

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<u>http://bmjopen.bmj.com</u>).

If you have any questions on BMJ Open's open peer review process please email <u>info.bmjopen@bmj.com</u>

BMJ Open

Minimal clinically important change and difference for knee osteoarthritis outcome measurement tools after nonsurgical interventions: a systematic review

Journal:	BMJ Open
Manuscript ID	bmjopen-2022-063026
Article Type:	Original research
Date Submitted by the Author:	30-Mar-2022
Complete List of Authors:	Silva, M Denika; University of Canberra Perriman, Diana; College of Medicine and Health Sciences, Australian National University; University of Canberra Fearon, A; University of Canberra Couldrick, Jacqui; University of Canberra Scarvell, Jennie; University of Canberra; Trauma and Orthopaedic Research Unit, Canberra Hospital
Keywords:	STATISTICS & RESEARCH METHODS, Knee < ORTHOPAEDIC & TRAUMA SURGERY, Musculoskeletal disorders < ORTHOPAEDIC & TRAUMA SURGERY





I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our <u>licence</u>.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which <u>Creative Commons</u> licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

terez oni

Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies



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

M Denika C Silva^{1,2,5}, Diana M Perriman^{1,2,3}, Angela M Fearon^{1,2,4}, Jacqui M Couldrick^{1,2}, Jennie M Scarvell^{1,2}

1. Faculty of Health, University of Canberra, Australia

2. Trauma and Orthopaedic Research Unit, Canberra Hospital, Canberra, Australia

3. College of Medicine and Health Sciences, Australian National University, Canberra, Australia

4. Research Institute for Sport and Exercise, University of Canberra, Australia

5. Department of Physiotherapy, General Sir John Kotelawala Defence University, Werahera, Sri Lanka

Corresponding author: M Denika C Silva (ORCID ID:0000-00018648-471X)Address: Faculty of Health, University of Canberra, Bruce ACT 2617 AustraliaTel.: +61 (0) 433857153Email: denika.silva@canberra.edu.au

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

Objectives: To systematically review and provide estimates of the minimal clinically

ABSTRACT

7 important change (MCIC) and difference (MCID) for outcome tools in people with knee
8 osteoarthritis (OA) after non-surgical interventions.

Design: A systematic review.

Data sources: MEDLINE, CINAHL, Web of Science, Scopus, and Cochrane 11 databases were searched up to September 21, 2021.

Study selection and data synthesis: We included studies that calculated MCIC and MCID using any calculation method including anchor, consensus and distribution methods, for any knee OA outcome tool after non-surgical interventions. We used quality assessment tools appropriate to the studies' methods to screen out low-quality studies. Values were pooled and expressed as median and range, for each method. Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Results: Forty-eight studies were eligible (anchor-k=12, consensus-k=1 and distribution-k=35). MCIC 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 five high-guality anchor studies. MCID values for 23 tools including KOOS-pain, ADL, QOL and WOMAC-function, stiffness and total were estimated using six high-quality anchor studies. One moderate quality consensus study reported MCIC for pain, function and global assessment. Minimum detectable

BMJ Open

change values (MDC, from distribution method estimates) for 126 tools including
 KOOS-QOL and WOMAC-total were estimated using 38 good-to-fair-quality studies.

Conclusion: Median MCIC, MCID and MDC estimates were reported for outcome tools in people with knee OA after non-surgical interventions. Clinicians and researchers may consider these MCIC and MCID values when designing or interpreting clinical trials.

31 PROSPERO registration number: CRD42020153962

Keywords: minimal clinically important change; minimal clinically important
 difference; minimum detectable change; knee osteoarthritis

1 ว		
2 3 4	34	Strengths and limitations of this study
5 6 7	35	• We estimated MCIC (within-group), MCID (between-groups), and minimum
7 8 9	36	detectable change values using anchor, consensus, or distribution methods
10 11	37	papers respectively.
12 13	38	• This systematic review included a defined population of people with knee OA,
14 15 16	39	after non-surgical interventions.
17 18	40	High-quality anchor studies were used to contribute to MCIC and MCID
19 20	41	estimates were assessed using a credibility tool specially designed to evaluate
21 22 23	42	anchor method papers.
24 25	43	Consensus and distribution methods papers were evaluated using quality
26 27	44	assessment tools suited to each method.
28 29 30	45	• Median estimates were used to reflect the pooled data due to data skewness.
31 32		
33 34		
35 36 37		
38 39		
40 41		
42 43 44		
45 46		
47 48		
49 50 51		
51 52 53		
54 55		
56 57		
58 59		
60		

INTRODUCTION

The efficacy of therapeutic interventions is commonly evaluated using statistical significance regardless of clinical importance¹. To understand whether differences in outcome measures after treatment are clinically important, 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 clinically important change (MCIC) and difference (MCID). There are numerous outcome measures for knee osteoarthritis (OA) and many estimates of MCIC and MCID. However, these estimates can arise from different methodologies leading to variability, confusion and misinterpretation²⁻⁵. 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 MCIC and MCID for knee OA outcome measurement tools in people with knee OA after non-surgical interventions.

> MCIC and MCID 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 MCIC is defined as the change in a clinical outcome measure within a single group or an individual over time. In contrast, MCID describes the difference between independent groups (for example, control versus treatment groups) or between individuals^{4,10-12}. However, the terminology of MCIC and MCID is used inconsistently¹³. The concept was first described by Jaeschke, who studied patients' perceptions of pre- and post-intervention beneficial change⁸. This concept later included both improvement¹⁴ and worsening¹⁵. Other commonly used terms are "minimal important change" for MCIC and "minimal important difference" for MCID.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open

Three methods are used to estimate MCIC and MCID: anchor, consensus and distribution^{6,16}. For the anchor method, MCIC or MCID values are estimated using patients' responses where values of the outcome measure are related to an externally validated scale or 'anchor'¹⁷. The "global rating of change" is commonly used as the anchor in this method. 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 achieved⁶. In the distribution method, values are estimated statistically, based on the variance of the outcome data using half the standard deviation (SD)¹⁸, one standard error of measurement (SEM)¹⁹ or minimum detectable change (MDC) which is based on SEM^{6,20}. The anchor method is widely considered to be the most valid because it is associated with a definition of minimal importance and is based on patient perception^{16,21}.

Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Knee OA is a common cause of pain and disability²². 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 studies²³. MCIC and MCID 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 bodies²⁻ 4.

BMJ Open

The primary objective of this systematic review was to estimate MCIC and MCID for knee OA outcome measurement tools based on pooled estimates from high-quality anchor studies only. Secondary objectives were to determine MCIC and MCID estimates based on consensus and distribution methods including MDC outcomes.

METHODS

This systematic review was designed and reported according to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement²⁴. The registered PROSPERO protocol (registration number: was on CRD42020153962).

Literature search

Five databases (MEDLINE, CINAHL, Web of Science, Scopus and Cochrane) were searched from each database's respective inception up to September 21, 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 version 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 (DS and JC) and conflicts were resolved by a third reviewer (JS). Included studies incorporated any design that calculated MCIC and MCID 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 method³. We considered studies with MCIC or MCID 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, e.g. hip OA or other knee pathologies. Studies of MCIC, MCID and MDC were included even if they used a different terminology e.g. minimal important change for MCIC, minimal important difference for MCID and smallest detectable change or difference or minimal detectable difference for MDC.

For consistency, we defined the MCIC as the pre-post change of one group i.e. threshold for those who responded that they had minimally improved. The MCID was

BMJ Open

defined as the difference (pre-post-change) between two groups i.e. "minimally improved" and "stayed the same" groups using the anchor response as defined in the previous studies^{10,12,25}. The MDC is the minimum change above the measurement error based upon a given level of confidence^{6,20}. MDC values for a 90 or 95 confidence interval are labelled as MDC90 or MDC95.

152 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 (2020)²⁶ which was specially designed to assess the anchor studies. This instrument includes five core criteria (the anchor is rated by the patient, the anchor is interpretable and relevant to the patient, the MCIC or MCID estimate is precise, the correlation between the anchor and the outcome measure reported by the patient is satisfactory, and the authors select a threshold on the anchor that reflects a small but important difference). 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" guality if not²⁷. Consensus studies were assessed using the Critical Appraisal Screening Program (CASP)-qualitative tool²⁸ 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 credibility²⁹. Distribution studies were evaluated using the National Heart, Lung and Blood Institute, National Institute of Health (NHLBI, NIH) quality assessment tool for before-after (pre-post) studies with no control group³⁰ and ratings included "good", "fair" or "poor" based on reliability and credibility³⁰. The quality assessment of included studies was performed

BMJ Open

2 3	170	by one reviewer (DS) and a random sample of 20% had an independent second review							
4 5 6	171	(AF or JC) to improve the accuracy ³¹ .							
7 8	172								
9 10 11	173	Data extraction and analysis							
12 13	174	We extracted study characteristics including sample size, participant							
14 15	175	demographics, details of the intervention, follow-up time, outcome measurement tools,							
16 17 18	176	calculation method and actual estimate reported based on the method (MCIC or MCID							
19 20	177	or MDC). Additionally, we extracted the details of the anchor used in each study.							
21 22	178								
23 24 25	179	We extracted reported MCIC, MCID and MDC values from each study. We							
26 27	180	normalised the values to a 0 to 100 scale. If a study reported MDC as a percentage of							
28 29	181	the grand mean (MDC divided by grand mean percentage) ³² , we converted the data							
30 31 32	182	into MDC90 or 95 for the pooled synthesis.							
33 34	183								
35 36	184	The median and range (minimum and maximum) of MCIC, MCID and MDC							
37 38 39	185	were calculated using multiple estimates from the included studies arising from							
40 41	186	different non-surgical interventions, calculation methods, time points, and anchors.							
42 43	187	Mean values were not calculated due to skewness of distributions ³³ . We excluded low-							
44 45	188	quality anchor studies from the median MCIC and MCID synthesis. Furthermore, we							
40 47 48	189	conducted a sub-analysis to determine median MCID based on the ROC method							
49 50	190	where available because the ROC estimates are considered to be more precise than							
51 52	191	mean-change estimates and recommended at both individual and group level							
53 54 55	192	analyses, and in clinical settings ^{9,15} . Though we planned to conduct a sub-analysis to							
56 57	193	determine the effect of follow-up time on MCIC and MCID, we were unable to do							
58 59 60	194	reliable rate estimates because of a limited number of studies. Therefore, we plotted							

Page 12 of 59

Erasmushogeschool Erasmushogeschool similar technologies.

3 4	195	the values against time including only studies where the outcome measures were
5 6 7	196	assessed at three-time points or more.
7 8 9	197	
10 11	198	Patient and public involvement
12 13	199	This is a systematic review. Patients or the public were not involved in this study.
14 15	200	
16 17 18	201	RESULTS
19 20	202	
21 22	203	Study selection
23 24 25	204	The search yielded 2376 studies and after duplicates were removed, 1059
25 26 27	205	records were screened. Two hundred and seventeen studies were screened in full-
28 29	206	text review resulting in 48 eligible studies (k=48) (Figure 1). No further studies were
30 31	207	identified after checking the reference lists of included studies.
32 33 34	208	
35 36	209	Included studies calculated MCIC and MCID by anchor method (k=12), by
37 38	210	consensus method (k=1) and MDC by distribution method (k=35).
39 40 41	211	
42 43	212	The methodological quality of included studies
44 45	213	Most anchor studies (k=10) were of high quality and two were of low quality ^{34,35}
46 47	214	(Supplement 1). The quality of the consensus study (k=1) was moderate
48 49 50	215	(Supplement 2). The quality of distribution studies ranged from good (k=11) to fair
51 52	216	(k=24) (Supplement 3).
53 54	217	
55 56 57 58 59 60	218	Study characteristics of included studies

Page 13 of 59

BMJ Open

All the anchor studies were observational prospective cohort studies^{14,15,34,36-42} and two of them were nested within randomised controlled trials^{35,43} (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/m² respectively. The interventions used in these studies were rehabilitation, exercise, physiotherapy and non-steroidal anti-inflammatory drugs. The follow-up time ranged from seven days to one year. Eleven studies used global rating of change as the external anchor while one study⁴³ used multiple anchors. Most anchor studies (k=6)^{35-37,38,41,43} reported MCID values only, four studies^{14,39,40,42} reported MCIC values only and two studies^{15,34,} reported both MCIC and MCID. Four of these studies^{34,36,40,41} also reported MDC values. Moreover, studies used different anchor questions and group classifications when calculating MCIC and MCID. For example, one MCIC study³⁴ considered the minimal improved response group as "my knee has got better" and another study³⁹ considered the response group as both "good" and "excellent" improvement groups (Supplement 4).

Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies

Page 14 of 59

BMJ	Open
-----	------

				BMJ Open BMJ	
1 2				'ight, ii	
³ 234 4 235	Table 1: Characteristics	of included studies		nclud	
5	Study	Number of participants; Age, years, Mean (SD)	Intervention (Follow-up)	Outcome tool 10 30 26 or c	Calculation method (Measure extracted)
3	Angst, 2018 ³⁶	190; 66.1 (10.2)	Rehabilitation (3 months)	WOMAC-pain, function, functional standing/walkieg and stiffness; SF-36-bodily pain, physical functioning physical, vitality, social functioning, mental and ஒற்று தி	Anchor (MCID); Distribution (MDC95)
10 11	Harris, 2013 ³⁴	134; 59 (11)	Non-operative (3 months)	OKS- pain, function, summary; ICOAP; KOOS	Anchor (MCIC, MCID); Distribution (MDC90)
12 13	Hmamouchi, 2012 ³⁷	173; 57.1 (10.1)	NSAID (6 weeks)	WOMAC-total 6 til 3 til 5 D	Anchor (MCID)
14 15	Klokker, 2016 ³⁵ Lee, 2017 ⁴³	41; NR 165; 61	exercises (12 weeks) Thai Chi or Physical therapy (12 weeks)	DAP کې کې PROMIS-Short Forms- physical function, pain inter depression, anxiety	Anchor (MCID) Anchor (MCID)
16 17	Mills, 2016 ¹⁵	272; NR	Non-surgical (26,52 weeks)	KOOS-pain, activities of daily living, quality of	Anchor (MCIC and MCID)
18 19	Mostafaee, 2021 ³⁸	142; 58.3 (7.6)	Physiotherapy (4 weeks)	KOOS-pain, symptoms, activities of daily living, sparts and recreation, and quality of life	Anchor (MCID)
20	Ornetti, 2011 ³⁹	881; 67.1 (10.9)	NSAID (4 weeks)	WOMAC-function; Patient-reported functional disability of an NRS; Physician reported functional disability on a NSS	Anchor (MCIC)
22 22 23	Perrot, 2013 ⁴² Singh, 2014 ⁴⁰	1606; 66.9 (9.0) 137; NR	Usual care (7 days) Conservative (2 weeks)	Pain on a 0-10 NRS at rest and movement ICOAP-pain, constant pain, intermittent pain; KOGS-P3; KOOS-quality of life	Anchor (MCIC) Anchor (MCIC); Distribution (MDC90)
24 25	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 (MCIC)
26	Williams, 2012 ⁴¹	159; NR	Exercise (2, 6, 12 months)	KOS-activities of daily living subscale; LEFS; WOMAC-total	Anchor (MCID); Distribution (MDC90, MDC95)
28	Salottolo, 2018 ⁴⁴ Alghadir, 2015 ⁴⁵	27 (clinicians); NR 65; 54.9 (9.9)	NR NR	Pain, Function and Global assessment 📷 🥄 Timed Up and Go test	Consensus (MCIC) Distribution (MDC95)
29 30	Alghadir, 2016 a ⁴⁶ Alghadir, 2016 b ⁴⁷	121; 52.9 (9.3) 121; 54.0 (9.3)	NR (48 hours) NR (48 hours)	WOMAC- Arabic version-pain, function, totat Arabic version 지 말 이 아이지 않는 것 같이 아이지 않는 것 않는 것 같이 아이지 않는 것 같이 아이지 않는 것 같이 아이지 않는 않는 것 같이 아이지 않는 것 않는 않는 것 않는	Distribution (MDC95) Distribution (MDC95)
31 32	Alghadir, 2017 ⁴⁸ Alghadir, 2018 ⁴⁹	97; 58 (11.5) 121; 52.9 (12.5)	NR (1 week) NR (24 hours)	OKS- Arabic version حلاج بع VAS for pain; NRS for pain; Verbal Rating Scale ہے، VAS for pain; NRS for pain; Verbal Rating Scale	Distribution (MDC95) Distribution (MDC95)
33	Baert, 2018 ⁵¹	8 (12 knees); 68 (11)	NR (3 seconds)	Knee joint position sense test using analogue incline ther; Knee force sense test using a handheld dynamometer	Distribution (MDC95)
34 35 36	Brisson, 2018 ⁵²	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 dynamorecter; Load frequency using a triaxial accelerometer; BMI	Distribution (MDC95)
37 38	Callghan, 2009 ⁵³ Hoglund, 2019 ⁵⁵ Hunter, 2006 ⁷⁷	55; 64 (14) 20; 58.3 (8.05) 72 [:] 58 9 (8.6)	NR (24- 72 hours) NR (2-7 days) Novel ther (6 months)	Quadriceps fatigue using surface electromyography 30-second fast-paced walk test	Distribution (MDC95) Distribution (MDC95) Distribution (MDC95)
39 40 41	ljima, 2019 ⁵⁶	59; 59.1 (6.1)	NR (1 month)	Stopwatch-based 11- Step stair climb test	Distribution (MDC90, MDC95)

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

e 15 of 59				BMJ Open BMJ Open	
	Jansen, 2021 ⁵⁷	103; NR	NR (1 month)	Knee image: bone density, joint space width, osteophyles, eminence beight and joint angle	Distribution (MDC95)
	Kanko, 2019 ⁷⁵ Kean, 2010 ⁵⁸	74; 57.7 (8.8) 20; 53.6 (9.1)	Exercises (1 week) NR (same day)	Quadriceps isokinetic strength, isometric strength and percent	Distribution (MDC95) Distribution (MDC90)
	Klokker, 2015 ⁵⁹ McCarthy, 2004 ³²	20; 64 (6.6) 15; 65.1 (11.3)	NR (1 week) Exercises (1 week)	of voluntary activation using a dynamometer g DAP Aggregated locomotor function score; Eight-meter walk the;	Distribution (MDC95) Distribution (MDC95)
	McCarthy, 2008 ⁶⁰	55; 64.9 (9.7)	NR (1 week)	Stair ascent and descent time; Transferring time Maximum isometric strength; Vastus medialis oblique, astus	Distribution (MDC95)
	Monticone, 2021 ⁶¹	102; 69.1 (9.0)	NR (10 days)	medialis oblique, Vastus lateralis, Rectus femoris fatigue sione Fremantle Knee Awareness Questionnaire, Italian are	Distribution (MDC95)
	Motyl, 2013 ⁷ ° Mutlu, 2015 ⁷⁶	15; 61.0 (7.8) 73 (141 knees);	days) Physiotherapy (4	20-meter walk test د م الحق ع ع 8 م Pressure pain threshold using a handheld pressure aggingeter	Distribution (MDC95)
	Nalbant, 2021 ⁶² Navlor 2014 ⁶³	56.2 (6.7) 25; 62.3 (9.8) 75: 67 6 (9 4)	weeks) NR (same day) NR (10 days)	L-test	Distribution (MDC95) Distribution (MDC95)
	Parveen, 2017 ⁶⁴	25; 50.5 (6.3)	NR (7 days)	KOOS-pain, symptoms, activities of daily living, quarty of life Tinetti Performance-Oriented Mobility Assessment scale-total,	Distribution (MDC95)
	Peter, 2018 ⁶⁵ Piva, 2004 ⁶⁶	135; NR 25; 57 (8)	NR (1 week) NR (same day)	Animated activity questionnaire	Distribution (MDC95) Distribution (MDC95)
	Pratheep, 2018 ⁶⁷ Ravaud, 1999 ⁷⁴ Subail and	20; 70.3 (5.8) 30; 65.4 (7.7) 82: 60 4 (6 7)	NR (2 weeks) NR (2 weeks) NR (48 hours)	Pressure pain threshold using algometer 5 Joint space width 2-minute walk test	Distribution (MDC95) Distribution (MDC95) Distribution (MDC95)
	Chaudha, 2021 ⁵⁴⁶ Suwit, 2020 ⁵⁰²	55; 69.0 (11)	NR (1 week)	30-seconds chair stand; 40-meter fast-paced test; 9-step climb	Distribution (MDC90)
	Takacs, 2014°° Tevald, 2016 ⁶⁹ Tse, 2021 ⁷⁰	20; 64.1 (7.9) 25; 61 (9) 38; 65.1 (7.0)	NR (14 days) NR (7 days) NR (same day)	Single-leg standing balance test = Hip abductor strength using a hand-held dynameter Frontal plane tibial alignment using inclinometer	Distribution (MDC95) Distribution (MDC95) Distribution (MDC95)
	Turcot, 2008 ⁷¹ Van der Straaten, 2020 ⁷²	25; 63.9 (7.6) 19; 65.1 (5.2)	NR (8 days) NR (20 days)	3D linear accelerations of the tibia and femur during waking Range of motion (trunk abduction-adduction, pelvis adduction- adduction, hip flexion-extension) and Center of mask	Distribution (MDC95) Distribution (MDC95)
226	Yuruk, 2014 ⁷³	40; NR	NR (One day)	displacement وي كي De Motion Mobility Index (Turkish version) ه	Distribution (MDC90, MDC95)
230 237 238 239 240 241 242	MCIC: Minimal Clinical Im Minimum Detectable Char Short Form health survey Function Short form, KOC Measurement Information adduction moment, 3-dim	portant Change, MCII nge based on 95% co , OKS: Oxford Knee S DS: Knee injury and Os o System, NRS: numer ensional: 3D, KFM: kr	D: Minimal Clinically Impo nfidence interval, NR: No core, ICOAP: Intermitten steoarthritis Outcome Sco ric rating scale, VAS: Visio nee flexion moment, MRI:	ortant Difference, MDC90: Minimum Detectable Change based on ot Reported, WOMAC: Western Ontario and McMaster Universitie t and Constant Osteoarthritis Pain, KOOS-PS: Knee injury and Os ore, DAP: Dynamic weight-bearing Assessment of Pain, PR MIS: ual Analog Scale, KOS: Knee Outcome Survey, LEFS: Low Extra Magnetic Resonance Imaging	90% confidence interval, MDC95: es Arthritis Index, SF-36: 36-item teoarthritis Outcome Score Physical Patient-Reported Outcome emity Functional Scale, KAM: knee
				14 T	
			For peer review only - h	http://bmjopen.bmj.com/site/about/guidelines.xhtml	

Page

36

BMJ Open

Page 16 of 59

Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

> The consensus study $(k=1)^{44}$ (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 MCIC values for pain, function and global assessment for severe knee OA. However, participants were not asked to consider time duration nor the interventions. MCIC was termed "minimal clinically important improvement" in this study.

Most distribution studies (k=30) were test-retest observational studies⁴⁵⁻⁷⁴ assessing the reliability of the outcome tool, and five used datasets from interventional cohort studies^{75,76} and randomised controlled trials^{32,77,78} (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/m², respectively. Studies estimated MDC90 and MDC95. The follow-up time ranged from the same day to one year.

The MCIC estimates derived using the anchor method

The median MCIC for 13 tools (with subscales) were calculated based on five high-quality anchor studies^{14,15,39,40,42} using 23 estimates (**Table 2**). These estimates were based on different underlying calculations, follow-up time and anchor questions. Methods for calculating MCIC included: mean change (pre and post mean change of the minimally improved group)^{8,15,40}, 75th centile value of the mean change of the group^{39,42} and 75th centile value adjusted with the baseline score, age, and disease duration¹⁴. Most studies included one follow-up time (range: 7 days to 4 weeks), but one study¹⁵ reported MCIC at two time-points (26 and 52 weeks). One anchor question was used in most studies, however, two studies^{15,39} reported different MCIC values

based on two different anchor questions (general health status and functional state).

All data were pooled for the median estimates regardless of differences in calculation

methods, follow-up time, and anchor questions.

Table 2: Minimal Clinical Important Change (MCIC) values of knee osteoarthritis outcome tools derived using the anchor method

Outcome tools	Score	Median	Minimum	Maximum	Number of ostimatos	Study
ICOAP-pain	100=worst	18.5			1	Singh, 2014 ⁴⁰
ICOAP-constant pain	100=worst	18.7			1	Singh, 2014 ⁴⁰
ICOAP-intermittent pain	100=worst	18.4			1	Singh, 2014 ⁴⁰
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 ¹⁵ ; Singh, 2014 ⁴⁰
KOOS-PS	100=worst	2.2			1	Singh, 2014 ⁴⁰
Pain on NRS at rest	100=worst	10.0	10	10	2	Perrot, 201342
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 ³⁹
Patient's global assessment of disease activity on VAS	100=worst	18.3 mm	4.		1	Tubach, 2005 ¹⁴
Physician-reported functional disability on an NRS	100=worst	25.3	25	25.5	2	Ornetti, 2011 ³⁹
WOMAC-function	100=worst	17.0	9.1	17.1	3	Tubach, 2005 ¹⁴ ; Ornetti, 2011 ³⁹
	Estim	ates based	l on low-qua	lity studies		
ICOAP-pain	100=worst	13.4			1	Harris, 2013 ³⁴
KOOS-PS	100=worst	12.0			1	Harris, 2013 ³⁴
OKS-pain	100=best	17.3			1	Harris, 2013 ³⁴
OKS-function	100=best	10.6			1	Harris, 2013 ³⁴
OKS-summary	100=best	7.1			1	Harris, 2013 ³⁴

Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

MCIC: Minimal Clinical Important Change, 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, NRS: numeric rating scale, VAS: Visual Analog Scale, WOMAC: Western Ontario and McMaster Universities Arthritis Index, OKS: Oxford Knee Score

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

286 The MCID estimates derived using the anchor method

The median MCID for 23 tools were calculated based on six high-quality anchor studies^{15,36-38,41,43,45} using 83 estimates (**Table 3**). These estimates were based on different underlying calculations, follow-up time, and anchor questions. Methods for calculating MCID included: ROC method⁷⁹ only (k=3)^{37,38,41}, mean change (Redilmier and Lorig) method only (pre-post mean change difference between two groups)⁸⁰ $(k=1)^{36}$, both ROC and mean change methods $(k=1)^{15}$, and mean change in T-scores with multiple anchors (k=1)⁴³. Most studies (k=4) included one follow-up (range: 4 to 12 weeks), but two studies included MCID at multiple time-points e.g. 2, 6 and 12 months^{15,41}. One anchor question was used in most studies, however, one study used multiple anchors⁴³. All data were pooled for the median estimates regardless of differences in calculation methods, follow-up time, and anchor questions.

review only

	Outcome tools	Score	Median	Minimum	Maximum	Number of estimates	Study
	KOOS-pain	100=best	11.8	4	17	13	Mills, 2010 Mostafaee, 2
	KOOS-activities of daily living	100=best	2.5	-1.5	15.5	7	Mills, 2010 Mostafaee, 2
	KOOS-quality of life	100=best	6.5	3	12.5	7	Mills, 2010 Mostafaee, 2
	KOOS-sports/recreation	100=best	17.5			1	Mostafaee, 2
	KOOS-symptoms	100=best	12.5			1	Mostafaee, 2
	KOS-activities of daily living	100=best	6.4	2.2	10.6	6	Williams, 20
	LEFS	100=best	6.9	0.6	15.6	6	Williams, 20
	PROMIS Short Forms- physical function*	100=wor st		4.5	5.3	NR	Lee, 201
	PROMIS Short Forms-	100=wor		4.2	4.2	NR	Lee, 201
	pain interference PROMIS Short Forms- depression*	st 100=wor st		6	6.2	NR	Lee, 201
	PROMIS Short Forms- anxiety*	100=wor st		4.5	6.6	NR	Lee, 201
	WOMAC-pain	100=best	8.7	7.1	21	3	Angst, 201
	WOMAC-function	100=best	14.5	11.3	14.9	3	Angst, 201
	WOMAC-stiffness	100=best	20.2	16.2	23.8	3	Angst, 201
	WOMAC- standing/walking	100=best	8.1	5.9	10.2	2	Angst, 201
	WOMAC-total	100=best	6.8	1.6	16.8	8	Williams, 20 Hmamouchi,
	SF-36-bodily pain	100=best	9.3	8.2	10.4	2	Angst, 201
	SF-36-physical function	100=best	4.0	3.8	4.2	2	Angst, 201
	SF-36-role-physical	100=best	8.5	7.5	9.5	2	Angst, 201
	SF-36-vitality	100=best	4.5	4.16	4.9	2	Angst, 201
	SF-36-social function	100=best	4.8	2.6	7.0	2	Angst, 201
	SF-36-mental health	100=best	4.1	2.9	5.2	2	Angst, 201
	SF-36-general health	100=best	6.6	6	7.2	2	Angst, 201
		Esti	imates bas	ed on low-qu	ality studies		
	DAP	100=wor	24			1	Klokker, 20
	ICOAP-pain	100=wor	7.8			1	Harris, 201
	KOOS-PS	100=wor st	7.8			1	Harris, 201
	OKS-pain	100=best	14.3			1	Harris, 201
	OKS-function	100=best	9.5			1	Harris, 20 ²
	OKS-summary	100=best	6.4			1	Harris, 20 ²
300 301 302 303 304	MCID: Minimal Clinically Impo Score, KOS: Knee Outcome S Outcome Measurement Informand McMaster Universities An Intermittent and constant oster	prtant Differer Survey, LEFS mation Syster thritis Index, coarthritis pair	nce, OA: os 1 Lower Ex 10 SF36: 36- DAP: Dyna 10 KOOS-PS	teoarthritis, K tremity Functi item Short Fo mic weight-be S: Knee injury	OOS: Knee inj onal Scale, PF rm health surv earing Assessn and Osteoarth	ury and Osteoa OMIS: Patient ey, WOMAC: Monent of Pain, IC paritis Outcome S	arthritis Outcom Reported Western Ontario OAP: Score Physical

*Estimates were reported as a range The median estimates are shaded in blue 309 310 59

60 Estimates based on low-quality studies are shaded in grey

2
4
5
5
7
/ 0
ð
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58

315

MCID estimates based on the ROC method were reported and compared with 311 pooled MCID estimates for all methods. Four^{15,37,38,41} of six studies (67%) used the 312 ROC method. The ROC estimates were the same as the overall pooled estimates in 313 most cases (Table 4). 314

Table 4: Comparison of receiver operating curve (ROC) method based Minimal Clinical Important Difference 316 317 (MCID) estimates with overall pooled estimates 318

Outcome tools	ROC method-based estimates, Median (range)	ROC method- based, Number of estimates (study)	Pooled estimates regardless of calculation method, Median (range)	Number of estimates- Pooled regardless of calculation method, (study)
KOOS-pain	11.8 (4.0 to 18.0)	8 (Mills, 2016 ¹⁵ ;	11.8 (4.0 to 17.0)	13 (Mills, 2016 ¹⁵ ;
		Mostafaee, 202138)	,	Mostafaee, 2021 ³⁸)
KOOS-activities of	12 (-1.5 to 155)	5	12.5	7
daily living		(Mills, 2016 ¹⁵ ; Mostafaee, 2021 ³⁸)	(-1.5 to 15.5)	(Mills, 2016 ¹⁵ ; Mostafaee, 2021 ³⁸)
KOOS-quality of life	6.5 (3.0 to 12.5)	6	6.5	7
		(Mills, 2016 ¹⁵ ; Mostafaee, 2021)	(3.0 to 12.5)	(Mills, 2016 ¹⁵ ; Mostafaee, 2021 ³⁸)
KOOS- sports/recreation	17.5	1 (Mostafaee, 2021 ³⁸)	17.5	1Mostafaee, 2021 ³⁸)
COOS-symptoms	12.5	1 (Mostafaee, 2021 ³⁸)	12.5	1 (Mostafaee, 2021 ³⁸)
(OS-ADL	6.4 (2.2 to 10.6)	6	6.4	6
	. ,	(Williams, 201241)	(2.2 to 10.6)	(Williams, 201241)
LEFS	6.9 (0.6 to 15.6)	6	6.9	6
		(Williams, 2012 ⁴¹)	(0.6 to 15.6)	(Williams, 2012 ⁴¹)
WOMAC-total	7.8 (1.6 to 16.8)	8	7.8	8
		(Williams, 2012 ⁴³ ;	(1.6 to 16.8)	(Williams, 2012 ⁴¹ ;
		Hmamouchi, 2012 ³⁷)		Hmamouchi, 2012 ³⁷ ;)

ROC: Receiver Operating Curve, MCID: Minimal Clinically Important Difference, KOOS: Knee injury and Osteoarthritis Outcome Score, KOS: Knee Outcome Survey, LEFS: Lower Extremity Functional Scale, WOMAC: Western Ontario and McMaster Universities Arthritis Index

All the estimates are out of 100.

325 326 The median estimates are shaded in blue

327 Shaded in green: pooled estimates of these MCID values were based on the ROC method only

319 320

321

322

323 324

328

The effect of follow-up time on MCIC and MCID 329

There were insufficient data to establish reliable rate estimates for the effect of 330

time. The MCIC of Knee injury and Osteoarthritis Outcome Score (KOOS)-pain and 331

59 KOOS-quality of life (QOL) and, the MCID of KOOS-pain, KOOS-QOL, Knee Outcome 332 60

Page 21 of 59

BMJ Open

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 MCIC of KOOS-QOL and, MCID of KOOS-QOL, LEFS and KOS-ADL appeared to increase with increasing follow-up time. However, MCIC of KOOS-pain and MCID of WOMAC-total appeared to reduce with follow-up time and KOOS-pain remained constant (Supplement 5). MCIC values derived using the consensus method One consensus study⁴⁴ reported that MCIC 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 (Supplement 6). These estimates were based on different calculation methods and follow-up times. Four included studies reported MDC90 values only^{40,50,58,76} and 29 studies MDC95 values only. Five studies^{36,41,56,63,73} reported both the MDC90 and MDC95 values. Most studies (k=37) reported unadjusted MDC, while one study reported both the adjusted and unadjusted estimates³⁶. Six studies separately reported inter/intra-rater MDC values^{45,51,55,59,66,72}. Furthermore, three studies reported distinct values for two patient groups in each study, for example, the placebo group and the treatment group⁷⁷, the most painful and the least painful groups⁶⁹, and the groups that reported moderate improvement ("great deal better") and MCID improvement ("somewhat better")⁴⁰. Most studies assessed the index (worst) knee, but one study based the estimate on all diseased knees (12 knees in 8 patients)⁵¹.

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)⁴¹. All data were pooled for the median estimates regardless of differences in calculation methods and follow-up time.

BMJ Open

DISCUSSION

This systematic review provided estimates for MCIC and MCID 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 pooled estimates for MCIC, MCID (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 MCIC (13) or MCID (23). MCID estimates based on the ROC method were similar to the overall pooled median estimates, however, the majority of MCID studies used the ROC method. Although we found that some MCIC and MCID appear to increase with follow-up time, this was not consistent.

The estimates for MCIC and MCID reported in this review are lower than those reported previously⁸¹⁻⁸³. Previous reviews which included knee replacement interventions⁸¹⁻⁸³ produced higher estimates suggesting that knee replacement cohorts need more improvement to be satisfied. The MCID 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 to reviews of total knee replacements which reported values ranging from 4.0 to 47.9 and 1.8 to 33.0 (out of 100)⁸¹⁻⁸³. This disparity may be due to Page 23 of 59

BMJ Open

ices
core
ng n
hav
hat I
pplic
MDC
froi
s' pe
) are
large
studi
ants
The
est
ls ⁷⁹ .
to k
ver, t
et the
inatio
ur m

differen in disease severity which has been previously reported based on baseline pain so ^{6,15,84}. Therefore, our data is more applicable to patients and cohorts on-surgical interventions. Furthermore, previous knee OA intervention receivir studies e used MCID estimates from studies with combined hip and knee OA⁸⁵⁻⁸⁶. MCID is sensitive to disease type⁶, our median estimates are likely to be Given t more a able to the knee OA population.

was reported for more outcome measures than MCIC or MCID. MDC is Ν derived m data distribution only, unlike MCIC and MCID which are related to erception^{17,21}. Researchers may use MDC estimates as an option if MCIC patients or MCIE e not reported. Yet, according to the results of this study and others, MDC can be er or smaller than MCID^{4,10}. Hence, researchers using MDC estimates from single s es to establish a sample size may over-or under-estimate the number of particip required for a given power.

Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

ROC estimates were similar to the pooled MCID estimates. Our pooled imates are based on a combination of the mean change⁸⁰ and ROC median There is strong support for the exclusive use of ROC because it can be method applied both individuals and groups and is recommended in clinical settings^{9,15}. Moreov the area under the curve of the ROC has the advantage of being able to interpre e level of confidence for the MCID estimate from acceptable to outstanding on between responders and non-responders¹⁷. Therefore, we recommend discrimi using o edian ROC based MCID estimates where possible.

- 56 406

BMJ Open

> Although we found that some MCIC and MCID appear to increase with followup time, this was not consistent. Two previous studies^{15,41} suggested that there may be an effect of time, due to changing of perceptions over time (response-shift), especially in patients with chronic conditions^{15,87}. Additionally, recall bias is affected by increased follow-up time and may also affect estimates^{6,26,88}. Therefore, although the consistency of follow-up time must be considered, more data is required to determine the effect of follow up time on MCIC and MCID.

One of the studies included in this review used the consensus method. They
reported that MCIC was 20% of the maximum score for pain, function and global
assessment⁴⁴, but, our anchor studies data suggest that MCIC is highly variable (2.2
to 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 MCIC, MCID and MDC but there are other measures of clinical improvement. While the MCIC and MCID are used to assess meaningful clinical effects, recent reports have guestioned 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 in 2005⁸⁹. 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 inconvenience⁹⁰ and the estimates are always compared to natural recovery⁹¹. However, only one study has reported SWE for people undergoing total knee replacement⁹². Other studies have evaluated "patients acceptable symptom state" (PASS) which is the symptom state

Page 25 of 59

BMJ Open

that patients consider acceptable or when they feel "well" after treatment^{83,93,94}. PASS estimates for WOMAC function are reported to be between 31 and 34.4⁹⁵. These values are much higher than our MCIC median estimate of 17 (9.1 to 17.1). Although MCIC and MCID 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. Median estimates were used to reflect the pooled data due to data skewness, but some of the ranges were wide, challenging the certainty of some of these estimates. Previous evidence suggests that data from follow-up times of less than one month are more reliable^{26,96,97}. 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 certainty of evidence using Grading of Recommendations Assessment, Development and Evaluation (GRADE) was not attempted due to the heterogeneity of the data. The grey literature was not searched for this review.

Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

This review presents median estimates for MCIC, MCID and MDC for people with knee OA following non-surgical interventions. A subset of MCID estimates is reported based on the ROC method. The pooled ROC data are recommended 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 MCIC, MCID and MDC in the knee OA population.

1 ว		
2 3 4	457	
5 6 7	458	Author contributions
7 8 9	459	Conception and design: M Denika C Silva, Diana M Perriman, Angela M Fearon,
10 11 12 13 14 15 16	460	Jennie M Scarvell
	461	Data acquisition: M Denika C Silva, Jacqui M Couldrick
	462	Data-analysis and Interpretation of the data: M Denika C Silva, Diana M Perriman,
17 18	463	Angela M Fearon, Jennie M Scarvell
19 20	464	Drafting of the article revising it critically for important intellectual content: M Denika C
21 22 23	465	Silva, Diana M Perriman, Angela M Fearon, Jacqui M Couldrick Jennie M Scarvell
24 25	466	Final approval of the version: M Denika C Silva, Diana M Perriman, Angela M Fearon,
26 27 28 29 30	467	Jacqui M Couldrick, Jennie M Scarvell
	468	Role of the funding source
31 32	469	This research received no specific grant from any funding agency in the public,
33 34 25	470	commercial or not-for-profit sectors.
35 36 37	471	Conflict of interest
38 39	472	All the authors declare that there are no conflicts of interest in this work.
40 41	473	Ethics approval
42 43	474	This study is a systematic review and does not involve human participants. Therefore,
44 45 46	475	ethics approval is not applicable.
47 48 49 50 51 52 53	476	Data Sharing
	477	No additional data are available.
	478	
54 55 56 57 58 59 60	479	REFERENCES

BMJ Open

2			
3 4	480	1.	Harris JD, Brand JC, Cote MP, Faucett SC, Dhawan A. Research Pearls: The
5 6	481		significance of statistics and perils of pooling. part 1: clinical versus statistical
7 8	482		significance. Arthroscopy 2017; 33: 1102-1112.
9 10 11	483	2.	Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et
12 13	484		al. Mind the MIC: large variation among populations and methods. J Clin
14 15	485		Epidemiol 2010; 63: 524-534.
16 17 18	486	3.	de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal
19 20	487		changes in health status questionnaires: distinction between minimally
21 22	488		detectable change and minimally important change. Health Qual Life Outcomes
23 24	489		2006; 4: 54-54.
25 26 27	490	4.	de Vet HC, Terwee CB. The minimal detectable change should not replace the
28 29	491		minimal important difference. J Clin Epidemiol 2010; 63: 804-805; author reply
30 31	492		806.
32 33 34	493	5.	de Vet HCW, Beckerman H, Terwee CB, Terluin B, Bouter LM. Definition of
35 36	494		clinical differences. J Rheumatol 2006; 33: 434.
37 38	495	6.	Katz NP, Paillard FC, Ekman E. Determining the clinical importance of
39 40	496		treatment benefits for interventions for painful orthopedic conditions. J Orthop
41 42 43	497		Surg Res 2015; 10: 24.
44 45	498	7.	Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: a closer
46 47	499		look at the minimal clinically important difference (MCID). J Man Manip Ther
48 49 50	500		2012; 20: 160-166.
50 51 52	501	8.	Jaeschke R, Singer J, Guyatt GH. Measurement of health status: Ascertaining
53 54	502		the minimal clinically important difference. Control Clin Trials 1989; 10: 407-
55 56	503		415.
57 58 59			
60			

BMJ Open

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: 452-454. 10. Beard DJ, Harris K, Dawson J, Doll H, Murray DW, Carr AJ, et al. Meaningful

- beard DJ, Harris K, Dawson J, Doll H, Murray DW, Carr AJ, et al. Meaningful
 changes for the Oxford hip and knee scores after joint replacement surgery. J
 Clin Epidemiol 2015; 68: 73-79.
- ¹⁷ 510 11. King MT, Dueck AC, Revicki DA. Can methods developed for interpreting
 ¹⁹ 511 group-level patient-reported outcome data be applied to individual patient
 ²¹ 22 512 management? Med Care 2019; 57.
- 513 12. Clement ND, Bardgett M, Weir D, Holland J, Gerrand C, Deehan DJ. What is
 514 the minimum clinically important difference for the WOMAC index after TKA?
 515 Clin Orthop Relat Res 2018; 476: 2005-2014.
- King MT. A point of minimal important difference (MID): a critique of terminology
 and methods. Expert Rev Pharmacoecon Outcomes Res 2014; 11: 171-184.
- 14. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, 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.
- Mills KA, Naylor JM, Eyles JP, Roos EM, Hunter DJ. Examining the minimal 15. 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.
- 53
 54 526 16. Make B. How can we assess outcomes of clinical trials: The MCID approach.
 55
 56 527 COPD: J Chronic Obstr Pulm Dis 2007; 4: 191-194.

Page 29 of 59

BMJ Open

1 2			
3 4 5 6 7 8 9	528	17.	Copay AG, Subach BR, Glassman SD, Polly DW, Jr., Schuler TC.
	529		Understanding the minimum clinically important difference: a review of
	530		concepts and methods. Spine J 2007; 7: 541-546.
10 11	531	18.	Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related
12 13	532		quality of life: the remarkable universality of half a standard deviation. Med Care
14 15	533		2003; 41: 582-592.
16 17 18	534	19.	Wyrwich KW. Minimal important difference thresholds and the standard error of
19 20	535		measurement: is there a connection? J. Biopharm. Stat. 2004; 14: 97-110.
21 22	536	20.	Martin Bland J, Altman D. Statistical methods for assessing agreement between
23 24 25 26 27 28 29 30 31 32 33 34 35 36	537		two methods of clinical measurement. Lancet 1986; 327: 307-310.
	538	21.	Angst F, Aeschlimann A, Angst J. The minimal clinically important difference
	539		raised the significance of outcome effects above the statistical level, with
	540		methodological implications for future studies. J Clin Epidemiol 2017; 82: 128-
	541		136.
	542	22.	Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, et al. The global
37 38	543		burden of hip and knee osteoarthritis: estimates from the global burden of
39 40 41	544		disease 2010 study. Ann Rheum Dis 2014; 73: 1323.
42 43	545	23.	Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, et al.
44 45	546		Recommendations for a core set of outcome measures for future phase III
46 47	547		clinical trials in knee, hip, and hand osteoarthritis. Consensus development at
48 49 50	548		OMERACT III. J Rheumatol 1997; 24: 799-802.
51 52	549	24.	Moher D, Liberati A, Tetzlaff J, Altman DG, Group atP. Preferred Reporting
53 54	550		Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement.
55 56	551		Ann Intern Med 2009; 151: 264-269.
58 59 60			

552 25. Dawson J, Boller I, Doll H, et al. Minimally important change was estimated for
 553 the Manchester-Oxford Foot Questionnaire after foot/ankle surgery. J Clin
 554 Epidemiol 2014;67(6):697-705.

- Devji T, Carrasco-Labra A, Qasim A, Phillips M, Johnston BC, Devasenapathy 26. N, 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.
- Hao Q, Devji T, Zeraatkar D, Wang Y, Qasim A, Siemieniuk RAC, et al. Minimal 27. important differences for improvement in shoulder condition patient-reported outcomes: a systematic review to inform a BMJ Rapid Recommendation. BMJ Open 2019; 9: e028777.
- ²⁸₂₉ 563 28. Programme CAS. CASP (Qualitative) Checklist. [online] 2018.
- 564 29. Long HA, French DP, Brooks JM. Optimising the value of the critical appraisal
 565 skills programme (CASP) tool for quality appraisal in qualitative evidence
 566 synthesis. Research Methods in Medicine & Health Sciences 2020; 1: 31-42.
- NHLBI N-NH, Lung, and Blood Institute, National Institute of Health quality 30. assessment tool for before-after (pre-post) studies with no control group. 2013. Haby MM, Chapman E, Clark R, Barreto J, Reveiz L, Lavis JN. What are the 31. 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.
- 51
52
5357332.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: 514-517.

BMJ Open

576	33.	Higgins JPT CJ, Cumpston M, Li T, Page MJ, Welch VA. Cochrane handbook
577		for systematic reviews of interventions. version 6.0 edition2019.
578	34.	Harris KK, Dawson J, Jones LD, Beard DJ, Price AJ. Extending the use of
579		PROMs in the NHSusing the Oxford Knee Score in patients undergoing non-
580		operative management for knee osteoarthritis: a validation study. BMJ Open
581		2013; 3: e003365.
582	35.	Klokker L, Christensen R, Waehrens EE, Bandak E, Bartholdy C, Bliddal H, et
583		al. Dynamic weight-bearing assessment of pain in knee osteoarthritis: construct
584		validity, responsiveness, and interpretability in a research setting. Health Qual.
585		Life Outcomes 2016; 14: 91.
586	36.	Angst F, Benz T, Lehmann S, Aeschlimann A, Angst J. Multidimensional
587		minimal clinically important differences in knee osteoarthritis after
588		comprehensive rehabilitation: a prospective evaluation from the Bad Zurzach
589		Osteoarthritis Study. RMD Open 2018; 4: e000685.
590	37.	Hmamouchi I, Allali F, Tahiri L, et al. Clinically important improvement in the
591		WOMAC and predictor factors for response to non-specific non-steroidal anti-
592		inflammatory drugs in osteoarthritic patients: a prospective study. BMC Res
593		Notes 2012;5:58
594	38.	Mostafaee N, Nourollahi F, Mostamand J, Negahban H. Responsiveness and
595		the minimal important change of Knee injury and Osteoarthritis Outcome Score
596		in Persian patients with knee osteoarthritis following physiotherapy intervention.
597		Physiother Theory Pract 2021: 1-10.
598	39.	Ornetti P, Dougados M, Paternotte S, Logeart I, Gossec L. Validation of a
599		numerical rating scale to assess functional impairment in hip and knee
	576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 591 592 593 591 592 593 594 595 594 595 596 595	57633.57734.57834.57934.57935.58135.58335.58436.58536.58737.59337.59137.59238.59538.59539.59739.59839.59739.59839.

BMJ Open

600osteoarthritis: comparison with the WOMAC function scale. Ann Rheum Dis6012011; 70: 740-746.

- 602 40. Singh JA, Luo R, Landon GC, Suarez-Almazor M. Reliability and clinically
 603 important improvement thresholds for osteoarthritis pain and function scales: a
 604 multicenter study. J Rheumatol 2014; 41: 509-515.
- Williams VJ, Piva SR, Irrgang JJ, Crossley C, Fitzgerald GK. Comparison of 41. reliability and responsiveness of patient-reported clinical outcome measures in knee osteoarthritis rehabilitation. J Orthop Sports Phys Ther 2012; 42: 716-723. 42. 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: 248-256.
- Lee AC, Driban JB, Price LL, Harvey WF, Rodday AM, Wang 43. C. 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: 1096-1110.
- 618
 618
 618
 618
 618
 619
 619
 619 patients with severe osteoarthritis of the knee: Short report from a survey of
 620
 620
 610
 611
 612
 612
 613
 614
 614
 615
 616
 617
 618
 618
 619
 619
 610
 610
 610
 611
 611
 612
 612
 613
 614
 614
 615
 614
 615
 616
 617
 618
 618
 618
 619
 610
 610
 610
 611
 611
 612
 612
 612
 612
 612
 614
 614
 615
 616
 617
 618
 618
 619
 610
 610
 610
 610
 610
 611
 611
 612
 612
 612
 612
 612
 614
 614
 614
 614
 614
 615
 614
 614
 615
 614
 614
 615
 614
 614
 615
 614
 614
 615
 614
 614
 614
 614
 614
 614
 614
 614
 614
 614
 614
 614
 614
 614
 614
 614
 614
 614
 614
 614
 614
 614
 614
 614
 <li
- 621 45. Alghadir A, Anwer S, Brismée J-M. The reliability and minimal detectable
 622 change of Timed Up and Go test in individuals with grade 1-3 knee
 623 osteoarthritis. BMC Musculoskelet Disord 2015; 16: 174-174.

Page 33 of 59

1 2 BMJ Open

3 4	624	46.	Alghadir A, Anwer S, Iqbal ZA, Alsanawi HA. Cross-cultural adaptation,
5 6	625		reliability and validity of the Arabic version of the reduced Western Ontario and
7 8 9 10 11 12 13	626		McMaster Universities Osteoarthritis index in patients with knee osteoarthritis.
	627		Disabil Rehabil 2016; 38: 689-694.
	628	47.	Alghadir AH, Anwer S, Iqbal ZA. The psychometric properties of an Arabic
14 15	629		numeric pain rating scale for measuring osteoarthritis knee pain. Disabil
16 17 18 19 20 21 22 23 24 25	630		Rehabil 2016; 38: 2392-2397.
	631	48.	Alghadir AH, Al-Eisa ES, Anwer S. Cross-cultural adaptation and psychometric
	632		analysis of the Arabic version of the oxford knee score in adult male with knee
	633		osteoarthritis. BMC Musculoskelet Disord 2017; 18: 190.
23 26 27	634	49.	Alghadir AH, Anwer S, Iqbal A, Iqbal ZA. Test-retest reliability, validity, and
28 29	635		minimum detectable change of visual analog, numerical rating, and verbal
30 31 22	636		rating scales for measurement of osteoarthritic knee pain. J Pain Res 2018; 11:
32 33 34	637		851-856.
35 36	638	50.	Suwit A, Rungtiwa K, Nipaporn T. Reliability and validity of the Osteoarthritis
37 38	639		Research Society International minimal core set of recommended
39 40 41	640		performance-based tests of physical function in knee osteoarthritis in
42 43	641		community-dwelling adults. Malays J Med Sci 2020; 27: 77-89.
44 45	642	51.	Baert IAC, Lluch E, Struyf T, Peeters G, Van Oosterwijck S, Tuynman J, et al.
46 47 48	643		Inter- and intrarater reliability of two proprioception tests using clinical
49 50	644		applicable measurement tools in subjects with and without knee osteoarthritis.
51 52	645		Musculoskelet Sci Pract 2018; 35: 105-109.
53 54	646	52.	Brisson NM, Stratford PW, Maly MR. Relative and absolute test-retest
55 56 57	647		reliabilities of biomechanical risk factors for knee osteoarthritis progression:
58 59 60	648		benchmarks for meaningful change. Osteoarthr Cartil 2018; 26: 220-226.

BMJ Open

3 4	649	53.	Callaghan MJ, McCarthy CJ, Oldham JA. The reliability of surface
5 6	650		electromyography to assess quadriceps fatigue during multi joint tasks in
7 8 0	651		healthy and painful knees. J Electromyogr Kinesiol 2009; 19: 172-180.
9 10 11	652	54.	Suhail A, Chaudhary S. Test-retest reliability and minimum detectable change
12 13	653		of 2-minute walk test among individuals with knee osteoarthritis. J Clin
14 15	654		Diagnostic Res 2021; 15: 4-7.
16 17 18	655	55.	Hoglund LT, Folkins E, Pontiggia L, Knapp MW. The validity, reliability,
19 20	656		measurement error, and minimum detectable change of the 30-second fast-
21 22	657		paced walk test in persons with knee osteoarthritis: a novel test of short-
23 24 25	658		distance walking ability. ACR Open Rheumatol 2019; 1: 279-286.
25 26 27	659	56.	lijima H, Shimoura K, Eguchi R, Aoyama T, Takahashi M. Concurrent validity
28 29	660		and measurement error of stair climb test in people with pre-radiographic to
30 31	661		mild knee osteoarthritis. Gait Posture 2019; 68: 335-339.
32 33 34	662	57.	Jansen MP, Welsing PMJ, Vincken KL, Mastbergen SC. Performance of knee
35 36	663		image digital analysis of radiographs of patients with end-stage knee
37 38	664		osteoarthritis. Osteoarthr Cartil 2021; 29: 1530-1539.
39 40 41	665	58.	Kean CO, Birmingham TB, Garland SJ, Bryant DM, Giffin JR. Minimal
42 43	666		detectable change in quadriceps strength and voluntary muscle activation in
44 45	667		patients with knee osteoarthritis. Arch Phys Med Rehabil 2010; 91: 1447-1451.
46 47 48	668	59.	Klokker L, Christensen R, Osborne R, Ginnerup E, Waehrens EE, Bliddal H, et
49 50	669		al. Dynamic weight-bearing assessment of pain in knee osteoarthritis: a
51 52	670		reliability and agreement study. Qual Life Res 2015; 24: 2985-2992.
53 54	671	60.	McCarthy CJ, Callaghan MJ, Oldham JA. The reliability of isometric strength
56 57	672		and fatigue measures in patients with knee osteoarthritis. Man Ther 2008; 13:
58 59 60	673		159-164.
Page 35 of 59

1

BMJ Open

2 3	674	61	Monticone M Sconza C Portoghese I Nishigami T Wand BM Sorrentino G
4 5	675	011	at al. Cross sultural adaptation, reliability and validity of the Fromantia Know
6 7	675		et al. Cross-cultural adaptation, reliability and validity of the Fremantie Knee
8 9	676		Awareness Questionnaire in Italian subjects with painful knee osteoarthritis.
10 11	677		Health Qual Life Outcomes 2021; 19: 114.
12 13	678	62.	Nalbant A, Unver B, Karatosun V. Test-retest reliability of the L-test in patients
14 15	679		with advanced knee osteoarthritis. Physiother Theory Pract 2021: 1-5.
16 17 18	680	63.	Naylor J M, Hayen A, Davidson E, Hackett D, Harris I A, Kamalasena G, et al.
19 20	681		Minimal detectable change for mobility and patient-reported tools in people with
21 22	682		osteoarthritis awaiting arthroplasty. BMC Musculoskelet Disord 2014: 15, 235.
23 24 25	683	64.	Parveen H, Noohu MM. Evaluation of psychometric properties of Tinetti
23 26 27	684		performance-oriented mobility assessment scale in subjects with knee
28 29	685		osteoarthritis. Hong Kong Physiother J 2017; 36: 25-32.
30 31	686	65.	Peter WF, de Vet HCW, Terwee CB. Reliability of the Animated Activity
32 33 34	687		Questionnaire for assessing activity limitations of patients with hip and knee
35 36	688		osteoarthritis. Musculoskeletal Care 2018; 16: 363-369.
37 38	689	66.	Piva SR, Fitzgerald GK, Irrgang JJ, Bouzubar F, Starz TW. Get up and go test
39 40 41	690		in patients with knee osteoarthritis. Arch Phys Med Rehabil. 2004;85(2):284-
41 42 43	691		289.
44 45	692	67.	Srimurugan Pratheep N, Madeleine P, Arendt-Nielsen L. Relative and absolute
46 47	693		test-retest reliabilities of pressure pain threshold in patients with knee
48 49 50	694		osteoarthritis. Scand J Pain 2018; 18: 229-236.
51 52	695	68.	Takacs J, Carpenter MG, Garland SJ, Hunt MA. Test re-test reliability of centre
53 54	696		of pressure measures during standing balance in individuals with knee
55 56 57	697		osteoarthritis. Gait Posture 2014; 40: 270-273.
57 58			
60			

Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

BMJ Open

698
69. Tevald MA, Murray A, Luc BA, Lai K, Sohn D, Pietrosimone B. Hip abductor
699 strength in people with knee osteoarthritis: A cross-sectional study of reliability
700 and association with function. Knee 2016; 23: 57-62.

701 70. Tse CTF, Charlton JM, Lam J, Ho J, Bears J, Serek A, et al. Smartphone
 702 inclinometry is a valid and reliable tool for measuring frontal plane tibial
 703 alignment in healthy and osteoarthritic knees. Phys Ther 2021; 101: pzab091.

- 71. Turcot K, Aissaoui R, Boivin K, Hagemeister N, Pelletier M, de Guise JA. 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: 732-737.
- 72. van der Straaten R, Wesseling M, Jonkers I, Vanwanseele B, Bruijnes A, Malcorps J, 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.
- 712 73. Yuruk Z O, Aytar A, Tüzün E H, Eker L, Yüksel I, Morton N A. Acceptability,
 713 reliability and validity of the turkish version of the De morton mobility index in
 714 elderly patients with knee osteoarthritis. Turk Geriatri Dergisi 2014; 17.
- Ravaud P, Giraudeau B, Auleley GR, Edouard-Noël R, Dougados M, Chastang 74. C. Assessing smallest detectable change over time in continuous structural outcome measures: application to radiological change in knee osteoarthritis. J Clin Epidemiol 1999; 52: 1225-1230.
- 719 75. Kanko LE, Birmingham TB, Bryant DM, Gillanders K, Lemmon K, Chan R, et
 720 al. The star excursion balance test is a reliable and valid outcome measure for
 721 patients with knee osteoarthritis. Osteoarthritis Cartilage 2019; 27: 580-585.

1 2 BMJ Open

3 4	722	76.	Mutlu EK, Ozdincler AR. Reliability and responsiveness of algometry for
5 6	723		measuring pressure pain threshold in patients with knee osteoarthritis. J Phys
7 8 0	724		Ther Sci. 2015;27(6):1961-1965.
9 10 11	725	77.	Hunter DJ, Conaghan PG, Peterfy CG, Bloch D, Guermazi A, Woodworth T, et
12 13	726		al. Responsiveness, effect size, and smallest detectable difference of Magnetic
14 15	727		Resonance Imaging in knee osteoarthritis. Osteoarthritis Cartilage 2006; 14
16 17 18	728		Suppl A: A112-115.
19 20	729	78.	Motyl JM, Driban JB, McAdams E, Price LL, McAlindon TE. Test-retest
21 22	730		reliability and sensitivity of the 20-meter walk test among patients with knee
23 24 25	731		osteoarthritis. BMC Musculoskelet Disord 2013; 14: 166
26 27	732	79.	Perkins NJ, Schisterman EF. The Inconsistency of "optimal" cutpoints obtained
28 29	733		using two criteria based on the receiver operating characteristic curve. Am J
30 31	734		Epidemiol 2006; 163: 670-675.
32 33 34	735	80.	Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic
35 36	736		improvements. An illustration in rheumatology. Arch Intern Med 1993; 153:
37 38	737		1337-1342.
39 40 41	738	81.	Celik D, Coban O, Kilicoglu O. Minimal clinically important difference of
42 43	739		commonly used hip-, knee-, foot-, and ankle-specific questionnaires: a
44 45	740		systematic review. J Clin Epidemiol 2019; 113: 44-57.
46 47	741	82.	Maredupaka S, Meshram P, Chatte M, Kim WH, Kim TK. Minimal clinically
48 49 50	742		important difference of commonly used patient-reported outcome measures in
51 52	743		total knee arthroplasty: review of terminologies, methods and proposed values.
53 54	744		Knee Surg Relat Res 2020; 32: 19.
55 56	745	83.	MacKay C, Clements N, Wong R, Davis AM. A systematic review of estimates
57 58 59 60	746		of the minimal clinically important difference and patient acceptable symptom

Érasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

BMJ Open

1

2 3 4	747		state of the Western Ontario and McMaster Universities Osteoarthritis Index in
5 6	748		patients who underwent total hip and total knee replacement. Osteoarthritis
7 8	749		Cartilage 2019; 27: 1408-1419.
9 10 11	750	84.	Escobar A, Riddle DL. Concordance between important change and acceptable
12 13	751		symptom state following knee arthroplasty: the role of baseline scores.
14 15	752		Osteoarthritis Cartilage 2014; 22: 1107-1110.
16 17 19	753	85.	Angst F, Aeschlimann A, Steiner W, Stucki G. Responsiveness of the WOMAC
19 20	754		osteoarthritis index as compared with the SF-36 in patients with osteoarthritis
21 22	755		of the legs undergoing a comprehensive rehabilitation intervention. Ann Rheum
23 24 25	756		Dis 2001; 60: 834-840.
25 26 27	757	86.	Bellamy N. Pain assessment in osteoarthritis: experience with the WOMAC
28 29	758		osteoarthritis index. Semin. Arthritis Rheum 1989; 18: 14-17.
30 31	759	87.	Schwartz CE, Andresen EM, Nosek MA, Krahn GL. Response shift theory:
32 33 34	760		important implications for measuring quality of life in people with disability. Arch
35 36	761		Phys Med Rehabil 2007; 88: 529-536.
37 38	762	88.	Guyatt GH, Norman GR, Juniper EF, Griffith LE. A critical look at transition
39 40 41	763		ratings. J Clin Epidemiol 2002; 55: 900-908.
42 43	764	89.	Barrett B, Brown D, Mundt M, Brown R. Sufficiently important difference:
44 45	765		expanding the framework of clinical significance. Med Decis Making 2005; 25:
46 47 48	766		250-261.
49 50	767	90.	Ferreira M. Research Note: The smallest worthwhile effect of a health
51 52	768		intervention. J Physiother 2018; 64: 272-274.
53 54	769	91.	Ferreira ML, Herbert RD, Ferreira PH, Latimer J, Ostelo RW, Nascimento DP,
56 57	770		et al. A critical review of methods used to determine the smallest worthwhile
58 59 60	771		effect of interventions for low back pain. J Clin Epidemiol 2012; 65: 253-261.

BMJ Open

92. Wofford N. Smallest worthwhile effect values for pain and function after a total knee replacement. health related sciences, vol. PhD: Virginia Commonwealth University 2020. 93. Bellamy N, Hochberg M, Tubach F, Martin-Mola E, Awada H, Bombardier C, et

al. Development of multinational definitions of minimal clinically important
 improvement and patient acceptable symptomatic state in osteoarthritis.
 Arthritis Care Res (Hoboken) 2015; 67: 972-980.

TK. KB. Minimal clinically important 94. Kvien Heiberg Τ. Hagen improvement/difference (MCII/MCID) and patient acceptable symptom state (PASS): what do these concepts mean?. Ann Rheum Dis. 2007;66 Suppl 3(Suppl 3):iii40-iii41. doi:10.1136/ard.2007.079798

95. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, et al. Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state. Ann Rheum Dis 2005; 64: 34-37.

Turner D, Schünemann HJ, Griffith LE, Beaton DE, Griffiths AM, Critch JN, et 96. al. Using the entire cohort in the receiver operating characteristic analysis maximizes precision of the minimal important difference. J Clin Epidemiol 2009; 62: 374-379.

97. 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: 1589-1593.

1 2		
3 4	795	List of tables
5 6	796	
7 8 0	797	Table 1: Characteristics of included studies
9 10 11	798	Table 2: Minimal Clinical Important Change (MCIC) values of knee osteoarthritis
12 13	799	outcome tools derived using the anchor method
14 15	800	Table 3: Minimal Clinical Important Difference (MCID) values of knee osteoarthritis
16 17 18	801	outcome tools derived using the anchor method
19 20	802	Table 4: Comparison of receiver operating characteristic (ROC) method based
21 22	803	Minimal Clinical Important Difference (MCID) estimates with overall pooled estimates
$\begin{array}{c} 25\\ 26\\ 27\\ 28\\ 29\\ 30\\ 31\\ 32\\ 33\\ 34\\ 35\\ 36\\ 37\\ 38\\ 39\\ 40\\ 41\\ 42\\ 43\\ 44\\ 45\\ 46\\ 47\\ 48\\ 49\\ 50\\ 51\\ 52\\ 53\\ 54\\ 55\\ 56\\ 57\\ 58\\ 59\\ 60\\ \end{array}$		

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

1 2		
3 4	804	Figure legends
5 6	805	
7 8 9	806	Figure 1: Flow diagram of study selection (PRISMA, Preferred Reporting Items for
10 11	807	Systematic reviews and Meta-Analyses) ²⁴
12 13		List of Supplements
14 15		
16 17 18		Supplement 1: Quality of anchor studies assessed using the Devji (2020) credibility
19 20		instrument ²⁶
21 22		Supplement 2: Quality of consensus studies assessed using the Critical Appraisal
23 24 25		Screening Program (CASP)-qualitative checklist ²⁸
26 27		Supplement 3: Quality of distribution studies assessed using the National Institute
28 29		of Health (NIH) quality assessment tool for before-after (pre-post) studies with no
30 31 32		control group ³⁰
33 34		Supplement 4: Anchor properties of included anchor studies
35 36		Supplement 5: The effect of follow-up time on A. Minimal Clinically Important
37 38 30		Change (MCIC) and B. Minimal Clinically Important Difference (MCID)
40 41		†: multiple MCIC estimates using two different anchor questions; *: multiple MCID
42 43		estimates using two different anchor questions and different calculation methods
44 45 46		KOOS: Knee injury and Osteoarthritis Outcome Score, QOL: Quality of Life, KOS:
40 47 48		Knee Outcome Survey, ADL: Activities of Daily Living, LEFS: Lower Extremity
49 50		Functional Scale, WOMAC: Western Ontario and McMaster Universities Arthritis
51 52		Index
53 54 55		Supplement 6: Minimum Detectable Change (MDC) values of knee osteoarthritis
56 57		outcome tools derived using the distribution method
58 59		



Angst, 2018 ³⁶ Yes Impossible to tell Definitely No To a greater extent Harris, 2013 ³⁴ Yes To a great extent Impossible to tell Not so much Hmamouchi, 2012 ³⁷ Yes Not so much Impossible to tell To a greater extent Klokker, 2016 ¹⁵ Yes Definitely Yes Impossible to tell To a greater extent Klokker, 2016 ¹⁵ Yes Definitely Yes Not so much To a greater extent Mills, 2016 ¹⁵ Yes To a great extent Impossible to tell To a greater extent Or a greater extent To a greater extent To a greater extent To a greater extent To a greater extent To a greater extent Or a greater extent To a greater extent Or a greater extent To a greater extent To a greater extent To a greater extent To a greater extent To a greater extent To a greater extent To a greater extent Or a greater extent To a greater extent To a greater extent To a greater extent To a greater extent To a greater extent To a greater extent To a greater extent Impossible to tell Not so much Perrot, 201 ³⁴² Yes To a great extent Impossible to tell Not so much Singh, 2014 ⁴⁰ Yes To a great extent Impossible to tell Not so much Tubach, 2005 ¹⁴ Yes Not so much Impossible to tell To a greater extent Tubach, 2005 ¹⁴ Yes Not so much Impossible to tell To a greater extent Tubach, 2005 ¹⁴ Yes Not so much Impossible to tell To a greater extent Tubach som a good correlation with the PROM? Item 1: Is the anchor seasily understandable and relevant for patients or a necessary proxy? Item 3: Has the anchor shown a good correlation with the PROM? Item 4: Is the MID precise? Item 5: Does the threshold or difference between groups on the anchor used to estimate the MID reflect a small but important difference? * The responses to items: Yes, No, Impossible to tell # The responses to items: Definitely yes, To a great extent, Not so much, Definitely no, Impossible to tell * "to verall quality: three of the five criteria were met "Yes" or "definitely yes" or "to a great extent", the paper is of "high" quality if for, the paper is of "low" quality (credibility) studies are shaded.	Study	Item 1*	Item 2 [#]	Item 3#	Item 4 [#]	<u> </u>	The ove
Angst, 2018 ³⁶ Yes Impossible to tell Definitely No To a greate extent To a greate extent <t< th=""><th>,</th><th></th><th></th><th></th><th></th><th>for -</th><th>quality of</th></t<>	,					for -	quality of
Angs, 2013TesImpossible tollDefinitely NoTo a great extentImpossible to tellTo a great extentImpossible to tellTo a great extentHarris, 2013YesNot so muchImpossible to tellTo a great extentImpossible to tellNot so muchImpossible to tellTo a great extentKlokker, 2016YesDefinitely YesImpossible to tellTo a great extentDefinitely NoLee, 2017YesDefinitely YesImpossible to tellTo a great extentTo a great extentKlokker, 2016YesDefinitely YesImpossible to tellTo a great extentTo a great extentMills, 2016YesDefinitely YesNot so muchTo a great extentTo a great extentOrnetti, 2013YesTo a great extentImpossible to tellTo a great extentTo a great extentOrnetti, 2013YesYesTo a great extentImpossible to tellTo a great extentPerot, 2013YesTo a great extentImpossible to tellNot so muchSingh, 2014YesTo a great extentImpossible to tellNot so muchWilliams, 2012YesTo a great extentImpossible to tellNot so muchWilliams, 2012YesNot so muchImpossible to tellTo a greater extentWilliams, 2012YesNot so muchImpossible to tellTo a great extentWilliams, 2012YesNot so muchImpossible to tellTo a greater extentWilliams, 2012YesNot so much<	Angst 2018 ³⁶	Voc	Impossible to tell	Definitely No	To a greater extent	<u> </u>	studies
Hamapouchi, 2012 ³⁷ Yes Not so much Impossible to tell To a greater extent Klokker, 2016 ³⁵ Yes To a great extent Impossible to tell Definitely No Lee, 2017 ⁴³ Yes Definitely Yes Impossible to tell To a greater extent Mostafaee, 2021 ³⁸ Yes To a great extent To a great extent To a greater extent Ornetti, 2011 ³⁹ Yes To a great extent Impossible to tell To a greater extent Impossible to tell To a greater extent To a great extent Impossible to tell Not so much Definitely Yes Not so much Singh, 2014 ⁴⁰ Yes To a great extent Impossible to tell Not so much Impossible to tell Not so much Impossible to tell Not so much Impossible to tell To a greater extent Impossible to tell Not so much Impossible to tell To a greater extent Impossible to tell To a greater extent Impossible to tell Not so much Impossible to tell Not so much Impossible to tell Not so much Impossible to tell Impossible to	Harris 2013 ³⁴	Yes		Impossible to tell	Not so much		Low
Mokker, 2016 ³⁵ Yes To a great extent Impossible to tell Definitely No Lee, 2017 ⁴³ Yes Definitely Yes Impossible to tell To a greater extent To a great extent Definitely Yes To a great extent	Hmamouchi, 2012 ³⁷	Yes	Not so much	Impossible to tell	To a greater extent	a greater extent	High
Lee, 2017 ⁴³ Yes Definitely Yes Impossible to tell To a greate extent Mills, 2016 ¹⁵ Yes Definitely Yes Not so much To a greate extent Mostafaee, 2021 ³⁸ Yes To a great extent To a greate extent To a greate extent Ornetti, 2013 ³⁹ Yes To a great extent Impossible to tell To a greate extent Singh, 2014 ⁴⁰ Yes To a great extent Impossible to tell Not so much Singh, 2014 ⁴⁰ Yes To a great extent Impossible to tell Definitely Yes and great extent Tubach, 2005 ¹⁴ Yes To a great extent Impossible to tell Definitely Yes and great extent Williams, 2012 ⁴¹ Yes Not so much Impossible to tell Definitely Yes and great extent Impossible to tell To a greater extent of the parent or necessary proxy responding directly to both the PROM and the anchor? tem 1: Is the patient or necessary proxy responding directly to both the PROM and the anchor? tem 2: Is the anchor easily understandable and relevant for patients or a necessary proxy? tem 3: Has the anchor shown a good correlation with the PROM? tem 4: Is the MID precise? tem 5: Does the threshold or difference between groups on the anchor used to estimate the MID reflect a small but important difference? 'The responses to items: Yes, No, Impossible to tell * The responses to items: Definitely yes, To a great extent, Not so much, Definitely no, Impossible to tell * The responses to items: Definitely yes, To a great extent, Not so much, Definitely no, Impossible to tell * The responses to items: Definitely yes, To a great extent, Not so much, Definitely no, Impossible to tell * The responses to items: Def	Klokker, 2016 ³⁵	Yes	To a great extent	Impossible to tell	Definitely No	Not so much	Low
Mills, 2016 ¹⁵ Yes Definitely Yes Not so much To a greater extent To a great extent	Lee, 2017 ⁴³	Yes	Definitely Yes	Impossible to tell	To a greater extent	ο ές μπροssible to tell	High
Mostafaee, 2021 ³⁸ Yes To a great extent To a great extent To a great extent To a great extent Impossible to tell To a great extent Impossible to tell Definitely Yes To a great extent Impossible to tell Definitely Yes To a great extent Impossible to tell To a great extent To a great extent To a great extent Impossible to tell To a great extent To a great extent Impossible to tell To a great extent To a great extent To a great extent Impossible to tell To a great extent To a great extent To a great extent Impossible to tell To a great extent To a great extent To a great extent Impossible to tell To a great extent To a great extent Impossible to tell To a great extent To a great extent To a great extent Impossible to tell To a great extent To a great extent To a great extent Impossible to tell To a great extent To a great extent To a great extent Impossible to tell To a great extent To a great extent To a great extent Impossible to tell To a great extent Impossible to tell To a great extent Impossible to tell To a great extent To a great ext	Mills, 2016 ¹⁵	Yes	Definitely Yes	Not so much	To a greater extent	Definitely Yes	High
Ornetti, 2011 ³⁹ Yes To a great extent Impossible to tell To a greater extent Definitely Yes Perrot, 2013 ⁴² Yes To a great extent Not so much Definitely Yes To a great extent Definitely Yes Singh, 2014 ⁴⁰ Yes To a great extent Impossible to tell Not so much Definitely Yes To a great extent Definitely Yes Tubach, 2005 ¹⁴ Yes To a great extent Impossible to tell Definitely Yes To a great extent Definitely Yes Williams, 2012 ⁴¹ Yes Not so much Impossible to tell To a greater extent Definitely Yes tem 1: Is the patient or necessary proxy responding directly to both the PROM and the anchor? To a great extent for patients or a necessary proxy? To a great extent for patients or a necessary proxy? To a great extent for patients or a necessary proxy? tem 4: Is the MID precise? To a great extent, Not so much, Definitely no, Impossible to tell To a great extent, Not so much, Definitely no, Impossible to tell To a great extent, the paper is of "high" quality for ot, mat the paper is of "low" quality (credibility) studies are shaded.	Mostafaee, 2021 ³⁸	Yes	To a great extent	To a great extent	To a great extent	ar <u>g n</u> Not so much	High
Perrot, 2013 ⁴² Yes To a great extent Not so much Definitely Yes to a great extent Singh, 2014 ⁴⁰ Yes To a great extent Impossible to tell Not so much To a great extent To a great extent To a great extent Not so much To a great extent To a great exten	Ornetti, 2011 ³⁹	Yes	To a great extent	Impossible to tell	To a greater extent	d S a Definitely No	High
Singh, 2014 ⁴⁰ Yes To a great extent Impossible to tell Not so much To a great extent Tubach, 2005 ¹⁴ Yes To a great extent Impossible to tell Definitely Yes Not so much Definitely Yes Williams, 2012 ⁴¹ Yes Not so much Impossible to tell To a greater extent Definitely Yes tem 1: Is the patient or necessary proxy responding directly to both the PROM and the anchor? To a greater extent To a greater extent To a greater extent tem 1: Is the patient or necessary proxy responding directly to both the PROM and the anchor? To a greater extent To a greater extent To a greater extent tem 3: Has the anchor shown a good correlation with the PROM? The responses to items: Yes, No, Impossible to tell The responses to items: Yes, No, Impossible to tell The responses to items: Definitely yes, To a great extent, Not so much, Definitely no, Impossible to tell The responses to items: Definitely yes, To a great extent, Not so much, Definitely no, Impossible to tell To a great extent", the paper is of "high" quality The paper is of "low" quality (credibility) studies are shaded.	Perrot, 201342	Yes	To a great extent	Not so much	Definitely Yes	Not so much	High
Tubach, 2005 ¹⁴ Yes To a great extent Impossible to tell Definitely Yes Williams, 2012 ⁴¹ Yes Not so much Impossible to tell Definitely Yes tem 1: Is the patient or necessary proxy responding directly to both the PROM and the anchor? To a greater extent Impossible to tell Definitely Yes tem 1: Is the patient or necessary proxy responding directly to both the PROM and the anchor? Impossible to tell To a greater extent Impossible to tell To a greater extent Impossible to tell tem 1: Is the patient or necessary proxy responding directly to both the PROM? Impossible to reaction with the PROM? Impossible to reaction with the PROM? Impossible to tell Imposs	Singh, 2014 ⁴⁰	Yes	To a great extent	Impossible to tell	Not so much	$\mathbf{x} = \mathbf{x}$ a great extent	High
Williams, 2012 ⁴¹ Yes Not so much Impossible to tell To a greater extent Definitely Yes tem 1: Is the patient or necessary proxy responding directly to both the PROM and the anchor? The anchor easily understandable and relevant for patients or a necessary proxy? Impossible to tell Impossible to	Tubach, 2005 ¹⁴	Yes	To a great extent	Impossible to tell	Definitely Yes	Not so much	High
tem 1: Is the patient or necessary proxy responding directly to both the PROM and the anchor? tem 2: Is the anchor easily understandable and relevant for patients or a necessary proxy? tem 3: Has the anchor shown a good correlation with the PROM? tem 4: Is the MID precise? tem 5: Does the threshold or difference between groups on the anchor used to estimate the MID reflect a small but important difference? The responses to items: Yes, No, Impossible to tell # The responses to items: Definitely yes, To a great extent, Not so much, Definitely no, Impossible to tell * overall quality: three of the five criteria were met "Yes" or "definitely yes" or "to a great extent", the paper is of "high" quality Ifform, at the paper is of "low" qua- Low quality (credibility) studies are shaded.	Williams, 2012 ⁴¹	Yes	Not so much	Impossible to tell	To a greater extent	B EDefinitely Yes	High
.ow quality (credibility) studies are shaded.	em 4: Is the MID precise? em 5: Does the threshold The responses to items: Υ The responses to items: I ' overall quality: three of th	or difference betwe /es, No, Impossible Definitely yes, To a ne five criteria were	een groups on the anchor us e to tell a great extent, Not so much, e met "Yes" or "definitely yes	ed to estimate the MID r Definitely no, Impossible " or "to a great extent", th	eflect a small but importa to tell he paper is of "high" qualit	unt dafference? similar toom/on ty lfect, too ty lfect, too age to the paper is of "	low" quality ²⁷
gi 20		ties are shaded				9, 2 19, 2	
ም እ PROM: Patient-Reported Outcome Measure, MID-Minimal Important Difference (In this study minimal clinically important change/difference)	ow quality (credibility) stuc						

BMJ	Open
-----	------

					BM.	l Open		by copyright, i	136/bmjopen-2(
pplement Study Author, Year)	2: Quality of c Item 1 [#]	tem 2 *	dies assessed u Item 3 [#]	sing the Critica Item 4 #	al Appraisal Sci Item 5 [#]	reening Prograr Item 6 [#]	n (CASP)-qualit Item 7 [#]	ative chedelis Item 80 for u	5tt- 61tem 3026 on	Item 10	The overa quality of the studie
alottolo, 2018 ⁴⁴	Yes	Yes	Can't tell	Can't tell	Yes	Yes	Can't tell	ses related to text and Yes	18 May 2023. Downloa	Researcher discusses the the contribution the study makes to existing knowledge or understanding	Moderate
Item Item Item Item Item Item # Th μ Ον	3: Was the re 4: Was the re 5: Was the da 6: Has the re 7: Have ethic 8: Was the da 9: Is there a c 10: How value e responses to erall quality of	search design cruitment strat ata collected in elationship betw al issues been ata analysis su clear statemen able is the resu o items: Yes, 0 f study: high",	appropriate to a tegy appropriate a way that add ween researche t taken into cons ifficiently rigorou t of findings? earch? Can't Tell, No "moderate" or "I	address the air to the aims of ressed the res r and participal sideration? us?	ns of the resear the research? earch issue? nts been adequ	urch? uately consider iew team based	ed? d on how reliable	ining, Al training, and similable and creditechnologies.	m http://bmjopen.bmj.com/ en May 9, 2025 at Do	tudy was, without sp	ecific rules

Page 45 of 59

BMJ Open

Э	BMJ Open S D	
	py njopen righ	
	Supplement 3: Quality of distribution studies assessed using the National Institute of Health (NIH) quality assessment tool for before after (pre-post) studies with no control aroup ³⁰	I

Study (Author, Year)	Criteria 1 [#]	Criteria 2 [#]	Criteria 3 [#]	Criteria 4 [#]	Criteria 5 [#]	Criteria 6 [#]	Criteria 7 [#]	Criteria 8 [#]	Criteria 9 [#]	Criteria 2 10 [#]	Criteria 11 [#]	Criteria 12 [#]	Overall quality ^µ
Alghadir, 2017 ⁵⁰	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	us Yes	Yes	Other	Fair
Alghadir, 2018 ⁵¹	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	s ve	Yes	Other	Fair
Alghadir, 2016 b	Yes	No	Yes	Yes	Yes	Other	Yes	Other	Yes	ein¶a aj¶a	Yes	Other	Fair
Alghadir, 2017	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes		Yes	Other	Fair
Baert, 2018 ⁵³	Yes	No	Yes	Yes	Other	Other	Yes	Other	Yes		Yes	Other	Fair
Baert, 2018	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes		Yes	Other	Fair
Brisson, 2018 ⁵⁴	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	anges	Yes	Other	Good
Callaghan, 2009 ⁵⁵	Yes	No	Yes	Yes	Yes	Other	Yes	Other	Other		Yes	Other	Fair
Hoglund, 2019 ⁵⁷	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes		Yes	Other	Good
Hunter, 2006 ⁷⁹	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Other	ni Yes	Yes	Other	Fair
ljima, 2019 ⁵⁸	Yes	Yes	Yes	Yes	Other	Yes	Yes	Other	Yes	ing Y	Yes	Other	Fair
Jansen, 2021 ⁵⁹	Yes	Yes	Yes	Yes	Other	Yes	Yes	Other	Yes	, ≥Y	Yes	Other	Fair
Kanko, 2019 ⁷⁷	Yes	Other	Yes	ta Y	Yes	Other	Good						
Kean, 2010 ⁶⁰	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	ini y	Yes	Other	Fair
Klokker, 2015 ⁶¹	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	Ģ Yes	Yes	Other	Good
McCarthy, 2004 ³²	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	and Yes	Yes	Other	Fair
McCarthy, 2008 ⁶²	Yes	Yes	No	No	Other	Other	Yes	Other	Yes	sin Y 🍪	Yes	Other	Fair
Monticone, 2021 63	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes		Yes	Other	Good
Motyl, 2013 ⁸⁰	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes		Yes	Other	Fair
Mutlu, 2015 ⁷⁸	Yes	Other	Yes	H Y	Yes	Other	Good						
Nalbant, 2021 ⁶⁴	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	o Yes	Yes	Other	Good
Naylor, 2014 ⁶⁵	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	gie Yes	Yes	Other	Fair
Parveen, 201766	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	y″ on Y@as	Yes	Other	Fair
Peter, 201867	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	Y	Yes	Other	Fair
Piva, 2004 ⁶⁸	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	Yes	Yes	Other	Fair
Pratheep, 2018 ⁶⁹	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	Y	Yes	Other	Fair
Ravaud, 1999 ⁷⁶	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	Yes	Yes	Other	Fair

BMJ Open

d by copyrig 136/bmjope

<u>a</u>

artment G

EZ-LTA

	Criteria	Criteria	Criteria	Criteria	Criteria	Criteria	Criteria	Criteria	Criteria	Eriteria	Criteria	Criteria	Overall
	1 *	2 *	3 *	4 *	5 *	6 *	7*	8 *	9 *	≗.105* <u> </u>	11 *	12 *	quality ^p
Subail and	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	ug Yeng T	Yes	Other	Good
Chaudhary, 2021 ³⁰					0.1	0.4		0.1					
Suwit, 2020 ³²	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes		Yes	Other	Fair
Takacs, 2014	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	ĭ Y∰	Yes	Other	Fair
Tevald, 2016	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	en ŭe	Yes	Other	Fair
Takacs, 2014 ⁷⁰	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	ted as n	Yes	Other	Good
Tevald, 2016 ⁷¹	Yes	No	Yes	Yes	Other	Other	Yes	Other	Yes	to t	Yes	Other	Fair
Tse, 2021 ⁷²	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	ie Age	Yes	Other	Fair
Turcot, 2008 ⁷³	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	an By A	Yes	Other	Good
										$\omega \ge \omega$			
the description of the criteria iteria 1: Was the study que iteria 2: Were eligibility/sele iteria 3: Were the participal iteria 4: Were all eligible pa iteria 5: Was the sample siz iteria 6: Was the test/servic iteria 7: Ware the outcome	a is given b estion or ob ection crite nts in the s articipants t ze sufficier ce/interven	elow. bjective clea ria for the s study represent that met the ntly large to tion clearly prespecifie	arly stated? study popu sentative o e prespecif provide co described	lation presp f those who ied entry cl onfidence in and delive	pecified and p would be riteria enro n the findin red consist	d clearly de eligible for lled? gs? tently acros	escribed? the test/se	rvice/interv / population	rention in th	ata miningg Al training	or clinical po	pulation of ir	nterest?
ne description of the criteria iteria 1: Was the study que iteria 2: Were eligibility/sele iteria 3: Were the participan iteria 4: Were all eligible pa iteria 5: Was the sample siz iteria 6: Was the test/servic iteria 7: Were the outcome iteria 8: Were the people as	a is given b estion or ob ection crite nts in the s articipants f ze sufficier ce/interven measures	elow. pjective clea ria for the s study represent that met the ntly large to tion clearly prespecifie prespecifie prespecifie	arly stated? study popu sentative o prespecif provide ca described ed, clearly shinded t	lation presp f those who ried entry cl onfidence in and delive defined, va	pecified and b would be riteria enro n the findin red consist lid, reliable binants' exr	d clearly de eligible for lled? gs? tently acros e, and asse	escribed? the test/se s the study ssed consisterventions?	rvice/interv / population stently acro	rention in th n? oss all stud	ed from http://bmjopeiip.b ool . ata mininggAl trainingartion.art	or clinical po ts?	pulation of ir	nterest?

Criteria 10: Did the statistical methods examine changes in outcome measures from before to after the intervention? Were statistical methods examine changes in outcome measures from before to after the intervention? pre-to-post changes?

Criteria 11: Were outcome measures of interest taken multiple times before the intervention and multiple times after the intervention $\frac{1}{3}$ i.e., did they use an interrupted timeseries design)?

Criteria 12: If the intervention was conducted at a group level (e.g., a whole hospital, a community, etc.) did the statistical analysis take into account the use of individual-level data to determine effects at the group level?

The responses to items: Yes, No, Other (CD, NR, NA)*, *CD, cannot determine; NA, not applicable; NR, not reported

µ Overall quality: Good, Fair, or Poor was evaluated by the review team based on how reliable and credible each study was, withou specific rules as recommended by the tool

			BMJ Open	36/bmjopen-20 y copyright, ii	
Supplement 4: Anchor Study	properties of included anchor Description of anchor	studies Anchor question	Anchor	Definitions of MacIo	Reported terminolog
Angst, 2018 ³⁶	GRC transition (5 points)	NR	properties/responses much worse, slightly worse, almost equal, slightly better, much better	transition question Difference betହeer the "slightly better" group an େthe almost equal" groupନ MତାD ତ _ ୍ଲି	in the study Minimal Clinically Important Difference
Harris, 2013 ³⁴	GRC responses (3 points)	Compared to one week before your clinical visit, please indicate how much your knee problem has changed?	1. My knee has got better, 2. My knee has stayed the same, 3. My knee has got worse	Mean change in Begoup "my knee has got better" By CIC and the difference in the generative generati	Minimal Important Change, Minimal Important Difference
Hmamouchi, 2012 ³⁷	GRC transition (5 points)	How do you feel in general today as compared to six weeks earlier as far as your osteoarthritis is compared?	much better, slightly better, no change, slightly worse, much worse	Difference between the mean effects of "slight better' group and "no change" group s= MCID	Minimal Clinically Important Difference improvement
Klokker, 2016 ³⁵	modified GRC (3 responses and a GRC spanning from -7 to +7)	Did your knee pain change since you entered this project?	unchanged, better, worse: if it is better or worse, bring up a scale spanning from -7 to +7 (-7 (worst) to +7 (best) scale)	Difference of a score of at least 2 (+2: little better: 2: little worse) and no change (score of 0, +1 (almost the same or had ly any better), -1 (worse at all)) MCID	Minimal Important Change
Lee, 2018 ⁴³	Multiple anchors: 36 item Short Form subscales- physical functioning, social role functioning, energy and vitality, bodily pain, mental health, physical role functioning, emotional role functioning, general	NR	NR	Prospective change for people achieving preversitive established MCID on legacy comparators: MCID in a single item anchor e.g. Patient Global Assessment was defined as range= MCID to 1+MgID, MCID in a multi-teem anchor e.g. SF36 range= MCID to 2XDCID	Minimally Importan Difference

8 9

16

18

20

42

d by copyright, in Definitions of MCID using Study **Description of anchor** Anchor question Anchor **Reported terminology** transition question properties/responses in the study 3026 on 18 May 2023. Downloaded from http://bmjopen Erasmushogeschool . ng for uses related to text and data mining, Al training, health perception (0-100 scale, 0: best, 100: worse); WOMAC-pain and function (0-100 scales,0: best, 100: worse); Back depression (0-63, 0: best, 63: worse); Perceived Stress scale (0-40, 0: best, 40: worse); Patient Global eer revie Assessment in a 0-10 cm visual analogue scale (higher number= greater perception of disease activity); Six-minute walk test (in meters); 20-meter walk test (in seconds) Mean change within slightly Mills, 2016¹⁵ Global transition Two anchor question much improved, Minimal Important improved ground MCIC GRC-7-point Likert scale 1. Compared with when moderately improved, Differencenand o slightly improved, not I started this program, Mean change my walking on level changed, slightly worse, (Jaeschke) method Difference between perticipants who ground has (walking Minimal Important moderately worse, much responded Differenceanchor) and worse slightly, moderately, much improved 2. Compared with when Mean change as the "improved geo up" and no I started this program, (Redelmeier and Lorig) change or worse "non-improved my knee had (knee method, group" =Mc ROC method (Youden health anchor): method and 80% specific method) The difference between improved Mostafee, 2021³⁸ GRC- 7points Likert scale How did your knee 1. very much worse; 2. Minimal Important group (6 and 7) and not improved status change much worse; 3. slightly Change group (1,2,3,4 an 25)=MCID compared to the worse; 4. no change; 5. Z-LTA

BMJ Open

1 2 3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

44 45 46 Page 48 of 59

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Page 49 of 59				BMJ Open	l136/bmjop d by copyri	
1 2 3 - 4	Study	Description of anchor	Anchor question	Anchor properties/responses	연구 주 구: 은 Definitions of MCIO/MCID using transitio을 quastion	Reported terminology in the study
5 6 7 8 9 10	Ornetti, 2011 ³⁹	GRC- (3points Likert scale, then again in a 4- point Likert scale	beginning of the physiotherapy intervention? specific question NR. Two anchor questions. 1. The degree of	slightly better; 6. much better; and 7. very much better The degree of improvement of global state (global MCII) on a	75 th centile of the absolute change in score amologication of the solute state responded the man by the solute state in score amologication of the solute state s	Minimal Clinically Important Improvement
11 12 13 14 15 16 17 18 19			improvement of global state 2. The degree of improvement of functional state	3-point Likert scale (worsened function, no change, improved function). Then, among the patients who improved, the degree of improvement was scored on a scale (poor, fair, good, excellent)	good or exceller代摘取oved group) = text and data mining	
20 21 22 23 24 25 26 27 28 29 30 31 32	Perrot, 2013 ⁴²	GRC (5 points)	Taking into account the pain due to your arthritis in the last 24 hours and the pain you experienced initially, how has your pain changed?	Patients who said that their pain had improved were asked to rate the level of improvement on a 5-point Likert scale extending from very large improvement to no improvement at all	75 th percentile of the pain intensity difference (day 0 to 7) for patients considering their improvement to be at least moderate = MCIC similar technologi, 20	Minimal Clinically Important Improvement
33 34 35 36 37 38 39 40 41 42	Singh, 2014 ⁴⁰	GRC (5 points)	Since the last time you completed the survey 2 weeks ago, would you say your knee arthritis is:	A great deal better, somewhat better, about the same, somewhat worse, a great deal worse	Mean change between baseline and follow-up of the group who said somewhat better GEZ-LTT	Minimal Clinically Important Difference for improvement
43 44 45		For p	eer review only - http://b	mjopen.bmj.com/site/about	/guidelines.xhtml	

Page 50 of 59

			BMJ Open	136/bmjopen-20 by copyright, i	
Study	Description of anchor	Anchor question	Anchor properties/responses	ے ج Definitions of McIOMCID using transition وللھstion	Reported terminolog in the study
Tubach, 2005 ¹⁴	GRC (5 points)	Response to NSAID treatment	None= no good at all, ineffective drug; poor= some effect but unsatisfactory; fair= reasonable effect but could be better; good= satisfactory effect with occasional episodes of pain or stiffness; excellent= ideal response, virtually pain- free	75 th centile of the group who reported as "good" using logistic regression=MCIC Erasmushogeschool	Minimal Clinically Important Improvemer
Williams, 2012 ⁴¹	GRC questionnaire on a 15-level scale (15 points)	Not specified To rate the extent to which they perceive their condition as having changed over time	GRC on a 15- level scale. The GRC ranges from 1= a very great deal better to 8=about the same to 15= very great deal worse	Difference between ubjects who perceived "improved" (those with a GRC between (very great deal better) and 5 (somewhat better)) from subjects who perceived "not improved" (those with between 6 (a little better) and 15 a very great deal worge)	Minimal Clinically Important Difference
ICIC: Minimal Clinically VOMAC: Western Onta	y Important Change, MCID: Mi ario and McMaster Universities	inimal Clinically Important E Arthritis Index, SF36: Sho	Difference, GRC: Global Rati	ing of Change, NRENot Reported, CI: C eroid Anti-Inflammatory Or May 9, 2025 at Department GEZ	confidence Interval,
				LTA	



†: multiple MCIC estimates using two different anchor questions; *: multiple MCID estimates using two different anchor questions and different calculation methods

KOOS: Knee injury and Osteoarthritis Outcome Score, QOL: Quality of Life, KOS: Knee Outcome Survey, ADL: Activities of Daily Living, LEFS: Lower Extremity Functional Scale, WOMAC: Western Ontario and McMaster Universities Arthritis Index

-			s derived using the		00	
Outcome tool	Score	Median	Minimum	Maximuਸ਼ੁਰੋ ਤੋ	Solumber of	Study
Aggregated locomotor function	N/A	2.3 seconds		<u>v</u>	9 1	McCarthy,2004 ³²
Animated Activity Questionnaire	100=best	11.2		se	a 1	Peter, 201865
BMI using a scale and a stadiometer	N/A	2.6 kg/m ²		s re	≤ 1	Brisson, 201852
Bone density-femur mean lateral	N/A	0.5 mm	0.4 mm	0.6 mm 🖉 🖫	a 2	Jansen, 202157
Bone density-femur mean medial	N/A	0.6 mm Al eq	0.5 mm Al eq	0.7 mm Al	2 2	Jansen, 202157
Bone density-tibia mean lateral	N/A	2.6 mm Al eq	2.4 mm Al eq	2.8 mm Al 6	23 . 2	Jansen, 202157
Bone density-tibia mean medial	N/A	0.8 mm Al eq	0.7 mm Al eq	1 mm Al 🚭 🏅	D 2	Jansen, 202157
Cartilage volume-lateral tibia	N/A	0.6ml	0.5 ml	0.6 ml 👬 🦉	¥ 2	Hunter, 200677
Cartilage volume-medial Tibia	N/A	0.6 ml	0.4 ml	0.7 ml a š	2	Hunter, 200677
Cartilage volume-femur	N/A	1.1 ml	0.8 ml	1.3 ml 👷 💆	de 2	Hunter, 200677
Cartilage volume-patella	N/A	0.9 ml	0.7 ml	1.1 ml 🗟 🔍	<u>å</u> 2	Hunter, 200677
Center of mass mediolateral displacement during	N/A	0.01°	0.01°	0.02 ° Ξ .	ro 3	VandeStraaten, 2020
unipodal stance using 3D motion analysis				nin	3	,
De Motion Mobility Index	100=best	8.0	7.3	8.7	2	Yuruk, 2014 ⁷³
Eight-meter walk time	N/A	1.4 seconds		A	1	McCarthy,2004 ³²
Eminence lateral (knee image)	mm	2.2 mm	2.2 mm	2.2 mm 🖣	2	Jansen, 202157
Eminence medial (knee image)	mm	1.8 mm	1.7 mm	1.8 mm	2	Jansen, 202157
Get up and Go test	N/A	1.4 seconds	1.2 seconds	1.5 secon	S 2	Piva, 2004 ⁶⁶
Fremantle Knee Awareness Questionnaire	100=worst	14.4		an	9 1	Monticone,2021 ⁶¹
Frontal plane tibial alignment using smartphone inclinometer	N/A	3.7°		d sim	nj.con	Tse, 2021 ⁷⁰
Frontal plane tibial alignment using a manual inclinometer	N/A	3.2°		ilar te	n 1	Tse, 2021 ⁷⁰
Hip abductor strength using a hand-held dynameter	N/A	0.3 Nm/ka	0.3 Nm/ka	0.3 Nm/k	Ma 2	Tevald, 2016 ⁶⁹
Hip flexion-extension during unipodal stance using 3D	N/A	2.7 °	2.2 °	4.6 ° 9	 ✓ ✓	VandeStraaten, 2020
motion analysis				go	20	
ICOAP-constant pain	100=worst	51.7	49.6	53.8 G	25 2	Singh, 2014 ⁴⁰
ICOAP-intermittent pain	100=worst	49.8	48.7	50.8	at 2	Singh, 2014 ⁴⁰
ICOAP-pain	100=worst	46.6	23.0	49.6	Depa	Singh, 2014 ⁴⁰ , Harris 2013 ³⁴
KAM impulse in 3D motion analysis	N/A	4.9 Nm*s			f 1	Brisson, 201852
KAM impulse in 3D motion analysis	N/A	0.4 %BW*HT*s			ien 1	Brisson, 201852

Page 53 of 59

		BMJ Open		y copyright	36/bmjopen	
Outcome tool	Score	Median	Minimum	Maximune G	Number of	Study
Knee force sense test using a handheld	N/A	20.6 N	13.2 N	26.7 NG	8 3	Baert, 2018 ⁵¹
dynamometer-reposition error 20°				for	.6 0	
Knee force sense test using a handheld	N/A	38.0 N	27.8 N	43.3 N ⋤	<u>я</u> 3	Baert, 2018 ⁵¹
dynamometer-reposition error 45°				es	18 N	
Knee force sense test using a handheld	N/A	32.2 N	22.5 N	49.5 N 🗟 🗖	May 3	Baert, 2018 ⁵¹
dynamometer-reposition error 70°	N1/A				20	D 1 001051
Knee force sense test using a handheid	N/A	32.7 N	30.0 N	33.7 N G E	23	Baert, 2018 ³¹
Knee joint ande	dearees	0.40	4.00	shc		lansen 2021 ⁵⁷
Knee joint space width	mm	2.1°	1.9°			Jansen, 2021
Knee joint space width-lateral	mm	1 mm	0.8 mm			Jansen, 2021
Knee joint space width medial	mm	1.7 mm	1.4 mm	2.0 mm at 0		Jansen, 2021
	mm	0.6 mm	0.5 mm	0.8 mm 3 .	fro	
Knee joint space width-minimum	mm	0.8 mm	0.6 mm	0.9mm <u>=</u>	ž 2	Jansen, 20213
KOOS-activities of daily living	100=best	19.1	17.4	20.8 ය	2	Naylor, 201463
KOOS-pain	100=best	18.6	17	20.2 >	2	Naylor, 2014 ⁶³
KOOS-quality of life	100=best		22.4	39.0 a	<u> </u>	Naylor, 2014 ⁶⁵ ; Sin
		27.8		nin ·	8	2014 ⁴²
KOOS-symptoms	100=best	20.2	2.9	24.1 🤤	3	Naylor, 201463
KOOS-PS	100=worst	28.3	16.0	35.5 nd	3	Harris, 2013 ³⁴ ; Sin
	N1/A	0.0.0	100	sin		2014 ⁴⁰
reposition error 70°	IN/A	8.0 °	4.0 °	8.0 ° <u>ni</u> a	J 3	Baen, 2018
Knee joint position sense test-analogue inclinometer-	N/A	4 0 °	300	80° F	on ₃	Baert 2018 ⁵¹
reposition error 45°		1.0	0.0	°.°	Maj	Baon, 2010
Knee joint position sense test-analogue inclinometer-	N/A	4.0 °	3.0 °	4.0 ° ol	9 3	Baert, 2018 ⁵¹
reposition error 20°				ogi	20	
Knee joint position sense test-analogue inclinometer-	N/A	3.0 °	3.0 ⁰	4.0 ° .	25 3	Baert, 2018 ⁵¹
reposition error all					at [
KOS-activities of daily living	100=best	15.8*	10.5	21.0	Gep 6	Williams, 201241
L-test	N/A	5.28 seconds			arti 1	Nalbant, 202162
Load frequency using a triaxial accelerometer	N/A	4.3 steps/day			Be 1	Brisson, 2018 ⁵²
LEFS	100=best	18.3*	14.8	22.6	R 6	Williams, 201241

		BMJ Open		y copyright,	36/bmjopen-;	
Outcome tool	Score	Median	Minimum	<u>جن</u> Maximur	Number of Sestimates	Study
NRS- pain	100=worst	16.5	13.3	19.6 g for	30 26	Alghadir, 2016b ⁴⁷ , Alghadir, 2018 ⁴⁹
Osteophytes-Femur lateral (using knee image)	N/A	7.8 mm ²	5.4 mm ²	10.3 mm ട്ട്	5 2	Jansen, 2021 ⁵⁷
Osteophytes-Femur medial (using knee image)	N/A	12.3 mm ²	11.2 mm ²	13.4 mm	∞ 2 ≤	Jansen, 202157
Osteophytes-Tibia lateral (using knee image)	N/A	10.5 mm ²	9.4 mm ²	11.6 mm	ay 2	Jansen, 2021 ⁵⁷
Osteophytes-Tibia medial (using knee image) 📐	N/A	11.6 mm ²	11 mm ²	12.2 mm 2 3	202 2	Jansen, 202157
OKS-summary	100=best	6.1	6.0	6.2 to tex	3. 1 Do	Alghadir, 2017 ⁴⁸ , Ha 2013 ³⁴
Peak KAM in 3D motion analysis	N/A	0.2 Nm/kg		ges tar	vn 1	Brisson, 201852
Peak KAM in 3D motion analysis	N/A	1.3 %BW*HT		id d	oad 1	Brisson, 201852
Peak KFM in 3Dmotion analysis	N/A	0.4 Nm/kg		lata	ed 1	Brisson, 201852
Peak KFM in 3D motion analysis	N/A	2.3 %BW*HT		<u> </u>	fro 1	Brisson, 201852
Pelvic abduction-adduction during unipodal stance using 3D motion analysis	N/A	3.1 °	2.8 °	3.5 ° ning , /	m 3	VandeStraaten, 202
Per cent of voluntary muscle activation using a dynamometer	N/A	6.6%		Al trai	1 1	Kean, 2010 ⁵⁸
Pressure pain threshold-knee- using an algometer	N/A	131.8 kPa	92.9 kPa	196.3 kP	<mark>ខ</mark> ្ល 10	Pratheep, 201867
Pressure pain threshold-medial heel using a handheld pressure algometer	N/A	1.3 lb	Ch.	g, and	1 bm	Mutlu, 2015 ⁷⁶
Pressure pain threshold- medial knee using a handheld pressure algometer	N/A	1.2 lb		simila		Mutlu, 2015 ⁷⁶
Quadriceps fatigue using surface electromyography- VMO initial median frequency	N/A	11.0 Hz		ar tecl	on 1	Callghan, 2009 ⁵³
Quadriceps fatigue using surface electromyography- VL initial median frequency	N/A	10.0 Hz		hnolog	ay 1 9, 1	Callghan, 2009 ⁵³
Quadriceps fatigue using surface electromyography- RF initial median frequency	N/A	9.0 Hz		gies.	1 2025 a	Callghan, 2009 ⁵³
Quadriceps fatigue using surface electromyography- VMO final median frequency	N/A	7.4 Hz			rt 1 Dep	Callghan, 2009 ⁵³
Quadriceps fatigue using surface electromyography- VL final median frequency	N/A	6.5 Hz			ārtme	Callghan, 2009 ⁵³
Quadriceps fatigue using surface electromyography- RF final median frequency	N/A	10.5 Hz			nt 1 G	Callghan, 2009 ⁵³

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Page	55	of	59
------	----	----	----

Page 55 of 59			BMJ Open		d by copyrigi	136/bmjope	
2 3 4	Outcome tool	Score	Median	Minimum	nt, in Maximur du di	ף 20 20 20 20 20 20 20 20 20 20 20 20 20	Study
5	Quadriceps fatigue using surface electromyography-	N/A	2207.0 %/min*		ng	8 1	Callghan, 2009 ⁵³
6	VMO median frequency slope				for	26	3
7	Quadriceps fatigue using surface electromyography-	N/A	4000.0 %/min*		K	9 1	Callghan, 2009 ⁵³
8	VL median frequency slope				ies	18	
9	Quadriceps fatigue using surface electromyography-	N/A	2390.0 %/min*		re m	Ma 1	Callghan, 2009 ⁵³
10	RF median frequency slope				late	< 2	
11	Quadriceps isokinetic strength using a dynamometer-	N/A	33.9 Nm		ad t	0 2 1	Kean, 2010 ⁵⁸
12	absolute value				o t	ж. Г	
13	Quadriceps isometric strength using a dynamometer-	N/A	25.0 Nm	5.5	37.2 ຊິວ	o 3	McCarthy,2008 ⁶⁰ ; Brisson,
14	absolute value				an	<u>n</u>	2018 ⁵²
15	Quadriceps isometric strength using a dynamometer-	N/A	0.4 Nm/kg	0.3	0.5 0.5	a 2	Brisson, 2018 ⁵² ; Kean,
16	normalised to weight				latio	led	2010 ⁵⁸
17	Quadriceps isometric strength using a dynamometer-	N/A	1.5 %BW*height		<u> </u>	fr ¹	Kean, 2010 ⁵⁸
18	Normalised to body size				n.	ă	
19	Quadriceps power using a dynamometer	N/A	151.8 W/2.2 W/kg		ng,		Brisson, 2018 ⁵²
20	Rectus femoris fatigue slope	N/A	0.6 %/min		≥ .	1	McCarthy,2008 ⁶⁰
21	Rectus femoris initial median frequency	N/A	5.2 Hz		tra	1	McCarthy,2008 ⁶⁰
22	Single-leg standing balance test-mediolateral	N/A	0.3		ini		Takacs 2014 ⁶⁸
23	standard deviation				ng,	er ·	
24	Single-leg standing balance test-anteroposterior	N/A	0.5		an	y 1	Takacs, 2014 ⁶⁸
25	standard deviation				ā,	<u>ni</u> .	
20	Single-leg standing balance test-path length	N/A	20.2 cm		im	ğ 1	Takacs, 2014 ⁶⁸
27	Single-leg standing balance test-velocity	N/A	0.3 m/seconds		ilar	2 o 1	Takacs, 2014 ⁶⁸
20	Single-leg standing balance test-area	N/A	23.2 cm^2		teo		Takacs 2014 ⁶⁸
30		N// X	20.2 011		, in	hay '	
31	Six minute walk test	N/A	72.7 m	66.3 m	79.0 m č	9 2	Naylor, 2014 ⁶³
32	Stair ascent and descent time (seven steps (four of 15cm, three of 20cm))	N/A	2.6 seconds		ogies	1 2025	McCarthy,2004 ³²
22	Stopwatch-based 11- Step stair climb test	N/A	0.1 seconds	0.1 seconds	0.1 seconds	a 2	ljima, 2019 ⁵⁶
24 25	Star excursion balance test-Raw value	N/A	7.4 cm				Kanko, 2019 ⁷⁵
36	Star excursion balance test-Normalized (raw value/leg	N/A	85%		•	Da 1	Kanko 2019 ⁷⁵
30	length%)	1.077	0.0 /0				
38	Timed Up and Go test	N/A	1.1 seconds	1.1 seconds	1.1 seconds	ien 2	Alghadir. 201545
30						Ġ	
40						EZ	
41						ς.	
						×	

		BMJ Open		y copyright	36/bmjopen-	
Outcome tool	Score	Median	Minimum	, <u>j</u> Maximur t <u>d</u>	Number of	Study
Timed Up and Go test (as a ratio of the original	N/A	41.06%	37.5%	44.6% g	302 2	Naylor, 201463
measurement)				for	6 0	
Tinetti Performance-Oriented Mobility Assessment	NR	0.8		use	n 1	Parveen, 2017 ⁶⁴
scale- balance subscale	NP	0.6		is re	∞ ≤ 1	$Pan(een 2017^{64})$
scale- gait subscale		0.0		era Bat	ay '	1 alveen, 2017
Tinetti Performance-Oriented Mobility Assessment	NR	1.0		ed t	202 1	Parveen, 201764
scale- total				to t	3.	
Transferring time (distance of 2m to a chair and sit	N/A	1.7		ext	O W 1	McCarthy,2004 ³²
down, then, walk back to the start)	N 1/A	0.0.0		anc	nlo	
I runk-abduction-adduction during unipodal stance	N/A	2.9 °	2.6 °	2.9 ° d cho	ade 3	VandeStraaten, 2020
2-minute walk test	N/A	5.52m		ol.	-ĕ1	Subail and Chaudhar
		0102111		min	ron	2021 ⁵⁴
Vastus lateralis fatigue slope	N/A	0.5 %/min		ing	1	McCarthy,200860
Vastus lateralis initial median frequency	N/A	5.8 Hz		, <u>A</u>	1	McCarthy,200860
Vastus medialis oblique fatigue slope	N/A	0.8 %/min	•	tra	b 1	McCarthy,200860
Vastus medialis oblique initial median frequency	N/A	6.7 Hz		inir	j 1	McCarthy,200860
Verbal Rating Scale for pain		5.8		ıg,	P . 1	Alghadir, 201849
VAS for pain	100=worst	24.0 cm *	8.0 cm	28.0 cm	b 3	Alghadir. 2018 ⁴⁹ :
•				l sir	j.	Naylor,2014 ⁶³
WOMAC-pain	100=best	3.8	3.4	19.0 Di	ž 3	Angst, 2018 ³⁶ , Alghad
				ur te	on	2016 a ⁴⁶
WOMAC-function	100=best	3.1	3.1	18.7 č	Maj 3	Angst, 2018 ^{30,} Alghad
WOMAC-stiffness	100=best	52	4.8	56 00	9 , 2	Angst 2018 ³⁶
WOMAC-functional standing/walking	100=best	3.8	3.5	4.0 ig i	203 2	Angst, 2018 ³⁶
WOMAC-total	100-worst	18.7	11 7	20.8	15 6	Williams 2012 ^{43,} Alaba
	100-w013t	10.7	11.7	20.0	₽ ₽	2016 a ⁴⁸
SF-36-bodily pain	100=best	3.5	3.3	3.7	epa 2	Angst, 2018 ³⁶
SF-36-general health	100=best	2.1	2.0	2.2	1 2	Angst, 2018 ³⁶
SF-36-mental health	100=best	3.1	2.9	3.4	ient 2	Angst. 2018 ³⁶
SE-36-physical functioning	100=best	2.4	23	24	G 2	Angst 2018 ³⁶

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Page	57	of	59
------	----	----	----

	ьу сору	136/bmio				
Outcome tool	Score	Median	Minimum	Maximuntu di	Number of	Study
SF-36-role- physical	100=best	11.2	10.9	11.5	30 2	Angst, 2018 ³⁶
SF-36-social functioning	100=best	4.1	3.7	4.5	ი ი 2	Angst, 2018 ³⁶
SF-36-vitality	100=best	5.3	5.1	5.6 5 .6	n 2	Angst, 2018 ³⁶
20-meter walk test	N/A	0.9 seconds	0.2 seconds	1.6 secon é s	2	Motyl, 201378
30-second fast-paced walk test	N/A	3.2 m	0.8 m	5.7 m	2 2	Hoglund, 2019 ⁵⁵
40-meter fast-paced test	N/A	16.3		ed t	023 1	Suwit, 2020 ⁵⁰
30 seconds chair stand test	N/A	21.2		ush o te		Suwit, 2020 ⁵⁰
3D linear accelerations of the tibia during comfortable walking	N/A	0.3 g	0.1 g	0.8 g Xt and	12 12	Turcot, 2008 ⁷¹
3D linear accelerations of the femur during comfortable walking	N/A	0.3 g	0.1 g	1.0 g data	ad 12	Turcot, 2008 ⁷¹
3D linear accelerations of the tibia at fast speed	N/A	0.4 g	0.1 g	0.9 g 🚊	To 12	Turcot, 2008 ⁷¹
3D linear accelerations of the femur at fast speed	N/A	0.2 g	0.0 g	0.6 g ni	1 12	Turcot, 200871
9-step stair climb test	N/A			у, А	Ē.	Suwit, 2020 ⁰²

MDC: Minimum Detectable Change, OA: osteoarthritis, BMI: Body Mass Index, MRI: Magnetic Resonance Imaging, 3D: 3-dimensional, N/A: Not Applicable, 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, NRS: Numeric Rating Scale, OKS Oxford Knee Score, KAM: Knee adduction moment, KFM: Knee Flexion Moment, VMO: Vastus Medialis Oblique, VL: Vastus Lateralis, RF: Rectus Femoris, VAS: Visual Analog Scale, WOMAC: Western Ontario and McMaster Universities Arthritis Index, SF-36: 36-item Short Form health survey,

All the estimates are out of 100.

The median estimates are shaded in blue

BMJ Open

Location

reported

Page 2 to 3

Page 5 to 6

to 99

Page 7, line 96

Page 8 to 9, line

Page 7, line 108

Page 7 to 8, line

Page 8. line 125

Page 10, line 173

Page 10. line 170

Page 10, line

Page 10, line

Page 9 to 10.

Page 10, line

Page 10, line

Page 10 to 11.

line 184 to 196

Page 10 to11,

line 184 to 196

Page 10. line

184 to 188

184 to 188

line 152 to 171

173 to 182

173 to 176

131 to 150

112 to 123

to 112

to 131

to 182:

to 172

Page 1

where item is



46

1

47

PRISMA 2020 Checklist

Pag	je 59 of 59		BMJ Open	
1 2 3	PRIS	MA 20	020 Checklist	
4 5	Section and Topic	ltem #	Checklist item	Location where item is reported
6 7			model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s)	184 to 188
/ 8 9		13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup adaly and state regression).	Page 10 to 11, line 188 to 196
10		13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Not applicable
11 12	Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting base).	Not applicable
13 14 15	Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Not done due to highly heterogeneous data
17	RESULTS			
18 19 20	Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to t	Page 11, line 203 to 207 and figure 1
21		16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Figure 1
22 23 24	Study characteristics	17	Cite each included study and present its characteristics.	Page 11 to 15, line 218 to 256 and table 1
25 26 27 28	Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Page 11, line 212 to 216 and supplements1,2 and 3
29 30	Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) and the stimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Tables 2,3 and 4
31 32 33 34 35	Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Page 11 to 15, line 212 to 256 and table 1, supplements1,2 and 3
36 37 38 39 40		20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Page 15 to 21, line 258 to 361 and tables 2,3,4 and supplements 5,6
41 42 43 44 45		20c	Present results of all investigations of possible causes of heterogeneity among study results.	Page 15 to 21, line 258 to 361 and tables 2,3,4 and supplements
46				



47

PRISMA 2020 Checklist

		BMJ Open	Page 60 of 5
PRIS	6MA 20	020 Checklist	
Section and Topic	ltem #	Checklist item	Location where item is reported
			5,6
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Not applicable
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis a second	Not applicable
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Not done due to highly heterogeneous data
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Page 21 to 24, line 365 to 437
	23b	Discuss any limitations of the evidence included in the review.	Pages 23 to 24, line 421 to 437
	23c	Discuss any limitations of the review processes used.	Page 24, line 439 to 448
	23d	Discuss implications of the results for practice, policy, and future research.	Page 24, line 443 to 445 and 450 to 456
OTHER INFORMA	TION		
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Page 7, line 103 to 106
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Page 7, line 103 to 106
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	None
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the seven and the seven and the role of the funders or sponsors in the seven and the seven and the seven and the role of the funders of sponsors in the seven and the seven and the seven and the role of the funders of sponsors in the seven and the seven	None
Competing interests	26	Declare any competing interests of review authors.	None
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	None
From: Page MJ, McKe	nzie JE, I	Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic provide the provide the provided statement of the provide the provided statement of the provided st	oi: 10.1136/bmj.n71
		For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml	

BMJ Open

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

Journal:	BMJ Open
Manuscript ID	bmjopen-2022-063026.R1
Article Type:	Original research
Date Submitted by the Author:	25-Mar-2023
Complete List of Authors:	Silva, M Denika; University of Canberra, Faculty of Health; General Sir John Kotelawala Defence University, Department of physiotherapy Perriman, Diana; University of Canberra, Faculty of Health; ACT, Trauma and Orthopaedic Research Unit, Canberra Hospital Fearon, Angela; University of Canberra, Faculty of Health; Research Institute for Sport and Exercise, University of Canberra, Australia Couldrick, Jacqui; University of Canberra, Faculty of Health Scarvell, Jennie; University of Canberra, Faculty of Health; Trauma and Orthopaedic Research Unit, Canberra Hospital
Primary Subject Heading :	Rehabilitation medicine
Secondary Subject Heading:	Rehabilitation medicine
Keywords:	STATISTICS & RESEARCH METHODS, Knee < ORTHOPAEDIC & TRAUMA SURGERY, Musculoskeletal disorders < ORTHOPAEDIC & TRAUMA SURGERY

SCHOLARONE[™] Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our <u>licence</u>.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which <u>Creative Commons</u> licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

terez oni

Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies

Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

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

M Denika C Silva^{1,2,5}, Diana M Perriman^{1,2,3}, Angela M Fearon^{1,2,4}, Jacqui M Couldrick^{1,2}, Jennie M Scarvell^{1,2}

1. Faculty of Health, University of Canberra, Australia

2. Trauma and Orthopaedic Research Unit, Canberra Hospital, Canberra, Australia

3. College of Medicine and Health Sciences, Australian National University, Canberra, Australia

4. Research Institute for Sport and Exercise, University of Canberra, Australia

5. Department of Physiotherapy, Faculty of Allied Health Sciences, General Sir John Kotelawala Defence University, Werahera, Sri Lanka

Corresponding author: M Denika C Silva (ORCID ID:0000-00018648-471X)Address: Faculty of Health, University of Canberra, Bruce ACT 2617 AustraliaTel.: +61 (0) 433857153Email: denika.silva@canberra.edu.au

BMJ Open

2 3 4	1	Minimal important change and difference for knee osteoarthritis outcome
5 6	2	measurement tools after non-surgical interventions: a systematic review
7 8 9	3	
10 11	4	ABSTRACT
12 13	5	Objectives: To systematically review and provide estimates of the minimal important
14 15 16	6	change (MIC) and difference (MID) for outcome tools in people with knee osteoarthritis
10 17 18	7	(OA) after non-surgical interventions.
19 20	8	Design: A systematic review.
21 22 22	9	Data sources: MEDLINE, CINAHL, Web of Science, Scopus, and Cochrane
25 24 25	10	databases were searched up to September 21, 2021.
26 27	11	Eligibility criteria: We included studies that calculated MIC and MID using any
28 29	12	calculation method including anchor, consensus and distribution methods, for any
30 31 32	13	knee OA outcome tool after non-surgical interventions.
33 34	14	Data extraction and synthesis: We extracted reported MIC, MID and MDC
35 36	15	estimates. We used quality assessment tools appropriate to the studies' methods to
37 38 30	16	screen out low-quality studies. Values were combined to produce a median and range,
40 41	17	for each method.
42 43	18	Results: Forty-eight studies were eligible (anchor-k=12, consensus-k=1 and
44 45	19	distribution-k=35). MIC values for 13 outcome tools including Knee injury and
40 47 48	20	Osteoarthritis Outcome Score (KOOS)-pain, activities of daily living (ADL), quality of
49 50	21	life (QOL) and Western Ontario and McMaster Universities Arthritis Index (WOMAC)-
51 52	22	function were estimated using five high-quality anchor studies. MID values for 23 tools
53 54 55	23	including KOOS-pain, ADL, QOL and WOMAC-function, stiffness and total were
56 57	24	estimated using six high-quality anchor studies. One moderate quality consensus
58 59 60	25	study reported MIC for pain, function and global assessment. Minimum detectable

Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies

BMJ Open

change values (MDC, 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

Keywords: minimal important change; minimal important difference; minimum
 detectable change; knee osteoarthritis

rezien onz

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

2		
3 4	36	Strengths and limitations of this study
5 6 7	37	• We estimated MIC (within-group), MID (between-groups), and minimum
7 8 9	38	detectable change values using anchor, consensus, or distribution methods
10 11	39	papers respectively.
12 13	40	• This systematic review included a defined population of people with knee OA,
14 15 16	41	after non-surgical interventions.
17 18	42	High-quality anchor studies were used to contribute to MIC and MID estimates
19 20	43	were assessed using a credibility tool specially designed to evaluate anchor
21 22 22	44	method papers.
23 24 25	45	Consensus and distribution methods papers were evaluated using quality
26 27	46	assessment tools suited to each method.
28 29 30	47	• Median estimates were used to reflect the synthesised data due to data
30 31 32	48	skewness.
33 34		
35 36		
37 38		
39 40		
41 42		
42		
44 45		
46 47		
47 48		
49 50		
50 51		
52		
53		
54 55		
56		
57		
58 50		
60		

Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

49 INTRODUCTION

The efficacy of therapeutic interventions is commonly evaluated using statistical significance regardless of patient importance¹. 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 misinterpretation²⁻⁵. 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 inconsistently¹³. The concept was first described by Jaeschke, who studied patients' perceptions of pre- and post-intervention beneficial change⁸. This concept later included both improvement¹⁴ and worsening¹⁵.

Page 7 of 62

BMJ Open

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 used¹⁸. The receiver operating characteristic (ROC) method for deriving an estimate from anchor guestions 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 achieved⁶. In the distribution method, values are estimated statistically, based on the variance of the outcome data using half the standard deviation (SD)¹⁹, one standard error of measurement (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 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.

Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Knee OA is a common cause of pain and disability²³. 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 studies²⁴. MIC and MID values have been estimated for knee OA outcome
measures in these domains using anchor, consensus, and distribution methods with

Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

variable results. The variability of the methods used makes the selection of an
 appropriate estimate confusing for clinicians, researchers and regulatory bodies²⁻⁴.

> 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.

107 METHODS

This systematic review was designed and reported according to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement²⁵. The protocol was registered on PROSPERO (registration number: CRD42020215952).

, 3 113

114 Literature search

Five databases (MEDLINE, CINAHL, Web of Science, Scopus and Cochrane) were searched from each database's respective inception up to September 21, 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*

Page 9 of 62

BMJ Open

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 version 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 (DS and JC) and conflicts were resolved by a third reviewer (JS). 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 method³. 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, e.g. hip OA or other knee pathologies. Studies

Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.
BMJ Open

of MIC, MID and MDC were included even if they used a different terminology e.g.
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 i.e. 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 i.e. "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 upon a given level of confidence^{6,21}. MDC values for a 90 or 95 confidence interval 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 (2020)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 (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

Page 11 of 62

BMJ Open

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 not²⁸. Consensus studies were assessed using the Critical Appraisal Screening Program (CASP)-qualitative tool²⁹ which is designed to assess qualitative studies and is well-suited to consensus studies. The guality was rated as "high", "moderate" or "low" based on reliability and credibility³⁰. Distribution studies were evaluated using the National Heart, Lung and Blood Institute, National Institute of Health (NHLBI, NIH) quality assessment tool for before-after (pre-post) studies with no control group³¹ and ratings included "good", "fair" or "poor" based on reliability and credibility³¹. The quality assessment of included studies was performed by one reviewer (DS) and a random sample of 20% had an independent second review (AF or JC) to improve the accuracy³².

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). Additionally, 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 to 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.

BMJ Open

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 distributions³⁴. We excluded low-quality anchor studies from the median MIC and MID synthesis. Furthermore, we conducted a sub-analysis 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 and group level analyses, and in clinical settings^{9, 15}. Though we planned to conduct a sub-analysis 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
 standardized mean difference (SMD) was not performed due to the skewness of the
 distributions. Therefore, we reported median and range for each measure without

Page 13 of 62

1 2		
2 3 4	223	comparison. The planned meta-analysis was prevented by the skewness and
5 6 7	224	homogeneity of the data and a decision was made to follow a simple descriptive
7 8 9	225	approach using median estimates that was more accessible ²⁸ .
10 11	226	
12 13 14	227	RESULTS
14 15 16	228	
17 18	229	Study selection
19 20 21	230	The search yielded 2376 studies and after duplicates were removed, 1059
21 22 23	231	records were screened. Two hundred and seventeen studies were screened in full-
24 25	232	text review resulting in 48 eligible studies (k=48) (Figure 1). No further studies were
26 27	233	identified after checking the reference lists of included studies.
28 29 30	234	
31 32	235	Included studies calculated MIC and MID by anchor method (k=12), by
33 34	236	consensus method (k=1) and MDC by distribution method (k=35).
35 36 27	237	
37 38 39	238	The methodological quality of included studies
40 41	239	Most anchor studies (k=10) were of high quality and two were of low quality ^{35,36}
42 43	240	(Supplement 1). The quality of the consensus study (k=1) was moderate
44 45 46	241	(Supplement 2). The quality of distribution studies ranged from good (k=11) to fair
47 48	242	(k=24) (Supplement 3).
49 50	243	
51 52	244	Study characteristics of included studies
55 54 55	245	All the anchor studies were observational prospective cohort studies ^{14,15,35,37-43}
56 57	246	and two of them were nested within randomised controlled trials ^{36,44} (Table 1). The
58 59 60	247	number of participants in each study ranged from 41 to 1606. The mean age and body

BMJ Open

mass index ranged from 57.1 to 67.9 years and from 28.1 to 33 kg/m² respectively. The interventions used in these studies were rehabilitation, exercise, physiotherapy and non-steroidal anti-inflammatory drugs. The follow-up time ranged from seven days to one year. Eleven studies used global rating of change as the external anchor while one study⁴⁴ 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 study³⁵ considered the minimal improved response group as "my knee has got better" and another study⁴⁰ considered the response group as both "good" and "excellent" improvement groups (Supplement **4)**. terez oni

Page	15 of 62				BMJ Open by copyr	
1 2 3 4	260 261	Table 1: Characteristics of	fincluded studies		ight, includi	
5 6 7		Study	Number of participants; Age, years, Mean (SD)	Intervention (Follow-up)	Outcome tool ng 30 for 6 or 0	Calculation method (Measure extracted)
8		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)
9 10 11		Mills, 2016 ¹⁵ McCarthy, 2004 ³³	272; NR 15; 65.1 (11.3)	Non-surgical (26,52 weeks) Exercises (1 week)	KOOS-pain, activities of daily living, wall of life Aggregated locomotor function score; Eigherneter walk time;	Anchor (MIC and MID) Distribution (MDC95)
12 13		Harris, 2013 ³⁵	134; 59 (11)	Non-operative (3 months)	OKS- pain, function, summary; ICOAR KOOS-PS	Anchor (MIC, MID); Distribution (MDC90)
14 15 16		Klokker, 2016 ³⁶ Angst, 2018 ³⁷	41; NR 190; 66.1 (10.2)	exercises (12 weeks) Rehabilitation (3 months)	DAP کی تصفیح WOMAC-pain, function, functional stand والمنتخبة stiffness; SF-36-bodily pain, physical functioning, role- physical, vitality, social functioning, men	Anchor (MID) Anchor (MID); Distribution (MDC95)
17 18		Hmamouchi, 2012 ³⁸	173; 57.1 (10.1)	NSAID (6 weeks)	WOMAC-total	Anchor (MID)
19		Mostafaee, 2021 ³⁹	142; 58.3 (7.6)	Physiotherapy (4 weeks)	KOOS-pain, symptoms, activities of daily Bing sports and	Anchor (MID)
20 21		Ornetti, 2011 ⁴⁰	881; 67.1 (10.9)	NSAID (4 weeks)	WOMAC-function; Patient-reported functional disability on an	Anchor (MIC)
22 23		Singh, 2014 ⁴¹	137; NR	Conservative (2 weeks)	ICOAP-pain, constant pain, intermittent pain; KOOS-PS; KOOS-quality of life	Anchor (MIC); Distribution (MDC90)
24 25 26		Williams, 2012 ⁴²	159; NR	Exercise (2, 6, 12 months)	KOS-activities of daily living subscale; LEFS; WOMAC-total	Anchor (MID); Distribution (MDC90, MDC95)
20 27 28		Perrot, 2013 ⁴³ Lee, 2017 ⁴⁴	1606; 66.9 (9.0) 165; 61	Usual care (7 days) Thai Chi or Physical therapy (12 weeks)	Pain on a 0-10 NRS at rest and mevernment PROMIS-Short Forms- physical function, pain interference, depression, anxiety	Anchor (MIC) Anchor (MID)
29 30		Salottolo, 2018 ⁴⁵ Alghadir, 2015 ⁴⁶	27 (clinicians); NR 65; 54.9 (9.9)	NR NR	Pain, Function and Global asses쌺meat Timed Up and Go test 로 온	Consensus (MIC) Distribution (MDC95)
31 32		Alghadir, 2016 a ⁴⁷ Alghadir, 2016 b ⁴⁸	121; 52.9 (9.3) 121; 54.0 (9.3)	NR (48 hours) NR (48 hours)	WOMAC- Arabic version-pain, fundon, eotal NRS for pain- Arabic version	Distribution (MDC95) Distribution (MDC95)
33 34		Alghadir, 2017 ⁴³ Alghadir, 2018 ⁵⁰ Suwit, 2020 ⁵¹	97, 58 (11.5) 121; 52.9 (12.5) 55; 69.0 (11)	NR (1 week) NR (24 hours) NR (1 week)	VAS for pain; NRS for pain; Verbal Rating Scale for pain 30-seconds chair stand; 40-meter fast-paced test 9-step climb	Distribution (MDC95) Distribution (MDC95) Distribution (MDC90)
35 36		Baert, 2018 ⁵²	8 (12 knees); 68 (11)	NR (3 seconds)	Knee joint position sense test using analogue pclinometer; Knee force sense test using a handheld dyng mometer	Distribution (MDC95)
37 38		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 acceleromoter: BMI	Distribution (MDC95)
39 40 41		Callghan, 2009 ⁵⁴	55; 64 (14)	NR (24- 72 hours)	Quadriceps fatigue using surface electrom	Distribution (MDC95)

Page 16 of 62

			BMJ Open BMJ Open	
			op mj	
			vri,	
			ght	
			; - 20	
Suhail and	82:60.4 (6.7)	NR (48 hours)	2-minute walk test	Distribution (MDC
Chaudha, 2021 ⁵⁵	0_,0011(011)			2104.104.001 (11.2.01
Hoalund, 2019 ⁵⁶	20: 58.3 (8.05)	NR (2-7 davs)	30-second fast-paced walk test	Distribution (MDC
ljima, 2019 ⁵⁷	59; 59.1 (6.1)	NR (1 month)	Stopwatch-based 11- Step stair cimb st	Distribution (MDC
		, , , , , , , , , , , , , , , , , , ,		MDC95)
Jansen, 2021 ⁵⁸	103; NR	NR (1 month)	Knee image: bone density, joint space width, osteophytes,	Distribution (MDC
		, , , , , , , , , , , , , , , , , , ,	eminence height, and joint awgle $\overline{\mathbf{\Sigma}}$, ,
Kean, 2010 ⁵⁹	20; 53.6 (9.1)	NR (same day)	Quadriceps isokinetic strength, isometric streingth and percent	Distribution (MDC
			of voluntary activation using a dynamic ter	,
Klokker, 2015 ⁶⁰	20; 64 (6.6)	NR (1 week)		Distribution (MDC
McCarthy, 2008 ⁶¹	55; 64.9 (9.7)	NR (1 week)	Maximum isometric strength; Vastus media	Distribution (MDC
			lateralis, Rectus femoris-initial median freating y, Vastus	,
			medialis obligue, Vastus lateralis, Rectus ferende is fatigue slope	
Monticone, 202162	102; 69.1 (9.0)	NR (10 days)	Fremantle Knee Awareness Questionnair	Distribution (MDC
Nalbant, 202163	25: 62.3 (9.8)	NR (same day)	L-test C S	Distribution (MDC
Navlor, 2014 ⁶⁴	75: 67.6 (9.4)	NR (10 days)	VAS for pain: Six Minute Walk Test: Time and So for test:	Distribution (MDC
,			KOOS-pain, symptoms, activities of daily living, quality of life	2101110011011 (1112 0
Parveen, 201765	25: 50.5 (6.3)	NR (7 days)	Tinetti Performance-Oriented Mobility Assessment scale- total	Distribution (MDC
	20, 0010 (010)	rur (rudyb)	balance subscale gait subscale	
Peter 201866	135' NR	NR (1 week)	Animated activity questionnare	Distribution (MDC)
$Piva 2004^{67}$	25: 57 (8)	NR (same day)	Get up and Go test	Distribution (MDC)
Pratheen 201868	20, 27 (0)	NR (2 weeks)		Distribution (MDC)
Takace 201469	20, 70.0 (0.0)	NR (14 days)		Distribution (MDC)
Tevald 2016 ⁷⁰	25: 61 (9)	NR(7 days)	Hin abductor strength using a hand-herd domameter	Distribution (MDC)
Teo 202171	38: 65 1 (7 0)	NR (same day)	Frontal plane tibial alignment using inclingmenter	Distribution (MDC)
Turcot 2008 ⁷²	25: 63 9 (7.6)	NR (8 days)	3D linear accelerations of the tibia and femer during walking	Distribution (MDC)
Van der Straaten	19:65 1 (5 2)	NR (20 days)	Bange of motion (trunk abduction-adduction	Distribution (MDC)
202073	19, 00.1 (0.2)	Nix (20 days)	adduction bin flexion extension) and Canter of mass	
2020				
Vuruk 201474		NR (One day)	Do Motion Mobility Index (Turkish Yorson)	Distribution (MDC)
Turuk, 2014	40, NK	NR (One day)		
Payaud 100075	30:65 4 (7 7)	NP (2 wooks)	loint space width	Distribution (MDC)
Kanko 201076	30, 03.4(7.7)	Exercises (1 weeks)	Star oxcursion balance to 9	Distribution (MDC)
Muthu 201577	74, 57.7 (0.0)	Dhysiothoropy (4 wooks)	Brossure pain throshold using a handhold reserve algometer	Distribution (MDC)
Wutiu, 2015	73 (141 KIEES), 50.2	Filysiotherapy (4 weeks)		
	(0.7)	Noval that (6 months)	Cartilago volumo uging X rov. MPL (Equir. Datallo, Tibio)	Distribution (MDC)
Huptor 2006^{78}	//). 60 (1 /0 61	NOVELINEL (O HIOHIIS)	Calliage volume using A-ray, with (Femul, Fagelia, mola)	
Hunter, 2006 ⁷⁸	72; 58.9 (8.6) 15: 61 0 (7.8)	Storoid injection (20 days)	20 motor walk toot	Distribution (MDC

3 4

Page 17 of 62

BMJ Open

Protected by copyright, including for uses related to text and data mining, Al training, and similar	Érasmushogeschool .
ng, and similar technologies.	

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 studies⁴⁶⁻⁷⁵ 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/m², respectively. Studies estimated MDC90 and MDC95. The follow-up time ranged from the same day to one year. The MIC estimates derived using the anchor method The median MIC for 13 tools (with subscales) were calculated based on five 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 duration¹⁴. Most studies included one follow-up time (range: 7 days to 4 weeks), but one study¹⁵ reported MIC at two time-points (26 and 52 weeks). One anchor question

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

4
5
6
7
/
ð
9
10
11
12
13
14
15
16
17
10
10
19
20
21
22
23
24
25
26
27
28
20
29
50 21
51
32
33
34
35
36
37
38
39
40
<u>41</u>
יד 12
42
43
44
45
46
47
48
49
50
51
52
52
57
+

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).

294

1 2 3

> 295 Table 2: Minimal Important Change (MIC) values of knee osteoarthritis outcome tools derived using the anchor 296 method

Outcome tools	Score	Median	Minimum	Maximum	Number	Study
					estimates	
ICOAP-pain	100=worst	18.5			1	Singh, 2014 ⁴¹
ICOAP-constant pain	100=worst	18.7			1	Singh, 2014 ⁴²
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 ¹⁵ ; Singh, 2014 ⁴¹
KOOS-PS	100=worst	2.2			1	Singh, 201441
Pain on NRS at rest	100=worst	10.0	10	10	2	Perrot, 201343
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	2.		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 ¹⁴ ; Ornetti, 2011 ⁴⁰
	Estim	ates based	l on low-qua	lity studies		
ICOAP-pain	100=worst	13.4			1	Harris, 2013 ³⁵
KOOS-PS	100=worst	12.0			1	Harris, 201335
OKS-pain	100=best	17.3			1	Harris, 201335
OKS-function	100=best	10.6			1	Harris, 201335
OKS-summary	100=best	7.1			1	Harris, 2013 ³⁵

298

MIC: Minimal Important Change, 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, NRS: numeric rating scale, VAS: Visual Analog Scale, WOMAC: Western Ontario and McMaster
 Universities Arthritis Index, OKS: Oxford Knee Score

303304 All the scores are from 0 to 100

305306 The median estimates are shaded in blue

307 Estimates based on low-quality studies are shaded in grey308

57 58 59

55

|--|

The median MID for 23 tools were calculated based on six 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 method⁸⁰ 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 to 12 weeks), but two studies included MID at multiple time-points e.g. 2, 6 and 12 months^{15,42}. One anchor question was used in most studies, however, one study used multiple anchors⁴⁴.

Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies

Outcome tools	Score	Medi	Minimum	Maximum	Number of	Study
KOOS-nain	100=best	11 8	4	17	13	Mills 2016 ¹⁵
Nooo-pain	100-5031	11.0	-		10	Mostafaee, 2021 ³⁹
KOOS-activities of d	laily 100=best	2.5	-1.5	15.5	7	Mills, 2016 ¹⁵ ; Mostafaee, 2021 ³⁹
KOOS-quality of life	100=best	6.5	3	12.5	7	Mills, 2016 ¹⁵ ; Mostafaee, 2021 ³⁹
KOOS-sports/recrea	tion 100=best	17 5			1	Mostafaee, 2021 ³⁹
KOOS-symptoms	100=best	12.5			1	Mostafaee, 2021 ³⁹
KOS-activities of da	ily 100=best	6.4	2.2	10.6	6	Williams, 201242
living	,					,
LEFS	100=best	6.9	0.6	15.6	6	Williams, 201242
PROMIS Short Form	is- 100=worst		4.5	5.3	NR	Lee, 201744
physical function [*] PROMIS Short Form	is- 100=worst		4.2	4.2	NR	Lee, 201744
pain interference [*] PROMIS Short Form	is- 100=worst		6	6.2	NR	Lee, 2017 ⁴⁴
depression*						
PROMIS Short Form	is- 100=worst		4.5	6.6	NR	Lee, 201744
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-	100=best	81	59	10.2	2	Angst 2018 ³⁷
standing/walking	100 5000	•	0.0	10.2	-	, angot, 2010
WOMAC-total	100=best	6.8	1.6	16.8	8	Williams, 2012 ⁴² ; Hmamouchi, 2012 ³⁸
SF-36-bodily pain	100=best	9.3	8.2	10.4	2	Angst, 2018 ³⁷
SF-36-physical func	tion 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 functio	n 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	n 100=best	6.6	6	7.2	2	Angst, 2018 ³⁷
	Fstim	ates has	sed on low-a	uality studies		
			···· •			
DAP	100=worst	24			1	Klokker, 2016 ³⁶
ICOAP-pain	100=worst	7.8			1	Harris, 201335
KOOS-PS	100=worst	7.8			1	Harris, 201335
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
MID: Minimal Importan KOS: Knee Outcome S Measurement Informat McMaster Universities and constant osteoarth	t Difference, OA: oste Survey, LEFS: Lower I ion SystemSF36: 36-i Arthritis Index, DAP: I ıritis pain,KOOS-PS: I	oarthritis Extremity tem Shoi Dynamic Knee inju	, KOOS: Knee Functional So tf Form health weight-bearin ry and Osteoa	e injury and Os cale, PROMIS survey, WOM g Assessment irthritis Outcor	teoarthritis Out Patient-Repor AC: Western O of Pain, ICOAF ne Score Physic	come Score, ted Outcome ontario and P: Intermittent cal Function

331 *Estimates were reported as a range

The median estimates are shaded in blue

Estimates based on low-quality studies are shaded in grey

1	
2	333
4	334
5	335
6 7	226
8	330
9 10	337
11	
12	338
13 14	339
15 16	340
16 17	341
18	342
19 20	
20 21	
22	
23 24	
2 4 25	
26	
27 28	
29	
30	
31 32	
33	
34 25	
35 36	
37	
38 30	
40	
41	
42 43	
44	242
45	343 344
46 47	345
48	340
49 50	348 349
50 51	350
52	351
53 54	352
54 55	0.54
56	353
57 58	354
50 59	554

333 334 MID estimates based on the ROC method were reported and compared with 335 MID estimates for all methods. Four^{15,38,39,42} of six studies (67%) used the ROC 336 method. The ROC estimates were the same as the overall estimates in most cases 337 (Table 4). 338

Table 4: Comparison of receiver operating curve (ROC) method based Minimal Important Difference (MID) estimates with overall estimates

4	ROC method-based estimates, Median (range)	ROC method- based, Number of estimates (study)	Estimates regardless of calculation method	Number of estimates- Regardless of calculation
		estimates (study)	Median (range)	method, (study)
KOOS-pain	11.8 (4.0 to 18.0)	8	11.8	13
		(Mills, 2016 ¹⁵ ;	(4.0 to 17.0)	(Mills, 2016 ¹⁵ ;
		Mostafaee, 2021 ³⁹)		Mostafaee, 202139)
KOOS-activities of	12 (-1.5 to 155)	5	12.5	7
daily living		(Mills, 2016 ¹⁵ ;	(-1.5 to 15.5)	(Mills, 2016 ¹⁵ ;
		Mostafaee, 2021 ³⁹)		Mostafaee, 2021 ³⁹)
KOOS-quality of life	6.5 (3.0 to 12.5)	6	6.5	7
		(Mills, 2016 ¹⁵ ;	(3.0 to 12.5)	(Mills, 2016 ¹⁵ ;
		Mostafaee, 2021 ³⁹)		Mostafaee, 2021 ³⁹)
(OOS-	17.5	1	17.5	1Mostafaee,
sports/recreation		(Mostafaee, 2021 ³⁹)		2021 ³⁹)
KOOS-symptoms	12.5	1	12.5	1
		(Mostafaee, 2021 ³⁹)		(Mostafaee, 2021 ³⁹)
KOS-ADL	6.4 (2.2 to 10.6)	6	6.4	6
		(Williams, 2012 ⁴²)	(2.2 to 10.6)	(Williams, 201242)
LEFS	6.9 (0.6 to 15.6)	6	6.9	6
	, , , , , , , , , , , , , , , , , , ,	(Williams, 201242)	(0.6 to 15.6)	(Williams, 201242)
WOMAC-total	7.8 (1.6 to 16.8)	8	7.8	8
	· · · · · · · · · · · · · · · · · · ·	(Williams, 201242;	(1.6 to 16.8)	(Williams, 201242;
		Hmamouchi,	, , ,	Hmamouchi,
		,		,

ROC: Receiver Operating Curve, MID: Minimal Important Difference, KOOS: Knee injury and Osteoarthritis Outcome Score, KOS: Knee Outcome Survey, LEFS: Lower Extremity Functional Scale, WOMAC: Western Ontario and McMaster Universities Arthritis Index

348 All the estimates are out of 100.

The median estimates are shaded in blue 350

351 Shaded in green: Estimates of these MID values were based on the ROC method only

The effect of follow-up time on MIC and MID 353

- There were insufficient data to establish reliable rate estimates for the effect of 354
- 60 355 time. The MIC of Knee injury and Osteoarthritis Outcome Score (KOOS)-pain and

> KOOS-guality 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 (Supplement 5).

MIC values derived using the consensus method

One consensus study⁴⁵ 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 (Supplement 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 estimates³⁷. Six studies separately reported inter/intra-rater 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 group⁷⁸, the most painful and the least painful groups⁷⁰, and the groups that reported moderate improvement ("great deal better") and MCID improvement ("somewhat better")⁴¹. Most studies assessed the index (worst) knee,

Page 23 of 62

1

BMJ Open

2		
3 4	381	but one study based the estimate on all diseased knees (12 knees in 8 patients) ⁵² .
5 6	382	Regarding the time point of MDC estimation, most studies (k=39) reported MDC
7 8	383	estimates at one time-point only, but one study reported MDC estimates at three-time
9 10 11	384	points (2, 6 and 12 months) ⁴² .
12 13	385	
14 15	386	DISCUSSION
16 17	387	
18 19 20	388	This systematic review provided estimates for MIC and MID of knee OA
20 21 22	389	outcome tools after non-surgical interventions derived using anchor, consensus and
23 24	390	distribution methods respectively. This review is unique in that it provides estimates
25 26	391	for MIC, MID (based on high-quality studies) and MDC (from good to fair-quality
27 28 20	392	studies) of knee OA outcome tools after non-surgical interventions. MDC was reported
29 30 31	393	for a greater number of outcome measures (126) than for MIC (13) or MID (23) MID
32 33	39/	estimates based on the ROC method were similar to the overall median estimates
34 35	205	between the majority of MID studies used the DOC method. Although we found that
36 37	395	nowever, the majority of MiD studies used the ROC method. Although we found that
38 39	396	some MIC and MID appear to increase with follow-up time, this was not consistent.
40 41	397	
42 43	398	The estimates for MIC and MID reported in this review are lower than those
44 45	399	reported previously ⁸²⁻⁸⁴ . Previous reviews which included knee replacement
46 47 48 49 50	400	interventions82-84 produced higher estimates suggesting that knee replacement
	401	cohorts need more improvement to be satisfied. The MID values for WOMAC-pain and
51 52	402	function in this review ranged from 7.1 to 21 and 11.3 to 14.9 (out of 100) respectively;
53 54	403	compared to reviews of total knee replacements which reported values ranging from
55 56 57	404	4.0 to 47.9 and 1.8 to 33.0 (out of 100) ⁸²⁻⁸⁴ . This disparity may be due to differences
58 59	405	in disease severity which has been previously reported based on baseline pain
60		

BMJ Open

score^{6,15,85}. Therefore, our data is 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 OA⁸⁶⁻⁸⁷. Given that MID
is sensitive to disease type⁶, 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 over-or under-estimate the number of participants required for a given power.

⁹ 426

427 The ROC estimates were similar to the synthesised MID estimates which used 428 all calculation methods. Our synthesised estimates were based on a combination of 429 both the mean change⁸¹ and ROC methods⁸⁰. However, the ROC estimates are 430 reported to be more precise, and can be applied to both individuals and groups and

BMJ Open

2	
3	
4	
5	
6	
7	
8	
0	
9 10	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
∠ı วา	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
22	
22	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
-TU // T	
47	
4ð	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
50	
72	
υU	

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 nonresponders¹⁷. Therefore, we recommend using our median ROC based MID estimates
where possible.

436

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}. Additionally, 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 is required to determine the effect of follow up time on MIC and MID.

444

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 assessment⁴⁵, but, our anchor studies data suggest that MIC is highly variable (2.2 to 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.

7 450

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-

Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

harm trade-off method, described by Barrett in 2005⁹⁰. 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 inconvenience⁹¹ and the estimates are always compared to natural recovery⁹². However, only one study has reported SWE for people undergoing total knee replacement⁹³. 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%. These values are much higher than our MIC median estimate of 17 (9.1 to 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 one 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.

BMJ Open

2 3	481	
4 5	402	This review presents modian estimates for MIC, MID and MDC of people with
6 7	482	
8 9	483	knee OA following non-surgical interventions. A subset of MID estimates based on the
10 11	484	ROC method is reported and, where available, this estimate is recommended as the
12 13	485	most precise for both individual and group analyses and clinical settings. MDC
14 15 16	486	estimates are available for more outcome measures but are purely statistical and
17 18	487	arguably less applicable. This review clarifies the current understanding of MIC, MID
19 20	488	and MDC in the knee OA population. However, some estimates suggest considerable
21 22	489	heterogeneity and require careful interpretation.
23 24 25	490	
26 27	491	Author contributions
28 29 30 31 32 33 34	492	MDCS, DMP, AMF, and JMS conceptualised and designed the study. MDCS ran the
	493	database search. MDCS and JMC did the screening, full-text review and data
	494	acquisition. MDCS, DMP, AMF and JMS were involved in data analysis and
35 36	495	Interpretation of the data. MDCS wrote the first draft of the manuscript. All authors
37 38	496	reviewed the draft, read and approved the final manuscript.
39 40 41	497	
42 43	498	Role of the funding source
44 45	499	This research received no specific grant from any funding agency in the public,
46 47	500	commercial or not-for-profit sectors.
48 49 50	501	Conflict of interest
51 52	502	All the authors declare that there are no conflicts of interest in this work.
53 54	503	Ethics approval
55 56 57	504	This study is a systematic review and does not involve human participants. Therefore,
58 59	505	ethics approval is not applicable.
60		

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

2 3 4	506	Data	Sharing
5 6	507	All da	ta relevant to the study are included in the article or uploaded as supplementary
7 8	508	inforn	nation.
9 10 11	509		
12 13	510	REFE	RENCES
14 15	511	1.	Harris JD, Brand JC, Cote MP, Faucett SC, Dhawan A. Research Pearls: The
16 17 18	512		significance of statistics and perils of pooling. part 1: clinical versus statistical
19 20	513		significance. Arthroscopy 2017; 33: 1102-1112.
21 22	514	2.	Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et
23 24 25	515		al. Mind the MIC: large variation among populations and methods. J Clin
23 26 27 28 29 30	516		Epidemiol 2010; 63: 524-534.
	517	3.	de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal
30 31 22	518		changes in health status questionnaires: distinction between minimally
32 33 34 35 36 37 38 39 40 41	519		detectable change and minimally important change. Health Qual Life Outcomes
	520		2006; 4: 54-54.
	521	4.	de Vet HC, Terwee CB. The minimal detectable change should not replace the
	522		minimal important difference. J Clin Epidemiol 2010; 63: 804-805; author reply
42 43	523		806.
44 45	524	5.	de Vet HCW, Beckerman H, Terwee CB, Terluin B, Bouter LM. Definition of
46 47 48	525		clinical differences. J Rheumatol 2006; 33: 434.
49 50	526	6.	Katz NP, Paillard FC, Ekman E. Determining the clinical importance of
51 52	527		treatment benefits for interventions for painful orthopedic conditions. J Orthop
53 54 55 56 57 58 59	528		Surg Res 2015; 10: 24.
60			

Page 29 of 62

1 2 BMJ Open

3 4	529	7.	Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: a closer
5 6 7	530		look at the minimal clinically important difference (MCID). J Man Manip Ther
7 8 9	531		2012; 20: 160-166.
10 11	532	8.	Jaeschke R, Singer J, Guyatt GH. Measurement of health status: Ascertaining
12 13	533		the minimal clinically important difference. Control Clin Trials 1989; 10: 407-
14 15 16	534		415.
10 17 18	535	9.	Wells G, Anderson J, Beaton D, Bellamy N, Boers M, Bombardier C, et al.
19 20	536		Minimal clinically important difference module: summary, recommendations,
21 22 23 24	537		and research agenda. J Rheumatol 2001; 28: 452-454.
23 24 25	538	10.	Beard DJ, Harris K, Dawson J, Doll H, Murray DW, Carr AJ, et al. Meaningful
26 27	539		changes for the Oxford hip and knee scores after joint replacement surgery. J
28 29 30	540		Clin Epidemiol 2015; 68: 73-79.
30 31 32	541	11.	King MT, Dueck AC, Revicki DA. Can methods developed for interpreting
33 34	542		group-level patient-reported outcome data be applied to individual patient
35 36 37 38 39 40 41	543		management? Med Care 2019; 57.
	544	12.	Clement ND, Bardgett M, Weir D, Holland J, Gerrand C, Deehan DJ. What is
	545		the minimum clinically important difference for the WOMAC index after TKA?
42 43	546		Clin Orthop Relat Res 2018; 476: 2005-2014.
44 45	547	13.	King MT. A point of minimal important difference (MID): a critique of terminology
46 47 48	548		and methods. Expert Rev Pharmacoecon Outcomes Res 2014; 11: 171-184.
49 50	549	14.	Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, et al.
51 52	550		Evaluation of clinically relevant changes in patient reported outcomes in knee
53 54 55	551		and hip osteoarthritis: the minimal clinically important improvement. Ann
55 56 57 58 59 60	552		Rheum Dis 2005; 64: 29.

BMJ Open

15. Mills KA, Naylor JM, Eyles JP, Roos EM, Hunter DJ. 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: J Chronic Obstr Pulm Dis 2007; 4: 191-194. 17. Copay AG, Subach BR, Glassman SD, Polly DW, Jr., Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. Spine J 2007; 7: 541-546. Carrasco-Labra A, Devji T, Qasim A, Phillips MR, Wang Y, Johnston BC, 18. Devasenapathy N, Zeraatkar D, Bhatt M, Jin X, Brignardello-Petersen R. Minimal important difference estimates for patient-reported outcomes: a systematic survey. Journal of Clinical Epidemiology. 2021; 133:61-71. 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: 582-592. Wyrwich KW. Minimal important difference thresholds and the standard error of 20.

21. Martin Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 327: 307-310.

measurement: is there a connection? J. Biopharm. Stat. 2004; 14: 97-110.

- 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: 128-136.

Page 31 of 62

BMJ Open

- 3
457723.Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, et al. The global5
6578burden of hip and knee osteoarthritis: estimates from the global burden of
disease 2010 study. Ann Rheum Dis 2014; 73: 1323.
- 24. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, 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.
- 19
2058425.Moher D, Liberati A, Tetzlaff J, Altman DG, Group atP. Preferred Reporting21
22585Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement.23
24586Ann Intern Med 2009; 151: 264-269.
- ²⁶ 587
 ²⁷ 26. Dawson J, Boller I, Doll H, et al. Minimally important change was estimated for
 ²⁸ 588 the Manchester-Oxford Foot Questionnaire after foot/ankle surgery. J Clin
 ³⁰ 589 Epidemiol 2014;67(6):697-705.
- Devji T, Carrasco-Labra A, Qasim A, Phillips M, Johnston BC, Devasenapathy 27. N, 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.
- Hao Q, Devji T, Zeraatkar D, Wang Y, Qasim A, Siemieniuk RAC, et al. Minimal 28. important differences for improvement in shoulder condition patient-reported outcomes: a systematic review to inform a BMJ Rapid Recommendation. BMJ Open 2019; 9: e028777.
- 52 598 29. Programme CAS. CASP (Qualitative) Checklist. [online] 2018.
- 5459930.Long HA, French DP, Brooks JM. Optimising the value of the critical appraisal5556600skills programme (CASP) tool for quality appraisal in qualitative evidence5758601synthesis. 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. Haby MM, Chapman E, Clark R, Barreto J, Reveiz L, Lavis JN. What are the 32. 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: 514-517. 34. Higgins JPT CJ, Cumpston M, Li T, Page MJ, Welch VA. Cochrane handbook for systematic reviews of interventions. version 6.0 edition2019. 35. Harris KK, Dawson J, Jones LD, Beard DJ, Price AJ. 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. Klokker L, Christensen R, Waehrens EE, Bandak E, Bartholdy C, Bliddal H, et 36. 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. 37. Angst F, Benz T, Lehmann S, Aeschlimann A, Angst J. Multidimensional minimal clinically important differences in knee osteoarthritis after comprehensive rehabilitation: a prospective evaluation from the Bad Zurzach Osteoarthritis Study. RMD Open 2018; 4: e000685. Hmamouchi I, Allali F, Tahiri L, et al. Clinically important improvement in the 38. WOMAC and predictor factors for response to non-specific non-steroidal anti-

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

1 2			
3 4 5 6 7 8 9 10 11 12 13 14 15 16	627		inflammatory drugs in osteoarthritic patients: a prospective study. BMC Res
	628		Notes 2012;5:58
	629	39.	Mostafaee N, Nourollahi F, Mostamand J, Negahban H. Responsiveness and
	630		the minimal important change of Knee injury and Osteoarthritis Outcome Score
	631		in Persian patients with knee osteoarthritis following physiotherapy intervention.
	632		Physiother Theory Pract 2021: 1-10.
16 17 18	633	40.	Ornetti P, Dougados M, Paternotte S, Logeart I, Gossec L. Validation of a
19 20	634		numerical rating scale to assess functional impairment in hip and knee
21 22	635		osteoarthritis: comparison with the WOMAC function scale. Ann Rheum Dis
23 24 25	636		2011; 70: 740-746.
25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41	637	41.	Singh JA, Luo R, Landon GC, Suarez-Almazor M. Reliability and clinically
	638		important improvement thresholds for osteoarthritis pain and function scales: a
	639		multicenter study. J Rheumatol 2014; 41: 509-515.
	640	42.	Williams VJ, Piva SR, Irrgang JJ, Crossley C, Fitzgerald GK. Comparison of
	641		reliability and responsiveness of patient-reported clinical outcome measures in
	642		knee osteoarthritis rehabilitation. J Orthop Sports Phys Ther 2012; 42: 716-723.
	643	43.	Perrot S, Bertin P. "Feeling better" or "feeling well" in usual care of hip and knee
42 43	644		osteoarthritis pain: determination of cutoff points for patient acceptable
44 45	645		symptom state (PASS) and minimal clinically important improvement (MCII) at
46 47 48	646		rest and on movement in a national multicenter cohort study of 2414 patients
49 50	647		with painful osteoarthritis. J Pain 2013; 154: 248-256.
51 52	648	44.	Lee AC, Driban JB, Price LL, Harvey WF, Rodday AM, Wang C.
53 54	649		responsiveness and minimally important differences for 4 Patient-Reported
55 56 57	650		Outcomes Measurement Information System short forms: physical function,
58 59			
60			

Page 34 of 62

pain interference, depression, and anxiety in knee osteoarthritis. J Pain 2017;
18: 1096-1110.

 653 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: 424-425.

- 1565646.Alghadir A, Anwer S, Brismée J-M. The reliability and minimal detectable1617657change of Timed Up and Go test in individuals with grade 1-3 knee1819658osteoarthritis. BMC Musculoskelet Disord 2015; 16: 174-174.
- 47. Alghadir A, Anwer S, Igbal ZA, Alsanawi HA. 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. Disabil Rehabil 2016; 38: 689-694.
- 3166348.Alghadir AH, Anwer S, Iqbal ZA. The psychometric properties of an Arabic3233664numeric pain rating scale for measuring osteoarthritis knee pain. Disabil35665Rehabil 2016; 38: 2392-2397.
- Alghadir AH, Al-Eisa ES, Anwer S. Cross-cultural adaptation and psychometric
 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, Igbal A, Igbal ZA. Test-retest reliability, validity, and minimum detectable change of visual analog, numerical rating, and verbal rating scales for measurement of osteoarthritic knee pain. J Pain Res 2018; 11: 851-856.
- 54 673 51. Suwit A, Rungtiwa K, Nipaporn T. Reliability and validity of the Osteoarthritis
 55
 56 674 Research Society International minimal core set of recommended
 57

Page 35 of 62

1

BMJ Open

2 3 4	675		performance-based tests of physical function in knee osteoarthritis in
5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20	676		community-dwelling adults. Malays J Med Sci 2020; 27: 77-89.
	677	52.	Baert IAC, Lluch E, Struyf T, Peeters G, Van Oosterwijck S, Tuynman J, et al.
	678		Inter- and intrarater reliability of two proprioception tests using clinical
	679		applicable measurement tools in subjects with and without knee osteoarthritis.
	680		Musculoskelet Sci Pract 2018; 35: 105-109.
	681	53.	Brisson NM, Stratford PW, Maly MR. Relative and absolute test-retest
	682		reliabilities of biomechanical risk factors for knee osteoarthritis progression:
21 22 22	683		benchmarks for meaningful change. Osteoarthr Cartil 2018; 26: 220-226.
23 24 25	684	54.	Callaghan MJ, McCarthy CJ, Oldham JA. The reliability of surface
26 27 28 29 30 31 32 33 34 35 36 37	685		electromyography to assess quadriceps fatigue during multi joint tasks in
	686		healthy and painful knees. J Electromyogr Kinesiol 2009; 19: 172-180.
	687	55.	Suhail A, Chaudhary S. Test-retest reliability and minimum detectable change
	688		of 2-minute walk test among individuals with knee osteoarthritis. J Clin
	689		Diagnostic Res 2021; 15: 4-7.
37 38 39	690	56.	Hoglund LT, Folkins E, Pontiggia L, Knapp MW. The validity, reliability,
40 41	691		measurement error, and minimum detectable change of the 30-second fast-
42 43	692		paced walk test in persons with knee osteoarthritis: a novel test of short-
44 45 46	693		distance walking ability. ACR Open Rheumatol 2019; 1: 279-286.
40 47 48	694	57.	lijima H, Shimoura K, Eguchi R, Aoyama T, Takahashi M. Concurrent validity
49 50 51 52	695		and measurement error of stair climb test in people with pre-radiographic to
	696		mild knee osteoarthritis. Gait Posture 2019; 68: 335-339.
53 54 55	697	58.	Jansen MP, Welsing PMJ, Vincken KL, Mastbergen SC. Performance of knee
56 57	698		image digital analysis of radiographs of patients with end-stage knee
58 59 60	699		osteoarthritis. Osteoarthr Cartil 2021; 29: 1530-1539.

BMJ Open

3 4	700	59.	Kean CO, Birmingham TB, Garland SJ, Bryant DM, Giffin JR. Minimal
5 6 7 8 9	701		detectable change in quadriceps strength and voluntary muscle activation in
	702		patients with knee osteoarthritis. Arch Phys Med Rehabil 2010; 91: 1447-1451.
10 11	703	60.	Klokker L, Christensen R, Osborne R, Ginnerup E, Waehrens EE, Bliddal H, et
12 13	704		al. Dynamic weight-bearing assessment of pain in knee osteoarthritis: a
14 15 16	705		reliability and agreement study. Qual Life Res 2015; 24: 2985-2992.
17 18 19 20	706	61.	McCarthy CJ, Callaghan MJ, Oldham JA. The reliability of isometric strength
	707		and fatigue measures in patients with knee osteoarthritis. Man Ther 2008; 13:
21 22	708		159-164.
23 24 25	709	62.	Monticone M, Sconza C, Portoghese I, Nishigami T, Wand BM, Sorrentino G,
23 26 27 28 29 30 31 32 33 34 35 36	710		et al. Cross-cultural adaptation, reliability and validity of the Fremantle Knee
	711		Awareness Questionnaire in Italian subjects with painful knee osteoarthritis.
	712		Health Qual Life Outcomes 2021; 19: 114.
	713	63.	Nalbant A, Unver B, Karatosun V. Test-retest reliability of the L-test in patients
	714		with advanced knee osteoarthritis. Physiother Theory Pract 2021: 1-5.
37 38 30	715	64.	Naylor J M, Hayen A, Davidson E, Hackett D, Harris I A, Kamalasena G, et al.
39 40 41	716		Minimal detectable change for mobility and patient-reported tools in people with
42 43	717		osteoarthritis awaiting arthroplasty. BMC Musculoskelet Disord 2014: 15, 235.
44 45	718	65.	Parveen H, Noohu MM. Evaluation of psychometric properties of Tinetti
46 47 48	719		performance-oriented mobility assessment scale in subjects with knee
49 50	720		osteoarthritis. Hong Kong Physiother J 2017; 36: 25-32.
51 52	721	66.	Peter WF, de Vet HCW, Terwee CB. Reliability of the Animated Activity
53 54 55	722		Questionnaire for assessing activity limitations of patients with hip and knee
56 57	723		osteoarthritis. Musculoskeletal Care 2018; 16: 363-369.
58 59 60			

BMJ Open

Piva SR, Fitzgerald GK, Irrgang JJ, Bouzubar F, Starz TW. Get up and go test
in patients with knee osteoarthritis. Arch Phys Med Rehabil. 2004;85(2):284289.

727 68. Srimurugan Pratheep N, Madeleine P, Arendt-Nielsen L. Relative and absolute
 728 test-retest reliabilities of pressure pain threshold in patients with knee
 729 osteoarthritis. Scand J Pain 2018; 18: 229-236.

- 730 69. Takacs J, Carpenter MG, Garland SJ, Hunt MA. Test re-test reliability of centre
 731 of pressure measures during standing balance in individuals with knee
 732 osteoarthritis. Gait Posture 2014; 40: 270-273.
- 733 70. Tevald MA, Murray A, Luc BA, Lai K, Sohn D, Pietrosimone B. Hip abductor
 734 strength in people with knee osteoarthritis: A cross-sectional study of reliability
 735 and association with function. Knee 2016; 23: 57-62.
- 736
 736
 71. Tse CTF, Charlton JM, Lam J, Ho J, Bears J, Serek A, et al. Smartphone
 737
 737
 738
 738
 738
- Turcot K, Aissaoui R, Boivin K, Hagemeister N, Pelletier M, de Guise JA. Test-72. 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: 732-737.
- 73. van der Straaten R, Wesseling M, Jonkers I, Vanwanseele B, Bruijnes A, Malcorps J, 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.

1 2

3 4	747	74.	Yuruk Z O, Aytar A, Tüzün E H, Eker L, Yüksel I, Morton N A. Acceptability,
5 6	748		reliability and validity of the turkish version of the De morton mobility index in
7 8	749		elderly patients with knee osteoarthritis. Turk Geriatri Dergisi 2014; 17.
9 10 11	750	75.	Ravaud P, Giraudeau B, Auleley GR, Edouard-Noël R, Dougados M, Chastang
12 13	751		C. Assessing smallest detectable change over time in continuous structural
14 15	752		outcome measures: application to radiological change in knee osteoarthritis. J
16 17 18	753		Clin Epidemiol 1999; 52: 1225-1230.
19 20	754	76.	Kanko LE, Birmingham TB, Bryant DM, Gillanders K, Lemmon K, Chan R, et
21 22	755		al. The star excursion balance test is a reliable and valid outcome measure for
23 24 25	756		patients with knee osteoarthritis. Osteoarthritis Cartilage 2019; 27: 580-585.
25 26 27	757	77.	Mutlu EK, Ozdincler AR. Reliability and responsiveness of algometry for
28 29	758		measuring pressure pain threshold in patients with knee osteoarthritis. J Phys
30 31	759		Ther Sci. 2015;27(6):1961-1965.
32 33 34	760	78.	Hunter DJ, Conaghan PG, Peterfy CG, Bloch D, Guermazi A, Woodworth T, et
35 36	761		al. Responsiveness, effect size, and smallest detectable difference of Magnetic
37 38	762		Resonance Imaging in knee osteoarthritis. Osteoarthritis Cartilage 2006; 14
39 40 41	763		Suppl A: A112-115.
42 43	764	79.	Motyl JM, Driban JB, McAdams E, Price LL, McAlindon TE. Test-retest
44 45	765		reliability and sensitivity of the 20-meter walk test among patients with knee
46 47 48	766		osteoarthritis. BMC Musculoskelet Disord 2013; 14: 166
49 50	767	80.	Perkins NJ, Schisterman EF. The Inconsistency of "optimal" cutpoints obtained
51 52	768		using two criteria based on the receiver operating characteristic curve. Am J
53 54	769		Epidemiol 2006; 163: 670-675.
55 56 57			
58			

BMJ Open

- 770 81. Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic
 771 improvements. An illustration in rheumatology. Arch Intern Med 1993; 153:
 772 1337-1342.
- 773 82. Celik D, Coban O, Kilicoglu O. Minimal clinically important difference of
 774 commonly used hip-, knee-, foot-, and ankle-specific questionnaires: a
 775 systematic review. J Clin Epidemiol 2019; 113: 44-57.
- 83. Maredupaka S, Meshram P, Chatte M, Kim WH, Kim TK. Minimal clinically important difference of commonly used patient-reported outcome measures in total knee arthroplasty: review of terminologies, methods and proposed values. Knee Surg Relat Res 2020; 32: 19.
- 84. MacKay C, Clements N, Wong R, Davis AM. 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 Cartilage 2019; 27: 1408-1419.
- 785 85. Escobar A, Riddle DL. Concordance between important change and acceptable
 786 symptom state following knee arthroplasty: the role of baseline scores.
 787 Osteoarthritis Cartilage 2014; 22: 1107-1110.
- Angst F, Aeschlimann A, Steiner W, Stucki G. Responsiveness of the WOMAC 86. 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: 834-840.
- 792 87. Bellamy N. Pain assessment in osteoarthritis: experience with the WOMAC
 793 osteoarthritis index. Semin. Arthritis Rheum 1989; 18: 14-17.

3 4	794	88.	Schwartz CE, Andresen EM, Nosek MA, Krahn GL. Response shift theory:
5 6	795		important implications for measuring quality of life in people with disability. Arch
7 8 9	796		Phys Med Rehabil 2007; 88: 529-536.

- 797 89. Guyatt GH, Norman GR, Juniper EF, Griffith LE. A critical look at transition
 798 ratings. J Clin Epidemiol 2002; 55: 900-908.
- 1579990.Barrett B, Brown D, Mundt M, Brown R. Sufficiently important difference:1617800expanding the framework of clinical significance. Med Decis Making 2005; 25:1819801250-261.
- 802 91. Ferreira M. Research Note: The smallest worthwhile effect of a health
 803 intervention. J Physiother 2018; 64: 272-274.
- 804 92. Ferreira ML, Herbert RD, Ferreira PH, Latimer J, Ostelo RW, Nascimento DP,
 805 et al. A critical review of methods used to determine the smallest worthwhile
 806 effect of interventions for low back pain. J Clin Epidemiol 2012; 65: 253-261.
- ³³
 ³³
 ³⁴
 ³⁵
 ³⁶
 ³⁶
 ³⁷
 ³⁸
 ³⁹
 ³¹
 ³⁵
 ³⁶
 ³⁷
 ³⁷
 ³⁸
 ³⁹
 ³¹
 ³¹
 ³²
 ³³
 ³⁴
 ³⁵
 ³⁶
 ³⁷
 ³⁷
 ³⁸
 ³⁹
 ³⁷
 ³⁹
 ³¹
 ³¹
 ³²
 ³³
 ³⁴
 ³⁵
 ³⁶
 ³⁷
 ³⁷
 ³⁸
 ³⁹
 ³⁷
 ³⁷
 ³⁸
 ³⁷
 ³⁸
 ³⁹
 ³¹
 ³¹
 ³¹
 ³²
 ³⁴
 ³⁵
 ³⁶
 ³⁷
 ³⁶
 ³⁷
 ³⁷
 ³⁸
 ³⁹
 ³¹
 ³¹
 ³¹
 ³²
 ³⁴
 ³⁵
 ³⁵
 ³⁶
 ³⁷
 ³⁶
 ³⁷
 ³⁷
 ³⁸
 ³⁹
 ³¹
 ³¹
 ³¹
 ³²
 ³³
 ³⁴
 ³⁵
 ³⁵
 ³⁶
 ³⁶
 ³⁷
 ³⁶
 ³⁷
 ³⁷
 ³⁸
 ³⁹
 ³¹
 ³¹
 ³²
 ³⁵
 ³⁵
 ³⁶
 ³⁷
 ³⁶
 ³⁷
 ³⁷
 ³⁸
 ³⁹
 ³⁹
 ³⁹
 ³¹
 ³¹
 ³¹
 ³²
 ³³
 ³⁴
 ³⁵
 ³⁵
 ³⁶
 ³⁶
 <li
- 40 Bellamy N, Hochberg M, Tubach F, Martin-Mola E, Awada H, Bombardier C, et 94. 810 41 42 al. Development of multinational definitions of minimal clinically important 811 43 44 improvement and patient acceptable symptomatic state in osteoarthritis. 812 45 46 47 Arthritis Care Res (Hoboken) 2015; 67: 972-980. 813 48
- 49 TK. 95. **Kvien** Heibera Τ. Hagen KB. Minimal clinically important 814 50 51 improvement/difference (MCII/MCID) and patient acceptable symptom state 815 52 53 (PASS): what do these concepts mean?. Ann Rheum Dis. 2007;66 Suppl 54 816 55 56 3(Suppl 3):iii40-iii41. doi:10.1136/ard.2007.079798 817 57
- 58 59 60

1 2

14

32

96. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, et al. Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state. Ann Rheum Dis 2005; 64: 34-37.

Turner D, Schünemann HJ, Griffith LE, Beaton DE, Griffiths AM, Critch JN, et 97. al. Using the entire cohort in the receiver operating characteristic analysis maximizes precision of the minimal important difference. J Clin Epidemiol 2009; 62: 374-379.

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: 1589-1593.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

1 2		
3 4	830	List of tables
5 6	831	
7 8	832	Table 1: Characteristics of included studies
9 10 11	833	Table 2: Minimal Important Change (MIC) values of knee osteoarthritis outcome
12 13	834	tools derived using the anchor method
14 15	835	Table 3: Minimal Important Difference (MID) values of knee osteoarthritis outcome
16 17 18	836	tools derived using the anchor method
19 20	837	Table 4: Comparison of receiver operating characteristic (ROC) method based
21 22	838	Minimal Important Difference (MID) estimates with overall estimates
24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60		

1 2		
2 3 4	839	Figure legends
5 6	840	
7 8	841	Figure 1: Flow diagram of study selection (PRISMA, Preferred Reporting Items for
9 10 11	842	Systematic reviews and Meta-Analyses) ²⁵
12 13		List of Supplements
14 15		
16 17 18		Supplement 1: Quality of anchor studies assessed using the Devji (2020) credibility
19 20		instrument ²⁷
21 22		Supplement 2: Quality of consensus studies assessed using the Critical Appraisal
23 24 25		Screening Program (CASP)-qualitative checklist ²⁹
25 26 27		Supplement 3: Quality of distribution studies assessed using the National Institute
28 29		of Health (NIH) quality assessment tool for before-after (pre-post) studies with no
30 31 32		control group ³¹
33 34		Supplement 4: Anchor properties of included anchor studies
35 36		Supplement 5: The effect of follow-up time on A. Minimal Important Change (MIC)
37 38 20		and B. Minimal Important Difference (MID)
39 40 41		†: multiple MIC estimates using two different anchor questions; *: multiple MID
42 43		estimates using two different anchor questions and different calculation methods
44 45		KOOS: Knee injury and Osteoarthritis Outcome Score, QOL: Quality of Life, KOS:
46 47 48		Knee Outcome Survey, ADL: Activities of Daily Living, LEFS: Lower Extremity
49 50		Functional Scale, WOMAC: Western Ontario and McMaster Universities Arthritis
51 52		Index
53 54 55		Supplement 6: Minimum Detectable Change (MDC) values of knee osteoarthritis
56 57		outcome tools derived using the distribution method
58 59		

for peer teriew only

6

7

8

9

10 11

12

13 14

20 21

22 23

30

31

32

37 38

39

40 41

42 43

44

45 46

47

48 49

50 51

52

53 54

55 56

Records identified through Duplicate records removed database search (n=1317) (n=2376) dentification Covidence identified n=1283, MEDLINE n=595, CINAHL n=290, Web of Science n=677, manually identified: n=34 Scopus n=664, Cochrane n=150 Records after duplicates removal (n=1059) Reports screened Records excluded (n=1059) (n=842) Screening Reports excluded (n=169) Reasons: Reports assessed for eligibility Abstract only (n=31) (n=217) Research protocol (n=6) No original data (n=3) Book chapter (n=1) Wrong population (n=36) Studies included in review Wrong outcomes (n=66) Included (n=48) Unable to extract knee OA data as combined Anchor papers=12 with other population Consensus paper=1 (n=26) **Distribution papers=35** •
Page 46 of 62

		essed using the Devji (2020)			ď	ģ	
Study	Item 1*	Item 2#	Item 3#	Item 4 [#]	ling for u	00000000000000000000000000000000000000	Th qua
Angst, 2018	Yes	Impossible to tell	Definitely No	To a greater extent		a great extent	
Harris, 2013	Yes	To a great extent	Impossible to tell	Not so much	ſe_	SDefinitely No	
Hmamouchi, 2012	Yes	Not so much	Impossible to tell	To a greater extent	at	a greater extent	
Klokker, 2016	Yes	To a great extent	Impossible to tell	Definitely No	ed	Not so much	
Lee, 2017	Yes	Definitely Yes	Impossible to tell	To a greater extent	to	impossible to tell	
Mills, 2016	Yes	Definitely Yes	Not so much	To a greater extent	sho tex	Definitely Yes	
Mostafaee, 2021	Yes	To a great extent	To a great extent	To a great extent	t a	S Not so much	
Ornetti, 2011	Yes	To a great extent	Impossible to tell	To a greater extent	nd	Definitely No	
Perrot, 2013	Yes	To a great extent	Not so much	Definitely Yes	da	Son Not so much	
Singh 2014	Yes	To a great extent	Impossible to tell	Not so much	÷0.	n a great extent	
Sillyii, 2014	100				<u>a</u> _ :	<u>i o</u> a great extern	
Tubach, 2005	Yes	To a great extent	Impossible to tell	Definitely Yes	a mir	Not so much	
Tubach, 2014 Tubach, 2005 Williams, 2012 Item 1: Is the patient or r Item 2: Is the anchor eas Item 3: Has the anchor s	Yes Yes necessary proxy responsily understandable an hown a good correlati	To a great extent Not so much anding directly to both the P and relevant for patients or a sion with the PROM?	Impossible to tell Impossible to tell ROM and the anchor? necessary proxy?	Definitely Yes To a greater extent	a mining, Al training	Not so much Definitely Yes	
Item 1: Is the patient or r Item 2: Is the anchor eas Item 3: Has the anchor eas Item 4: Is the MID precis Item 5: Does the thresho * The responses to items # The responses to items	Yes Yes Necessary proxy respondent ily understandable and hown a good correlation e? Id or difference betwee s: Yes, No, Impossible s: Definitely yes, To a	To a great extent Not so much anding directly to both the P and relevant for patients or a ion with the PROM? een groups on the anchor us to tell great extent, Not so much, mat "Yoo" or "definitely you	Impossible to tell Impossible to tell ROM and the anchor? necessary proxy? sed to estimate the MID Definitely no, Impossible	Definitely Yes To a greater extent	a mining, Al training, ant and similar	Not so much Definitely Yes	w" quolity
Tubach, 2014 Tubach, 2005 Williams, 2012 Item 1: Is the patient or r Item 2: Is the anchor eas Item 3: Has the anchor s Item 4: Is the MID precis Item 5: Does the thresho * The responses to items # The responses to items ** overall quality: three o Low quality (credibility) s	Yes Yes Necessary proxy responsible willy understandable and hown a good correlation hown a good correlation e? Id or difference betwee s: Yes, No, Impossible s: Definitely yes, To a f the five criteria were tudies are shaded.	To a great extent Not so much and relevant for patients or a ion with the PROM? een groups on the anchor us to tell great extent, Not so much, met "Yes" or "definitely yes	Impossible to tell Impossible to tell ROM and the anchor? necessary proxy? sed to estimate the MID Definitely no, Impossible s' or "to a great extent", th	Definitely Yes To a greater extent	a mining, Al training, ant and similar techno	The paper is of "lo	w" quality
Tubach, 2014 Tubach, 2005 Williams, 2012 Item 1: Is the patient or r Item 2: Is the anchor eas Item 3: Has the anchor s Item 4: Is the MID precis Item 5: Does the thresho * The responses to items # The responses to items * overall quality: three of Low quality (credibility) s PROM: Patient-Reported	Yes Yes Yes hecessary proxy respo- sily understandable an hown a good correlati e? Id or difference betwe s: Yes, No, Impossible s: Definitely yes, To a f the five criteria were tudies are shaded.	To a great extent Not so much and relevant for patients or a tion with the PROM? een groups on the anchor us to tell great extent, Not so much, met "Yes" or "definitely yes	Impossible to tell Impossible to tell ROM and the anchor? necessary proxy? sed to estimate the MID Definitely no, Impossible " or "to a great extent", the erence (In this study mini-	Definitely Yes To a greater extent	a mining, Al training, ant on similar tot, ity If the second seco	To a great extent Not so much Definitely Yes To Definitely Yes To Definitely Yes To May 9, 2025	w" quality
Tubach, 2014 Tubach, 2005 Williams, 2012 Item 1: Is the patient or r Item 2: Is the anchor eas Item 3: Has the anchor s Item 4: Is the MID precis Item 5: Does the thresho * The responses to items # The responses to items ** overall quality: three o Low quality (credibility) s PROM: Patient-Reported	Yes Yes Yes hecessary proxy respo- hown a good correlati e? Id or difference betwe s: Yes, No, Impossible s: Definitely yes, To a f the five criteria were tudies are shaded.	To a great extent Not so much and relevant for patients or a ion with the PROM? een groups on the anchor us to tell great extent, Not so much, met "Yes" or "definitely yes	Impossible to tell Impossible to tell ROM and the anchor? necessary proxy? sed to estimate the MID Definitely no, Impossible s' or "to a great extent", the erence (In this study mini	Definitely Yes To a greater extent	a mining, Al training, ant did similar technologies.	To a great extent Not so much Definitely Yes mice? That the paper is of "lo May 9, 2025 at De	w" quality
Tubach, 2014 Tubach, 2005 Williams, 2012 Item 1: Is the patient or r Item 2: Is the anchor eas Item 3: Has the anchor s Item 4: Is the MID precis Item 5: Does the thresho * The responses to items # The responses to items ** overall quality: three o Low quality (credibility) s PROM: Patient-Reported	Yes Yes Necessary proxy responsily with understandable ar hown a good correlative and or difference betwee so diff	To a great extent Not so much anding directly to both the P ad relevant for patients or a ion with the PROM? een groups on the anchor us to tell great extent, Not so much, met "Yes" or "definitely yes MID-Minimal Important Diffe	Impossible to tell Impossible to tell ROM and the anchor? necessary proxy? sed to estimate the MID Definitely no, Impossible s" or "to a great extent", the erence (In this study mini	Definitely Yes To a greater extent	a mining, Al training, and similar technologies.	To a great extent Not so much Definitely Yes more management of the paper is of "lo May 9. 2025 at Departme	w" quality

				BMJ	Open		by copyrigh	136/bmjopen		
blement 2: Quality of c	onsensus stud	ies assessed u	sing the Critica	l Appraisal Scr	ening Program	n (CASP)-qualit	ative chec	-2022-		
udy Item 1 # uthor, ear)	Item 2 #	Item 3 #	Item 4 #	Item 5 #	ltem 6 #	ltem 7 #	ittem 8 Item for us	063026 on	Item 10	Tł q th
ottolo, Yes 018	Yes	Can't tell	Can't tell	Yes	Yes	Can't tell	Yess related to tex	18 May 2023. Do	Researcher discusses the the contribution the study makes to existing knowledge or	N
Item 7: Have ethic Item 8: Was the da Item 9: Is there a d Item 10: How valu	al issues been ata analysis suf lear statement able is the rese o items: Yes, C f study: high", "	taken into cons ficiently rigorou of findings? arch? an't Tell, No moderate" or "le	sideration? is? ow" was evalua	ated by the revi	ew team based	on how reliable	training, and simted	bmjopen.bmj.comach si	tudy was, without sp	ecific
# The response μ Overall quality								achnologies.	May 9, 2025 at Dej	on May 9, 2025 at De r technologies.

Jioup										± 6			
Study (Author, Year)	Criteria	C rit e ria	Criteria	Criteria	Overall								
	1 #	2 #	3 #	4 #	5 #	6 #	7 #	8 #	9 #	ວຼີ 10 ⁰	11 #	12 #	quality ^µ
Alghadir, 2017	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	r Yes	Yes	Other	Fair
Alghadir, 2018	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	S Yes	Yes	Other	Fair
Alghadir, 2016 b	Yes	No	Yes	Yes	Yes	Other	Yes	Other	Yes	e mes	Yes	Other	Fair
Alghadir, 2017	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	te as eg	Yes	Other	Fair
Baert, 2018	Yes	No	Yes	Yes	Other	Other	Yes	Other	Yes	10 H 48	Yes	Other	Fair
Baert, 2018	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	tex 50	Yes	Other	Fair
Brisson, 2018	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	v gæs gæs	Yes	Other	Good
Callaghan, 2009	Yes	No	Yes	Yes	Yes	Other	Yes	Other	Other		Yes	Other	Fair
Hoglund, 2019	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	ata Yes	Yes	Other	Good
Hunter, 2006	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Other	∃. Yegs	Yes	Other	Fair
ljima, 2019	Yes	Yes	Yes	Yes	Other	Yes	Yes	Other	Yes		Yes	Other	Fair
Jansen, 2021	Yes	Yes	Yes	Yes	Other	Yes	Yes	Other	Yes	Ţ PY	Yes	Other	Fair
Kanko, 2019	Yes	Other	Yes	tra Yes	Yes	Other	Good						
Kean, 2010	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	ii Y	Yes	Other	Fair
Klokker, 2015	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	ġ Ye	Yes	Other	Good
McCarthy, 2004	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	and Yes	Yes	Other	Fair
McCarthy, 2008	Yes	Yes	No	No	Other	Other	Yes	Other	Yes	sin Yes	Yes	Other	Fair
Monticone, 2021	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes		Yes	Other	Good
Motyl, 2013	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	r yes	Yes	Other	Fair
Mutlu, 2015	Yes	Other	Yes	ch Ye	Yes	Other	Good						
Nalbant, 2021	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	O Yes	Yes	Other	Good
Naylor, 2014	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	gie Yes	Yes	Other	Fair
Parveen, 2017	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	v, on Yengs	Yes	Other	Fair

Other

Other

Other

Other

Yes

Yes

Yes

Yes

Other

Other

Other

Other

Yes

Yes

Yes

Yes

Departmented Y Y Y Y

Yes

Yes

Yes

Yes

Other

Other

Other

Other

Fair

Fair

Fair

Fair

BMJ Open **Supplement 3:** Quality of distribution studies assessed using the National Institute of Health (NIH) quality assessment tool for the form-after (pre-post) studies with no control group³¹

29

30

31

32

33

34

35

36 37

38

44 45 46 Peter, 2018

Piva, 2004

Pratheep, 2018

Ravaud, 1999

Yes

Other

Other

Other

Other

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

d by copyrigh 136/bmjoper

and

EZ-LTA

Study (Author, Year)	Criteria 1 #	Criteria 2 #	Criteria 3 #	Criteria 4 #	Criteria 5 #	Criteria 6 #	Criteria 7 #	Criteria 8 #	Criteria 9 #	¥Sritevia ⊡.1085	Criteria 11 #	Criteria 12 #	Overall quality ۲
Suhail and Chaudhary, 2021	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	ng Yes	Yes	Other	Good
Suwit, 2020	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	r v Yess	Yes	Other	Fair
Takacs, 2014	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	ŠYes	Yes	Other	Fair
Tevald, 2016	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	reation ≦	Yes	Other	Fair
Takacs, 2014	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes		Yes	Other	Good
Tevald, 2016	Yes	No	Yes	Yes	Other	Other	Yes	Other	Yes		Yes	Other	Fair
Tse, 2021	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	te xt og	Yes	Other	Fair
Turcot, 2008	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	an Se	Yes	Other	Good

Criteria 2: Were eligibility/selection criteria for the study population prespecified and clearly described? p://bmjo

Criteria 4: Were all eligible participants that met the prespecified entry criteria enrolled? Criteria 5: Was the sample size sufficiently large to provide confidence in the findings? Criteria 6: Was the test/service/intervention clearly described and delivered consistently across the study population? Criteria 7: Were the outcome measures prespecified, clearly defined, valid, reliable, and assessed consistently across all study articipants?

Criteria 8: Were the people assessing the outcomes blinded to the participants' exposures/interventions?

Criteria 9: Was the loss to follow-up after baseline 20% or less? Were those lost to follow-up accounted for in the analysis?

.bmj Criteria 10: Did the statistical methods examine changes in outcome measures from before to after the intervention? Were stati pre-to-post changes?

Criteria 11: Were outcome measures of interest taken multiple times before the intervention and multiple times after the intervention gi.e., did they use an interrupted time-

Criteria 12: If the intervention was conducted at a group level (e.g., a whole hospital, a community, etc.) did the statistical analysis take into account the use of individual-level data to determine effects at the group level? # The responses to items: Yes, No, Other (CD, NR, NA)*, *CD, cannot determine; NA, not applicable; NR, not reported

The responses to items: Yes, No, Other (CD, NR, NA)*, "CD, cannot determine; NA, not applicable; NR, not reported P up overall quality: Good, Fair, or Poor was evaluated by the review team based on how reliable and credible each study was, without provide the tool of tool of the tool of tool of

			BMJ Open	136/bmjopen-20; 4 by copyright, in	
Supplement 4: Anchor	properties of included anchor	studies	Anobor		Poported terminology
Study	Description of anchor	Anchor question	properties/responses		in the study
Angst, 2018	GRC transition (5 points)	NR	much worse, slightly worse, almost equal, slightly better, much better	Difference betvererRhe "slightly better" group an∉the⊐almost equal" groug= MBD a _ ≦	Minimal Clinically Important Difference
Harris, 2013	GRC responses (3 points)	Compared to one week before your clinical visit, please indicate how much your knee problem has changed?	1. My knee has got better, 2. My knee has stayed the same, 3. My knee has got worse	Mean change in the difference in the difference in the between group and the between group and the difference in the din	Minimal Important Change,
				"my knee has free free same" and "my knee has free free free same" and	Difference
Hmamouchi, 2012	GRC transition (5 points)	How do you feel in general today as compared to six weeks earlier as far as your osteoarthritis is compared?	much better, slightly better, no change, slightly worse, much worse	Difference between effects of "slights" better' group and "no change" groeps= MID	Minimal Clinically Important Difference for improvement
Klokker, 2016	modified GRC (3 responses and a GRC spanning from -7 to +7)	Did your knee pain change since you entered this project?	unchanged, better, worse: if it is better or worse, bring up a scale spanning from -7 to +7 (-7 (worst) to +7 (best) scale)	Difference of a core of at least 2 (+2: little better: 2: little worse) and no change (score of 0, +1 (almost the same or hagely by better), -1 (worse a all) MID	Minimal Important Change
Lee, 2018	Multiple anchors: 36 item Short Form subscales- physical functioning, social role functioning, energy and vitality, bodily pain, mental health, physical role functioning, emotional role functioning, general health perception (0-100	NR	NR	Prospective changes for people achieving previously established MID on legacy comparators: MID in a single item anchoge.g. Patient Global Assessment was defined as range= MID to 1+NDD, MID in a multi-teem anchor e.g. SF36 range= MID to 2X	Minimally Important Difference

Page 51 of 62				BMJ Open	136/bm 1 by cop	
1					jopen-' yright,	
2 3 - 4	Study	Description of anchor	Anchor question	Anchor properties/responses	ਤ ਠੁੱ Definitions oਊMIQMID using transition guestion	Reported terminology
5 -		scale, 0: best, 100:		h	ng 1	
6		worse);			6 O	
/		WOMAC-pain and			n 1 use	
8		host 100 worso):			*S 7 8 N	
9 10		Back depression (0-63, 0:			elay Er	
10		best, 63: worse):			ted 20	
17		Perceived Stress scale			l to	
13		(0-40, 0: best, 40: worse);			tey Do	
14		Patient Global			ct a	
15		Assessment in a 0-10 cm			nd	
16		visual analogue scale			dat	
17		(higher number= greater				
18		perception of disease			nin om	
19		Six-minute walk test (in			ing	
20		meters).			× P	
21		20-meter walk test (in			l tra	
22		seconds)			aini <mark>D</mark> o	
23	Mills, 2016	Global transition	Two anchor question	much improved,	Mean change within slightly	Minimal Important
24		GRC-7-point Likert scale	1. Compared with when	moderately improved,	improved group= MIC	Difference-
25			I started this program,	slightly improved, not	And z	Mean change
20			my walking on level	changed, slightly worse,	in g	(Jaeschke) method
27			ground has (walking	moderately worse, much	Difference between participants who	Minimal Important
20			anchor) and	worse		Difference-
30			2. Compared with when		slightly, moderately, reuch improved	(Padalmaiar and Laria)
31			my knee had (knee		change or worse "pon-improved	(Redelineer and Long) method
32			health anchor).		aroura = MPD	ROC method (Youden
33			noulli anonorj.		910 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	method and 80%
34					at	specific method)
35	Mostafee, 2021	GRC- 7points Likert scale	How did your knee	1. very much worse; 2.	The difference between improved	Minimal Important
36			status change	much worse; 3. slightly	group (6 and 7) and to improved	Change
37			compared to the	worse; 4. no change; 5.	group (1,2,3,4 ar	
38			beginning of the	slightly better; 6. much	nt	
39					ា ក្រុ	
40					Ž.	
41					ТА	
42						
43		For p	eer review only - http://br	mjopen.bmj.com/site/abou	t/guidelines.xhtml	
44 45						
46						

Page 52 of 62

		BMJ Open	136/bmjopen-20 I by copyright, ir	
Description of anchor	Anchor question	Anchor properties/responses	Definitions or MICMID using transition guastion	Reported terminology in the study
GRC- (3points Likert scale, then again in a 4- point Likert scale	physiotherapy intervention? specific question NR. Two anchor questions. 1. The degree of improvement of global state 2. The degree of improvement of functional state	better; and 7. very much better The degree of improvement of global state (global MCII) on a 3-point Likert scale (worsened function, no change, improved function). Then, among the patients who improved, the degree of	75 th centile of the absolute change in score amogg patients who responded the mperiod werent as good or exceller (The roved group) = 0100000000000000000000000000000000000	Minimal Clinically Important Improvement
GRC (5 points)	Taking into account the pain due to your arthritis in the last 24 hours and the pain you experienced initially, how has your pain changed?	improvement was scored on a scale (poor, fair, good, excellent) Patients who said that their pain had improved were asked to rate the level of improvement on a 5-point Likert scale extending from very large improvement to no improvement at all	75 th percentile of the distribution of the pain intensite difference (day 0 to 7) for patient≵considering their improvement to be at least moderate=ofIIC g, and simila	Minimal Clinically Important Improvement
GRC (5 points)	Since the last time you completed the survey 2 weeks ago, would you say your knee arthritis is:	A great deal better, somewhat better, about the same, somewhat worse, a great deal worse	Mean change beaveeo baseline and follow-up of the group who said somewhat better = MIC	Minimal Clinically Important Difference for improvement
	Description of anchor GRC- (3points Likert scale, then again in a 4- point Likert scale GRC (5 points) GRC (5 points)	Description of anchorAnchor questionGRC- (3points Likert scale, then again in a 4- point Likert scalephysiotherapy intervention? specific question NR. Two anchor questions. 1. The degree of improvement of global state 2. The degree of improvement of functional stateGRC (5 points)Taking into account the pain due to your arthritis in the last 24 hours and the pain you experienced initially, how has your pain changed?GRC (5 points)Since the last time you completed the survey 2 weeks ago, would you say your knee arthritis is:	BMJ OpenDescription of anchorAnchor questionGRC (3points Likert scale, then again in a t point Likert scalephysiotherapy intervention? specific question NR. Two anchor questions. 1. The degree of improvement of global state (2) The degree of improvement of global state (2) The degree of improvement of global state (2) The degree of improvement of functional statebetter; and 7. very much betterGRC (5 points)The degree of improvement of global state (2) The degree of improvement of functional statebetter; and 7. very much betterGRC (5 points)Taking into account th pain due to your arthritis in the last 24 hours and the pain guesbetter; and 7. very much betterMC (5 points)Taking into account th pain due to your arthritis in the last 24 hours and the pain guesbetter; and 7. very much betterMC (5 points)Since the last time your or hanged?Detter; and 7. very much the degree of improved function, no change, improved function). Then, among the patients who improved the scale (point Likert scale to a scale (poor, fair, good, excellent)MC (5 points)Since the last time you werk sago, would you ary your knee arthritis its:A great deal better, scale the same, somewhat yours)	BMJ Open To specify the server properties/responses Definitions of properties/responses GRC- (3points Likert scale point Likert scale point Likert scale physiotherapy intervention? better; and 7, very much better 75 th centile of the gates of improvement of global state (global MCII) on a 3-point Likert scale (worsened function, no functional state 75 th centile of the gates of improvement a state GRC (5 points) Taking into account the pain due to your arthritis in the last 24 hours and the pain you experienced initially, how has your pain changed? Taking into account the pain due to your arthritis in the last 24 hours and the pain you experienced initially, how has your pain changed? A great deal better, somewhat better, about the same, somewhat better, and an improvement at all Mean change between to function of the same, somewhat better, about the same, somewhat better, and a worse Mean change between the sate of the same addition of the same addition the same, somewhat better, about the same, somewhat better, account the same, somewhat better, about

Page 53 of 62				BMJ Open	1136/bmjc 1 by copy	
1 2 3	Study	Description of anchor	Anchor question	Anchor	ri: pp ght n n, 20 Definitions o£MIC/MID using	Reported terminology
4				properties/responses	transitio	in the study
5	Tubach, 2005	GRC (5 points)	Response to NSAID	None= no good at all,	75 th centile of the charge in score of	Minimal Clinically
6			treatment	ineffective drug; poor=	the group who region redeat as "good"	Important Improvement
7				some effect but	using logistic gegression=MIC	
8				unsatisfactory; fair=	8 N 8	
9				reasonable effect but	rela Elay	
10				satisfactory effect with	antec	
11				occasional episodes of		
12				pain or stiffness:	b te	
13 14				excellent= ideal		
15				response, virtually pain-	and	
16				free	da	
17					ta .	
18	Williams, 2012	GRC questionnaire on a	Not specified	GRC on a 15- level scale.	Difference betveen Subjects who	Minimal Clinically
19		15-level scale (15 points)	To rate the extent to	The GRC ranges from 1=	perceived "imprevention" (those with a	Important Difference
20			which they perceive	a very great deal better to	GRC between 🕇 (very great deal	
21			their condition as	8=about the same to 15=	better) and 5 (somewhat better))	
22			having changed over	very great deal worse	from subjects who perceived "not	
23			time		improved (thoses with between 6 (a	
24					deal worke) EMID	
25						
26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41	MIC: Minimal Important and McMaster Universi	Change, MID: Minimal Importaties Arthritis Index, SF36: Shor	ant Difference, GRC: Glob t Form-36, NSAID: Non-S	al Rating of Change, NR: No teroid Anti-Inflammatory Drug	t Reported, CI: Coaling on May 9, 2025 at Department GEZ-LTA	DMAC: Western Ontario
43		Farm	oor rovious only http://b	mionon hmi com/site/shaw	t/quidalinas vhtml	
44		For p	eer review only - http://b	mjopen.omj.com/site/abou	guidennes.xntmi	
45						
46						

Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.









†: multiple MIC estimates using two different anchor questions; *: multiple MID estimates using two different anchor questions and different calculation methods

KOOS: Knee injury and Osteoarthritis Outcome Score, QOL: Quality of Life, KOS: Knee Outcome Survey, ADL: Activities of Daily Living, LEFS: Lower Extremity Functional Scale, WOMAC: Western Ontario and McMaster Universities Arthritis Index

Page 55 of 62

1	
2	
3	
4	
5	
6	
7	
, 0	
0	
9	
10	
11	
12	
13	
14	
15	
16	
1/	
18	
19	
20	
21	
22	
23	
24	
25	
20	
27	
28	
29	
30	
31	
32	
33 24	
34 25	
35	
30	
3/	
30 20	
39 40	
40 41	
41 42	
4∠ ⊿2	
45 11	
44 15	
40	

Supplement 6: Minimum Detectab	e Change (MDC) values of knee osteoarthritis outcome tools derived using the distribution	on r P eth
--------------------------------	---	-----------------------

Supplement 6: Minimum Detectable Change (MDC) valu	ies of knee os	teoarthritis outcome tools	s derived using the	distribution reeth	り ンダー 06	
Outcome tool	Score	Median	Minimum	Maximu	Number of	Study
Aggregated locomotor function	N/A	2.3 seconds				McCarthy,2004
Animated Activity Questionnaire	100=best	11.2		ISe	1	Peter, 2018
BMI using a scale and a stadiometer	N/A	2.6 kg/m ²		Ű,	z 1	Brisson, 2018
Bone density-femur mean lateral	N/A	0.5 mm	0.4 mm	0.6 mm 🛱 🖽	2	Jansen, 2021
Bone density-femur mean medial	N/A	0.6 mm Al eq	0.5 mm Al eq	0.7 mm Al 🗟 🕉 🖁	2	Jansen, 2021
Bone density-tibia mean lateral	N/A	2.6 mm Al eq	2.4 mm Al eq	2.8 mm Al g 2	2 2	Jansen, 2021
Bone density-tibia mean medial	N/A	0.8 mm Al eq	0.7 mm Al eq	1 mm Alena	J 2	Jansen, 2021
Cartilage volume-lateral tibia	N/A	0.6ml	0.5 ml	0.6 ml 🛱 🖁	2	Hunter, 2006
Cartilage volume-medial Tibia	N/A	0.6 ml	0.4 ml	0.7 ml	2	Hunter, 2006
Cartilage volume-femur	N/A	1.1 ml	0.8 ml	1.3 ml	2	Hunter, 2006
Cartilage volume-patella	N/A	0.9 ml	0.7 ml	1.1 ml 🔐 🖸 🤅	2	Hunter, 2006
Center of mass mediolateral displacement during unipodal stance using 3D motion analysis	N/A	0.01°	0.01°	0.02 ° m .	from 3	VandeStraaten, 20
De Motion Mobility Index	100=best	8.0	7.3	8.7 ng	2	Yuruk, 2014
Eight-meter walk time	N/A	1.4 seconds		, A	1	McCarthy,2004
Eminence lateral (knee image)	mm	2.2 mm	2.2 mm	2.2 mm 式	2	Jansen, 2021
Eminence medial (knee image)	mm	1.8 mm	1.7 mm	1.8 mm	2	Jansen, 2021
Get up and Go test	N/A	1.4 seconds	1.2 seconds	1.5 secon	2	Piva, 2004
Fremantle Knee Awareness Questionnaire	100=worst	14.4		a g	1	Monticone,202
Frontal plane tibial alignment using smartphone inclinometer	N/A	3.7°	~V,	Ind si	1	Tse, 2021
Frontal plane tibial alignment using a manual inclinometer	N/A	3.2°	C	nilar t	1	Tse, 2021
Hip abductor strength using a hand-held dynameter	N/A	0.3 Nm/kg	0.3 Nm/kg	0.3 Nm/k	s 2	Tevald, 2016
Hip flexion-extension during unipodal stance using 3D motion analysis	N/A	2.7 °	2.2 °	4.6 ° hnolo	3 9	VandeStraaten, 20
ICOAP-constant pain	100=worst	51.7	49.6	53.8 9	2	Singh, 2014
ICOAP-intermittent pain	100=worst	49.8	48.7	50.8 %	й 2	Singh, 2014
ICOAP-pain	100=worst	46.6	23.0	49.6	4 3	Singh, 2014, Harris,
KAM impulse in 3D motion analysis	N/A	4.9 Nm*s			1	Brisson, 2018
KAM impulse in 3D motion analysis	N/A	0.4 %BW*HT*s			1	Brisson, 2018
Knee force sense test using a handheld dynamometer-reposition error 20°	N/A	20.6 N	13.2 N	26.7 N	3	Baert, 2018

		BMJ Open		y copyright	36/hmiopen	
Outcome tool	Score	Median	Minimum	Maximur	Number of	Study
Knee force sense test using a handheld	N/A	38.0 N	27.8 N	43.3 N G	3	Baert, 2018
dynamometer-reposition error 45°		00 0 N			ภั	D
Knee force sense test using a handheid	N/A	32.2 N	22.5 N	49.5 N L S	5 3	Baert, 2018
Knee force sense test using a handheld	N/A	32 7 N	30 0 N	337N 7	x Z 3	Baert 2018
dynamometer-reposition error all		02.11	00.011	elat		20010, 2010
Knee joint angle	degrees	2.1°	1.9°	2.3° ed as	2	Jansen, 2021
Knee joint space width	mm	1 mm	0.8 mm	5 ق 5 1.2 mm ج 2	2 2	Jansen, 2021
Knee joint space width-lateral	nm 🚬	1.7 mm	1.4 mm	2.0 mm 2.0	2	Jansen, 2021
Knee joint space width-medial	mm	0.6 mm	0.5 mm		2	Jansen, 2021
Knee joint space width-minimum	mm	0.8 mm	0.6 mm	0.9mm d 0	2	Jansen, 2021
KOOS-activities of daily living	100=best	19.1	17.4	20.8	± ∎ 2	Naylor, 2014
KOOS-pain	100=best	18.6	17	20.2 n i	2	Naylor, 2014
KOOS-quality of life	100=best	27.8	22.4	39.0 in g	4	Naylor, 2014; Singh,
KOOS-symptoms	100=best	20.2	2.9	24.1 ≥	3	Naylor, 2014
KOOS-PS	100=worst	28.3	16.0	35,5 fra	3	Harris, 2013 Singh.
Knee joint position sense test-analogue inclinometer- reposition error 70°	N/A	8.0 °	4.0 °	8.0 ° ini g,	3	Baert, 2018
Knee joint position sense test-analogue inclinometer- reposition error 45°	N/A	4.0 °	3.0 °	8.0 ° and s	3	Baert, 2018
Knee joint position sense test-analogue inclinometer- reposition error 20°	N/A	4.0 °	3.0 °	4.0 ° mila	3	Baert, 2018
Knee joint position sense test-analogue inclinometer- reposition error all	N/A	3.0 °	3.0 ⁰	4.0 ° tech	3 4	Baert, 2018
KOS-activities of daily living	100=best	15.8*	10.5	21.0 0	6	Williams, 2012
L-test	N/A	5.28 seconds		ogić	3 1	Nalbant, 2021
Load frequency using a triaxial accelerometer	N/A	4.3 steps/day		š.	5 1	Brisson, 2018
LEFS	100=best	18.3*	14.8	22.6		Williams, 2012
NKS- pain	100=worst	16.5	13.3	19.6	2 2	Aignadir, 2016b, Alg 2018
Osteophytes-Femur lateral (using knee image)	N/A	7.8 mm ²	5.4 mm ²	10.3 mm ²	2	Jansen, 2021
Osteophytes-Femur medial (using knee image)	N/A	12.3 mm ²	11.2 mm ²	13.4 mm ²	2	Jansen, 2021
Osteophytes-Tibia lateral (using knee image)	N/A	10.5 mm ²	9.4 mm ²	11.6 mm ²	ק ד 2	Jansen, 2021

Page 57 of 62			BMJ Open		d by copyrig	136/bmjope	
2					ht, in	n-20	
3 4	Outcome tool	Score	Median	Minimum		Number of Sestimates	Study
5	Osteophytes-Tibia medial (using knee image)	N/A	11.6 mm ²	11 mm ²	12.2 mm ²	2	Jansen, 2021
6 7	OKS-summary	100=best	6.1	6.0	6.2 for us	26 on	Alghadir, 2017, Harris, 2013
8	Peak KAM in 3D motion analysis	N/A	0.2 Nm/kg		ës	8 1	Brisson, 2018
9	Peak KAM in 3D motion analysis	N/A	1.3 %BW*HT				Brisson, 2018
10	Peak KFM in 3Dmotion analysis	N/A	0.4 Nm/kg		atec	20 1	Brisson, 2018
12	Peak KFM in 3D motion analysis	N/A	2.3 %BW*HT		d to	123 . 1	Brisson, 2018
13 14	Pelvic abduction-adduction during unipodal stance using 3D motion analysis	N/A	3.1 °	2.8 °	3.5 ° text a	Do 3	VandeStraaten, 2020
15 16	Per cent of voluntary muscle activation using a dynamometer	N/A	6.6%		nd dat	load 1	Kean, 2010
17	Pressure pain threshold-knee- using an algometer	N/A	131.8 kPa	92.9 kPa	196.3 kPag	1 10	Pratheep, 2018
18 19	Pressure pain threshold-medial heel using a handheld pressure algometer	N/A	1.3 lb		nining		Mutlu, 2015
20 21	Pressure pain threshold- medial knee using a handheld pressure algometer	N/A	1.2 lb		, Al tr	:tp://b	Mutlu, 2015
22 23	Quadriceps fatigue using surface electromyography- VMO initial median frequency	N/A	11.0 Hz		aining	mjope	Callghan, 2009
24 25	Quadriceps fatigue using surface electromyography- VL initial median frequency	N/A	10.0 Hz	Ch.	g, and	n. bm	Callghan, 2009
26 27	Quadriceps fatigue using surface electromyography- RF initial median frequency	N/A	9.0 Hz		simil	j.com	Callghan, 2009
28 29	Quadriceps fatigue using surface electromyography- VMO final median frequency	N/A	7.4 Hz		ar tec		Callghan, 2009
30 31	Quadriceps fatigue using surface electromyography- VL final median frequency	N/A	6.5 Hz		hnolo	lay 9,	Callghan, 2009
32 33	Quadriceps fatigue using surface electromyography- RF final median frequency	N/A	10.5 Hz		ogies.	1 2025	Callghan, 2009
34 35	Quadriceps fatigue using surface electromyography- VMO median frequency slope	N/A	2207.0 %/min*			at De	Callghan, 2009
36	Quadriceps fatigue using surface electromyography- VL median frequency slope	N/A	4000.0 %/min*			1 partm	Callghan, 2009
37 38 39	Quadriceps fatigue using surface electromyography- RF median frequency slope	N/A	2390.0 %/min*			1 Ient GEZ	Callghan, 2009
40 41						:-LTA	

		BMJ Open		y copyright	36/bmjopen-	
Outcome tool	Score	Median	Minimum	Maximur <mark>e</mark>	Number of	Study
Quadriceps isokinetic strength using a dynamometer- absolute value	N/A	33.9 Nm		ing fo	3026	Kean, 2010
Quadriceps isometric strength using a dynamometer- absolute value	N/A	25.0 Nm	5.5	37.2 uses	on 3	McCarthy,2008; Bri 2018
Quadriceps isometric strength using a dynamometer- normalised to weight	N/A	0.4 Nm/kg	0.3	0.5 relat	May 2	Brisson, 2018; Ke 2010
Quadriceps isometric strength using a dynamometer- Normalised to body size	N/A	1.5 %BW*height		ed to	1 2023.	Kean, 2010
Quadriceps power using a dynamometer	N/A	151.8 W/2.2 W/kg		sho tex		Brisson, 2018
Rectus femoris fatigue slope	N/A	0.6 %/min		t ges	M 1	McCarthy,2008
Rectus femoris initial median frequency	N/A	5.2 Hz		nd o	oad 1	McCarthy,2008
Single-leg standing balance test-mediolateral standard deviation	N/A	0.3		bol . lata m	led fro	Takacs, 2014
Single-leg standing balance test-anteroposterior standard deviation	N/A	0.5		ining.		Takacs, 2014
Single-leg standing balance test-path length	N/A	20.2 cm		, ≥		Takacs, 2014
Single-leg standing balance test-velocity	N/A	0.3 m/seconds	•	tra	b 1	Takacs, 2014
Single-leg standing balance test-area	N/A	23.2 cm ²	0.	ining,	joper 1	Takacs, 2014
Six minute walk test	N/A	72.7 m	66.3 m	79.0 m g	g 2	Naylor, 2014
Stair ascent and descent time (seven steps (four of 15cm, three of 20cm))	N/A	2.6 seconds		d sim		McCarthy,2004
Stopwatch-based 11- Step stair climb test Star excursion balance test-Raw value	N/A N/A	0.1 seconds 7.4 cm	0.1 seconds	0.1 seconotici to	2 91 1	ljima, 2019 Kanko, 2019
Star excursion balance test-Normalized (raw value/leg length%)	N/A	8.5 %		chnol	May 9	Kanko, 2019
Timed Up and Go test	N/A	1.1 seconds	1.1 seconds	1.1 secono	, 20 2	Alghadir, 2015
Timed Up and Go test (as a ratio of the original measurement)	N/A	41.06%	37.5%	44.6% 9	25 2 25 at	Naylor, 2014
Tinetti Performance-Oriented Mobility Assessment scale- balance subscale	NR	0.8			Depai	Parveen, 2017
Tinetti Performance-Oriented Mobility Assessment scale- gait subscale	NR	0.6			tmen 1	Parveen, 2017
Tinetti Performance-Oriented Mobility Assessment scale- total	NR	1.0			t GEZ	Parveen, 2017

Page 59 of 62

		BMJ Open		у сору	36/bmj	
				rright, in	open-20	
Outcome tool	Score	Median	Minimum	Maximur <mark>e</mark>	Number of 8 estimates	Study
Transferring time (distance of 2m to a chair and sit	N/A	1.7		ng fo	3 026	McCarthy,200
Trunk-abduction-adduction during unipodal stance using 3D motion analysis	N/A	2.9 °	2.6 °	2.9 ° use	on 3	VandeStraaten,
2-minute walk test	N/A	5.52m		s relat	May	Suhail and Chauc 2021
Vastus lateralis fatigue slope	N/A	0.5 %/min		asm ed t	202 1	McCarthy,200
Vastus lateralis initial median frequency	N/A	5.8 Hz		to to	<u>ω</u> 1	McCarthy,200
Vastus medialis oblique fatigue slope	N/A	0.8 %/min		ext	OW 1	McCarthy,200
Vastus medialis oblique initial median frequency	N/A	6.7 Hz		and		McCarthy,200
Verbal Rating Scale for pain		5.8		l da	ade 1	Alghadir, 201
VAS for pain	100=worst	24.0 cm *	8.0 cm	28.0 cm mi.	d fron	Alghadir, 201 Naylor,2014
WOMAC-pain	100=best	3.8	3.4	19.0 g ,	n http	Angst, 2018, Alg 2016 a
WOMAC-function	100=best	3.1	3.1	18.7 H trai		Angst, 2018 [,] Algl 2016 a
WOMAC-stiffness	100=best	5.2	4.8	5.6 nin	<mark>8</mark> 2	Angst, 2018
WOMAC-functional standing/walking	100=best	3.8	3.5	4.0 ق	2	Angst, 2018
WOMAC-total	100=worst	18.7	11.7	20.8 nd ,	<u> </u>	Williams, 2012 [,] Al 2016 a
SF-36-bodily pain	100=best	3.5	3.3	3.7 nila	2 2	Angst, 2018
SF-36-general health	100=best	2.1	2.0	2.2 Tr	n 2	Angst, 2018
SF-36-mental health	100=best	3.1	2.9	3.4 chn	May 2	Angst, 2018
SF-36-physical functioning	100=best	2.4	2.3	2.4 <mark>e</mark> .	9 2	Angst, 2018
SF-36-role- physical	100=best	11.2	10.9	11.5 g	202 2	Angst, 2018
SF-36-social functioning	100=best	4.1	3.7	4.5	5 2 at 2	Angst, 2018
SF-36-vitality	100=best	5.3	5.1	5.6	D 2	Angst, 2018
20-meter walk test	N/A	0.9 seconds	0.2 seconds	1.6 seconds	2	Motyl, 2013
30-second fast-paced walk test	N/A	3.2 m	0.8 m	5.7 m	2 2	Hoglund, 201
40-meter fast-paced test	N/A	16.3			nt 1 G	Suwit, 2020
30 seconds chair stand test	N/A	21.2			Ë 1	Suwit, 2020

				ght, ir	5 5 5	
Outcome tool	Score	Median	Minimum	Maximur	Number of estimates	Study
3D linear accelerations of the tibia during comfortable walking	N/A	0.3 g	0.1 g	0.8 g g for	1 2	Turcot, 2008
3D linear accelerations of the femur during comfortable walking	N/A	0.3 g	0.1 g	1.0 g uses	12	Turcot, 2008
3D linear accelerations of the tibia at fast speed	N/A	0.4 g	0.1 g	0.9 g e g	5 12	Turcot, 2008
3D linear accelerations of the femur at fast speed	N/A	0.2 g	0.0 g	0.6 g te	5 12	Turcot, 2008
9-step stair climb test	N/A			d to		Suwit, 2020

MDC: Minimum Detectable Change, OA: osteoarthritis, BMI: Body Mass Index, MRI: Magnetic Resonance Imaging, 3D: 3-dimensional, N/A: Not Applicable, ICOAP: Intermittent and constant osteoarthritis pain, KOOS: Knee injury and Osteoarthritis Outcome Score, KOOS-PS: Knee injury and Osteoarthritis Outcome Score Physical MDC: Minimum Detectable Change, OA: osteoarthritis, BMI: Body Mass Index, MRI: Magnetic Resonance Imaging, 3D: 3-dimetry ben and constant obseconthritis Juckows Escore Nyosical Events and constant obseconthritis Juckows Escore Nyosical Function Short form, KOS: Knee Outcome Survey, LEFS: Lower Extremity Functional Scale, NRS: Numeric Rating Scale, OKS and the source of the sour

d by copyr l 136/bmjop

Pa	ge 61 of 62		BMJ Open Cfed C 10.11	
1 2 2	PRIS	SMA 2	2020 Checklist	
4 5	Section and Topic	ltem #	Checklist item	Location where item is reported
6	TITLE		ng 00	
7	Title	1	Identify the report as a systematic review.	Page 2
9	ABSTRACT			
10	Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Page 2 to 3
11	INTRODUCTION			
12	Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Page 5 to 7
13 14	Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Page 7, line 103 to 107
16	METHODS			
17	Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Page 8 to 9, line 133 to 153
19 20	Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to perify studies. Specify the date when each source was last searched or consulted.	Page 7, line 116 to 121
21 22	Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Page 7 to 8, line 121 to 131
23 24	Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how mane were screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation to see in the process.	Page 8, line 125 to 131
25 26 27 28	Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Page 8 -9, line 133 to 153;
29 30	Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Page 10, line 173 to 182
31 32		10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Page 10, line 173 to 176
33 34	Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, hoge make reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Page 9 to 10, line 163 to 183
35 36	Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Page 10, line 184 to 188
37	Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Page 10, line 184 to 188
39 40 ⊿1		13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Page 10 to 11, line 184 to 196
41 42 43		13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Page 10 to11, line 184 to 196
44 45		13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent/of statistical heterogeneity band/software package(s) used.	Page 10, line 184 to 188



1

PRISMA 2020 Checklist

Page 62 of 62

PRISMA 2020 Checklist opposite Section and topic tem 136 Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup and/ysec on the synthesized results). meta-regression). 137 Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup and/ysec on the synthesized results). meta-regression). 138 Describe any methods used to assess robustness of the synthesized results. meta-regression). 139 Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome. meta-regression). 140 Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome. meta-regression). 150 Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome. meta-regression. 161 Certainty 16a Describe the results of the search and selection process, from the number of records identified in the search to the foregreating of the included in the review, ideally using a flow diagram. meta-regrease of the search and selection process. 170 Cite each included study and present its characteristics. meta-regrease of the search and selection process. meta-regrease of the search and selection process. 171 Cite each included study and present its characteristics.	Location where item is reported Not done Not applicable Not applicable Not applicable Page 11, line 203 to 207 and figure 1 Figure 1 Page 11 to 15, line 218 to 256
Section and Topic Item Checklist item 13e Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis) meta-regression). 13f Describe any methods used to assess robustness of the synthesized results. meta-regression). 13f Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting big sees) meta-regression). 14f Describe any methods used to assess cretainty (or confidence) in the body of evidence for an outcome. meta-regression). 15 ctribut 15 Describe any methods used to assess cretainty (or confidence) in the body of evidence for an outcome. meta-regression). 16 assessment Describe the results of the search and selection process, from the number of records identified in the search to the feed of the dise diagram. meta-regression. 17 Cite actudies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were distributed. meta-regression. 18 Study criteristics 17 Cite each included study and present its characteristics. meta-regression. 19 For all outcomes, present, for each included study. meta-regression (e.g. confidence/credible interval), ideally using structured tables or plots. meta-regression (e.g. confidence/credible inte	Location where item is reported Not done Not applicable Not applicable Not applicable Page 11, line 203 to 207 and figure 1 Figure 1 Page 11 to 15, line 218 to 256
13e Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup and ysig meta-regression). 13f Describe any sensitivity analyses conducted to assess robustness of the synthesized results. assessment Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting bias assessment) 15 Describe any methods used to assess critainty (or confidence) in the body of evidence for an outcome. RESULTS The search and selection process, from the number of records identified in the search to serve included in the review, ideally using a flow diagram. 16b Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were serve included. 18b Cite each included study and present its characteristics. 17 Cite each included study and present its characteristics. 18b Present assessments of risk of bias for each included study. 17 Cite each included study. 28 Results of individual studies 29 Present assessments of risk of bias for each included study. 20 For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) and precision (e.g. confidence/credible interval), ideally using structured tables or plots. 29 Results of syntheses 20a For each synthesis	Not done Not applicable Not applicable Not applicable Page 11, line 203 to 207 and figure 1 Figure 1 Page 11 to 15, line 218 to 256
B 13f Describe any sensitivity analyses conducted to assess robustness of the synthesized results. C P Reporting bias assessment 14 Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting bigses) Image: Constraint of the synthesis (arising from reporting bigses) I Certainty 15 Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome. Image: Constraint of the synthesis (arising from reporting bigses) I Certainty assessment 15 Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome. Image: Constraint of the synthesis (arising from reporting bigses) I Certainty assessment 16 Describe the results of the search and selection process, from the number of records identified in the search to the foregrad from reporting bigses of the search and selection process, from the number of records identified in the search to the foregrad from reporting bigses of the search and selection process. Image: Constraint of the search and selection process. 16 Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were grad from reporting bigses of the search and selection process. Image: Constraint of the search and selection process. 2 Risk of bias in study 17 Cite estudies that might appear to meet the inclusion criteria, but which were excluded, and expla	Not applicable Not applicable Not applicable Page 11, line 203 to 207 and figure 1 Figure 1 Page 11 to 15, line 218 to 256
9 Reporting bias 10 14 Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting bias escent assessment 15 Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome. Image: Control of the search of the search of the search and selection process, from the number of records identified in the search to the results of the search and selection process, from the number of records identified in the search to the record of the search and selection process, from the number of records identified in the search to the record of the search and selection process, from the number of records identified in the search to the record of the search and selection process, from the number of records identified in the search to the record of the recor	Not applicable Not applicable Page 11, line 203 to 207 and figure 1 Figure 1 Page 11 to 15, line 218 to 256
1 Certainty assessment 15 Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome. Image: Configure to the search and selection process, from the number of records identified in the search to the search of studies included in the review, ideally using a flow diagram. 1 Study selection 16a Describe the results of the search and selection process, from the number of records identified in the search to the search of the search and selection process, from the number of records identified in the search to the search of the search and selection process, from the number of records identified in the search to the search of the review, ideally using a flow diagram. 16b Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were the included. 17 Cite each included study and present its characteristics. Image: Configure to the search of the search and selection process. 2 Risk of bias in studies 18 Present assessments of risk of bias for each included study. 2 Results of individual studies 19 For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) and the precision (e.g. confidence/credible interval), ideally using structured tables or plots. Image: Confidence/credible interval), ideally using structure tables or plots. 2 Results of syntheses 20a For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.<	Not applicable Page 11, line 203 to 207 and figure 1 Figure 1 Page 11 to 15, line 218 to 256
RESULTS Study selection 16a Describe the results of the search and selection process, from the number of records identified in the search to included in the review, ideally using a flow diagram. 16b Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they we flow flow flow flow flow flow flow flow	Page 11, line 203 to 207 and figure 1 Figure 1 Page 11 to 15, line 218 to 256
14 Study selection 16a Describe the results of the search and selection process, from the number of records identified in the search to from the included in the review, ideally using a flow diagram. 16 Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were for the search of the search and selection process, from the number of records identified in the search to for the search of studies 18 Study 17 Cite each included study and present its characteristics. 17 2 Risk of bias in studies 18 Present assessments of risk of bias for each included study. 18 2 Results of studies 19 For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an of precision (e.g. confidence/credible interval), ideally using structured tables or plots. 18 2 Results of synthesis, briefly summarise the characteristics and risk of bias among contributing studies. 19 2 For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. 10 2 10 For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. 10 2 10 10 10 10 10 10 2 10 10 10 10	Page 11, line 203 to 207 and figure 1 Figure 1 Page 11 to 15, line 218 to 256
17 16b Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded. 18 Study 17 Cite each included study and present its characteristics. 19 characteristics 17 Cite each included study and present its characteristics. 20 Risk of bias in studies 18 Present assessments of risk of bias for each included study. 21 Results of individual studies 19 For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots. 22 Results of syntheses 20a For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. 50 23 10 For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. 50 24 20a For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. 50 24 20a For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. 50 25 26 20a For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. 50	Figure 1 Page 11 to 15, line 218 to 256
18 Study characteristics 17 Cite each included study and present its characteristics. 20 Risk of bias in studies 18 Present assessments of risk of bias for each included study. 21 Results of individual studies 19 For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an offfeet estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots. 22 Results of syntheses 20a For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Page 11 to 15, line 218 to 256
22 Risk of bias in studies 18 Present assessments of risk of bias for each included study. 23 18 Present assessments of risk of bias for each included study. 24 18 Present assessments of risk of bias for each included study. 25 Results of individual studies 19 For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots. 26 Results of syntheses 20a For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. sinitian of syntheses 29 20a For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. sinitian of syntheses 29 20a For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. sinitian of syntheses 29 20a For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. sinitian of syntheses 20a For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. sinitian of synthese	and table 1
25 Results of 19 For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an office estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots. 27 Results of syntheses 20a For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. since of the synthesis 28 30 30 and the synthesis since of the synthesis since of the synthesis 30 30 and the synthesis since of the synthesis since of the synthesis since of the synthesis	Page 11, line 212 to 216 and supplements1,2 and 3
 ²⁷ Results of syntheses ²⁹ Solution ²⁰ Por each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. ³⁰ Por each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. 	Tables 2,3 and 4
	Page 14 to 18, and table 1, supplements1,2 and 3
20b Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary statigete and its precision 32 (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect. 33 ••••••••••••••••••••••••••••••••••••	Page 18 to 24; tables 2,3,4 and supplements 5,6
20c Present results of all investigations of possible causes of heterogeneity among study results.	Page 18 to 24;, tables 2,3,4 and supplements 5,6
Y 20d Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results. G	Not applicable
42 Reporting biases 21 Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Not applicable
43 Certainty of 22 Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed. 44 evidence	Not applicable
4 ⁴ DISCUSSION For peer review only - http://bmjopen.bmj.com/site/about/guidennes.xhtmi	

Pa	age 63 of 62		BMJ Open	cted by	10 113	
1 2 3	PRIS	SMA 2	2020 Checklist	v copyright		
4 5	Section and Topic	ltem #	Checklist item	includ	00-200	Location where item is reported
6 7	Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	na for	30006	Page 26 to 30, line 399 to 503
0 9 1(23b	Discuss any limitations of the evidence included in the review.	uses	an 18	Pages 29 to 30, line 483 to 491
11		23c	Discuss any limitations of the review processes used.	Eras	May 20	Page 29 to 30, line 491 to 494
13 14		23d	Discuss implications of the results for practice, policy, and future research.	mush d to te	23 D	Page 30, line 496 to 503
15	OTHER INFORMA	TION		xt og	WC	
16 17	Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the	estcho and da	w was not registered.	Page 7, line 110 to 114
18 19	8	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	ol . Ita min	ad from	Page 7, line 110 to 114
20	þ	24c	Describe and explain any amendments to information provided at registration or in the protocol.	i i	3	None
21	Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the terms of the funders or sponsors in the terms of the funders or sponsors in the terms of te	e Ne re	view.	None
22	Competing interests	26	Declare any competing interests of review authors.	train		None
25 26 26	Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms included studies; data used for all analyses; analytic code; any other materials used in the review.	no dat and	extracted from	None
28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44	From: Page MJ, Mi 5 From: Page MJ, Mi 10.1136/bmj.n71 2 3 4 5 5 7 3 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9	cKenzie	JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for re For more information, visit: <u>http://www.prisma-statement.org/</u>	similar technologies.	g systematic reviews. BMJ	2021;372:n71. doi:
45 46 47	5 5 7		For peer review only - http://bmJopen.bmJ.com/site/about/guidelines.xhtml			