BMJ Open

Minimal important change and difference for knee osteoarthritis outcome measurement tools after non-surgical interventions: a systematic review

M Denika C Silva ,[1,2] Diana M Perriman,[1,3] Angela M Fearon ,[1,4] Jacqui M Couldrick,[1] Jennie M Scarvell[1,3]

To cite: Silva MDC, Perriman DM, Fearon AM, *et al*. Minimal important change and difference for knee osteoarthritis outcome measurement tools after non-surgical interventions: a systematic review. *BMJ Open* 2023;**13**:e063026. doi:10.1136/bmjopen-2022-063026

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2022-063026).

Received 30 March 2022
Accepted 21 April 2023

For numbered affiliations see end of article.

Correspondence to
Mrs M Denika C Silva;
denika.silva@canberra.edu.au

ABSTRACT

Objectives To systematically review and provide estimates of the minimal important change (MIC) and difference (MID) for outcome tools in people with knee osteoarthritis (OA) after non-surgical interventions.

Design
A systematic review.

Data sources MEDLINE, CINAHL, Web of Science, Scopus and Cochrane databases were searched up to 21 September 2021.

Eligibility criteria We included studies that calculated MIC and MID using any calculation method including anchor, consensus and distribution methods, for any knee OA outcome tool after non-surgical interventions.

Data extraction and synthesis We extracted reported MIC, MID and minimum detectable change (MDC) estimates. We used quality assessment tools appropriate to the studies' methods to screen out low-quality studies. Values were combined to produce a median and range, for each method.

Results Forty-eight studies were eligible (anchor-k=12, consensus-k=1 and distribution-k=35). MIC values for 13 outcome tools including Knee injury and Osteoarthritis Outcome Score (KOOS)-pain, activities of daily living (ADL), quality of life (QOL) and Western Ontario and McMaster Universities Arthritis Index (WOMAC)-function were estimated using 5 high-quality anchor studies. MID values for 23 tools including KOOS-pain, ADL, QOL and WOMAC-function, stiffness and total were estimated using 6 high-quality anchor studies. One moderate quality consensus study reported MIC for pain, function and global assessment. MDC values from distribution method estimates for 126 tools including KOOS-QOL and WOMAC-total were estimated using 38 good-to-fair-quality studies.

Conclusion Median MIC, MID and MDC estimates were reported for outcome tools in people with knee OA after non-surgical interventions. The results of this review clarify the current understanding of MIC, MID and MDC in the knee OA population. However, some estimates suggest considerable heterogeneity and require careful interpretation.

PROSPERO registration number CRD42020215952.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ We estimated minimal important change (MIC) (within-group), minimal important difference (MID) (between-groups) and minimum detectable change values using anchor, consensus or distribution methods papers, respectively.
⇒ This systematic review included a defined population of people with knee osteoarthritis, after non-surgical interventions.
⇒ High-quality anchor studies were used to contribute to MIC and MID estimates were assessed using a credibility tool specially designed to evaluate anchor method papers.
⇒ Consensus and distribution methods papers were evaluated using quality assessment tools suited to each method.
⇒ Median estimates were used to reflect the synthesised data due to data skewness.

INTRODUCTION

The efficacy of therapeutic interventions is commonly evaluated using statistical significance regardless of patient importance.[1] To understand whether differences in outcome measures after treatment are important to patients, it is necessary to know what constitutes a minimum important change or difference for the individual or cohort. These changes and differences are called the minimal important change (MIC) and difference (MID). There are numerous outcome measures for knee osteoarthritis (OA) and many estimates of MIC and MID. However, these estimates can arise from different methodologies leading to variability, confusion and misinterpretation.[2–5] Achieving clarity in this space is crucial as these values are used in regulatory and clinical decision-making.[6 7] This systematic review aimed to provide estimates of MIC and MID for knee OA outcome

measurement tools in people with knee OA after non-surgical interventions.

MIC and MID are defined as the minimum value of an outcome measure that the patient, clinician or relevant others perceive as an important change or difference.[4 8 9] The MIC considers the change in a clinical outcome measure within a single group or an individual over time. In contrast, MID considers the difference between independent groups or between individuals.[4 10–12] However, the terminology of MIC and MID is used inconsistently.[13] The concept was first described by Jaeschke, who studied patients' perceptions of preintervention and postintervention beneficial change.[8] This concept later included both improvement[14] and worsening.[15]

Three methods are used to estimate MIC and MID: anchor, consensus and distribution.[6 16] For the anchor method, MIC or MID values are usually estimated by referencing the patients' responses against an externally validated scale ('anchor').[17] The 'global rating of change' is most commonly used as the anchor but other methods (proxy responses or performance based measures) are also used.[18] The receiver operating characteristic (ROC) method for deriving an estimate from anchor questions has been suggested to be more precise for clinical settings than the mean change method.[9 15] In the consensus method, values are directly estimated by a group of experienced clinicians or patients until a consensus is achieved.[6] In the distribution method, values are estimated statistically, based on the variance of the outcome data using half the SD,[19] 1 SEM[20] or minimum detectable change (MDC) which is based on SEM.[6 21] The anchor method is widely considered to be the most valid because it is based on patient perception of what constitutes minimal change or difference.[16 22] In this paper, we have included MIC and MID estimates from anchor-based and consensus-based papers as well as MDC estimates. MDC estimates, as a distribution measure, are less meaningful because they do not reflect patient perception, but they do estimate instrument error which is of value to researchers.[6 21] For this reason, we included MDC as well as the anchor and consensus estimates.

Knee OA is a common cause of pain and disability.[23] Outcome measurement tools that include the domains of pain, physical function, patient global assessment and imaging are recommended to determine the efficacy of therapeutic interventions in knee OA studies.[24] MIC and MID values have been estimated for knee OA outcome measures in these domains using anchor, consensus and distribution methods with variable results. The variability of the methods used makes the selection of an appropriate estimate confusing for clinicians, researchers and regulatory bodies.[2–4]

The primary objective of this systematic review was to estimate MIC and MID for knee OA outcome measurement tools based on estimates from high-quality anchor studies only. The secondary objectives were to determine MIC and MID estimates based on consensus method and to synthesis MDC values derived from distribution methods.

METHODS

This systematic review was designed and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement.[25] The protocol was registered on PROSPERO (registration number: CRD42020215952).

Literature search

Five databases (MEDLINE, CINAHL, Web of Science, Scopus and Cochrane) were searched from each database's respective inception to 21 September 2021. A comprehensive search strategy was developed to capture all relevant articles, and database-specific MESH terms were used. The search strategy was as follows. (*knee OR genu OR tibiofemoral OR patellofemoral) AND (osteoarthr* OR degenerat*) AND ((“MCIC” OR “MCID” OR “MCII” OR “MIC” OR “MII” OR “MPCC” OR “MPCD” OR “MPCI” OR “MDC” OR “SDC” OR “SDD” OR “CIC” OR “CID”) OR (“minim* clinical* important change*” OR “minim* clinical* important difference*” OR “minim* clinical* important improvement*” OR “minim* important change*” OR “minim* important difference*” OR “minim* important improvement*” OR “minim* perceptible clinical* change*” OR “minim* perceptible clinical* difference*” OR “minim* perceptible clinical* improvement*” OR “minim* detectable change*” OR “small* detectable change* " OR “small* detectable difference* " OR “clinical* important change*” OR “clinical* important difference*")). The records were exported to EndNote V.X9.2 for reference management.

Study screening and selection criteria

Covidence software (Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia (www.covidence.org) was used to manage the selection process. Records identified in the search were uploaded and duplicates were removed. Screening of titles and abstracts, then full texts, were performed independently by two reviewers (MDCS and JMC) and conflicts were resolved by a third reviewer (JMS). Included studies incorporated any design that calculated MIC and MID for any knee OA outcome measurement tool considering improvement after non-surgical intervention for adults with knee OA, and using any calculation method: anchor, consensus or distribution methods. We included studies that reported MDC because MDC is considered as an estimate from the distribution method.[3] Though distribution-based approaches such as MDC do not reflect the patients perception, MDC values are important for researchers to get some idea about instrument error.[6 21] We considered studies with MIC or MID values for improvement only and excluded values for deterioration because improvement values are used to evaluate the efficacy of treatment. Studies were excluded if the data from participants with

knee OA could not be separated from other conditions, for example, hip OA or other knee pathologies. Studies of MIC, MID and MDC were included even if they used a different terminology for example, minimal clinically important change for MIC, minimal clinically important difference for MID and smallest detectable change or difference or minimal detectable difference for MDC.

For consistency, we defined the MIC as the pre–post change of one group that is, threshold for those who responded that they had minimally improved on the anchor measure. The MID is defined as the difference (pre–post change) between two groups, that is, 'minimally improved' and 'stayed the same' groups using the anchor response as defined in previous studies.[10 12 26] The MDC is the minimum change above the measurement error based on a given level of confidence.[6 21] MDC values for a 90 or 95 CI are labelled as MDC90 or MDC95.

Quality assessment

The quality of the included studies was assessed according to their methodology. The quality of the anchor studies was assessed using the credibility instrument developed by Devji *et al*[27] which was designed to assess the credibility of anchor studies assessing MIC and MID of patient-reported outcome measures. However, we adapted this tool in the following ways. The credibility instrument includes five core criteria, namely: (1) The anchor is rated by the patient, (2) The anchor is interpretable and relevant to the patient, (3) The MIC or MID estimate of patient-reported outcome measure is precise, (4) The correlation between the anchor and the outcome measure reported by the patient is satisfactory and (5) The authors select a threshold on the anchor that reflects a small but important difference). We adapted criteria 1, 3 and 4. For criteria 1 and 4, we included both patient and clinician as relevant anchor respondents. For criteria 3, we included performance measures as well as patient-reported outcome measures .We considered the paper to be 'high' quality if at least three of the five criteria were 'yes', 'definitely yes' or 'to a great extent' and of 'low' quality if not.[28] Consensus studies were assessed using the Critical Appraisal Screening Programme qualitative tool[29] which is designed to assess qualitative studies and is well suited to consensus studies. The quality was rated as 'high', 'moderate' or 'low' based on reliability and credibility.[30] Distribution studies were evaluated using the National Heart, Lung and Blood Institute, National Institute of Health) quality assessment tool for before-and-after (pre–post) studies with no control group[31] and ratings included 'good', 'fair' or 'poor' based on reliability and credibility.[31] The quality assessment of included studies was performed by one reviewer (MDCS) and a random sample of 20% had an independent second review (AMF or JMC) to improve the accuracy.[32]

Data extraction and analysis

We extracted study characteristics including sample size, participant demographics, details of the intervention, follow-up time, outcome measurement tools, calculation method and actual estimate reported based on the method (MIC or MID or MDC). In addition, we extracted the details of the anchor used in each study.

We extracted reported MIC, MID and MDC values from each study. We normalised the values to a 0–100 scale. If a study reported MDC as a percentage of the grand mean (MDC divided by grand mean percentage),[33] we converted the data into MDC90 or 95 for the synthesis.

All data were synthesised and described as the median estimates. The median and range (minimum and maximum) of MIC, MID and MDC were calculated using multiple estimates from the included studies arising from different non-surgical interventions, calculation methods, time points and anchors. Mean values were not calculated due to skewness of distributions.[34] We excluded low-quality anchor studies from the median MIC and MID synthesis. Furthermore, we conducted a subanalysis to determine median MID based on the ROC method where available because the ROC estimates are considered to be more precise than mean-change estimates and recommended at both individual-level and group-level analyses, and in clinical settings.[9 15] Though we planned to conduct a subanalysis to determine the effect of follow-up time on MIC and MID, we were unable to do reliable rate estimates because of a limited number of studies. Therefore, we plotted the values against time including only studies where the outcome measures were assessed at three time points or more.

Patient and public involvement

This is a systematic review. Patients or the public were not involved in this study.

Deviations from the protocol

The protocol registered in PROSPERO lists searches in MEDLINE, Embase, CENTRAL (Cochrane Central Register of Controlled Trial), Web of Science and CINAHL. However, Embase ceased to be available to the research team, so, Scopus was substituted. The data synthesis plan to assess MIC and MID in terms of standardised mean difference was not performed due to the skewness of the distributions. Therefore, we reported median and range for each measure without comparison. The planned meta-analysis was prevented by the skewness and homogeneity of the data and a decision was made to follow a simple descriptive approach using median estimates that was more accessible.[28]

RESULTS

Study selection

The search yielded 2376 studies and after duplicates were removed, 1059 records were screened. Two hundred and seventeen studies were screened in full-text review resulting in 48 eligible studies (k=48) (figure 1). No further studies were identified after checking the reference lists of included studies.

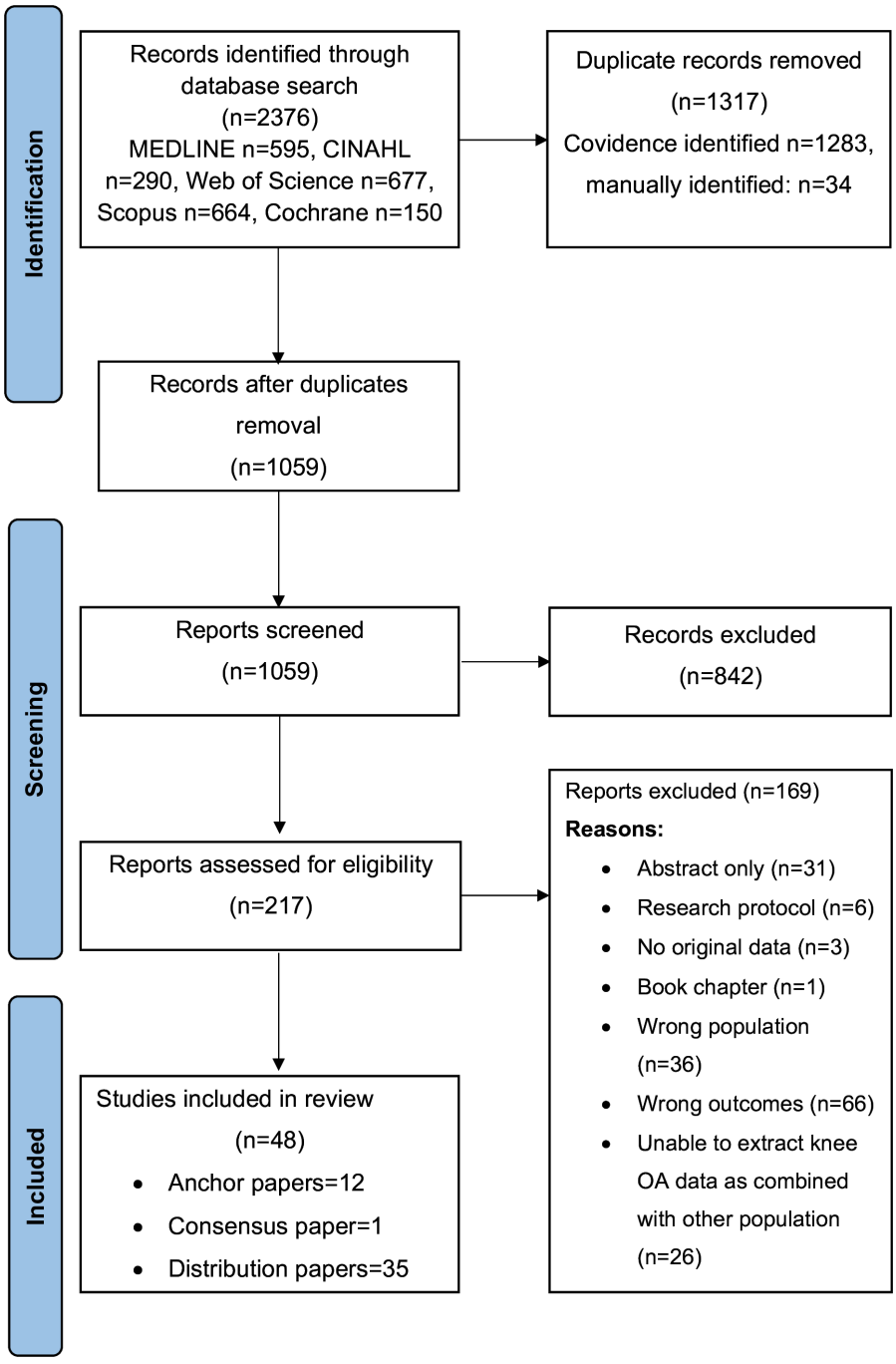

Figure 1 Flow diagram of study selection (PRISMA).[25] PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Included studies calculated MIC and MID by anchor method (k=12), by consensus method (k=1) and MDC by distribution method (k=35).

The methodological quality of included studies

Most anchor studies (k=10) were of high quality and two were of low quality[35 36] (online supplemental file 1). The quality of the consensus study (k=1) was moderate (online supplemental file 2). The quality of distribution studies ranged from good (k=11) to fair (k=24) (online supplemental file 3).

Study characteristics of included studies

All the anchor studies were observational prospective cohort studies[14 15 35 37–43] and two of them were nested within randomised controlled trials[36 44](table 1). The number of participants in each study ranged from 41 to 1606. The mean age and body mass index ranged from 57.1 to 67.9 years and from 28.1 to 33 kg/m^2, respectively. The interventions used in these studies were rehabilitation, exercise, physiotherapy and non-steroidal anti-inflammatory drugs. The follow-up time ranged from 7 days to 1 year. Eleven studies used global rating

Table 1 Characteristics of included studies

Study	No of participants; age, years, mean (SD)	Intervention (Follow-up)	Outcome tool	Calculation method (Measure extracted)
Tubach, 2005[14]	603; 67.9 (10.2)	NSAID (4 weeks)	Pain on VAS on movement; Patient's global assessment of disease activity on a VAS; WOMAC-function subscale	Anchor (MIC)
Mills, 2016[15]	272; NR	Non-surgical (26,52 weeks)	KOOS-pain, activities of daily living, quality of life	Anchor (MIC and MID)
McCarthy, 2004[33]	15; 65.1 (11.3)	Exercises (1 week)	Aggregated locomotor function score; 8 m walk time; Stair ascent and descent time; Transferring time	Distribution (MDC95)
Harris, 2013[35]	134; 59 (11)	Non-operative (3 months)	OKS-pain, function, summary; ICOAP; KOOS-PS	Anchor (MIC, MID); Distribution (MDC90)
Klokker, 2016[36]	41; NR	exercises (12 weeks)	DAP	Anchor (MID)
Angst, 2018[37]	190; 66.1 (10.2)	Rehabilitation (3 months)	WOMAC-pain, function, functional standing/walking and stiffness; SF-36-bodily pain, physical functioning, role-physical, vitality, social functioning, mental and general	Anchor (MID); Distribution (MDC95)
Hmamouchi, 2012[38]	173; 57.1 (10.1)	NSAID (6 weeks)	WOMAC-total	Anchor (MID)
Mostafaee, 2021[39]	142; 58.3 (7.6)	Physiotherapy (4 weeks)	KOOS-pain, symptoms, activities of daily living, sports and recreation, and quality of life	Anchor (MID)
Ornetti, 2011[40]	881; 67.1 (10.9)	NSAID (4 weeks)	WOMAC-function; Patient-reported functional disability on an NRS; Physician reported functional disability on an NRS	Anchor (MIC)
Singh, 2014[41]	137; NR	Conservative (2 weeks)	ICOAP-pain, constant pain, intermittent pain; KOOS-PS; KOOS-quality of life	Anchor (MIC); Distribution (MDC90)
Williams, 2012[42]	159; NR	Exercise (2, 6, 12 months)	KOS-activities of daily living subscale; LEFS; WOMAC-total	Anchor (MID); Distribution (MDC90, MDC95)
Perrot, 2013[43]	1606; 66.9 (9.0)	Usual care (7 days)	Pain on a 0–10 NRS at rest and movement	Anchor (MIC)
Lee, 2017[44]	165; 61	Thai Chi or physical therapy (12 weeks)	PROMIS-Short Forms—physical function, pain interference, depression, anxiety	Anchor (MID)
Salottolo, 2018[45]	27 (clinicians); NR	NR	Pain, Function and Global assessment	Consensus (MIC)
Alghadir, 2015[46]	65; 54.9 (9.9)	NR	Timed Up and Go test	Distribution (MDC95)
Alghadir, 2016 a[47]	121; 52.9 (9.3)	NR (48 hours)	WOMAC-Arabic version-pain, function, total	Distribution (MDC95)
Alghadir, 2016 b[48]	121; 54.0 (9.3)	NR (48 hours)	NRS for pain-Arabic version	Distribution (MDC95)
Alghadir, 2017[49]	97; 58 (11.5)	NR (1 week)	OKS-Arabic version	Distribution (MDC95)
Alghadir, 2018[50]	121; 52.9 (12.5)	NR (24 hours)	VAS for pain; NRS for pain; Verbal Rating Scale for pain	Distribution (MDC95)
Suwit, 2020[51]	55; 69.0 (11)	NR (1 week)	30 s chair stand; 40 m fast-paced test; 9-step climb	Distribution (MDC90)

Continued

Table 1 Continued

Study	No of participants; age, years, mean (SD)	Intervention (Follow-up)	Outcome tool	Calculation method (Measure extracted)
Baert, 2018[52]	8 (12 knees); 68 (11)	NR (3 s)	Knee joint position sense test using analogue inclinometer; Knee force sense test using a handheld dynamometer	Distribution (MDC95)
Brisson, 2018[53]	46; 61.0 (6.6)	NR (6 months, 24 months)	Peak KAM, KAM impulse and Peak KFM using 3D motion analysis; Quadriceps strength and power using dynamometer; Load frequency using a triaxial accelerometer; BMI	Distribution (MDC95)
Callghan, 2009[54]	55; 64 (14)	NR (24–72 hours)	Quadriceps fatigue using surface electromyography	Distribution (MDC95)
Suhail and Chaudha, 2021[55]	82; 60.4 (6.7)	NR (48 hours)	2 min walk test	Distribution (MDC95)
Hoglund, 2019[56]	20; 58.3 (8.05)	NR (2–7 days)	30 s fast-paced walk test	Distribution (MDC95)
Ijima, 2019[57]	59; 59.1 (6.1)	NR (1 month)	Stopwatch-based 11-sep stair climb test	Distribution (MDC90, MDC95)
Jansen, 2021[58]	103; NR	NR (1 month)	Knee image: bone density, joint space width, osteophytes, eminence height and joint angle	Distribution (MDC95)
Kean, 2010[59]	20; 53.6 (9.1)	NR (same day)	Quadriceps isokinetic strength, isometric strength and percent of voluntary activation using a dynamometer	Distribution (MDC90)
Klokker, 2015[60]	20; 64 (6.6)	NR (1 week)	DAP	Distribution (MDC95)
McCarthy, 2008[61]	55; 64.9 (9.7)	NR (1 week)	Maximum isometric strength; Vastus medialis oblique, Vastus lateralis, Rectus femoris-initial median frequency; Vastus medialis oblique, Vastus lateralis, Rectus femoris fatigue slope	Distribution (MDC95)
Monticone, 2021[62]	102; 69.1 (9.0)	NR (10 days)	Fremantle Knee Awareness Questionnaire, Italian version	Distribution (MDC95)
Nalbant, 2021[63]	25; 62.3 (9.8)	NR (same day)	L-test	Distribution (MDC95)
Naylor, 2014[64]	75; 67.6 (9.4)	NR (10 days)	VAS for pain; 6 min Walk Test; Timed Up and Go test; KOOS-pain, symptoms, activities of daily living, quality of life	Distribution (MDC95)
Parveen, 2017[65]	25; 50.5 (6.3)	NR (7 days)	Tinetti Performance-Oriented Mobility Assessment scale—total, balance subscale, gait subscale	Distribution (MDC95)
Peter, 2018[66]	135; NR	NR (1 week)	Animated activity questionnaire	Distribution (MDC95)
Piva, 2004[67]	25; 57 (8)	NR (same day)	Get up and Go test	Distribution (MDC95)
Srimurugan Pratheep, 2018[68]	20; 70.3 (5.8)	NR (2 weeks)	Pressure pain threshold using algometer	Distribution (MDC95)
Takacs, 2014[69]	20; 64.1 (7.9)	NR (14 days)	Single-leg standing balance test	Distribution (MDC95)
Tevald, 2016[70]	25; 61 (9)	NR (7 days)	Hip abductor strength using a hand-held dynameter	Distribution (MDC95)

Continued

Table 1 Continued

Study	No of participants; age, years, mean (SD)	Intervention (Follow-up)	Outcome tool	Calculation method (Measure extracted)
Tse, 2021[71]	38; 65.1 (7.0)	NR (same day)	Frontal plane tibial alignment using inclinometer	Distribution (MDC95)
Turcot, 2008[72]	25; 63.9 (7.6)	NR (8 days)	3D linear accelerations of the tibia and femur during walking	Distribution (MDC95)
van der Straaten, 2020[73]	19; 65.1 (5.2)	NR (20 days)	Range of motion (trunk abduction-adduction, pelvis abduction-adduction, hip flexion-extension) and Centre of mass displacement	Distribution (MDC95)
Yuruk, 2014[74]	40; NR	NR (1 day)	De Motion Mobility Index (Turkish version)	Distribution (MDC90, MDC95)
Ravaud, 1999[75]	30; 65.4 (7.7)	NR (2 weeks)	Joint space width	Distribution (MDC95)
Kanko, 2019[76]	74; 57.7 (8.8)	Exercises (1 week)	Star excursion balance test	Distribution (MDC95)
Mutlu, 2015[77]	73 (141 knees); 56.2 (6.7)	Physiotherapy (4 weeks)	Pressure pain threshold using a handheld pressure algometer	Distribution (MDC90)
Hunter, 2006[78]	72; 58.9 (8.6)	Novel ther (6 months)	Cartilage volume using X-ray, MRI (Femur, Patella, Tibia)	Distribution (MDC95)
Motyl, 2013[79]	15; 61.0 (7.8)	Steroid injection (20 days)	20 m walk test	Distribution (MDC95)

.3D, three-dimension; DAP, Dynamic weight-bearing Assessment of Pain; ICOAP, Intermittent and Constant Osteoarthritis Pain; KAM, knee adduction moment; KFM, knee flexion moment; KOOS, Knee injury and Osteoarthritis Outcome Score; KOOS-PS, Knee injury and Osteoarthritis Outcome Score Physical Function Short form; KOS, Knee Outcome Survey; LEFS, Lower Extremity Functional Scale; MDC90, Minimum Detectable Change based on 90% CI; MDC95, minimum detectable change based on 95% CI; MIC, minimal important change; MID, minimal important difference; NR, not reported; NRS, Numeric Rating Scale; NSAID, Non-steroidal anti-inflammatory drugs ; OKS, Oxford Knee Score; PROMIS, Patient-Reported Outcome Measurement Information System; SF-36, 36-item Short Form health survey; VAS, Visual Analogue Scale; WOMAC, Western Ontario and McMaster Universities Arthritis Index.

of change as the external anchor while one study[44] used multiple anchors. Most anchor studies (k=6)[36–39 42 44] reported MID values only, four studies[14 40 41 43] reported MIC values only and two studies,[15 35] reported both MIC and MID. Four of these studies[35 37 41 42] also reported MDC values. Moreover, studies used different anchor questions and group classifications when calculating MIC and MID. For example, one MIC study[35] considered the minimal improved response group as 'my knee has got better' and another study[40] considered the response group as both 'good' and 'excellent' improvement groups (online supplemental file 4).

The consensus study (k=1)[45] (table 1) used a questionnaire to survey 27 clinicians from a range of specialities (orthopaedic (38%), rheumatology (33%), internal medicine (19%) and other (9%)). The clinicians were asked about MIC values for pain, function and global assessment for severe knee OA. However, participants were not asked to consider time duration nor the interventions. MIC was termed 'minimal clinically important improvement' in this study.

Most distribution studies (k=30) that reported MDC, were test–retest observational studies[46–75] assessing the reliability of the outcome tool, and five used datasets from interventional cohort studies[76 77] and randomised controlled trials[33 78 79] (table 1). In the distribution studies (k=35), the number of participants in each study ranged from 8 to 135. The mean age and body mass index ranged from 50.5 to 70.3 years and from 22.7 to 35 kg/m^2, respectively. Studies estimated MDC90 and MDC95. The follow-up time ranged from the same day to 1 year.

The MIC estimates derived using the anchor method

The median MIC for 13 tools (with subscales) were calculated based on 5 high-quality anchor studies[14 15 40 41 43] using 23 estimates (table 2). These estimates were based on different underlying calculations, follow-up time and anchor questions. Methods for calculating MIC included: mean change (pre and post mean change of the minimally improved group),[8 15 41] 75th centile value of the mean change of the group[40 43] and 75th centile value adjusted with the baseline score, age and disease duration.[14] Most studies included one follow-up time (range: 7 days to 4 weeks), but one study[15] reported MIC at two time-points (26 and 52 weeks). One anchor question was used in most studies, however, two studies[15 40] reported different MIC values based on two different anchor questions (general health status and functional state).

The MID estimates derived using the anchor method

The median MID for 23 tools were calculated based on 6 high-quality anchor studies[15 37–39 42 44 46] using 83 estimates (table 3). These estimates were based on different underlying calculations, follow-up time and anchor questions. Methods for calculating MID included: ROC method[80] only (k=3),[38 39 42] mean change (Redilmier and Lorig) method only (pre–post mean change difference between two groups)[81] (k=1),[37] both ROC and mean change methods (k=1),[15] and mean change in T-scores with multiple anchors (k=1).[44] Most studies (k=4) included one follow-up (range: 4–12 weeks), but two studies included MID at multiple time points for example, 2, 6 and 12 months.[15 42] One anchor question was used in most studies, however, one study used multiple anchors.[44]

MID estimates based on the ROC method were reported and compared with MID estimates for all methods. Overall, 4 of 6 studies[15 38 39 42] (67%) used the ROC method. The ROC estimates were the same as the overall estimates in most cases (table 4).

The effect of follow-up time on MIC and MID

There were insufficient data to establish reliable rate estimates for the effect of time. The MIC of Knee injury and Osteoarthritis Outcome Score (KOOS)-pain and KOOS-quality of life (QOL) and, the MID of KOOS-pain, KOOS-QOL, Knee Outcome Score (KOS)-activity of daily living (ADL), Lower Extremity Functional Scale (LEFS) and Western Ontario and McMaster Universities Arthritis Index (WOMAC)-total were assessed at more than three different time points. The MIC of KOOS-QOL and, MID of KOOS-QOL, LEFS and KOS-ADL appeared to increase with increasing follow-up time. However, MIC of KOOS-pain and MID of WOMAC-total appeared to reduce with follow-up time and KOOS-pain remained constant (online supplemental file 5).

MIC values derived using the consensus method

One consensus study[45] reported that MIC for pain, function and global assessment were 20% of the maximum score.

The MDC estimates derived using the distribution method

The median MDC was calculated for 126 tools based on 38 studies (35 good-to-fair distribution and three high-quality anchor studies) using 308 estimates (online supplemental file 6). These estimates were based on different calculation methods and follow-up times. Four included studies reported MDC90 values only[41 51 59 77] and 29 studies reported MDC95 values only. Five studies[37 42 57 64 74] reported both the MDC90 and MDC95 values. Most studies (k=37) reported unadjusted MDC, while one study reported both the adjusted and unadjusted estimates.[37] Six studies separately reported inter-rater/intrarater MDC values.[46 52 56 60 67 73] Furthermore, three studies reported distinct values for two patient groups in each study, for example, the placebo group and the treatment group,[78] the most painful and the least painful groups,[70] and the groups that reported moderate improvement ('great deal better') and MCID improvement ('somewhat better').[41] Most studies assessed the index (worst) knee, but one study based the estimate on all diseased knees (12 knees in 8 patients).[52] Regarding the time point of MDC estimation, most studies (k=39) reported MDC estimates at one time point only, but one study reported MDC estimates at three time points (2, 6 and 12 months).[42]

Table 2 MIC values of knee osteoarthritis outcome tools derived using the anchor method

Outcome tools	Score	Median	Minimum	Maximum	No of estimates	Study
ICOAP-pain	100=worst	**18.5**			1	Singh, 2014[41]
ICOAP-constant pain	100=worst	**18.7**			1	Singh, 2014[41]
ICOAP-intermittent pain	100=worst	**18.4**			1	Singh, 2014[41]
KOOS-pain	100=best	**12.4**	4.3	20.1	4	Mills, 2016[15]
KOOS-activities of daily living	100=best	**8.4**	8.2	8.7	2	Mills, 2016[15]
KOOS-quality of life	100=best	**9.8**	8.0	11.6	2	Mills, 2016; Singh, 2014[15 41]
KOOS-PS	100=worst	**2.2**			1	Singh, 2014[41]
Pain on NRS at rest	100=worst	**10.0**	10	10	2	Perrot, 2013[43]
Pain on VAS on movement	100=worst	**19.9**mm			1	Tubach, 2005[14]
Patient-reported functional disability on an NRS	100=worst	**27.6**	27.2	27.9	2	Ornetti, 2011[40]
Patient's global assessment of disease activity on VAS	100=worst	**18.3**mm			1	Tubach, 2005[14]
Physician-reported functional disability on an NRS	100=worst	**25.3**	25	25.5	2	Ornetti, 2011[40]
WOMAC-function	100=worst	**17.0**	9.1	17.1	3	Tubach, 2005; Ornetti, 2011[14 40]
Estimates based on low-quality studies						
ICOAP-pain	100=worst	**13.4**			1	Harris, 2013[35]
KOOS-PS	100=worst	**12.0**			1	Harris, 2013[35]
OKS-pain	100=best	**17.3**			1	Harris, 2013[35]
OKS-function	100=best	**10.6**			1	Harris, 2013[35]
OKS-summary	100=best	**7.1**			1	Harris, 2013[35]

All the scores are from 0 to 100.
The median estimates are shaded in blue.
Estimates based on low-quality studies are shaded in grey.
.ICOAP, Intermittent and Constant Osteoarthritis Pain; KOOS, Knee injury and Osteoarthritis Outcome Score; KOOS-PS, Knee injury and Osteoarthritis Outcome Score Physical Function Short form; MIC, minimal important change; NRS, Numeric Rating Scale; OKS, Oxford Knee Score; VAS, Visual Analogue Scale; WOMAC, Western Ontario and McMaster Universities Arthritis Index.

Table 3 MID values of knee osteoarthritis outcome tools derived using the anchor method

Outcome tools	Score	Median	Minimum	Maximum	Number of estimates	Study
KOOS-pain	100=best	**11.8**	4	17	13	Mills, 2016; Mostafaee, 2021[15 39]
KOOS-activities of daily living	100=best	**2.5**	−1.5	15.5	7	Mills, 2016; Mostafaee, 2021[15 39]
KOOS-quality of life	100=best	**6.5**	3	12.5	7	Mills, 2016; Mostafaee, 2021[15 39]
KOOS-sports/recreation	100=best	**17.5**			1	Mostafaee, 2021[39]
KOOS-symptoms	100=best	**12.5**			1	Mostafaee, 2021[39]
KOS-activities of daily living	100=best	**6.4**	2.2	10.6	6	Williams, 2012[42]
LEFS	100=best	**6.9**	0.6	15.6	6	Williams, 2012[42]
PROMIS Short Forms-physical function*	100=worst		4.5	5.3	NR	Lee, 2017[44]
PROMIS Short Forms-pain interference*	100=worst		4.2	4.2	NR	Lee, 2017[44]
PROMIS Short Forms-depression*	100=worst		6	6.2	NR	Lee, 2017[44]
PROMIS Short Forms-anxiety*	100=worst		4.5	6.6	NR	Lee, 2017[44]
WOMAC-pain	100=best	**8.7**	7.1	21	3	Angst, 2018[37]
WOMAC-function	100=best	**14.5**	11.3	14.9	3	Angst, 2018[37]
WOMAC-stiffness	100=best	**20.2**	16.2	23.8	3	Angst, 2018[37]
WOMAC-standing/walking	100=best	**8.1**	5.9	10.2	2	Angst, 2018[37]
WOMAC-total	100=best	**6.8**	1.6	16.8	8	Williams, 2012; Hmamouchi, 2012[38 42]
SF-36-bodily pain	100=best	**9.3**	8.2	10.4	2	Angst, 2018[37]
SF-36-physical function	100=best	**4.0**	3.8	4.2	2	Angst, 2018[37]
SF-36-role-physical	100=best	**8.5**	7.5	9.5	2	Angst, 2018[37]
SF-36-vitality	100=best	**4.5**	4.16	4.9	2	Angst, 2018[37]
SF-36-social function	100=best	**4.8**	2.6	7.0	2	Angst, 2018[37]
SF-36-mental health	100=best	**4.1**	2.9	5.2	2	Angst, 2018[37]
SF-36-general health	100=best	**6.6**	6	7.2	2	Angst, 2018[37]
Estimates based on low-quality studies						
DAP	100=worst	**24**			1	Klokker, 2016[36]
ICOAP-pain	100=worst	**7.8**			1	Harris, 2013[35]
KOOS-PS	100=worst	**7.8**			1	Harris, 2013[35]
OKS-pain	100=best	**14.3**			1	Harris, 2013[35]
OKS-function	100=best	**9.5**			1	Harris, 2013[35]
OKS-summary	100=best	**6.4**			1	Harris, 2013[35]

All the scores are from 0 to 100.
The median estimates are shaded in blue.
Estimates based on low-quality studies are shaded in grey.
*Estimates were reported as a range.
.DAP, Dynamic Weight-Bearing Assessment of Pain; ICOAP, Intermittent and Constant Osteoarthritis Pain; KOOS, Knee injury and Osteoarthritis Outcome Score; KOOS-PS, Knee injury and Osteoarthritis Outcome Score Physical Function Short form; KOS, Knee Outcome Survey; LEFS, Lower Extremity Functional Scale; MID, minimal important difference; NR, not reported; OA, osteoarthritis; OKS, Oxford Knee Score; PROMIS, Patient-Reported Outcome Measurement Information SystemS; SF-36, 36-item Short Form health survey; WOMAC, Western Ontario and McMaster Universities Arthritis Index.

Table 4 Comparison of ROC method-based MID estimates with overall estimates

	ROC method-based estimates, median (range)	ROC method-based, number of estimates (study)	Estimates regardless of calculation method, median (range)	No of estimates—regardless of calculation method, (study)
KOOS-pain	**11.8 (4.0 to 18.0)**	8 (Mills, 2016[15]; Mostafaee, 2021[39])	11.8 (4.0 to 17.0)	13 (Mills, 2016[15]; Mostafaee, 2021[39])
KOOS-activities of daily living	**12 (−1.5 to 155)**	5 (Mills, 2016[15]; Mostafaee, 2021[39])	12.5 (−1.5 to 15.5)	7 (Mills, 2016[15]; Mostafaee, 2021[39])
KOOS-quality of life	**6.5 (3.0 to 12.5)**	6 (Mills, 2016[15]; Mostafaee, 2021[39])	6.5 (3.0 to 12.5)	7 (Mills, 2016[15]; Mostafaee, 2021[39])
KOOS-sports/recreation	**17.5**	1 (Mostafaee, 2021[39])	17.5	1 (Mostafaee, 2021[39])
KOOS-symptoms	**12.5**	1 (Mostafaee, 2021[39])	12.5	1 (Mostafaee, 2021[39])
KOS-ADL	**6.4 (2.2 to 10.6)**	6 (Williams, 2012[42])	6.4 (2.2 to 10.6)	6 (Williams, 2012[42])
LEFS	**6.9 (0.6 to 15.6)**	6 (Williams, 2012[42])	6.9 (0.6 to 15.6)	6 (Williams, 2012[42])
WOMAC-total	**7.8 (1.6 to 16.8)**	8 (Williams, 2012[42]; Hmamouchi, 2012[38])	7.8 (1.6 to 16.8)	8 (Williams, 2012[42]; Hmamouchi, 2012[38])

All the estimates are out of 100.
The median estimates are shaded in blue.
Shaded in green: Estimates of these MID values were based on the ROC method only.
.ADL, activity of daily living; KOOS, Knee injury and Osteoarthritis Outcome Score; KOS, Knee Outcome Survey; LEFS, Lower Extremity Functional Scale; MID, minimal important difference; ROC, receiver operating curve; WOMAC, Western Ontario and McMaster Universities Arthritis Index.

DISCUSSION

This systematic review provided estimates for MIC and MID of knee OA outcome tools after non-surgical interventions derived using anchor, consensus and distribution methods respectively. This review is unique in that it provides estimates for MIC, MID (based on high-quality studies) and MDC (from good-to-fair-quality studies) of knee OA outcome tools after non-surgical interventions. MDC was reported for a greater number of outcome measures (126) than for MIC (13) or MID (23). MID estimates based on the ROC method were similar to the overall median estimates, however, the majority of MID studies used the ROC method. Although we found that some MIC and MID appear to increase with follow-up time, this was not consistent.

The estimates for MIC and MID reported in this review are lower than those reported previously.[82–84] Previous reviews which included knee replacement interventions[82–84] produced higher estimates suggesting that knee replacement cohorts need more improvement to be satisfied. The MID values for WOMAC-pain and function in this review ranged from 7.1 to 21 and 11.3 to 14.9 (out of 100), respectively; compared with reviews of total knee replacements which reported values ranging from 4.0 to 47.9 and 1.8 to 33.0 (out of 100).[82–84] This disparity may be due to differences in disease severity which has been previously reported based on baseline pain score.[6 15 85] Therefore, our data are more applicable to patients and cohorts receiving non-surgical interventions. Furthermore, previous knee OA intervention studies have used MID estimates from studies with combined hip and knee OA.[86 87] Given that MID is sensitive to disease type,[6] our median estimates are likely to be more applicable to the knee OA population.

Some of the median estimates presented in this study suggest considerable heterogeneity. For example, the MID for WOMAC-pain was 8.7, but, the range extended from 7.1 to 21. These wide ranges are seen for other estimates including MIC for KOOS-pain, KOOS-QOL and LEFS. The median estimate was used because it is robust when data is skewed. However, the uncertainty which accompanies the wide ranges reported must be acknowledged.

MDC was reported for more outcome measures than MIC or MID. MDC is derived from data distribution only, unlike MIC and MID which are related to patients' perception.[17 22] Researchers may use MDC estimates as an option if MIC or MID are not reported. Yet, according to the results of this study and others, MDC can be larger or smaller than MID.[4 10] Hence, researchers using MDC estimates from single studies to establish a sample size may overestimate or underestimate the number of participants required for a given power.

The ROC estimates were similar to the synthesised MID estimates which used all calculation methods. Our synthesised estimates were based on a combination of both the mean change[81] and ROC methods.[80] However, the ROC estimates are reported to be more precise and can be applied to both individuals and groups and are recommended in clinical settings.[9 15] Moreover, the area under the curve of the ROC has the advantage of being able to interpret the level of confidence for the MID estimate from acceptable to outstanding discrimination between responders and non-responders.[17] Therefore, we recommend using our median ROC based MID estimates where possible.

Although we found that some MIC and MID appear to increase with follow-up time, this was not consistent.

Two previous studies[15 42] suggested that there may be an effect of time, due to changing of perceptions over time (response-shift), especially in patients with chronic conditions.[15 88] In addition, recall bias is affected by increased follow-up time and may also affect estimates.[6 27 89] Therefore, although the consistency of follow-up time must be considered, more data are required to determine the effect of follow-up time on MIC and MID.

One of the studies included in this review used the consensus method. They reported that MIC was 20% of the maximum score for pain, function and global assessment,[45] but, our anchor studies data suggest that MIC is highly variable (2.2–27.6 out of 100) depending on the outcome measurement. Therefore, the blanket application of 20% may not be suggested regardless of the tool used.

In this review, we considered only MIC, MID and MDC but there are other measures of clinical improvement. While the MIC and MID are used to assess meaningful clinical effects, recent reports have questioned the applicability of these concepts as they do not consider the costs, risks, benefits and inconvenience of the treatment. The smallest worthwhile effect (SWE) was developed using the benefit-harm trade-off method, described by Barrett *et al*.[90] The SWE is defined as the smallest amount of improvement which is identified by the patient as worthwhile when considering the improvement outweighing risks and inconvenience[91] and the estimates are always compared with natural recovery.[92] However, only one study has reported SWE for people undergoing total knee replacement.[93] Other studies have evaluated 'patients acceptable symptom state' (PASS) which is the symptom state that patients consider acceptable or when they feel 'well' after treatment.[84 94 95] PASS estimates for WOMAC function are reported to be between 31 and 34.4.[96] These values are much higher than our MIC median estimate of 17 (9.1–17.1). Although MIC and MID are still commonly used, the development of this field of research will enable value judgements as well as clinical judgements to be considered in the interpretation of clinical trials of interventions.

This systematic review should be considered in light of its limitations. The results of this review have been affected by heterogeneity of the included studies including: sample size, participant demographics, severity of knee OA, varied interventions, follow-up time and calculation methods. Median estimates were used because of the data skewness, but some of the ranges were wide, challenging the certainty of some of these estimates. The reader is encouraged to take the range of the data into account when interpreting the results. Previous evidence suggests that data from follow-up times of less than 1 month are more reliable.[27 97 98] However, we included all estimates regardless of follow-up time. Statistical analysis was not conducted to determine the effect of follow-up time due to limited available data, but this is an interesting area for further study. The grey literature was not searched for this review.

This review presents median estimates for MIC, MID and MDC of people with knee OA following non-surgical interventions. A subset of MID estimates based on the ROC method is reported and, where available, this estimate is recommended as the most precise for both individual and group analyses and clinical settings. MDC estimates are available for more outcome measures but are purely statistical and arguably less applicable. This review clarifies the current understanding of MIC, MID and MDC in the knee OA population. However, some estimates suggest considerable heterogeneity and require careful interpretation.

Author affiliations
[1]Faculty of Health, University of Canberra, Canberra, Australian Capital Territory, Australia
[2]Department of Physiotherapy, Faculty of Allied Health Sciences, General Sir John Kotelawala Defence University, Werahera, Sri Lanka
[3]Trauma and Orthopaedic Research Unit, Canberra Hospital, Canberra, Australian Capital Territory, Australia
[4]Research Institute for Sport and Exercise, University of Canberra, Canberra, Australian Capital Territory, Australia

Twitter M Denika C Silva @denika.silva and Jennie M Scarvell @KneeMotion

Contributors MDCS, DMP, AMF and JMS conceptualised and designed the study. MDCS ran the database search. MDCS and JMC did the screening, full-text review and data acquisition. MDCS, DMP, AMF and JMS were involved in data analysis and Interpretation of the data. MDCS wrote the first draft of the manuscript. MDCS is responsible for the overall content as the guarantor. All authors reviewed the draft, read and approved the final manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as online supplemental information.

ORCID iDs
M Denika C Silva http://orcid.org/0000-0001-8648-471X
Angela M Fearon http://orcid.org/0000-0001-5055-3074

REFERENCES

1 Harris JD, Brand JC, Cote MP, *et al*. Research pearls: the significance of statistics and perils of pooling. Part 1: clinical versus statistical significance. *Arthroscopy* 2017;33:1102–12.

2 Terwee CB, Roorda LD, Dekker J, *et al*. Mind the MIC: large variation among populations and methods. *Journal of Clinical Epidemiology* 2010;63:524–34.
3 de Vet HC, Terwee CB, Ostelo RW, *et al*. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes* 2006;4:54.
4 de Vet HCW, Terwee CB. The minimal detectable change should not replace the minimal important difference. *Journal of Clinical Epidemiology* 2010;63:804–5.
5 de Vet HCW, Beckerman H, Terwee CB, *et al*. Definition of clinical differences. *J Rheumatol* 2006;33:434; author reply 435.
6 Katz NP, Paillard FC, Ekman E. Determining the clinical importance of treatment benefits for interventions for painful orthopedic conditions. *J Orthop Surg Res* 2015;10:24.
7 Wright A, Hannon J, Hegedus EJ, *et al*. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). *J Man Manip Ther* 2012;20:160–6.
8 Jaeschke R, Singer J, Guyatt GH. Measurement of health status. ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407–15.
9 Wells G, Anderson J, Beaton D, *et al*. Minimal clinically important difference module: summary, recommendations, and research agenda. *J Rheumatol* 2001;28:452–4.
10 Beard DJ, Harris K, Dawson J, *et al*. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. *J Clin Epidemiol* 2015;68:73–9.
11 King MT, Dueck AC, Revicki DA. Can methods developed for interpreting group-level patient-reported outcome data be applied to individual patient management? *Med Care* 2019;57 Suppl 5 Suppl 1(Suppl 5 1):S38–45.
12 Clement ND, Bardgett M, Weir D, *et al*. What is the minimum clinically important difference for the WOMAC index after TKA? *Clin Orthop Relat Res* 2018;476:2005–14.
13 King MT. A point of minimal important difference (mid): a critique of terminology and methods. *Expert Review of Pharmacoeconomics & Outcomes Research* 2011;11:171–84.
14 Tubach F, Ravaud P, Baron G, *et al*. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. *Ann Rheum Dis* 2005;64:29–33.
15 Mills KAG, Naylor JM, Eyles JP, *et al*. Examining the minimal important difference of patient-reported outcome measures for individuals with knee osteoarthritis: a model using the knee injury and osteoarthritis outcome score. *J Rheumatol* 2016;43:395–404.
16 Make B. How can we assess outcomes of clinical trials: the MCID approach. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2007;4:191–4.
17 Copay AG, Subach BR, Glassman SD, *et al*. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J* 2007;7:541–6.
18 Carrasco-Labra A, Devji T, Qasim A, *et al*. Minimal important difference estimates for patient-reported outcomes: a systematic survey. *J Clin Epidemiol* 2021;133:S0895-4356(20)31203-8:61–71.:.
19 Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003;41:582–92.
20 Wyrwich KW. Minimal important difference thresholds and the standard error of measurement: is there a connection? *J Biopharm Stat* 2004;14:97–110.
21 Martin Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* 1986;327:307–10.
22 Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *J Clin Epidemiol* 2017;82:128–36.
23 Cross M, Smith E, Hoy D, *et al*. The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. *Ann Rheum Dis* 2014;73:1323–30.
24 Bellamy N, Kirwan J, Boers M, *et al*. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. consensus development at omeract III. *J Rheumatol* 1997;24:799–802.
25 Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;151:264.
26 Dawson J, Boller I, Doll H, *et al*. Minimally important change was estimated for the manchester–oxford foot questionnaire after foot/ankle surgery. *Journal of Clinical Epidemiology* 2014;67:697–705.
27 Devji T, Carrasco-Labra A, Qasim A, *et al*. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. *BMJ* 2020;369:m1714.
28 Hao Q, Devji T, Zeraatkar D, *et al*. Minimal important differences for improvement in shoulder condition patient-reported outcomes: a systematic review to inform a BMJ rapid recommendation. *BMJ Open* 2019;9:e028777.
29 CAS Programme. Casp (qualitative) checklist [online]. 2018.
30 Long HA, French DP, Brooks JM. Optimising the value of the critical appraisal skills programme (CASP) tool for quality appraisal in qualitative evidence synthesis. *Research Methods in Medicine & Health Sciences* 2020;1:31–42.
31 NHLBI. *N-NH Lung, and blood Institute, National Institute of health quality assessment tool for before-after (pre-post) studies with no control group*. 2013.
32 Haby MM, Chapman E, Clark R, *et al*. What are the best methodologies for rapid reviews of the research evidence for evidence-informed decision making in health policy and practice: a rapid review. *Health Res Policy Syst* 2016;14:83.
33 McCarthy CJ, Oldham JA. The reliability, validity and responsiveness of an aggregated locomotor function (ALF) score in patients with osteoarthritis of the knee. *Rheumatology (Oxford)* 2004;43:514–7.
34 Higgins JPT, Thomas J, Chandler J, *et al*. *Cochrane handbook for systematic reviews of interventions*. 23 September 2019.
35 Harris KK, Dawson J, Jones LD, *et al*. Extending the use of PROMs in the NHS -- using the Oxford knee score in patients undergoing non-operative management for knee osteoarthritis: a validation study. *BMJ Open* 2013;3:e003365.
36 Klokker L, Christensen R, Wæhrens EE, *et al*. Dynamic weight-bearing assessment of pain in knee osteoarthritis: construct validity, responsiveness, and interpretability in a research setting. *Health Qual Life Outcomes* 2016;14:91:91.:.
37 Angst F, Benz T, Lehmann S, *et al*. Multidimensional minimal clinically important differences in knee osteoarthritis after comprehensive rehabilitation: a prospective evaluation from the bad zurzach osteoarthritis study. *RMD Open* 2018;4:e000685e000685.
38 Hmamouchi I, Allali F, Tahiri L, *et al*. Clinically important improvement in the WOMAC and predictor factors for response to non-specific non-steroidal anti-inflammatory drugs in osteoarthritic patients: a prospective study. *BMC Res Notes* 2012;5:58.
39 Mostafaee N, Nourollahi F, Mostamand J, *et al*. Responsiveness and the minimal important change of knee injury and osteoarthritis outcome score in Persian patients with knee osteoarthritis following physiotherapy intervention. *Physiotherapy Theory and Practice* 2022;38:2185–94.
40 Ornetti P, Dougados M, Paternotte S, *et al*. Validation of a numerical rating scale to assess functional impairment in hip and knee osteoarthritis: comparison with the WOMAC function scale. *Ann Rheum Dis* 2011;70:740–6.
41 Singh JA, Luo R, Landon GC, *et al*. Reliability and clinically important improvement thresholds for osteoarthritis pain and function scales: a multicenter study. *J Rheumatol* 2014;41:509–15.
42 Williams VJ, Piva SR, Irrgang JJ, *et al*. Comparison of reliability and responsiveness of patient-reported clinical outcome measures in knee osteoarthritis rehabilitation. *J Orthop Sports Phys Ther* 2012;42:716–23.
43 Perrot S, Bertin P. "feeling better'' or "feeling well'' in usual care of hip and knee osteoarthritis pain: determination of cutoff points for patient acceptable symptom state (pass) and minimal clinically important improvement (MCII) at rest and on movement in a national multicenter cohort study of 2414 patients with painful osteoarthritis. *J Pain* 2013;154:248–56.
44 Lee AC, Driban JB, Price LL, *et al*. Responsiveness and minimally important differences for 4 patient-reported outcomes measurement information system short forms: physical function, pain interference, depression, and anxiety in knee osteoarthritis. *J Pain* 2017;18:1096–110.
45 Salottolo K, Stahl E. Minimal clinically important improvement response in patients with severe osteoarthritis of the knee: short report from a survey of clinicians. *J Orthop* 2018;15:424–5.
46 Alghadir A, Anwer S, Brismée J-M. The reliability and minimal detectable change of timed up and go test in individuals with grade 1-3 knee osteoarthritis. *BMC Musculoskelet Disord* 2015;16:174.
47 Alghadir A, Anwer S, Iqbal ZA, *et al*. Cross-Cultural adaptation, reliability and validity of the Arabic version of the reduced Western Ontario and McMaster universities osteoarthritis index in patients with knee osteoarthritis. *Disability and Rehabilitation* 2016;38:689–94.
48 Alghadir AH, Anwer S, Iqbal ZA. The psychometric properties of an Arabic numeric pain rating scale for measuring osteoarthritis knee pain. *Disabil Rehabil* 2016;38:2392–7.

49 Alghadir AH, Al-Eisa ES, Anwer S. Cross-Cultural adaptation and psychometric analysis of the Arabic version of the Oxford knee score in adult male with knee osteoarthritis. *BMC Musculoskelet Disord* 2017;18:190.
50 Alghadir AH, Anwer S, Iqbal A, *et al*. . JPR 2018; volume 11:851–6. 10.2147/JPR.S158847
51 Suwit A, Rungtiwa K, Nipaporn T. Reliability and validity of the osteoarthritis research Society international minimal core set of recommended performance-based tests of physical function in knee osteoarthritis in community-dwelling adults. *Malays J Med Sci* 2020;27:77–89.
52 Baert IAC, Lluch E, Struyf T, *et al*. Inter- and intrarater reliability of two proprioception tests using clinical applicable measurement tools in subjects with and without knee osteoarthritis. *Musculoskelet Sci Pract* 2018;35:S2468-7812(17)30178-9:105–9.:.
53 Brisson NM, Stratford PW, Maly MR. Relative and absolute test–retest reliabilities of biomechanical risk factors for knee osteoarthritis progression: benchmarks for meaningful change. *Osteoarthritis and Cartilage* 2018;26:220–6.
54 Callaghan MJ, McCarthy CJ, Oldham JA. The reliability of surface electromyography to assess quadriceps fatigue during multi joint tasks in healthy and painful knees. *J Electromyogr Kinesiol* 2009;19:172–80.
55 ChaudHary S, Suhail A. Test-Retest reliability and minimum detectable change of 2-minute walk test among individuals with knee osteoarthritis. *JCDR* 2021;15:4–7.
56 Hoglund LT, Folkins E, Pontiggia L, *et al*. The validity, reliability, measurement error, and minimum detectable change of the 30-second fast-paced walk test in persons with knee osteoarthritis: a novel test of short-distance walking ability. *ACR Open Rheumatol* 2019;1:279–86.
57 Iijima H, Shimoura K, Eguchi R, *et al*. Concurrent validity and measurement error of stair climb test in people with pre-radiographic to mild knee osteoarthritis. *Gait & Posture* 2019;68:335–9.
58 Jansen MP, Welsing PMJ, Vincken KL, *et al*. Performance of knee image digital analysis of radiographs of patients with end-stage knee osteoarthritis. *Osteoarthritis and Cartilage* 2021;29:1530–9.
59 Kean CO, Birmingham TB, Garland SJ, *et al*. Minimal detectable change in quadriceps strength and voluntary muscle activation in patients with knee osteoarthritis. *Arch Phys Med Rehabil* 2010;91:1447–51.
60 Klokker L, Christensen R, Osborne R, *et al*. Dynamic weight-bearing assessment of pain in knee osteoarthritis: a reliability and agreement study. *Qual Life Res* 2015;24:2985–92.
61 McCarthy CJ, Callaghan MJ, Oldham JA. The reliability of isometric strength and fatigue measures in patients with knee osteoarthritis. *Man Ther* 2008;13:159–64.
62 Monticone M, Sconza C, Portoghese I, *et al*. Cross-Cultural adaptation, reliability and validity of the Fremantle knee awareness questionnaire in Italian subjects with painful knee osteoarthritis. *Health Qual Life Outcomes* 2021;19:114.
63 Nalbant A, Unver B, Karatosun V. Test-Retest reliability of the L-test in patients with advanced knee osteoarthritis. *Physiother Theory Pract* 2022;38:2983–7.
64 Naylor JM, Hayen A, Davidson E, *et al*. Minimal detectable change for mobility and patient-reported tools in people with osteoarthritis awaiting arthroplasty. *BMC Musculoskelet Disord* 2014;15:235.
65 Parveen H, Noohu MM. Evaluation of psychometric properties of tinetti performance-oriented mobility assessment scale in subjects with knee osteoarthritis. *Hong Kong Physiother J* 2017;36:25–32.
66 Peter WF, de Vet HCW, Terwee CB. Reliability of the animated activity questionnaire for assessing activity limitations of patients with hip and knee osteoarthritis. *Musculoskeletal Care* 2018;16:363–9.
67 Piva SR, Fitzgerald GK, Irrgang JJ, *et al*. Get up and go test in patients with knee osteoarthritis. *Archives of Physical Medicine and Rehabilitation* 2004;85:284–9.
68 Srimurugan Pratheep N, Madeleine P, Arendt-Nielsen L. Relative and absolute test-retest reliabilities of pressure pain threshold in patients with knee osteoarthritis. *Scand J Pain* 2018;18:229–36.
69 Takacs J, Carpenter MG, Garland SJ, *et al*. Test re-test reliability of centre of pressure measures during standing balance in individuals with knee osteoarthritis. *Gait Posture* 2014;40:S0966-6362(14)00111-8:270–3.:.
70 Tevald MA, Murray A, Luc BA, *et al*. Hip abductor strength in people with knee osteoarthritis: a cross-sectional study of reliability and association with function. *The Knee* 2016;23:57–62.
71 Tse CTF, Charlton JM, Lam J, *et al*. Smartphone inclinometry is a valid and reliable tool for measuring frontal plane tibial alignment in healthy and osteoarthritic knees. *Phys Ther* 2021;101:pzab091.
72 Turcot K, Aissaoui R, Boivin K, *et al*. Test-Retest reliability and minimal clinical change determination for 3-dimensional tibial and femoral accelerations during treadmill walking in knee osteoarthritis patients. *Arch Phys Med Rehabil* 2008;89:732–7.
73 van der Straaten R, Wesseling M, Jonkers I, *et al*. Discriminant validity of 3D joint kinematics and centre of mass displacement measured by inertial sensor technology during the unipodal stance task. *PLoS One* 2020;15:e0232513.
74 Yuruk ZO, Aytar A, Eker L, *et al*. Acceptability, Reliability and validity of the Turkish version of the de Morton mobility index in elderly patients with knee osteoarthritis. *Turk Geriatri Derg* 2014:17.
75 Ravaud P, Giraudeau B, Auleley GR, *et al*. Assessing smallest detectable change over time in continuous structural outcome measures: application to radiological change in knee osteoarthritis. *J Clin Epidemiol* 1999;52:1225–30.
76 Kanko LE, Birmingham TB, Bryant DM, *et al*. The StAR excursion balance test is a reliable and valid outcome measure for patients with knee osteoarthritis. *Osteoarthritis and Cartilage* 2019;27:580–5.
77 Mutlu EK, Ozdincler AR. Reliability and responsiveness of algometry for measuring pressure pain threshold in patients with knee osteoarthritis. *J Phys Ther Sci* 2015;27:1961–5.
78 Hunter DJ, Conaghan PG, Peterfy CG, *et al*. Responsiveness, effect size, and smallest detectable difference of magnetic resonance imaging in knee osteoarthritis. *Osteoarthritis and Cartilage* 2006;14:112–5.
79 Motyl JM, Driban JB, McAdams E, *et al*. Test-Retest reliability and sensitivity of the 20-meter walk test among patients with knee osteoarthritis. *BMC Musculoskelet Disord* 2013;14:166.
80 Perkins NJ, Schisterman EF. The inconsistency of "optimal'' cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol* 2006;163:670–5.
81 Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic improvements. An illustration in rheumatology. *Arch Intern Med* 1993;153:1337–42.
82 Çelik D, Çoban Ö, Kılıçoğlu Ö. Minimal clinically important difference of commonly used hip-, knee-, foot-, and ankle-specific questionnaires: a systematic review. *J Clin Epidemiol* 2019;113:44–57.
83 Maredupaka S, Meshram P, Chatte M, *et al*. Minimal clinically important difference of commonly used patient-reported outcome measures in total knee arthroplasty: review of terminologies, methods and proposed values. *Knee Surg Relat Res* 2020;32:19.
84 MacKay C, Clements N, Wong R, *et al*. 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. *Osteoarthritis and Cartilage* 2019;27:1408–19.
85 Escobar A, Riddle DL. Concordance between important change and acceptable symptom state following knee arthroplasty: the role of baseline scores. *Osteoarthritis and Cartilage* 2014;22:1107–10.
86 Angst F, Aeschlimann A, Steiner W, *et al*. Responsiveness of the WOMAC osteoarthritis index as compared with the SF-36 in patients with osteoarthritis of the legs undergoing a comprehensive rehabilitation intervention. *Ann Rheum Dis* 2001;60:834–40.
87 Bellamy N. Pain assessment in osteoarthritis: experience with the WOMAC osteoarthritis index. *Semin Arthritis Rheum* 1989;18(4 Suppl 2):14–7.
88 Schwartz CE, Andresen EM, Nosek MA, *et al*. Response shift theory: important implications for measuring quality of life in people with disability. *Arch Phys Med Rehabil* 2007;88:529–36.
89 Guyatt GH, Norman GR, Juniper EF, *et al*. A critical look at transition ratings. *J Clin Epidemiol* 2002;55:900–8.
90 Barrett B, Brown D, Mundt M, *et al*. Sufficiently important difference: expanding the framework of clinical significance. *Med Decis Making* 2005;25:250–61.
91 Ferreira M. Research note: the smallest worthwhile effect of a health intervention. *J Physiother* 2018;64:272–4.
92 Ferreira ML, Herbert RD, Ferreira PH, *et al*. A critical review of methods used to determine the smallest worthwhile effect of interventions for low back pain. *J Clin Epidemiol* 2012;65:253–61.
93 Wofford N. *Smallest worthwhile effect values for pain and function after a total knee replacement health related sciences PhD: Virginia Commonwealth University*. 2020.
94 Bellamy N, Hochberg M, Tubach F, *et al*. Development of multinational definitions of minimal clinically important improvement and patient acceptable symptomatic state in osteoarthritis. *Arthritis Care & Research* 2015;67:972–80. 10.1002/acr.22538 Available: http://doi.wiley.com/10.1002/acr.v67.7
95 Kvien TK, Heiberg T, Hagen KB. Minimal clinically important improvement/difference (MCII/MCID) and patient acceptable symptom state (pass): what do these concepts mean? *Annals of the Rheumatic Diseases* 2007;66(Supplement 3):iii40–1.

96 Tubach F, Ravaud P, Baron G, *et al*. Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state. *Annals of the Rheumatic Diseases* 2005;64:34–7.
97 Turner D, Schünemann HJ, Griffith LE, *et al*. Using the entire cohort in the receiver operating characteristic analysis maximizes precision of the minimal important difference. *Journal of Clinical Epidemiology* 2009;62:374–9.
98 Lurie F, Kistner RL. In prospective study using specific quality of life & outcomes response-venous (SQOR-V) questionnaire the recall bias had the same magnitude as the minimally important difference. *Qual Life Res* 2011;20:1589–93.