on patients with OA if the patients included in the studies were not on a waiting list or undergoing hip arthroplasty. Though OA is by far the most common cause of total hip replacement, it should be noted that not all patients affected by OA are eligible candidates for an arthroplasty [50–53]. OA is a chronic condition that could affect people with very different characteristics, including young patients and patients with non-severe OA. Young patients able to perform their daily activities well should not be eligible for an arthroplasty, and especially not for a total hip replacement. For these reasons, we consider the outlined inclusion and exclusion criteria appropriate to address the aims of this study. In addition, we want to point out the limitation related to including only hip specific tools and not general tools. General instruments are widely studied and should be used as the first approach to PROMs assessment. Nevertheless, and despite their use for assessment of specific populations, these instruments were not specifically designed to assess PROMs, which means their responsiveness could be compromised. Furthermore, in some cases the constructs evaluated with general tools are not the same or not entirely equivalent than those evaluated by the specific tools. After taking the conceptual differences between general and specific tools into account, particularly when used in specific populations, we decided to exclude general tools. Finally, we should highlight the lack of validation studies developed from the Item Response Theory (IRT) perspective. Including this approach when evaluating the metric properties of questionnaires could improve the precision of their assessment and decision making when selecting a specific tool.

In conclusion, our study shows that currently in Spain, there are some specific questionnaires with adequate metric properties to assess PROMs in hip arthroplasty. These tools are: WOMAC, OAKHQOL, mini-OAKHQOL and PFH. While these tools might be considered suitable for use in a clinical context, their recommendation for use in an arthroplasty register is more questionable, mainly due to the lack of validation studies of the OHS and HOOS, the more widely used tools in this context.

## Supporting information

### S1 Checklist. PRISMA 2009 checklist.

(DOC)

### S1 Filter. PubMed filter. PubMed/MEDLINE filter. Psychometric properties of specific PROMs questionnaires in the Spanish population.

(DOCX)

## Acknowledgments

We acknowledge Kayla Smith for the technical support she gave to this revision.

## Author Contributions

**Conceptualization:** Jorge Arias-de la Torre, Elisa Puigdomenech, Jose M. Valderas, Jonathan P. Evans, Vicente Martín, Antonio J. Molina, Nuria Rodríguez, Mireia Espallargues.

**Data curation:** Jorge Arias-de la Torre, Jose M. Valderas, Jonathan P. Evans, Vicente Martín, Mireia Espallargues.

**Formal analysis:** Jorge Arias-de la Torre, Elisa Puigdomenech, Jose M. Valderas, Jonathan P. Evans, Antonio J. Molina, Mireia Espallargues.

**Funding acquisition:** Jorge Arias-de la Torre, Jose M. Valderas, Vicente Martín, Antonio J. Molina, Mireia Espallargues.

**Investigation:** Jorge Arias-de la Torre, Elisa Puigdomenech, Jonathan P. Evans, Vicente Martín, Nuria Rodríguez, Mireia Espallargues.

**Methodology:** Jorge Arias-de la Torre, Elisa Puigdomenech, Jose M. Valderas, Jonathan P. Evans, Vicente Martín, Antonio J. Molina, Mireia Espallargues.

**Project administration:** Jorge Arias-de la Torre, Elisa Puigdomenech, Jose M. Valderas, Jonathan P. Evans, Mireia Espallargues.

**Resources:** Jorge Arias-de la Torre, Mireia Espallargues.

**Software:** Jorge Arias-de la Torre, Jose M. Valderas, Mireia Espallargues.

**Supervision:** Jorge Arias-de la Torre, Elisa Puigdomenech, Jose M. Valderas, Jonathan P. Evans, Vicente Martín, Antonio J. Molina, Nuria Rodríguez, Mireia Espallargues.

**Validation:** Jorge Arias-de la Torre, Elisa Puigdomenech, Jose M. Valderas, Jonathan P. Evans, Vicente Martín, Antonio J. Molina, Nuria Rodríguez, Mireia Espallargues.

**Visualization:** Jorge Arias-de la Torre, Jose M. Valderas, Mireia Espallargues.

**Writing – original draft:** Jorge Arias-de la Torre, Mireia Espallargues.

**Writing – review & editing:** Jorge Arias-de la Torre, Mireia Espallargues.

## References

1. Räsänen P, Paavolainen P, Sintonen H, Koivisto A-M, Blom M, Ryyänen O-P, et al. Effectiveness of hip or knee replacement surgery in terms of quality-adjusted life years and costs. *Acta Orthop.* 2007; 78: 108–115. <https://doi.org/10.1080/17453670610013501> PMID: 17453401
2. Jones CA, Beaupre LA, Johnston DWC, Suarez-Almazor ME. Total joint arthroplasties: current concepts of patient outcomes after surgery. *Rheum Dis Clin North Am.* 2007; 33: 71–86. <https://doi.org/10.1016/j.rdc.2006.12.008> PMID: 17367693

3. Halawi MJ. Outcome Measures in Total Joint Arthroplasty: Current Status, Challenges, and Future Directions. *Orthopedics*. 2015; 38: e685–9. <https://doi.org/10.3928/01477447-20150804-55> PMID: 26270754
4. Serra-Sutton V, Allepuz A, Espallargues M, Labek G, Pons JM V. Arthroplasty registers: A review of international experiences. *Int J Technol Assess Health Care*. 2009; 25: 63–72. <https://doi.org/10.1017/S0266462309090096> PMID: 19126253
5. Rothman M, Burke L, Erickson P, Leidy NK, Patrick DL, Petrie CD. Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use of Existing Instruments and Their Modification PRO Task Force Report. *Value Health*. 12: 1075–83. <https://doi.org/10.1111/j.1524-4733.2009.00603.x> PMID: 19804437
6. Black N. Patient reported outcome measures could help transform healthcare. *BMJ*. 2013;346. <https://doi.org/10.1136/bmj.f167> PMID: 23358487
7. Valderas JM, Alonso J. Patient reported outcome measures: A model-based classification system for research and clinical practice. *Qual Life Res*. 2008; <https://doi.org/10.1007/s1136-008-9396-4> PMID: 18836850
8. Ethgen O, Bruyère O, Richy F, Dardennes C, Reginster J-Y. Health-related quality of life in total hip and total knee arthroplasty. A qualitative and systematic review of the literature. *J Bone Joint Surg Am*. 2004; 86–A: 963–74. Available: <http://www.ncbi.nlm.nih.gov/pubmed/15118039> PMID: 15118039
9. Gagnier JJ. Patient reported outcomes in orthopaedics. *J Orthop Res*. 2017; 35: 2098–2108. <https://doi.org/10.1002/jor.23604> PMID: 28513993
10. Rolfson O, Eresian Chenok K, Bohm E, Lübbecke A, Denissen G, Dunn J, et al. Patient-reported outcome measures in arthroplasty registries. *Acta Orthop*. 2016; 87: 3–8. <https://doi.org/10.1080/17453674.2016.1181815> PMID: 27168175
11. Rolfson O, Bohm E, Franklin P, Lyman S, Denissen G, Dawson J, et al. Patient-reported outcome measures in arthroplasty registries. *Acta Orthop*. 2016; 87: 9–23. <https://doi.org/10.1080/17453674.2016.1181816> PMID: 27228230
12. Lovelock TM, Broughton NS, Williams CM. The Popularity of Outcome Measures for Hip and Knee Arthroplasties. *J Arthroplasty*. 2017; <https://doi.org/10.1016/j.arth.2017.08.024> PMID: 28939034
13. Harris K, Dawson J, Gibbons E, Lim C, Beard D, Fitzpatrick R, et al. Systematic review of measurement properties of patient-reported outcome measures used in patients undergoing hip and knee arthroplasty. *Patient Relat Outcome Meas*. 2016; Volume 7: 101–108. <https://doi.org/10.2147/PROM.S97774> PMID: 27524925
14. Thorborg K, Roos E, Bartels E, Petersen J, Holmich P. Validity, reliability and responsiveness of patient-reported outcome questionnaires when assessing hip and groin disability: a systematic review. *Br J Sports Med*. 2010; 44: 1186–1196. <https://doi.org/10.1136/bjism.2009.060889> PMID: 19666629
15. Thorborg K, Tjissen M, Habets B, Bartels EM, Roos EM, Kemp J, et al. Patient-Reported Outcome (PRO) questionnaires for young to middle-aged adults with hip and groin disability: a systematic review of the clinimetric evidence. *Br J Sports Med*. 2015; 49: 812–812. <https://doi.org/10.1136/bjsports-2014-094224> PMID: 25586913
16. Veenhof C, Bijlsma JWJ, van den Ende CHM, Dijk GM van, Pisters MF, Dekker J. Psychometric evaluation of osteoarthritis questionnaires: A systematic review of the literature. *Arthritis Rheum*. 2006; 55: 480–492. <https://doi.org/10.1002/art.22001> PMID: 16739188
17. Alviar M, Olver J, Brand C, Tropea J, Hale T, Pirpiris M, et al. Do patient-reported outcome measures in hip and knee arthroplasty rehabilitation have robust measurement attributes? A systematic review. *J Rehabil Med*. 2011; 43: 572–583. <https://doi.org/10.2340/16501977-0828> PMID: 21607295
18. Nilsson 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; <https://doi.org/10.1186/1471-2474-4-10> PMID: 12777182
19. Dawson J, Fitzpatrick R, Murray D, Carr A. Comparison of measures to assess outcomes in total hip replacement surgery. *Qual Saf Heal Care*. 1996; <https://doi.org/10.1136/qshc.5.2.81>
20. Klässbo M, Larsson E, Mannevik E. Hip disability and osteoarthritis outcome score: An extension of the Western Ontario and McMaster Universities Osteoarthritis Index. *Scand J Rheumatol*. 2003; <https://doi.org/10.1080/03009740310000409>
21. Paulsen A. Patient reported outcomes in hip arthroplasty registries. *Dan Med J*. 2014; 61: B4845. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24814747> PMID: 24814747
22. Valderas JM, Ferrer M, Mendivil J, Garin O, Rajmil L, Herdman M, et al. Development of EMPRO: A Tool for the Standardized Assessment of Patient-Reported Outcome Measures. *Value Heal*. 2008; 11: 700–708. <https://doi.org/10.1111/j.1524-4733.2007.00309.x> PMID: 18194398

23. Mokkink, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res.* 2010; 19: 539–549. <https://doi.org/10.1007/s11136-010-9606-8> PMID: 20169472
24. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of clinical epidemiology.* 2009. pp. e1–34. <https://doi.org/10.1016/j.jclinepi.2009.06.006> PMID: 19631507
25. Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, et al. The nuts and bolts of PROSPERO: An international prospective register of systematic reviews. *Syst Rev.* 2012; 1. <https://doi.org/10.1186/2046-4053-1-2> PMID: 22587842
26. Valderas JM, Mendivil J, Parada A, Losada-Yáñez M, Alonso J. Construcción de un filtro geográfico para la identificación en PubMed de estudios realizados en España. *Rev Española Cardiol.* 2006; 59: 1244–1251. <https://doi.org/10.1157/13096592>
27. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res.* 2009; 18: 1115–1123. <https://doi.org/10.1007/s11136-009-9528-5> PMID: 19711195
28. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Heal.* 2005; 8: 94–104. <https://doi.org/10.1111/j.1524-4733.2005.04054.x> PMID: 15804318
29. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. *Spine (Phila Pa 1976).* 2000; <https://doi.org/10.1097/00007632-200012150-00014>
30. Herdman M, Fox-Rushby J, Badia X. “Equivalence” and the translation and adaptation of health-related quality of life questionnaires. *Qual Life Res.* 1997; <https://doi.org/10.1023/A:1026410721664>
31. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res.* 2002; 11: 193–205. Available: <http://www.ncbi.nlm.nih.gov/pubmed/12074258> PMID: 12074258
32. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res.* 2013; <https://doi.org/10.1007/s11136-012-0344-y> PMID: 23288613
33. Escobar A, Quintana JM, Bilbao A, Azkárate J, Güenaga JI, A. E, et al. Validation of the Spanish version of the WOMAC questionnaire for patients with hip or knee osteoarthritis. *Clin Rheumatol.* A. Escobar, Unidad de Investigación, Hospital de Basurto, Avenida de Montevideo 18, 48013 Bilbao, Spain. E-mail: [aescobar@hbas.osakidetza.net](mailto:aescobar@hbas.osakidetza.net); Springer London (The Guildway, Old Portsmouth Road, Artington, Guildford GU3 1LP, United Kingdom); 2002; 21: 466–471. <https://doi.org/10.1007/s100670200117> PMID: 12447629
34. Quintana JM, Escobar A, Bilbao A, Arostegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after hip joint replacement. *Osteoarthr Cartil.* 2005; 13: 1076–1083. <https://doi.org/10.1016/j.joca.2005.06.012> PMID: 16154777
35. Lopez Alonso SR, Martinez Sanchez CM, Romero Canadillas AB, Navarro Casado F, Gonzalez Rojo J. [Metric properties of WOMAC questionnaires-original and reduced versions-to measure symptoms and Physical Functional Disability]. *Aten primaria.* Spain; 2009; 41: 613–620. <https://doi.org/10.1016/j.aprim.2009.02.005> PMID: 19464081
36. Quintana JM, Aguirre U, Barrio I, Orive M, Garcia S, Escobar A. Outcomes after total hip replacement based on patients’ baseline status: What results can be expected? *Arthritis Care Res.* 2012; 64: 563–572. <https://doi.org/10.1002/acr.21570> PMID: 22183807
37. Bilbao A, Quintana JM, Escobar A, Las Hayas C, Orive M, A. B, et al. Validation of a proposed WOMAC short form for patients with hip osteoarthritis. *Health Qual Life Outcomes.* England: BioMed Central Ltd. (Floor 6, 236 Gray’s Inn Road, London WC1X 8HB, United Kingdom); 2011; 9: 75. <https://doi.org/10.1186/1477-7525-9-75> PMID: 21936886
38. Escobar A, Vrotsou K, Bilbao A, Quintana JM, A, García Pérez L, Herrera-Espiñeira C, et al. [Validation of a short form of the function dimension of the WOMAC questionnaire]. *Gac Sanit.* Spain: Ediciones Doyma, S.L. (Travesera de Gracia 17–21, Barcelona 08021, Spain); 2011; 25: 513–518. <https://doi.org/10.1016/j.gaceta.2011.06.003> PMID: 21945068
39. Gonzalez Sáenz de Tejada M, Escobar A, Herdman M, Herrera C, García L, Sarasqueta C. Adaptation and validation of the Osteoarthritis Knee and Hip Quality of Life (OAKHQOL) questionnaire for use in

- patients with osteoarthritis in Spain. *Clin Rheumatol*. 2011; 30: 1563–75. <https://doi.org/10.1007/s10067-011-1855-6> PMID: 21947701
40. Gonzalez Sáenz de Tejada M, Bilbao A, Herrera C, García L, Sarasqueta C, Escobar A. Validation of the Mini-OAKHQOL for use in patients with osteoarthritis in Spain. *Clin Rheumatol*. 2017; 36: 1855–1864. <https://doi.org/10.1007/s10067-017-3611-z> PMID: 28353088
  41. Valls JM, Alonso J. Clinical results and perceived health among patients with hip prosthesis. *Med Clin (Barc)*. 1997; 108: 691–695. PMID: 9324584
  42. Alonso J, Lamarca R, Marti-Valls J. The pain and function of the hip (PFH) scale: a patient-based instrument for measuring outcome after total hip replacement. *Orthopedics*. 2000; 23: 1273–1277. Available: [file:///u/Art?culosUnidad en PDF/2000/5948 Orthop Alonso.pdf](file:///u/Art?culosUnidad%20en%20PDF/2000/5948%20Orthop%20Alonso.pdf) PMID: 11144496
  43. Navarro Collado MJ, Peiró Moreno S, Ruiz Jareño L, Payá Rubio A, Hervás Juan MT, López Matéu P. Validez de la escala de cadera de Harris en la rehabilitación tras artroplastia de cadera. *Rehabilitación*. 2005; 39: 147–154. [https://doi.org/10.1016/S0048-7120\(05\)74337-1](https://doi.org/10.1016/S0048-7120(05)74337-1)
  44. Escobar A, Gonzalez M, Quintana JM, Vrotsou K, Bilbao A, Herrera-Espiñeira C, et al. Patient acceptable symptom state and OMERACT-OARSI set of responder criteria in joint replacement. Identification of cut-off values. *Osteoarthr Cartil*. 2012; 20: 87–92. <https://doi.org/10.1016/j.joca.2011.11.007> PMID: 22155074
  45. Marti-Valls J, Alonso J, Lamarca R, Pinto JL, Auleda J, Girvent R, et al. Efectividad y costes de la intervención de prótesis total de cadera en siete hospitales de Cataluña. *Med Clínica (Barcelona)*. 2000; 114 Suppl: 34–39.
  46. Martin-Fernandez J, Gray-Laymon P, Molina-Siguero A, Martinez-Martin J, Garcia-Maroto R, Garcia-Sanchez I, et al. Cross-cultural adaptation and validation of the Spanish version of the Oxford Hip Score in patients with hip osteoarthritis. *BMC Musculoskelet Disord*. J. Martin-Fernandez, Cdegree Villamanta (C.S. Navalcarnero), Gerencia Asistencial de Atención Primaria, Servicio Madrilenos de Salud, Madrid, Spain. E-mail: [jmfernandez@salud.madrid.org](mailto:jmfernandez@salud.madrid.org), England: BioMed Central Ltd. (E-mail: [info@biomed-central.com](mailto:info@biomed-central.com)); 2017; 18: 205. <https://doi.org/10.1186/s12891-017-1568-3> PMID: 28532445
  47. Rolfson O, Karrholm J, Dahlberg LE, Garellick G. Patient-reported outcomes in the Swedish Hip Arthroplasty Register: RESULTS OF A NATIONWIDE PROSPECTIVE OBSERVATIONAL STUDY. *Bone Joint J*. 2011; <https://doi.org/10.1302/0301-620X.93B7.25737> PMID: 21705555
  48. Ornetti P, Parratte S, Gossec L, Tavernier C, JN A, EM R, et al. Cross-cultural adaptation and validation of the French version of the Hip disability and Osteoarthritis Outcome Score (HOOS) in hip osteoarthritis patients. *Osteoarthr Cartil*. Dijon University Hospital, Department of Rheumatology, Dijon F-21078, France: Elsevier B.V.; 2010; 18: 522–529. <https://doi.org/10.1016/j.joca.2009.12.007> PMID: 20060086
  49. Collins NJ, Roos EM. Patient-reported outcomes for total hip and knee arthroplasty: commonly used instruments and attributes of a “good” measure. *Clin Geriatr Med*. 2012; 28: 367–94. <https://doi.org/10.1016/j.cger.2012.05.007> PMID: 22840304
  50. Polkowski GG, Callaghan JJ, Mont MA, Clohisy JC. Total hip arthroplasty in the very young patient. *Journal of the American Academy of Orthopaedic Surgeons*. 2012. <https://doi.org/10.5435/JAAOS-20-08-487> PMID: 22855851
  51. Hofstede SN, Vlieland TPMV, Van Den Ende CHM, Nelissen RGHH, Marang-Van De Mheen PJ, Van Bodegom-Vos L. Variation in use of non-surgical treatments among osteoarthritis patients in orthopaedic practice in the Netherlands. *BMJ Open*. 2015; <https://doi.org/10.1136/bmjopen-2015-009117> PMID: 26353874
  52. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, et al. OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. *Osteoarthr Cartil*. 2008; <https://doi.org/10.1016/j.joca.2007.12.013> PMID: 18279766
  53. Learmonth ID, Young C, Rorabeck C. The operation of the century: total hip replacement. *Lancet*. 2007. [https://doi.org/10.1016/S0140-6736\(07\)60457-7](https://doi.org/10.1016/S0140-6736(07)60457-7)