

Osteoarthritis and Cartilage



A systematic review of estimates of the minimal clinically important difference and patient acceptable symptom state of the Western Ontario and McMaster Universities Osteoarthritis Index in patients who underwent total hip and total knee replacement



C. MacKay ^{†‡*}, N. Clements ^{§||}, R. Wong [§], A.M. Davis ^{§¶#††}

[†] West Park Healthcare Centre, Toronto, Ontario, Canada

[‡] Toronto Rehabilitation Institute, University Health Network, Toronto, Ontario, Canada

[§] Division of Health Care and Outcomes Research, Krembil Research Institute, University Health Network, Toronto, Ontario, Canada

^{||} The National University of Ireland, Galway, Ireland

[¶] Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

[#] Department of Physical Therapy, University of Toronto, Toronto, Ontario, Canada

^{††} Rehabilitation Science Institute, University of Toronto, Toronto, Ontario, Canada

ARTICLE INFO

Article history:

Received 8 January 2019

Accepted 1 May 2019

Keywords:

Osteoarthritis

Systematic review

Minimal clinically important difference

Patient acceptable symptom state

Western Ontario and McMaster Universities

Osteoarthritis Index

SUMMARY

Objective: To systematically review the minimal clinically important difference (MCID) and patient acceptable symptom state (PASS) estimates in pain and function measured using the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) in patients who underwent primary total knee replacement (TKR) and primary total hip replacement (THR).

Design: The study was carried out following PRISMA recommendations. We searched five electronic databases. Two reviewers independently screened titles, abstracts and full-text papers using a priori inclusion/exclusion criteria. Data were extracted by two independent reviewers. Data were synthesized, with WOMAC values converted to 0–100 scores (0 = best, 100 = worst).

Results: Thirteen studies were included. Research methods used to calculate MCIDs and PASS varied across studies (e.g., using anchor-based or distribution methods, wording of anchor questions within anchor-based methods). Baseline WOMAC scores also varied across studies. Across studies and methods, MCIDs for the WOMAC in patients undergoing TKR ranged from 13.3 to 36.0 for pain and 1.8–33.0 for function; values for WOMAC in THR ranged from 8.3 to 41.0 for pain and from 9.7 to 34.0 for function. PASS cut-offs for TKR ranged from 25.0 to 28.6 for pain and 32.3–36.7 for function, and cut-offs for THR from 15.0 to 30.6 for pain and 28.0–42.0 for function.

Conclusion: Although the WOMAC is a commonly used measure for a single condition, the variability in methods used to calculate MCID and PASS estimates results in a range of values across studies making it unclear whether values reported in the literature can be applied with confidence. Future research is needed to refine methods used to calculate MCIDs and PASS.

© 2019 Osteoarthritis Research Society International. Published by Elsevier Ltd. All rights reserved.

Introduction

Patient reported outcome measures (PROMs) are measures developed to assess health outcomes from the patient's

perspective. They are increasingly used to measure the effectiveness of treatments in clinical research, to inform clinical decision making and patient care, and to inform health policies¹. However, the interpretability of PROMs can be difficult (e.g., interpreting the meaning of a pain reduction of two points). In order to interpret clinically important changes in outcomes, methods have been developed to determine if a medical intervention improves perceived outcomes in patients. The minimal clinically important difference (MCID) is defined according to the patient's perspective of what change is improvement². The MCID was first defined by Jaeschke as “the smallest difference in score in the domain of

* Address correspondence and reprint requests to: C. MacKay, West Park Healthcare Centre, Research and Evaluation, 82 Buttonwood Avenue, Toronto, ON, M6M 2J5, Canada.

E-mail addresses: crystal.mackay@utoronto.ca (C. MacKay), nicolle.clements@hotmail.com (N. Clements), rwong@uhnresearch.ca (R. Wong), adavis@uhnresearch.ca (A.M. Davis).

interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management"³. While a number of definitions of a MCID have been documented in the literature, the common thread is that the MCID represents a lower boundary of change that has been defined to be important⁴. MCIDs are calculated using anchor-based methods which link the change in the outcome to an external anchor that accounts for the patient's perspective, or distribution methods which are data driven approaches that define different statistical parameters to assess clinical significance⁵. While there is no consensus on how to develop MCIDs, the primary approach recommended by the US Food and Drug Administration (FDA) and Outcome Measures in Rheumatology (OMERACT) is an empirical anchor-based approach^{6–8}. Another complementary concept, the patient acceptable symptom state (PASS) has been defined as the highest level of symptoms beyond which patients consider themselves well. In other words, the PASS is the symptom state patients consider acceptable⁹. Despite advancement in the development of MCIDs and PASS for PROMs, there have been methodological challenges in defining clinically important change from the patients' perspective^{2,4}.

Osteoarthritis (OA) is a leading cause of disability world-wide and results in a significant economic burden due to health care expenditures^{10,11}. Total joint replacement (TJR) is a procedure which has been shown to improve outcomes in people with end-stage disease^{12–14}. A widely used PROM, assessing pain, function and stiffness in OA, is the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)¹⁵. A number of empirical studies across different countries have been conducted to estimate the MCID and PASS of the WOMAC in patients with OA. Yet, the variation in MCIDs and PASS between and within methodological approaches and countries is unclear. Earlier work comparing studies of a range of PROMs in rheumatology (focused on MCIDs) suggested that there is wide variation in values across studies^{16–18} limiting their utility and prompting critique of the methods used¹⁶. Our objective was to systematically review the evidence regarding reported MCID and PASS estimates in pain and function measured using the WOMAC in patients who underwent primary total knee replacement (TKR) and primary total hip replacement (THR).

Methods

Womac

The WOMAC has demonstrated reliability, validity and responsiveness in patients with OA^{12,15}. It is comprised of three sub-scales (24 items), including pain (5 items), stiffness (2 items) and physical function (17 items)¹⁵. Scores are converted to a 0–100 scale with higher scores representing worse pain and functional limitations.

Search strategy

We followed guidelines for conducting and reporting in systematic reviews including the *Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) Statement*^{19,20}. The search strategy was developed and performed by an information specialist in collaboration with research team members (AD, RW). In order to ensure a comprehensive search of articles on MCID/PASS, we incorporated keywords used or recommended by authors who have previously published reviews on MCIDs/PASS. Five electronic databases were searched from inception of the databases: Ovid MEDLINE/Ovid MEDLINE In-Process & Other Non-Indexed Citations (1946), Ovid EMBASE (1974), EBSCO CINAHL (1981), Ovid Cochrane Database of Systematic Reviews (2005), and LILACS

(unknown year). Three core sets of search terms were included: MCID/PASS, WOMAC and OA/joint arthroplasty with Boolean operators OR/AND used to link search terms within/between core sets, respectively. For example, key words were: ('minimally/minimal/minimum clinical(ly) important difference(s)' OR 'MCID' OR 'minimally/minimal/minimum important difference(s)' OR 'MID' OR 'clinical(ly) important difference(s)' OR 'CID' OR 'minimally/minimal/minimum clinical(ly) important improvement(s)' OR 'MCII' OR 'minimally/minimal/minimum clinical(ly) important change(s)' OR 'MCIC' OR 'minimally/minimal/minimum perceptible change(s)' OR 'meaningful change(s)' OR 'smallest worthwhile effect(s)' OR 'minimally/minimal/minimum clinically relevant state' OR 'low disease state' OR 'PASSpatient acceptable symptom state' etc.) AND ('Western Ontario and McMaster Universities Osteoarthritis Index' OR 'WOMAC' etc.) AND ('osteoarthritis' OR 'hip arthroplasty' OR 'knee arthroplasty' OR 'hip replacement' OR 'knee replacement' etc.). The full list of key words utilized in each database is reported in the Appendix. The search in databases was limited to English language and publication type only when pertinent (i.e., excluded book chapters/series, conference proceeding, and letters/notes). Due to variation in the search databases, all search terms and limits were tailored to the specific database. The initial search was conducted on 2 June 2016 and updated to 13 August 2018 to identify additional literature published after the initial search.

Study screening and selection criteria

Results of the database searches were imported into EndNote X7. Two reviewers (RW, NC/KS) independently screened titles, followed by abstracts, and then full-text papers using a priori inclusion/exclusion criteria. Full text papers were retrieved if they passed the preliminary screen and if the records did not contain sufficient information to establish eligibility. Papers were eligible for inclusion if the following criteria were met: 1) patients had OA of the hip or knee and undergoing primary THR or TKR and 2) MCID or PASS estimates were calculated for WOMAC pain and function for THR and TKR patients separately. Papers were excluded if: 1) patient diagnosis other than OA; 2) patient population other than adults; 3) unicompartmental, bilateral or revision TKR/THR or another surgery other than TKR/THR; 4) a calculated MCID/PASS for an outcome measure other than WOMAC; 5) a calculated MCID/PASS for WOMAC but could not isolate to TKR or THR patients; 6) language other than English; and, 7) not original research (reviews/systematic reviews editorials, commentaries, workshop summaries, protocols, etc.). For papers that did not meet eligibility for final inclusion but cited or discussed an MCID/PASS estimate, the reference lists were searched to identify the primary sources of the MCID/PASS values reported and other relevant citations not generated from the database searches. Any discrepancies throughout the screening process were resolved through discussion and consensus was achieved by consulting a third reviewer (AD) as necessary.

Data abstraction

A data abstraction form was developed and pilot tested by team members (AD, RW, NC) using a randomly selected set of three eligible papers. Two independent reviewers (NC, CM) extracted information including study and patient characteristics, details about the primary outcome of interest (e.g., WOMAC version and scoring), the MCID and PASS definition adopted, and approach used to compute MCID or PASS values (e.g., distribution-versus anchor-based) including anchor properties, approach criteria, etc. where applicable. When a third reviewer (AD) compared the two

extractions, minimal inconsistencies were found and were rectified through discussion.

We completed the Interpretability box of the *Consensus-based Standards for the selection of health Measurement INstruments (COSMIN)* checklist for all studies²¹. Interpretability is considered an important characteristic of a measurement instrument²¹ and the COSMIN checklist has been used to facilitate interpretation of measures in other research²². We added two columns pertaining to whether each study specified anchor or distribution methods in deriving MCID/PASS values (no/yes) and rational for the method/cut points used (no/yes). These items refer to the reporting of information to facilitate interpretation of scores, rather than standards to assess risk of bias of a study on interpretability²³.

Results

The five database searches yielded 13,840 results (Medline 4661, Embase 7421, Cinahl 1382, CDSR 286, and LILACS 90). After duplications were removed, 9196 citations remained for preliminary screening. Screening of titles/abstracts for relevancy identified 430 potential citations from the database searches. An additional 2 records were identified through reference checking. The 432 full-text articles were screened for eligibility leading to the exclusion of 419 papers (i.e., 83 were not original research; 119 did not investigate MCID/PASS or WOMAC was not a primary outcome; 160 only cited or discussed an MCID/PASS but did not derive estimates; 29 derived

MCID/PASS estimates but not for WOMAC; and, 28 calculated MCID/PASS for WOMAC but were not exclusively OA patients undergoing THR/TKR). Four studies were excluded as they did not calculate MCID/PASS for TKR and THR separately and one study was excluded as it was a secondary analysis of data from another included article. A total of 13 unique studies from final screening met the inclusion criteria and were kept for this review. The selection process is illustrated in Fig. 1. Data will be presented separately for TKR and THR. While there was some variability in use of terminology (MCID, minimally important difference, clinically important difference), we will use the term MCID.

COSMIN checklist: interpretability

Findings from the interpretability section of the Cosmin checklist are in Table 1. Of the 13 studies included, seven reported on 1–3 of seven included COSMIN interpretability criteria^{17,24–29}; five reported on 4–5 criteria^{25,30}; and one reported on 6 criteria³¹. All publications reported the rationale for the method and cut points and 11 publications specified anchor or distribution method^{17,24,25,27–29,31–35}.

MCID and PASS for TKR

A combined total of 7233 individuals who had TKR participated in 9 cohort studies (8 prospective and one retrospective

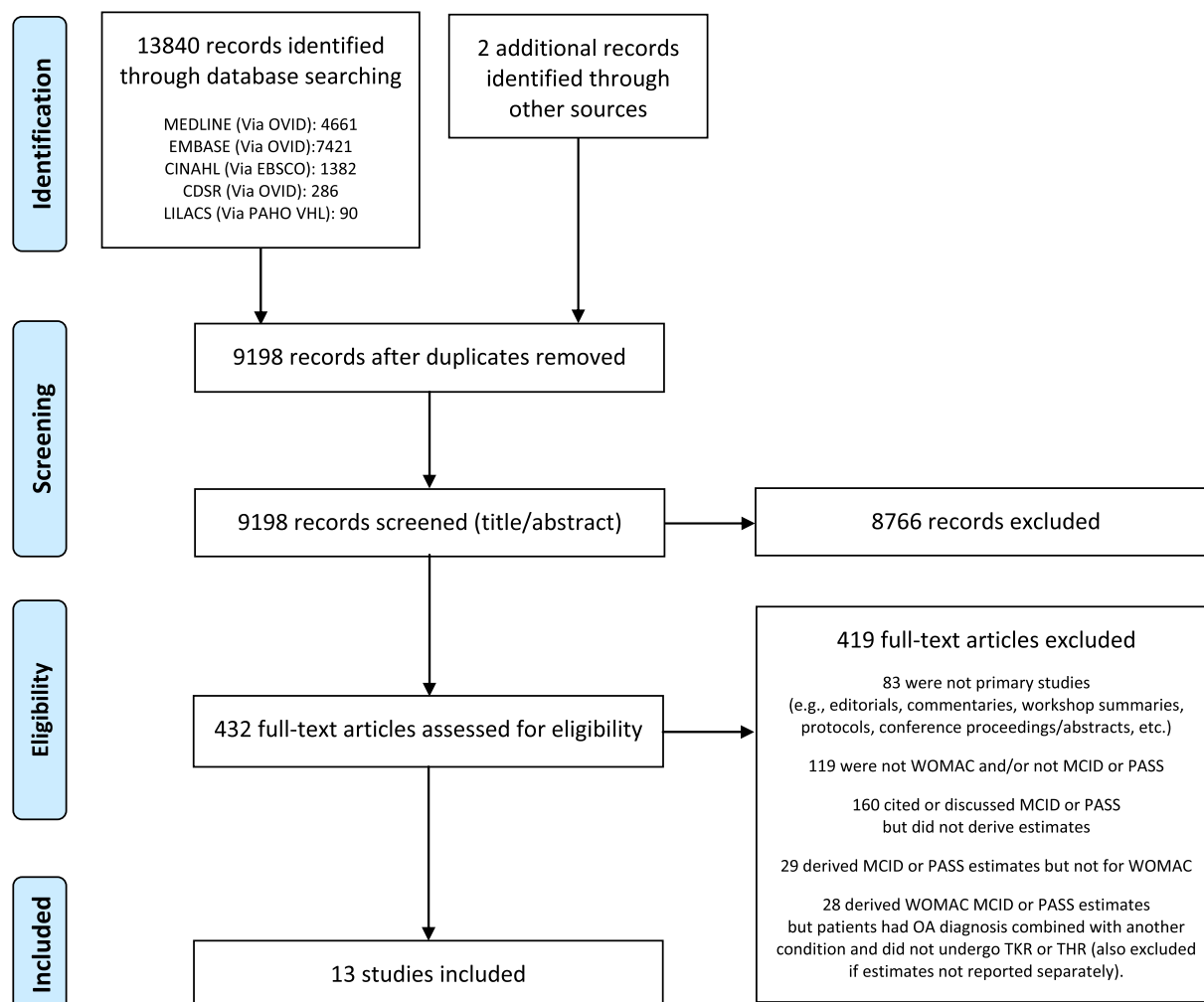


Fig. 1. PRISMA flow diagram summarizing systematic search results and screening process.

Table 1

Cosmin checklist – interpretability

Author	Percent of missing items	Description of how missing items were handled	Distribution of the (total) scores	Percent of responders who had the lowest (total) possible score	Percent of responders who had the highest (total) possible score	Scores and change scores (i.e., Means and SD) for relevant (sub) groups	Specifies anchor or distribution method	Rationale for method/cut points	MCID or PASS determined
Chesworth B, Mahomed N, Bourne R, Davis AM, OJRR Study group ³¹	no	Yes	Yes	Yes	Yes	yes	yes	yes	MCID
Escobar A, Quintana JM, Bilbao A, Arostegui I, Lafuente I, Vidaurreta I ²⁵	no	No	yes	yes	Yes	No	Yes	yes	MCID
Escobar A, Perez LG, Herrera-Espineira, Aizpuru F, Sarasqueta C, Conzalez Saenz de Tejada M, Quintana JM, Bilbao A ³²	no	no	yes	no	no	no	yes	yes	MCID
Escobar A, Gonzalez M, Quintana JM, Vrotsou K, Bilbao A, Herrera-Espineira, Garcia-Perez L, Aizpuru F, Sarasqueta C ²⁴	no	no	yes	no	no	yes	yes	yes	PASS
Vogl M, Wilkesmann R, Lausmann C, Hunger M, Plotz W ²⁹	no	no	yes	no	no	yes	yes	yes	PASS
Quintana JM, Aguirre U, Barrio I, Orive M, Garcia S, Escobar A ²⁷	no	no	yes	no	no	yes	yes	yes	MCID, PASS
Quintana J, Escobar A, Arostegui I, Bilbao A, Azkarate J, Goenaga J, Arenaza J ²⁸	no	no	no	no	no	no	yes	yes	MCID
Maratt JD, Lee YL, Lyman S, Westrich GH ³⁴	yes	yes	yes	no	no	no	yes	yes	MCID
Judge A, Cooper C, Williams S, Dreinhoefer K, Dieppe P ³³	yes	yes	yes	no	no	yes	yes	yes	MID and MCID
Quintana JM, Escobar A, Bilbao A, Arostegui I, Lafuente I, Vidaurreta I ³⁵	no	no	yes	yes	yes	no	yes	yes	MCID
Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jorda KP, Croft P, de Vet HCW ¹⁷	no	no	no	no	no	no	yes	yes	MIC
Naal F, Impellizzeri F, Lenze U, Wellauer V, Eisenhart-Rothe R, Leunig M ²⁶	no	no	no	no	no	no	no	yes	PASS
Vina ER, Hannon MJ, Kwok CK ³⁶	no	yes	yes	no	no	yes	no	yes	MID

Table II
Characteristics of studies by method (anchor-based and distribution) and country for total knee replacement (TKR)

Author	Year	Country	Participants	Diagnosis	Study Design	Setting	Time Points	Anchor Properties	Approach	Pre-op WOMAC Pain mean (SD)	Pre-op WOMAC Function mean (SD)
Studies Using Anchor-Based Methods by Country for TKR											
Chesworth <i>et al.</i> ³¹	2008	Canada	N = 1578; mean age 69.6; 62% female	OA	Prospective cohort study	Hospital setting (Ontario Joint Replacement Registry)	Decision for surgery and 1 year after surgery	15 point scale: –7 to +7 (0 = same; 7 = a very great deal better and –7 = a very great deal worse)	ROC curves	58 ¹⁷	62 ¹⁷
Escobar <i>et al.</i> ²⁵	2007	Basque Country, Spain	N = 516; mean age 71.6; 75% female	OA	Prospective study	3 hospitals	Pre-TKR, 6 months and 2 years	5 point scale: “a great deal better” to “a great deal worse”	Mean change score for those who were “somewhat better” at 6 months or 2 years	55.56 (18.53)	61.87 (17.75)
Escobar <i>et al.</i> ³²	2013	Spain	Cohort 1: N = 415; mean age 71.4; 71.3% female Cohort 2: N = 497; mean age 71.4; 69.4% female	OA	2 prospective cohorts	Cohort 1: public hospitals cohort 2: 15 hospitals	Baseline and 12 months after surgery	5 point scale: “a great deal better” to “a great deal worse”	Mean change score for those who were “somewhat better”; ROC approach and 2 questions about satisfaction	Cohort 1: 55.8 (17.3) Cohort 2: 55.4 (18.7)	Cohort 1: 61.7(17.3) Cohort 2: 61.8 (17.6)
Escobar <i>et al.</i> ²⁴	2012	Spain	N = 510; mean age 71.4; 68.4% female	OA	Secondary analysis of a prospective cohort study	15 hospitals	Baseline and 1 year data used	4 responses from “very satisfied” to “very dissatisfied”	Responder analysis (OMERCT OARS) and ROC analysis	55.3 (18.6)	61.0 (17.5)
Quintana <i>et al.</i> ²⁸	2006	Spain	N = 792; 71.9 years; 73% women	OA	Prospective study	5 large and 2 medium public teaching hospitals	Pre-operatively and 6 months post-operative	5 point scale: “a great deal better,” “somewhat better” to “a great deal worse”	Mean change scores for those who were “somewhat better”	Appropriate 57.94, uncertain 50.26, inappropriate 42.58	Appropriate 65.42, uncertain 54.36, inappropriate 46.24
Terwee <i>et al.</i> ¹⁷	2010	Netherlands	N = 73	OA	Cohort study	Hospitals	Pre-operatively and 6 months	Anchor properties not stated but used “a little better”	1. Mean change in score in the subgroup of patients who reported themselves as “a little better”; 2. ROC Receiver characteristic curve	Not reported	Not reported
Naal <i>et al.</i> ²⁶	2015	Germany and Switzerland	N = 233; mean age 70.6; 67.8% female	end stage OA	Prospective cohort study	One large teaching hospital and one university hospital in Germany	1–2 weeks before surgery and at 3, 6 and 12 months after surgery	Two 15-point global rating scales on change in pain and function: “very much worse” to “very much better”		Not reported	Not reported
Vina <i>et al.</i> ³⁶	2016	USA	N = 269; mean age 67.6 years; 61% female	OA/high risk for knee OA	Prospective cohort study	4 OAI clinical centres	Before TKR; t+1 (first follow up after TKR - any time after day 1 to 1 year after T0) and T+2 (second follow-up after TKR 1 year following T+1)	original research used anchor method	PASS defined as the value beyond which patients consider themselves satisfied with actual OA symptoms (WOMAC pain <32.4 WOMAC disability <31.0); MID: 1/2 SD of difference between change scores	64.3 ± 19.1	65.3 ± 17.7

Maratt <i>et al.</i> ³⁴	2015	USA	N = 2350; mean age 66.7; 57.1% female	OA	Retrospective review of prospectively collected data (from hospital's data for all TKA)	Hospital	Baseline and 2 years	6 point scale: "more improvement than I ever dreamed" to "the quality of my life is worse"	ROC analysis to determine cut points using Youden index	45.1 ± 17.3	45.7 ± 17.4
Studies Using Distribution-Based Methods by Country for TKR											
Vina <i>et al.</i> ³⁶	2016	USA	N = 269; mean age 67; 61% female	OA/high risk for knee OA	Prospective cohort study	4 OAI clinical centres	Before TKR; t+1 (first follow up after TKR - any time after day 1 to 1 year after TO) and T+2 (second follow-up after TKR 1 year following T+1)	Not applicable	1/2 SD of difference between change scores	64.3 ± 19.1	65.34 ± 17.7
Terwee <i>et al.</i> ¹⁷	2010	The Netherlands	N = 73	OA	Prospective cohort study	Hospital	Baseline and 6 months	No Applicable	1/2 SD of baseline score	Not reported	Not reported

cohort)^{17,24–26,28,31,32,34,36}. The mean age of individuals was greater than 65 in all studies. Overall mean pre-operative pain scores ranged from 45.1 (17.3) to 64.3 (19.1) and function ranged from 45.7 (17.4) to 65.3 (17.7) (0–100 scale in which 100 is worse). Studies reported on data from TKR cohorts in six countries: Spain ($n = 4$), Germany ($n = 1$), Netherlands ($n = 1$), Switzerland ($n = 1$), Canada ($n = 1$) and the United States ($n = 2$) (some studies included cohorts from > one country). Nine studies used anchor-based methods^{17,24–26,28,31,32,34,36} and two used distribution methods^{17,36} (Table II summarizes the characteristics of the studies using anchor-based methods and distribution methods).

Studies ($n = 9$) using anchor-based methods calculated mean change scores or receiver operating curves (ROCs) or both. There was variability in the anchor used in the studies (question, response options), with three studies using a 5-point scale^{25,28,32}, two studies using a 15-point scale^{26,31}, one study using a 4-point scale²⁴ and one using a 6-point scale³⁴ (2 studies not reported^{17,36}). The time frame (period to compare current health status to prior health status) was also variable: three studies collected data pre-surgery and 1 year^{24,31,32}, two at 6 months only^{17,28}, one at 2 years only³⁴, and three at multiple time points^{25,26,36}.

Using anchor-based methods, the MCID for WOMAC pain calculated using mean change scores was 13.3 (0.5–26.1) in one study conducted in Holland ($n = 73$; calculated as mean change in score in the subgroup of patients who reported themselves as "a little better" on the anchor)¹⁷ and ranged from 22.6 to 29.9 in the three studies conducted in Spain^{25,28,32}. Using the ROC method, the MCIDs for WOMAC pain ranged from 20.5 to 36.0. The MCID for WOMAC function calculated using the mean change method was 1.8 (–8.3–11.9) in the Dutch study¹⁷ and ranged from 17.67 to 33.5 in studies conducted in Spain^{25,28,32}. Using the ROC method, the MCIDs for WOMAC function ranged from 22.8 to 33.0. Two studies reported on distribution methods^{17,36}. The MCID calculated was 9.4 (summary score) in an American study³⁶ and 10.6 for pain and 10.0 for function in another study from Holland¹⁷. MCIDs are presented by method in Table III (anchor-based) and Table IV (distribution).

PASS cut-offs were calculated using the 75th centile or ROC analysis approach. Some researchers used a range of response options (4 items) to rate their satisfaction with joint replacement and others used a yes/no response option. Escobar *et al.*²⁴ reported cut-offs for TKR that ranged from 25.0 to 28.6 for pain and 32.3–36.7 for function. Naal *et al.*²⁶ reported PASS cut-offs for the summary WOMAC as 30.7 at 6 months and 12.2 at 12 months (Table III).

MCID and PASS for THR

A combined total of 5540 individuals who had undergone THR from 9 prospective cohort studies were included^{17,24,26–29,31,33,35}. The mean age of individuals was greater than 65 in all studies. Pre-operative pain scores ranged from a mean of 44.4 (1.1) to 58.0¹⁷; pre-operative function scores ranged from 48.7 (1.1) to 65.8 (16.97) (0–100 scale with higher scores indicating more pain and worse function). Studies were conducted in individuals undergoing THR in Canada ($n = 1$), Spain ($n = 4$), Netherlands ($n = 1$), Germany ($n = 2$), Switzerland ($n = 1$), and multiple countries in Europe ($n = 1$). Eight studies used anchor-based methods^{17,24,26–29,31,35} and 2 used distribution methods^{17,33}. The anchor question, response options and time frames varied across studies. Data were collected pre-surgery in all studies and at 6 months ($n = 4$)^{17,27–29}, 1 year ($n = 2$)^{24,31} and multiple time points ($n = 2$)^{26,35}. Table V summarizes the characteristics of the studies using anchor-based methods and distribution methods.

Using anchor-based methods, the MCID for WOMAC pain calculated using mean change scores was 8.3 in one study from Holland¹⁷. In Spain, values ranged from 24.55 to 29.26 in three

Table III

Estimates of minimal clinically important difference (MCID) and patient acceptable symptom state (PASS) for WOMAC Pain and Function in Studies of TKR using Anchor-based Methods

Country	Author	Year	MCID: Mean Change	MCID: ROC	PASS	MCID: Mean Change	MCID: ROC	PASS
			Pain			Function		
Canada	Chesworth <i>et al.</i> ³¹	2008		Improvement: 36; deterioration: 30			Improvement: 33; Deterioration: 25	
Basque Country	Escobar <i>et al.</i> ²⁵	2007	6 months: 22.87(18.13); 2 years: 27.98 (19.44)			6 months - 19.01 (17.48); 2 years: 20.84 (18.2)		
Spain	Escobar <i>et al.</i> ³²	2013	Cohort 1: Global: 29.9 (27.1 –32.6); Surgery worthwhile: 28 (23.6–32.4) Global satisfaction: 25.2 (21.9–28.5) Cohort 2: Global: 28.1 (25.1 –31.0); Surgery worthwhile: 25.6 (21.6 –29.7); Global satisfaction: 27.5 (23.7–31.4)	Cohort 1: ROC global: 20.5 (20.2 –20.9) Cohort 2: ROC global: 23.5 (23.1 –23.8)		Cohort 1: Global: 31.1 (28.3 –33.9); Surgery worthwhile: 20.6 (16.4 –24.8); Global satisfaction: 18.8 (15.7–21.9) Cohort 2: Global: 33.5 (30.9 –36.0); Surgery worthwhile: 22.5 (19.0 –26.1); Global satisfaction: 25.8 (22.4–29.2)	Cohort 1: ROC global: 24.2 (23.6–24.7) Cohort 2: ROC global: 23.0 (22.7–23.2)	
Spain	Escobar <i>et al.</i> ²⁴	2012			Centile Approach: 25.0 (20–30); ROC 28.6 (28.2–28.9)			Centile Approach: 32.3 (35.3–29.4); ROC 36.7 (36.4–37.1)
Spain	Quintana <i>et al.</i> ²⁸	2006	22.6			17.67		
The Netherlands	Terwee <i>et al.</i> ¹⁷	2010	13.3 (0.5–26.1)	29.4 (19.7–39.0)		1.8 (–8.3 to 11.9)	22.8 (6.2–39.4)	
Germany and Switzerland	Naal <i>et al.</i> ²⁶	2015			AUC*: 6 months: 0.77 (0.68–0.85); cut-off: 30.7 12 months: AUC: 0.89 (0.82–0.96); cut- off: 12.2 90.81% improved			
USA	Vina <i>et al.</i> ³⁶	2016						90.68% improved
USA	Maratt <i>et al.</i> ³⁴	2015		31.25			26.93	

* PASS for pain and function not presented separately.

Table IV

Estimates of MCID for WOMAC Pain and Function in Studies of TKR using Distribution Methods

Country of Conduct	Author	Year	Summary Score	Pain	Function
			½ standard deviation		
USA	Vina <i>et al.</i> ³⁶	2016	9.4*		
Netherlands	Terwee <i>et al.</i> ¹⁷	2010		10.6†	10.0†

* 0.5 SD of difference between change scores.

† 0.5 SD of the baseline score.

studies (with one study reporting cut-offs of 15, 23, 36 for 1, 2 and 3 baseline tertile scores, respectively)^{27,28,35}. Using the ROC method, the MCIDs for WOMAC pain ranged from 22.4 to 41.0. For function, the MCID, using the mean change method ranged from 20.8 to 26.54. Using the ROC method, the MCID ranged from 18.4 to 34.0. Using distribution methods, MCIDs were calculated as 10.5 for pain and 9.7 for function in one study¹⁷ and reported in another study as a difference in pain score of ≥ 24.5 and in function score of ≥ 23.9 ³³. MCIDs are presented by approach in [Tables VI and VII](#).

PASS cutoffs were calculated using the 75th centile or ROC method. The number of response options on the question of satisfaction varied (4 or 10 options). The cut-off for pain was reported as 20.0 and 30.0 in one study using different methods (percentile and ROC approach, respectively) and 30.9 and 31.2 for function²⁴. Two studies reported that PASS values for summary WOMAC scores were close to 15^{26,29} ([Table VI](#)).

Discussion

In this systematic review the estimates of the MCIDs and PASS for pain and function were examined for a commonly used measure, the WOMAC, and in the context of a specific procedure for end stage OA i.e., primary THR and TKR. The derived MCID and PASS values from the 13 studies included were variable within and across countries. Across studies and methods, MCIDs for the WOMAC in patients undergoing TKR ranged from 13.3 to 36.0 for pain and 1.8–33.0 for function; values for WOMAC pain in THR ranged from 8.3 to 41.0 for pain and from 9.7 to 34.0 for function.

Previous studies have reported significant heterogeneity in the calculation methods used for MCIDs in rheumatology^{16,18}. While we found variation in methodological approaches to determining the MCID (anchor and distribution approaches as expected) and the PASS, we also found variation within a given method (e.g., different wording of anchor questions and choice of cut-points). Additionally, there was variation in patient sample characteristics (i.e., baseline WOMAC scores) used to determine values and, given the known influence of baseline scores on MCID and PASS calculations^{2,4}, this likely also contributed to the variability in the values obtained.

Using the COSMIN interpretability questions, we found that seven of 13 studies reported on only 1–3 COSMIN criteria. While interpretability is not a measurement property, these questions are intended to facilitate interpretation of the scores^{21,23}. Lack of reporting in a number of studies highlight limitations in interpreting findings from these studies and points to the need for better reporting in MCID/PASS studies in the future (e.g., percentage of responders with lowest possible score, highest possible score and score in sub-groupings).

It is generally accepted that a PROM like the WOMAC would not have a single MCID and that values for the MCID would vary depending on the intervention and context (e.g., patient group)^{37,38}. Interestingly, even within the context of one intervention with a relatively predictable result in a population with end

stage OA, we found that there is heterogeneity in the approaches used to calculate MCIDs for the WOMAC in patients undergoing TJR and subsequent values for the MCID. In this review, more studies used anchor-based methods compared to distribution methods as recommended^{6–8}. However, there was variation within methods, which included: 1) variation in the wording of the anchor question and the response scale (e.g., in TKR, anchors ranged from a 4-point scale to a 15-point scale), 2) the time frame studied, and 3) the approach to calculating the MCID (mean change and ROC). First, MCID was more often calculated for patients reported to be “somewhat better” in the anchor response but cutoffs varied (e.g., patients reporting to be “a little better”). While the cutoff is paramount to defining small but important change, there is a lack of evidence and agreement to optimal cutoff levels². Some have suggested that the validity of anchors could be improved by querying about the importance of change rather than using the magnitude of change item¹⁷. Moreover, patients may need to be involved in specifying the cut-point to be used, rather than the investigator/clinician as often occurs. Variation also included the time frame for the study (e.g., baseline and 6 months, 12 months, 2 years). Since studies have demonstrated that much of the change in pain and function occurs within the first 6 months following TJR^{38,39} it is likely the studies reporting on 6 and 12 months are most relevant as change beyond this time period may be influenced by external contextual factors or events. In other research, 36% of patients who underwent THR reported one or more positive impact life events in the year following surgery, while 63% reported one or more negative life events. The number of positive life events was associated with engagement in life activities following THR⁴⁰. Finally, not all studies conducted a ROC analysis. This is despite recommendations that ROC methods be used for values that are to be applied at the individual level¹⁷. Overall, there appeared to be variability in values regardless of method used.

We found there was significant variation in the baseline WOMAC pain and function scores across studies for the same intervention. While TJR is a procedure for end-stage OA, this is not surprising as there are no agreed upon guidelines on surgical candidacy for TJR. There is evidence of large heterogeneity in patient status at the time of surgery and pain and function alone do not determine who may be a surgical candidate^{41–43}. Variation in baseline scores has been found across other studies of MCIDs and researchers have highlighted the influence of baseline scores on the value calculated in studies of MCIDs^{2,4,44}. Davis *et al.*³⁸ showed that failure to consider baseline scores could result in some patients being misclassified as having not benefitted from treatment (some patients could not achieve important improvement due to baseline scores). Some researchers have studied the use of item response theory-informed methods to correct for the variability in MCID based on baseline score⁴⁵ but the limited evidence to date does not provide information on what methods are superior in developing MCIDs.

Studies were included from a range of countries. Studies conducted in Spain were most common and used similar methods (i.e., anchor-based, mean change). While the variability within studies conducted in Spain was less than across methods and countries, there was heterogeneity in results particularly for THR (TKR: 22.6–29.9; THR: 17.67–33.5). These findings further highlight the challenges inherent when determining and interpreting important change.

Fewer studies were included which calculated a PASS for WOMAC pain and function for patient undergoing TJR. Studies used ROC curves or the 75th centile approach (75th percentile of the scores for improvement in patients who report an important improvement by the anchoring question). PASS cut-offs were similar across the studies of TKR (25.0–28.6 for pain; 32.3–36.7 for

Table V
Characteristics of studies by method (anchor-based and distribution) and country for total hip replacement (THR)

Author	Year	Country	Participants	Diagnosis	Study Design	Setting	Time Points	Anchor Properties	MCID Approach	Pre-op WOMAC Pain mean (SD)	Pre-op WOMAC Function mean (SD)
Studies Using Anchor-Based Methods by Country for THR											
Chesworth et al. ³¹	2008	Canada	N=1131; mean age 68.8; OA 57% female	OA	Prospective cohort study	Hospital	Decision for surgery and 1 year	15 point scale varied from –7 to +7 (0 = same; 7 = a very great deal better and –7 = a very great deal worse)	ROC curves	58.0(17)	62.0(17)
Escobar et al. ²⁴	2012	Spain	N=351; 44.3% female; mean age 65.4	OA	Secondary analysis of a prospective cohort study	15 hospitals	Baseline and 1 year	4 response options from “very satisfied” to “very dissatisfied”	PASS: Responder analysis (OMERCT OARS) and ROC analysis	55.2(17.4)	64.9(16.5)
Quintana et al. ²⁷	2012	Basque Country; Spain	Cohort 1: N=573; 51.8% female; 48.5% > age 70 Cohort 2: N=333; 48.1% female; 48.8% > age 70	OA	2 prospective cohorts	7 teaching hospitals	Preoperative and 6 months	“If you had to live the rest of your life with the hip symptoms you have now, how would you feel?” Response: “totally satisfied”, “slightly satisfied”, “not satisfied”, and “not at all satisfied”; Cohort 1 only: “Compared with status before you had a hip prosthesis how would you rate the status of your hip right now?” 7 responses from “a great deal better” to “a great deal worse”	PASS	Total: 55.62(19.05) Cohort 1: 54.3(18.57) Cohort 2: 57.8(19.58)	Total: 65.80(16.97) Cohort 1: 64.39(16.86) Cohort 2: 68.0(16.98)
Quintana et al. ²⁸	2006	Spain	N=784; 48% female; mean age 69.1	OA	Prospective cohort study	5 large and 2 medium public teaching hospitals	Preoperative and 6 months	Question about joint improvement; five response options: “a great deal better” to “a great deal worse.”	Change scores calculated for “somewhat better”	appropriate - 59.61; uncertain 42.23; inappropriate 27.08; knees appropriate 54.68 (18.71)	appropriate - 69.62, uncertain 55.58, inappropriate 42.92 64.73 (16.27)
Quintana et al. ³⁵	2005	Basque Country	N=379 at 6 months and 310 at 2 years; 50.7% female, mean age 69.4 (8.8)	OA	Prospective cohort study	3 teaching hospitals	Preoperative, 6 months and 2 years	Transitional question about joint improvement with five response options from “a great deal better” to “a great deal worse”	Change for those who responded “somewhat better”		
Terwee et al. ¹⁷	2010	The Netherlands	N=188	OA	Prospective cohort study	Hospital	Baseline and 6 months	Anchor properties not stated	1. Mean change in score in the subgroup of patients who reported themselves as “a little better” on the anchor; 2. ROC	Not reported	Not reported
Naal et al. ²⁶	2015	Germany and Switzerland	N=193; Mean age 66.1; 48.2% female	OA	Prospective cohort study	One large teaching hospital in Switzerland and one university hospital in Germany	Pre-surgery, 3, 6, 12 months	Two 15-point global rating scales, indicating whether they had changed in pain and physical functioning (“very much worse” to “very much better”)	PASS: ROC (cut off based on Youden index)	Not reported	Not reported
Vogl et al. ²⁹	2014	Germany	N=281; 58% female; mean age 68	OA and osteonecrosis	Prospective cohort study	Clinic	Baseline and 6 months	Were your expectations on THR fulfilled?	PASS (75%); ROC	44.4 (1.1)	48.7(1.1)
Studies Using Distribution-Based Methods by Country for THR											
Terwee et al. ¹⁷	2010	The Netherlands	N=188	OA	Prospective cohort study	Hospital	Baseline and 6 months	Not applicable	1/2 SD of baseline score	Not reported	Not reported
Judge et al. ³³	2010	12 countries in Europe	N=1327 at baseline; 908 completed 12 months; 56% female, Age:<50 = 77 (8.6), 50–69 = 486 (54.3), >=70 (37.1)	OA	Prospective cohort study	20 orthopaedic centres	Baseline and 12 months	Not applicable	½ SD mean difference in WOMAC measurements	54.5 ± 17.6	58.7 ± 16.3

Table VI
Estimates of MCID and PASS for WOMAC Pain and Function in Studies of THR using Anchor-based Methods

Country	Author	Year	Pain		Function	
			MCID: Mean change	MCID: ROC	PASS	PASS
Canada	Chesworth et al. ³¹	2008		CID improvement: 41; CID deterioration: 35		
Spain	Escobar et al. ²⁴	2012		Centile approach: 20.0; ROC: 30 AUC: 0.77 (0.70–0.84)		Centile approach: 30.9; ROC: 31.2 AUC: 0.81 (0.75–0.87)
Basque Country, Spain	Quintana et al. ²⁷	2012	15, 23, 36 for tertiles	19, 25, 25 Validation cohort: 20, 25, 25	20, 25 and 25 for tertiles Validation cohort: 20, 25, 25	28, 35 and 42 for tertiles Validation cohort: 32, 32, 40
Spain	Quintana et al. ²⁸	2006	24.55			
Basque Country, Spain	Quintana et al. ³⁵	2005	29.26 (16.9)			
Netherlands	Terwee et al. ¹⁷	2010	8.3 (–0.3 to 16.9)	22.4 (2.5–42.3)	9, 22 and 31 for tertiles	26.4, 39, 40 Validation cohort: 25, 39, 40
Germany and Switzerland	Naal et al. ²⁶	2015			20.8 26.54 (17.79) 23.7 (11.1–36.3)	
Germany	Vogl et al. ²⁹	2014		6 months AUC: 0.92 (0.85–0.98); cut-off: 15.4 12 months AUC: 0.84 (0.70–0.98); cut-off: 14.4* Estimates above 15 for global WOMAC; of patients who considered state satisfactory 75% had a change in WOMAC of >25*	18.4 (2.9–39.6)	

* PASS for pain and function not presented separately.

Table VII

Estimates of MCID for WOMAC Pain and Function in Studies of THR using Distribution Methods

Country	Author	Year	Pain	Function
			½ Standard Deviation	
Netherlands	Terwee et al. ^{17,*}	2010	10.5	9.7
Europe	Judge et al. ^{33,†}	2010	≥24.5‡	≥23.9‡

* 0.5 SD of the baseline score.

† 0.5 SD of mean difference in scores.

‡ As reported.

function). While there were few findings reported for PASS for THR, values ranged from 20.0 to 30.0 for pain and 30.9–31.2 for function. Vogl et al.²⁹ reported a summary WOMAC score of 15 and interpreted that their low threshold state compared to other studies conducted in THR was because the cohort had lower baseline and follow-up scores. Similar to MCIDs, researchers have suggested that the methodology for identification of PASS may influence the identified cut-off points. The ROC approach has provided estimates that were somewhat lower than the cut-off points identified with the 75th centile approach⁹. This was not evident from the few studies included in this systematic review.

The study has limitations. Publications in languages other than English were not included. Findings are specific to the study population and are nontransferable across patient groups or interventions. Despite a comprehensive search strategy to locate articles, it is possible studies meeting inclusion/exclusion criteria were missed.

The WOMAC is a commonly used PROM in patients undergoing TJR. Yet, our findings highlight variation in methodological approaches used to determine MCIDs and PASS for the WOMAC, variation in approaches within methods, and variation in patient sample characteristics used to determine values. Due to the heterogeneity in the research methods used across studies, we were unable to identify clear patterns (e.g., participant characteristics or methodological approaches) that may explain the heterogeneity in estimates. Given the variability in the values reported for MCID and PASS across studies it is unclear that the values reported in the current literature can be applied to new research with confidence. To be able to use PROMs and identify responders to interventions, more standardization of methodological approaches to estimating MCIDs and PASS (including methods within approaches i.e., anchor questions) may be required. This standardization will be critical in the era of personalized medicine in which therapies are targeted to sub-groups of patients with unique attributes. At present, careful consideration needs to be given to the applicability of a given MCID to the specific context in which the WOMAC will be used. Future research is needed to refine methods used to calculate MCIDs and PASS including comparative methods research.

Author contributions

Conception and design: Davis.

Data acquisition: Clements, Wong.

Analysis and interpretation of the data: MacKay, Clements, Wong, Davis.

Drafting of the article: MacKay.

Critical revision of the article for important intellectual content: Clements, Wong, Davis.

Final approval of the article: MacKay, Clements, Wong, Davis.

Dr. Davis (adavis@uhnresearch.ca) takes responsibility for the integrity of the work as a whole.

Competing interests

None of the authors have any conflicts of interests in relation to this work.

Role of the funding source

There was no role of a funding agency in this work.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.joca.2019.05.002>.

References

- Black N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013;346:f167.
- Engel L, Beaton DE, Touma Z. Minimal clinically important difference: a review of outcome measure score interpretation. *Rheum Dis Clin N Am* 2018;44(2):177–88.
- Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Contr Clin Trials* 1989;10(4):407–15.
- Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol* 2002;14(2):109–14.
- Copay AG, Subach BR, Glassman SD, Polly Jr DW, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J* 2007;7(5):541–6.
- McLeod LD, Coon CD, Martin SA, Fehnel SE, Hays RD. Interpreting patient-reported outcome results: US FDA guidance and emerging methods. *Expert Rev Pharmacoecon Outcomes Res* 2011;11(2):163–9.
- Wells G, Anderson J, Beaton D, Bellamy N, Boers M, Bombardier C, et al. Minimal clinically important difference module: summary, recommendations, and research agenda. *J Rheumatol* 2001;28(2):452–4.
- Wells G, Beaton D, Shea B, Boers M, Simon L, Strand V, et al. Minimal clinically important differences: review of methods. *J Rheumatol* 2001;28(2):406–12.
- Kvien TK, Heiberg T, Hagen KB. Minimal clinically important improvement/difference (MCII/MCID) and patient acceptable symptom state (PASS): what do these concepts mean? *Ann Rheum Dis* 2007;66(Suppl 3):iii40–1.
- Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;380(9859):2163–96.
- Menon J, Mishra P. Health care resource use, health care expenditures and absenteeism costs associated with osteoarthritis in US healthcare system. *Osteoarthritis Cartilage* 2018;26(4):480–4.
- Ethgen O, Bruyere O, Richy F, Dardennes C, Reginster JY. 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(5):963–74.
- Visser MM, Bussmann JB, Verhaar JA, Arends LR, Furlan AD, Reijman M. Recovery of physical functioning after total hip arthroplasty: systematic review and meta-analysis of the literature. *Phys Ther* 2011;91(5):615–29.
- Kane RL, Saleh KJ, Wilt TJ, Bershadsky B. The functional outcomes of total knee arthroplasty. *J Bone Joint Surg Am* 2005;87(8):1719–24.
- Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988;15(12):1833–40.
- Doganay Erdogan B, Leung YY, Pohl C, Tennant A, Conaghan PG. Minimal clinically important difference as applied in rheumatology: an OMERACT rasch working group systematic review and critique. *J Rheumatol* 2016;43(1):194–202.
- Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et al. Mind the MIC: large variation among populations and methods. *J Clin Epidemiol* 2010;63(5):524–34.
- Copay AG, Eyberg B, Chung AS, Zurcher KS, Chutkan N, Spanghehl MJ. Minimum clinically important difference: current trends in the orthopaedic literature, Part II: lower extremity: a systematic review. *JBJS Rev* 2018;6(9):e2.
- Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.
- Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6(7), e1000097.
- Mokkink LB, 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(4):539–49.
- Højgaard P, Klokke L, Orbaai AM, Holmsted K, Bartels EM, Leung YY, et al. A systematic review of measurement properties of patient reported outcome measures in psoriatic arthritis: a GRAPPA-OMERACT initiative. *Semin Arthritis Rheum* 2018;47(5):654–65.
- Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27(5):1171–9.
- Escobar A, Gonzalez M, Quintana JM, Vrotsou K, Bilbao A, Herrera-Espineira C, et al. Patient acceptable symptom state and OMERACT-OARSI set of responder criteria in joint replacement. Identification of cut-off values. *Osteoarthritis Cartilage* 2012;20(2):87–92.
- Escobar A, Quintana JM, Bilbao A, Arostegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. *Osteoarthritis Cartilage* 2007;15(3):273–80.
- Naal FD, Impellizzeri FM, Lenze U, Wellauer V, von Eisenhart-Rothe R, Leunig M. Clinical improvement and satisfaction after total joint replacement: a prospective 12-month evaluation on the patients' perspective. *Qual Life Res* 2015;24(12):2917–25.
- 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(4):563–72.
- Quintana JM, Escobar A, Arostegui I, Bilbao A, Azkarate J, Goenaga JJ, et al. Health-related quality of life and appropriateness of knee or hip joint replacement. *Arch Intern Med* 2006;166(2):220–6.
- Vogl M, Wilkesmann R, Lausmann C, Hunger M, Plotz W. The impact of preoperative patient characteristics on health states after total hip replacement and related satisfaction thresholds: a cohort study. *Health Qual Life Outcomes* 2014;12:108.
- Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol* 2009;62(10):1006–12.

31. Chesworth BM, Mahomed NN, Bourne RB, Davis AM, Group OS. Willingness to go through surgery again validated the WOMAC clinically important difference from THR/TKR surgery. *J Clin Epidemiol* 2008;61(9):907–18.
32. Escobar A, Garcia Perez L, Herrera-Espineira C, Aizpuru F, Sarasqueta C, Gonzalez Saenz de Tejada M, *et al.* Total knee replacement; minimal clinically important differences and responders. *Osteoarthritis Cartilage* 2013;21(12):2006–12.
33. Judge A, Cooper C, Williams S, Dreinhoefer K, Dieppe P. Patient-reported outcomes one year after primary hip replacement in a European Collaborative Cohort. *Arthritis Care Res* 2010;62(4):480–8.
34. Maratt JD, Lee YY, Lyman S, Westrich GH. Predictors of satisfaction following total knee arthroplasty. *J Arthroplast* 2015;30(7):1142–5.
35. 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. *Osteoarthritis Cartilage* 2005;13(12):1076–83.
36. Vina ER, Hannon MJ, Kwoh CK. Improvement following total knee replacement surgery: exploring preoperative symptoms and change in preoperative symptoms. *Semin Arthritis Rheum* 2016;45(5):547–55.
37. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res* 2011;11(2):171–84.
38. Davis AM, Perruccio AV, Lohmander LS. Minimally clinically important improvement: all non-responders are not really non-responders an illustration from total knee replacement. *Osteoarthritis Cartilage* 2012;20(5):364–7.
39. Lenguerrand E, Wylde V, Gooberman-Hill R, Sayers A, Brunton L, Beswick AD, *et al.* Trajectories of pain and function after primary hip and knee arthroplasty: the ADAPT cohort study. *PLoS One* 2016;11(2), e0149306.
40. MacKay C, Webster F, Venkataramanan V, Bytautas J, Perruccio AV, Wong R, *et al.* A prospective cohort study examining medical and social factors associated with engagement in life activities following total hip replacement. *Osteoarthritis Cartilage* 2017;25(7):1032–9.
41. Dieppe P, Judge A, Williams S, Ikwueke I, Guenther KP, Floeren M, *et al.* Variations in the pre-operative status of patients coming to primary hip replacement for osteoarthritis in European orthopaedic centres. *BMC Musculoskelet Disord* 2009;10:19.
42. Ackerman IN, Dieppe PA, March LM, Roos EM, Nilsson AK, Brown GC, *et al.* Variation in age and physical status prior to total knee and hip replacement surgery: a comparison of centers in Australia and Europe. *Arthritis Rheum* 2009;61(2):166–73.
43. Gademan MG, Hofstede SN, Vliet Vlieland TP, Nelissen RG, Marang-van de Mheen PJ. Indication criteria for total hip or knee arthroplasty in osteoarthritis: a state-of-the-science overview. *BMC Musculoskelet Disord* 2016;17(1):463.
44. de Vet HC, Terwee CB. The minimal detectable change should not replace the minimal important difference. *J Clin Epidemiol* 2010;63(7):804–5. author reply 6.
45. Rouquette A, Blanchin M, Sebille V, Guillemin F, Cote SM, Falissard B, *et al.* The minimal clinically important difference determined using item response theory models: an attempt to solve the issue of the association with baseline score. *J Clin Epidemiol* 2014;67(4):433–40.