Supplement 1: Quality of anchor studies assessed using the Devji (2020) credibility instrument

Study	Item 1*	Item 2#	*Item 3#*	Item 4#	Item 5#	The overall quality of the studies**
Angst, 2018	Yes	Impossible to tell	Definitely No	To a greater extent	To a great extent	**High**
Harris, 2013	Yes	To a great extent	Impossible to tell	Not so much	Definitely No	**Low**
Hmamouchi, 2012	Yes	Not so much	Impossible to tell	To a greater extent	To a greater extent	**High**
Klokker, 2016	Yes	To a great extent	Impossible to tell	Definitely No	Not so much	**Low**
Lee, 2017	Yes	Definitely Yes	Impossible to tell	To a greater extent	Impossible to tell	**High**
Mills, 2016	Yes	Definitely Yes	Not so much	To a greater extent	Definitely Yes	**High**
Mostafaee, 2021	Yes	To a great extent	To a great extent	To a great extent	Not so much	**High**
Ornetti, 2011	Yes	To a great extent	Impossible to tell	To a greater extent	Definitely No	**High**
Perrot, 2013	Yes	To a great extent	Not so much	Definitely Yes	Not so much	**High**
Singh, 2014	Yes	To a great extent	Impossible to tell	Not so much	To a great extent	**High**
Tubach, 2005	Yes	To a great extent	Impossible to tell	Definitely Yes	Not so much	**High**
Williams, 2012	Yes	Not so much	Impossible to tell	To a greater extent	Definitely Yes	**High**

Item 1: Is the patient or necessary proxy responding directly to both the PROM and the anchor?
Item 2: Is the anchor easily 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
** overall quality: three of the five criteria were met “Yes” or “definitely yes” or “to a great extent”, the paper is of “high” quality If not, that the paper is of “low” quality[27]

Low quality (credibility) studies are shaded.

PROM: Patient-Reported Outcome Measure, MID-Minimal Important Difference (In this study minimal important change/difference)

Supplement 2: Quality of consensus studies assessed using the Critical Appraisal Screening Program (CASP)-qualitative checklist[29]

Study (Author, Year)	Item 1 #	Item 2 #	Item 3 #	Item 4 #	Item 5 #	Item 6 #	Item 7 #	Item 8 #	Item 9 #	Item 10	The overall quality of the studies µ
Salottolo, 2018	Yes	Yes	Can't tell	Can't tell	Yes	Yes	Can't tell	Yes	Yes	Researcher discusses the the contribution the study makes to existing knowledge or understanding	**Moderate**

Item 1: Was there a clear statement of the aims of the research?
Item 2: Is a qualitative methodology appropriate?
Item 3: Was the research design appropriate to address the aims of the research?
Item 4: Was the recruitment strategy appropriate to the aims of the research?
Item 5: Was the data collected in a way that addressed the research issue?
Item 6: Has the relationship between researcher and participants been adequately considered?
Item 7: Have ethical issues been taken into consideration?
Item 8: Was the data analysis sufficiently rigorous?
Item 9: Is there a clear statement of findings?
Item 10: How valuable is the research?

The responses to items: Yes, Can't Tell, No
µ Overall quality of study: high", "moderate" or "low" was evaluated by the review team based on how reliable and credible each study was, without specific rules

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 group[31]

Study (Author, Year)	Criteria 1 #	Criteria 2 #	Criteria 3 #	Criteria 4 #	Criteria 5 #	Criteria 6 #	Criteria 7 #	Criteria 8 #	Criteria 9 #	Criteria 10 #	Criteria 11 #	Criteria 12 #	Overall quality µ
Alghadir, 2017	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	Yes	Yes	Other	**Fair**
Alghadir, 2018	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	Yes	Yes	Other	**Fair**
Alghadir, 2016 b	Yes	No	Yes	Yes	Yes	Other	Yes	Other	Yes	Yes	Yes	Other	**Fair**
Alghadir, 2017	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	Yes	Yes	Other	**Fair**
Baert, 2018	Yes	No	Yes	Yes	Other	Other	Yes	Other	Yes	Yes	Yes	Other	**Fair**
Baert, 2018	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	Yes	Yes	Other	**Fair**
Brisson, 2018	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	Yes	Yes	Other	**Good**
Callaghan, 2009	Yes	No	Yes	Yes	Yes	Other	Yes	Other	Other	Yes	Yes	Other	**Fair**
Hoglund, 2019	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	Yes	Yes	Other	**Good**
Hunter, 2006	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Other	Yes	Yes	Other	**Fair**
Ijima, 2019	Yes	Yes	Yes	Yes	Other	Yes	Yes	Other	Yes	Yes	Yes	Other	**Fair**
Jansen, 2021	Yes	Yes	Yes	Yes	Other	Yes	Yes	Other	Yes	Yes	Yes	Other	**Fair**
Kanko, 2019	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Other	Yes	Yes	Yes	Other	**Good**
Kean, 2010	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	Yes	Yes	Other	**Fair**
Klokker, 2015	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	Yes	Yes	Other	**Good**
McCarthy, 2004	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	Yes	Yes	Other	**Fair**
McCarthy, 2008	Yes	Yes	No	No	Other	Other	Yes	Other	Yes	Yes	Yes	Other	**Fair**
Monticone, 2021	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	Yes	Yes	Other	**Good**
Motyl, 2013	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	Yes	Yes	Other	**Fair**
Mutlu, 2015	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Other	Yes	Yes	Yes	Other	**Good**
Nalbant, 2021	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	Yes	Yes	Other	**Good**
Naylor, 2014	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	Yes	Yes	Other	**Fair**
Parveen, 2017	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	Yes	Yes	Other	**Fair**
Peter, 2018	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	Yes	Yes	Other	**Fair**
Piva, 2004	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	Yes	Yes	Other	**Fair**
Pratheep, 2018	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	Yes	Yes	Other	**Fair**
Ravaud, 1999	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	Yes	Yes	Other	**Fair**

Study (Author, Year)	Criteria 1 #	Criteria 2 #	Criteria 3 #	Criteria 4 #	Criteria 5 #	Criteria 6 #	Criteria 7 #	Criteria 8 #	Criteria 9 #	Criteria 10 #	Criteria 11 #	Criteria 12 #	Overall quality µ
Suhail and Chaudhary, 2021	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	Yes	Yes	Other	**Good**
Suwit, 2020	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	Yes	Yes	Other	**Fair**
Takacs, 2014	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	Yes	Yes	Other	**Fair**
Tevald, 2016	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	Yes	Yes	Other	**Fair**
Takacs, 2014	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	Yes	Yes	Other	**Good**
Tevald, 2016	Yes	No	Yes	Yes	Other	Other	Yes	Other	Yes	Yes	Yes	Other	**Fair**
Tse, 2021	Yes	Yes	Yes	Yes	Other	Other	Yes	Other	Yes	Yes	Yes	Other	**Fair**
Turcot, 2008	Yes	Yes	Yes	Yes	Yes	Other	Yes	Other	Yes	Yes	Yes	Other	**Good**

The description of the criteria is given below.
Criteria 1: Was the study question or objective clearly stated?
Criteria 2: Were eligibility/selection criteria for the study population prespecified and clearly described?
Criteria 3: Were the participants in the study representative of those who would be eligible for the test/service/intervention in the general or clinical population of interest?
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 participants?
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?
Criteria 10: Did the statistical methods examine changes in outcome measures from before to after the intervention? Were statistical tests done that provided p values for the pre-to-post changes?
Criteria 11: Were outcome measures of interest taken multiple times before the intervention and multiple times after the intervention (i.e., did they use an interrupted time-series 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, without specific rules as recommended by the tool

Supplement 4: Anchor properties of included anchor studies

Study	Description of anchor	Anchor question	Anchor properties/responses	Definitions of MIC/MID using transition question	Reported terminology in the study
Angst, 2018	GRC transition (5 points)	NR	much worse, slightly worse, almost equal, slightly better, much better	Difference between the "slightly better" group and the "almost equal" group= MID	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 group "my knee has got better"= MIC and the difference in the change score between groups responded with " my knee has stayed the same" and "my knee has got better"= MID	Minimal Important Change, Minimal Important 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 the mean effects of "slightly better' group and "no change" groups= 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 score of at least 2 (+2: little better: -2: little worse) and no change (score of 0, +1 (almost the same or hardly any better), -1 (worse at 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 change for people achieving previously established MID on legacy comparators: MID in a single item anchor e.g. Patient Global Assessment was defined as range= MID to 1+MID, MID in a multi-teem anchor e.g. SF36 range= MID to 2XMID	Minimally Important Difference

Study	Description of anchor	Anchor question	Anchor properties/responses	Definitions of MIC/MID using transition question	Reported terminology in the study
	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 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)				
Mills, 2016	Global transition GRC-7-point Likert scale	Two anchor question 1. Compared with when I started this program, my walking on level ground has (walking anchor) and 2. Compared with when I started this program, my knee had (knee health anchor):	much improved, moderately improved, slightly improved, not changed, slightly worse, moderately worse, much worse	Mean change within slightly improved group= MIC And Difference between participants who responded slightly, moderately, much improved as the "improved group" and no change or worse “non-improved group” =MID	Minimal Important Difference- Mean change (Jaeschke) method Minimal Important Difference- Mean change (Redelmeier and Lorig) method, ROC method (Youden method and 80% specific method)
Mostafee, 2021	GRC- 7points Likert scale	How did your knee status change compared to the beginning of the	1. very much worse; 2. much worse; 3. slightly worse; 4. no change; 5. slightly better; 6. much	The difference between improved group (6 and 7) and not improved group (1,2,3,4 and 5)=MID	Minimal Important Change

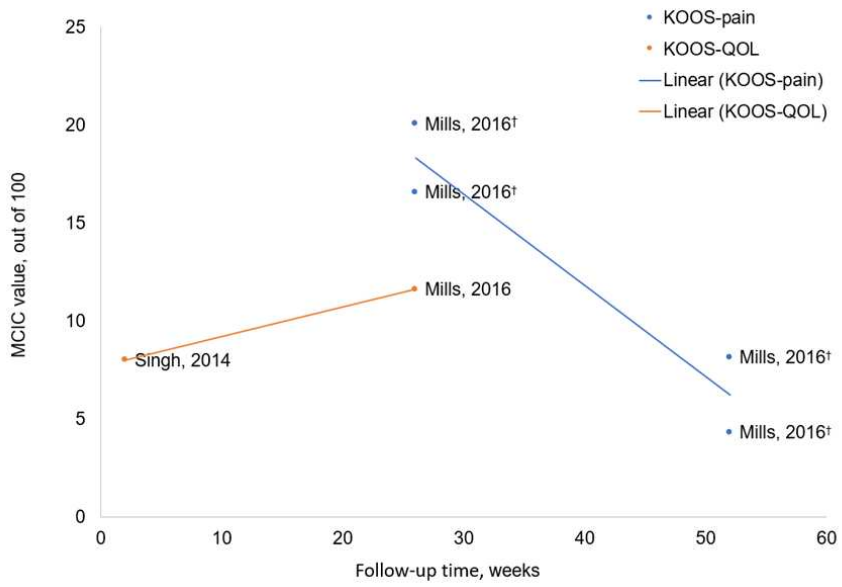
Study	Description of anchor	Anchor question	Anchor properties/responses	Definitions of MIC/MID using transition question	Reported terminology in the study
		physiotherapy intervention?	better; and 7. very much better		
Ornetti, 2011	GRC- (3points Likert scale, then again in a 4-point Likert scale	specific question NR. Two anchor questions. 1. The degree of improvement of global state 2. The degree of improvement of functional state	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 improvement was scored on a scale (poor, fair, good, excellent)	75th centile of the absolute change in score among patients who responded the improvement as good or excellent (improved group) = MIC	Minimal Clinically Important Improvement
Perrot, 2013	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	75th percentile of the distribution of the pain intensity difference (day 0 to 7) for patients considering their improvement to be at least moderate= MIC	Minimal Clinically Important Improvement
Singh, 2014	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= MIC	Minimal Clinically Important Difference for improvement

Study	Description of anchor	Anchor question	Anchor properties/responses	Definitions of MIC/MID using transition question	Reported terminology in the study
Tubach, 2005	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	75th centile of the change in score of the group who responded as “good” using logistic regression=MIC	Minimal Clinically Important Improvement
Williams, 2012	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 subjects who perceived "improved" (those with a GRC between 1 (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 worse))=MID	Minimal Clinically Important Difference

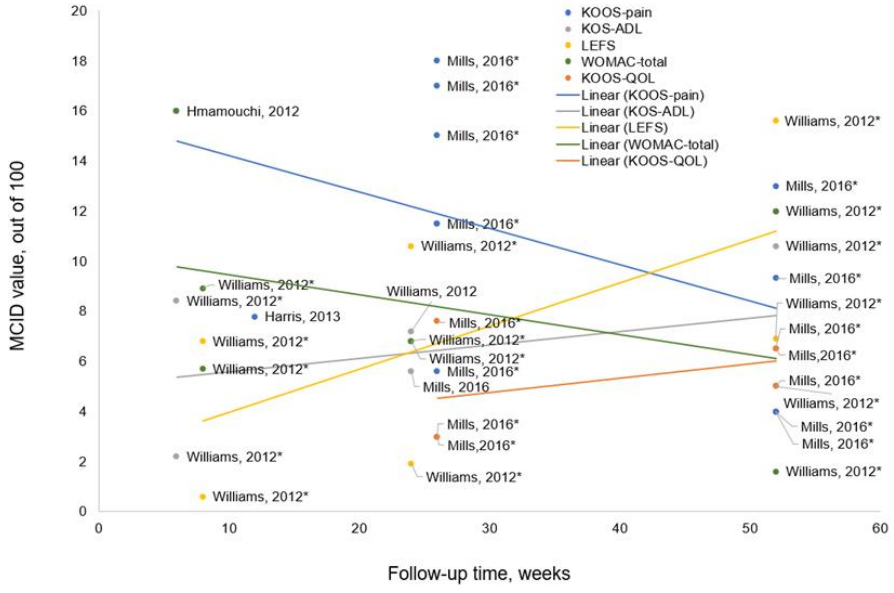
MIC: Minimal Important Change, MID: Minimal Important Difference, GRC: Global Rating of Change, NR: Not Reported, CI: Confidence Interval, WOMAC: Western Ontario and McMaster Universities Arthritis Index, SF36: Short Form-36, NSAID: Non-Steroid Anti-Inflammatory Drug

Supplement 5: The effect of follow-up time on **A.** Minimal Important Change (MIC) and **B.** Minimal Important Difference (MID)

A.



B.



†: 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

Supplement 6: Minimum Detectable Change (MDC) values of knee osteoarthritis outcome tools derived using the distribution method

Outcome tool	Score	Median	Minimum	Maximum	Number of estimates	Study
Aggregated locomotor function	N/A	2.3 seconds			1	McCarthy,2004
Animated Activity Questionnaire	100=best	11.2			1	Peter, 2018
BMI using a scale and a stadiometer	N/A	2.6 kg/m^2			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 eq	2	Jansen, 2021
Bone density-tibia mean lateral	N/A	2.6 mm Al eq	2.4 mm Al eq	2.8 mm Al eq	2	Jansen, 2021
Bone density-tibia mean medial	N/A	0.8 mm Al eq	0.7 mm Al eq	1 mm Al eq	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	0.01º	0.01º	0.02 º	3	VandeStraaten, 2020
De Motion Mobility Index	100=best	8.0	7.3	8.7	2	Yuruk, 2014
Eight-meter walk time	N/A	1.4 seconds			1	McCarthy,2004
Eminence lateral (knee image)	mm	2.2 mm	2.2 mm	2.2 mm	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 seconds	2	Piva, 2004
Fremantle Knee Awareness Questionnaire	100=worst	14.4			1	Monticone,2021
Frontal plane tibial alignment using smartphone inclinometer	N/A	3.7º			1	Tse, 2021
Frontal plane tibial alignment using a manual inclinometer	N/A	3.2º			1	Tse, 2021
Hip abductor strength using a hand-held dynameter	N/A	0.3 Nm/kg	0.3 Nm/kg	0.3 Nm/kg	2	Tevald, 2016
Hip flexion-extension during unipodal stance using 3D motion analysis	N/A	2.7 º	2.2 º	4.6 º	3	VandeStraaten, 2020
ICOAP-constant pain	100=worst	51.7	49.6	53.8	2	Singh, 2014
ICOAP-intermittent pain	100=worst	49.8	48.7	50.8	2	Singh, 2014
ICOAP-pain	100=worst	46.6	23.0	49.6	3	Singh, 2014, Harris, 2013
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 20º	N/A	20.6 N	13.2 N	26.7 N	3	Baert, 2018

Outcome tool	Score	Median	Minimum	Maximum	Number of estimates	Study
Knee force sense test using a handheld dynamometer-reposition error 45º	N/A	38.0 N	27.8 N	43.3 N	3	Baert, 2018
Knee force sense test using a handheld dynamometer-reposition error 70º	N/A	32.2 N	22.5 N	49.5 N	3	Baert, 2018
Knee force sense test using a handheld dynamometer-reposition error all	N/A	32.7 N	30.0 N	33.7 N	3	Baert, 2018
Knee joint angle	degrees	2.1º	1.9º	2.3º	2	Jansen, 2021
Knee joint space width	mm	1 mm	0.8 mm	1.2 mm	2	Jansen, 2021
Knee joint space width-lateral	mm	1.7 mm	1.4 mm	2.0 mm	2	Jansen, 2021
Knee joint space width-medial	mm	0.6 mm	0.5 mm	0.8 mm	2	Jansen, 2021
Knee joint space width-minimum	mm	0.8 mm	0.6 mm	0.9mm	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	2	Naylor, 2014
KOOS-quality of life	100=best	27.8	22.4	39.0	4	Naylor, 2014; Singh, 2014
KOOS-symptoms	100=best	20.2	2.9	24.1	3	Naylor, 2014
KOOS-PS	100=worst	28.3	16.0	35.5	3	Harris, 2013 Singh, 2014
Knee joint position sense test-analogue inclinometer-reposition error 70º	N/A	8.0 º	4.0 º	8.0 º	3	Baert, 2018
Knee joint position sense test-analogue inclinometer-reposition error 45º	N/A	4.0 º	3.0 º	8.0 º	3	Baert, 2018
Knee joint position sense test-analogue inclinometer-reposition error 20º	N/A	4.0 º	3.0 º	4.0 º	3	Baert, 2018
Knee joint position sense test-analogue inclinometer-reposition error all	N/A	3.0 º	3.0 ^{0}	4.0 º	3	Baert, 2018
KOS-activities of daily living	100=best	15.8*	10.5	21.0	6	Williams, 2012
L-test	N/A	5.28 seconds			1	Nalbant, 2021
Load frequency using a triaxial accelerometer	N/A	4.3 steps/day			1	Brisson, 2018
LEFS	100=best	18.3*	14.8	22.6	6	Williams, 2012
NRS- pain	100=worst	16.5	13.3	19.6	2	Alghadir, 2016b, Alghadir, 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	2	Jansen, 2021

Outcome tool	Score	Median	Minimum	Maximum	Number of estimates	Study
Osteophytes-Tibia medial (using knee image)	N/A	11.6 mm^2	11 mm^2	12.2 mm^2	2	Jansen, 2021
OKS-summary	100=best	6.1	6.0	6.2	1	Alghadir, 2017, Harris, 2013
Peak KAM in 3D motion analysis	N/A	0.2 Nm/kg			1	Brisson, 2018
Peak KAM in 3D motion analysis	N/A	1.3 %BW*HT			1	Brisson, 2018
Peak KFM in 3Dmotion analysis	N/A	0.4 Nm/kg			1	Brisson, 2018
Peak KFM in 3D motion analysis	N/A	2.3 %BW*HT			1	Brisson, 2018
Pelvic abduction-adduction during unipodal stance using 3D motion analysis	N/A	3.1 °	2.8 °	3.5 °	3	VandeStraaten, 2020
Per cent of voluntary muscle activation using a dynamometer	N/A	6.6%			1	Kean, 2010
Pressure pain threshold-knee- using an algometer	N/A	131.8 kPa	92.9 kPa	196.3 kPa	10	Pratheep, 2018
Pressure pain threshold-medial heel using a handheld pressure algometer	N/A	1.3 lb			1	Mutlu, 2015
Pressure pain threshold- medial knee using a handheld pressure algometer	N/A	1.2 lb			1	Mutlu, 2015
Quadriceps fatigue using surface electromyography-VMO initial median frequency	N/A	11.0 Hz			1	Callghan, 2009
Quadriceps fatigue using surface electromyography-VL initial median frequency	N/A	10.0 Hz			1	Callghan, 2009
Quadriceps fatigue using surface electromyography-RF initial median frequency	N/A	9.0 Hz			1	Callghan, 2009
Quadriceps fatigue using surface electromyography-VMO final median frequency	N/A	7.4 Hz			1	Callghan, 2009
Quadriceps fatigue using surface electromyography-VL final median frequency	N/A	6.5 Hz			1	Callghan, 2009
Quadriceps fatigue using surface electromyography-RF final median frequency	N/A	10.5 Hz			1	Callghan, 2009
Quadriceps fatigue using surface electromyography-VMO median frequency slope	N/A	2207.0 %/min*			1	Callghan, 2009
Quadriceps fatigue using surface electromyography-VL median frequency slope	N/A	4000.0 %/min*			1	Callghan, 2009
Quadriceps fatigue using surface electromyography-RF median frequency slope	N/A	2390.0 %/min*			1	Callghan, 2009

Outcome tool	Score	Median	Minimum	Maximum	Number of estimates	Study
Quadriceps isokinetic strength using a dynamometer-absolute value	N/A	33.9 Nm			1	Kean, 2010
Quadriceps isometric strength using a dynamometer-absolute value	N/A	25.0 Nm	5.5	37.2	3	McCarthy,2008; Brisson, 2018
Quadriceps isometric strength using a dynamometer-normalised to weight	N/A	0.4 Nm/kg	0.3	0.5	2	Brisson, 2018; Kean, 2010
Quadriceps isometric strength using a dynamometer-Normalised to body size	N/A	1.5 %BW*height			1	Kean, 2010
Quadriceps power using a dynamometer	N/A	151.8 W/2.2 W/kg			1	Brisson, 2018
Rectus femoris fatigue slope	N/A	0.6 %/min			1	McCarthy,2008
Rectus femoris initial median frequency	N/A	5.2 Hz			1	McCarthy,2008
Single-leg standing balance test-mediolateral standard deviation	N/A	0.3			1	Takacs, 2014
Single-leg standing balance test-anteroposterior standard deviation	N/A	0.5			1	Takacs, 2014
Single-leg standing balance test-path length	N/A	20.2 cm			1	Takacs, 2014
Single-leg standing balance test-velocity	N/A	0.3 m/seconds			1	Takacs, 2014
Single-leg standing balance test-area	N/A	23.2 cm^2			1	Takacs, 2014
Six minute walk test	N/A	72.7 m	66.3 m	79.0 m	2	Naylor, 2014
Stair ascent and descent time (seven steps (four of 15cm, three of 20cm))	N/A	2.6 seconds			1	McCarthy,2004
Stopwatch-based 11- Step stair climb test	N/A	0.1 seconds	0.1 seconds	0.1 seconds	2	Ijima, 2019
Star excursion balance test-Raw value	N/A	7.4 cm			1	Kanko, 2019
Star excursion balance test-Normalized (raw value/leg length%)	N/A	8.5 %			1	Kanko, 2019
Timed Up and Go test	N/A	1.1 seconds	1.1 seconds	1.1 seconds	2	Alghadir, 2015
Timed Up and Go test (as a ratio of the original measurement)	N/A	41.06%	37.5%	44.6%	2	Naylor, 2014
Tinetti Performance-Oriented Mobility Assessment scale- balance subscale	NR	0.8			1	Parveen, 2017
Tinetti Performance-Oriented Mobility Assessment scale- gait subscale	NR	0.6			1	Parveen, 2017
Tinetti Performance-Oriented Mobility Assessment scale- total	NR	1.0			1	Parveen, 2017

Outcome tool	Score	Median	Minimum	Maximum	Number of estimates	Study
Transferring time (distance of 2m to a chair and sit down, then, walk back to the start)	N/A	1.7			1	McCarthy,2004
Trunk-abduction-adduction during unipodal stance using 3D motion analysis	N/A	2.9 º	2.6 º	2.9 º	3	VandeStraaten, 2020
2-minute walk test	N/A	5.52m			1	Suhail and Chaudhary, 2021
Vastus lateralis fatigue slope	N/A	0.5 %/min			1	McCarthy,2008
Vastus lateralis initial median frequency	N/A	5.8 Hz			1	McCarthy,2008
Vastus medialis oblique fatigue slope	N/A	0.8 %/min			1	McCarthy,2008
Vastus medialis oblique initial median frequency	N/A	6.7 Hz			1	McCarthy,2008
Verbal Rating Scale for pain		5.8			1	Alghadir, 2018
VAS for pain	100=worst	24.0 cm *	8.0 cm	28.0 cm	3	Alghadir, 2018; Naylor,2014
WOMAC-pain	100=best	3.8	3.4	19.0	3	Angst, 2018, Alghadir, 2016 a
WOMAC-function	100=best	3.1	3.1	18.7	3	Angst, 2018, Alghadir, 2016 a
WOMAC-stiffness	100=best	5.2	4.8	5.6	2	Angst, 2018
WOMAC-functional standing/walking	100=best	3.8	3.5	4.0	2	Angst, 2018
WOMAC-total	100=worst	18.7	11.7	20.8	6	Williams, 2012, Alghadir, 2016 a
SF-36-bodily pain	100=best	3.5	3.3	3.7	2	Angst, 2018
SF-36-general health	100=best	2.1	2.0	2.2	2	Angst, 2018
SF-36-mental health	100=best	3.1	2.9	3.4	2	Angst, 2018
SF-36-physical functioning	100=best	2.4	2.3	2.4	2	Angst, 2018
SF-36-role- physical	100=best	11.2	10.9	11.5	2	Angst, 2018
SF-36-social functioning	100=best	4.1	3.7	4.5	2	Angst, 2018
SF-36-vitality	100=best	5.3	5.1	5.6	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	Hoglund, 2019
40-meter fast-paced test	N/A	16.3			1	Suwit, 2020
30 seconds chair stand test	N/A	21.2			1	Suwit, 2020

Outcome tool	Score	Median	Minimum	Maximum	Number of estimates	Study
3D linear accelerations of the tibia during comfortable walking	N/A	0.3 g	0.1 g	0.8 g	12	Turcot, 2008
3D linear accelerations of the femur during comfortable walking	N/A	0.3 g	0.1 g	1.0 g	12	Turcot, 2008
3D linear accelerations of the tibia at fast speed	N/A	0.4 g	0.1 g	0.9 g	12	Turcot, 2008
3D linear accelerations of the femur at fast speed	N/A	0.2 g	0.0 g	0.6 g	12	Turcot, 2008
9-step stair climb test	N/A					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 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