

BMJ Open Application of minimal important differences in degenerative knee disease outcomes: a systematic review and case study to inform *BMJ* Rapid Recommendations

Tahira Devji,¹ Gordon H Guyatt,^{1,2} Lyubov Lytvyn,³ Romina Brignardello-Petersen,^{1,4} Farid Foroutan,^{1,5} Behnam Sadeghirad,^{1,6} Rachelle Buchbinder,^{7,8} Rudolf W Poolman,⁹ Ian A Harris,^{10,11} Alonso Carrasco-Labra,^{1,4} Reed A C Siemieniuk,^{1,12} Per O Vandvik¹³

To cite: Devji T, Guyatt GH, Lytvyn L, *et al.* Application of minimal important differences in degenerative knee disease outcomes: a systematic review and case study to inform *BMJ* Rapid Recommendations. *BMJ Open* 2017;7:e015587. doi:10.1136/bmjopen-2016-015587

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2016-015587>).

Received 15 December 2016
Accepted 23 January 2017



► <http://dx.doi.org/10.1136/bmjopen-2017-016114>



CrossMark

For numbered affiliations see end of article.

Correspondence to
Dr Tahira Devji;
devjits@mcmaster.ca

ABSTRACT

Objectives: To identify the most credible anchor-based minimal important differences (MIDs) for patient important outcomes in patients with degenerative knee disease, and to inform *BMJ* Rapid Recommendations for arthroscopic surgery versus conservative management

Design: Systematic review.

Outcome measures: Estimates of anchor-based MIDs, and their credibility, for knee symptoms and health-related quality of life (HRQoL).

Data sources: MEDLINE, EMBASE and PsycINFO.

Eligibility criteria: We included original studies documenting the development of anchor-based MIDs for patient-reported outcomes (PROs) reported in randomised controlled trials included in the linked systematic review and meta-analysis and judged by the parallel *BMJ* Rapid Recommendations panel as critically important for informing their recommendation: measures of pain, function and HRQoL.

Results: 13 studies reported 95 empirically estimated anchor-based MIDs for 8 PRO instruments and/or their subdomains that measure knee pain, function or HRQoL. All studies used a transition rating (global rating of change) as the anchor to ascertain the MID. Among PROs with more than 1 estimated MID, we found wide variation in MID values. Many studies suffered from serious methodological limitations. We identified the following most credible MIDs: Western Ontario and McMaster University Osteoarthritis Index (WOMAC; pain: 12, function: 13), Knee injury and Osteoarthritis Outcome Score (KOOS; pain: 12, activities of daily living: 8) and EuroQol five dimensions Questionnaire (EQ-5D; 0.15).

Conclusions: We were able to distinguish between more and less credible MID estimates and provide best estimates for key instruments that informed evidence presentation in the associated systematic review and judgements made by the Rapid Recommendation panel.

Strengths and limitations of this study

- This is the first systematic review of minimal important differences (MIDs) for patient-reported outcomes measuring pain, function and health-related quality of life in patients with degenerative knee disease.
- We demonstrate how MIDs can inform presentation of findings in systematic reviews, and judgements in guideline development.
- There are no established credibility criteria for MIDs with measurement properties, particularly reliability, that have been formally tested.
- Even applying our credibility criterion of a sufficiently high correlation, the range of MIDs reported was very wide; credibility of the estimates may still be limited.

Trial registration number: CRD42016047912.

INTRODUCTION

Degenerative knee disease (osteoarthritis in the knee, which can involve the joint lining and/or menisci) is a chronic, progressively debilitating condition, affecting more than nine million people in the USA.¹ A number of randomised controlled trials (RCTs) have assessed the impact of arthroscopic surgery involving partial meniscectomy, debridement or both in patients with degenerative knee disease. These RCTs have reported effects of arthroscopy on patient-reported outcomes (PROs) of knee pain, function and health-related quality of life (HRQoL), which are critical outcomes in degenerative knee

disease trials.^{2 3} The RCTs have demonstrated that arthroscopic surgery results in a small improvement in pain and function over the short term, but guidance for clinicians and patients requires determining the importance of these benefits.⁴

Investigators are increasingly relying on PROs as key end points in clinical trials. Although PROs provide patients' experience of the impact of disease and treatment on their health status, challenges in interpreting changes in PRO scores can limit their usefulness in informing patient-centred care.⁵ For instance, does a 10 mm reduction in self-reported knee pain on a 0–100 mm visual analogue scale (VAS) reflect a trivial difference, a small but important difference, a moderate or even a large effect? The key issue for those making recommendations is how patients value the outcomes: in this case, where in the continuum between trivial and very important will patients place observed improvements in pain and function? The smallest change that patients perceive as important, either beneficial or harmful—the minimal important difference (MID)^{6 7}—reflects patients' values and preferences, and can therefore enhance the interpretation of PROs, facilitating understanding of the importance of intervention effects in RCTs.

Establishing the MID for an index instrument requires comparison of instrument scores with another instrument (typically referred to as an anchor) that is itself interpretable. The most popular approach uses a transition instrument (asking patients whether they have improved or deteriorated, and the magnitude of that improvement or deterioration) as the anchor, and relating change in instrument score to the patients' rating of change in status over time.⁸ In this method, patients complete the index instrument on two occasions. On the second occasion, they rate the extent to which they have improved or deteriorated; it is this transition rating that provides the anchor. Typically, patients who have experienced a small but important improvement or deterioration inform the MID estimate.

BMJ Rapid Recommendations⁹ is a new series of trustworthy recommendations published in response to potentially practice changing evidence. *BMJ* Rapid Recommendations panels, as in any guideline, require best current evidence to inform their recommendations, covered by one or more linked systematic reviews.^{9–11} Another requirement is appropriate interpretation of the importance of effects when moving from evidence to recommendations—judgements that should reflect patients' values and preferences.^{10 11} The panel responsible for creating the second *BMJ* Rapid Recommendations, addressing the impact of arthroscopic surgery versus conservative management in patients with degenerative knee disease, faced challenges in interpreting the significance of apparent treatment effects on the critical outcomes of interest: pain, function and HRQoL from the linked systematic review.^{4 12} To help address this challenge, we conducted an additional

linked systematic review to identify the most credible anchor-based MID estimates for the PROs used in trials comparing arthroscopic surgery to conservative management. In this paper, we describe our approach to gathering and interpreting the credibility of MID estimates, and note how our results informed the linked systematic review of treatment effectiveness⁴ and the subsequent development of the *BMJ* Rapid Recommendations.¹²

METHODS

Guideline panel and patient involvement

According to the *BMJ* Rapid Recommendations process,⁹ a guideline panel provided critical oversight to our systematic review addressing MID estimates as well as the linked systematic review of effectiveness. The panel, which included eight content experts and front-line clinicians (three orthopaedic surgeons, one rheumatologist, one epidemiologist, one general practitioner and two physiotherapists), four methodologists (three of whom are also front-line clinicians and general internists) and three patients with lived experience of degenerative knee disease, identified populations, subgroups and outcomes of interest.^{4 9} Patients received personal training and support throughout the guideline development process.

Patient values and preferences were incorporated in the guideline process through application of the MIDs from our systematic review of studies in which patients provided ratings of the magnitude of change they had experienced, and whether that change was trivial, small but important, or larger. Patients also led the interpretation of the results in the guideline panel based on their assessment of typical patient values and preferences, as well as the variation in values between patients.

Literature search and study identification

We updated our search from a systematic review of anchor-based MIDs¹³ that identified articles from 1989 up to 13 April 2015 (the MID concept was first introduced into the medical literature in 1989⁸) using MEDLINE, EMBASE and PsycINFO from 2 February 2015 to 15 September 2016.⁸ For the update of our initial search, we added filters for the specific PROs assessed in RCTs included in the linked systematic review and meta-analysis addressing benefits and harms of arthroscopy that informed the guideline panel in making their recommendation.⁴ There were no restrictions on language. Online supplementary appendix 1 presents the search strategy for MEDLINE, which we adapted for each of the selected databases.

Study selection

The parallel *BMJ* Rapid Recommendations panel identified pain, function and HRQoL as key patient-important outcomes in the management of degenerative knee disease.⁴ We included original reports of studies that empirically estimated an anchor-based MID in patients with degenerative arthritis of the knee for PRO

measures that informed the systematic review and meta-analysis of treatment effects for the Rapid Recommendation, that is, outcomes included in the eligible randomised trials.⁴ Studies comparing the results of the PRO instrument to an independent standard (the anchor), irrespective of the interpretability or the quality of the anchor, were eligible.

Two pairs of reviewers performed title and abstract and full-text screening independently and in duplicate. All studies included by either reviewer in the title and abstract stage were screened in full text. Reviewers resolved disagreements at the full-text screening stage through discussion.

Data abstraction

Two pairs of reviewers independently extracted data from eligible studies in duplicate using a standardised pilot-tested spreadsheet including the following: first author; publication year; country; participant demographics, including age, sex, condition under investigation; characteristics of the PRO, such as type (generic vs specific), domain(s) and construct(s) captured by the instrument; details pertaining to the method(s) of MID estimation, including number of participants used to estimate the MID, duration of follow-up from baseline, characteristics of the anchor, analysis method (mean change vs receiver operating characteristic (ROC) curves), and correlation between the anchor and PRO scores). We abstracted and report only MIDs for improvement, expressed as absolute estimates, along with the associated 95% CI. We did not include estimates in which the estimated MID for improvement was reported as a deterioration.

Credibility assessment

We defined credibility as the extent to which the design and conduct of studies measuring MIDs are likely to have protected against misleading estimates.¹³ Although there are numerous established risk of bias and quality grading instruments for use in systematic reviews, none are suited to assess the credibility of studies estimating an MID. We dealt with credibility by focusing on a single criterion that is clearly related to credibility and can be ascertained without judgement: the correlation between change in the index PRO under consideration and the global rating of change that constitutes the anchor. Our threshold for an acceptable correlation was 0.4 or greater.^{14–16}

Synthesis of results

We summarised the MID estimates, along with intervention, population characteristics and characteristics of the anchor. We provided the systematic review team with the median, minimum and maximum values across the range of plausible trustworthy MID estimates generated from the eligible studies for the PROs of interest. We pooled the estimates using inverse variance weights and a random-effects model.

To explore potential heterogeneity in MID estimates across studies, we conducted subgroup analyses for possible effect modifiers when we identified at least two studies or two cohorts within studies for each subgroup class (for instance, for nature of intervention, we required at least two surgical cohorts and two non-surgical cohorts). We considered a number of factors plausibly associated with credibility of estimates including the anchor estimate coming from the patients and interpretable to the patient and clinician, precision around the estimate, whether the anchor represents a minimal change, and the length of time between the initial visit and follow-up. The required number of cohorts in each subgroup class was available for only the last of these. We also performed subgroup analyses comparing MIDs estimated in patients undergoing surgical intervention versus those receiving conservative management, and in those using ROC curve analysis versus mean change methods. When more than one MID derived from a single study or cohort was provided, we took the median of the estimates. For instance, in our subgroup analysis exploring the effect of intervention type (surgical or non-surgical) on MID, when authors provided data for more than one time point, we used the median of the available data. To determine if there was a subgroup effect, we considered a test for interaction p value of <0.05 between the proposed variables and the MID to be significant.

STATA software V.12.0 provided software for all analyses.

Practical application of MID estimates in *BMJ* Rapid Recommendations development

Three content experts from the guideline panel with clinical experience with the measures participated in our systematic review of MID estimates, ensuring applicability to the process of developing the recommendations. We applied the MID estimates identified as credible from our review in the evidence summary presented in the linked systematic review addressing treatment effectiveness that informed the *BMJ* Rapid Recommendations panel in their development of recommendations.^{4 12}

The panel used the MID estimates in two ways. One was to intuitively relate the MID estimates to the magnitude of the effect (the smaller the effect in relation to the MID, the less important the effect). The second was to inform statistical techniques to estimate the proportion of patients in intervention and control groups that improved more than the MID, calculate a risk difference on the basis of these results and pool risk differences across studies.¹⁷ We performed sensitivity analyses using the minimum and maximum MIDs across the range of credible estimates for each PRO to test the robustness of our findings.

RESULTS

We screened 4730 unique citations, of which 1716 were judged potentially eligible on review of titles and

abstracts, and 15 deemed eligible on full-text review (figure 1). Two^{18 19} of the 15 eligible publications provided secondary reports of the same patients included in earlier empirical studies^{20 21} estimating MIDs for the same PRO measures. We used both sets of reports to obtain all relevant data for our review.

Table 1 presents the study characteristics. Thirteen studies reported anchor-based MIDs for eight candidate PROs and/or their domains assessing knee pain, function or HRQoL. All studies used a transition rating (a global rating of change) as the anchor to ascertain the MID. The number of patients informing the estimation of the MID ranged from 31 to 497. Table 1 highlights the studies from which the credible MIDs (those with a correlation of 0.4 or greater between change in the index instrument and the global rating of change) were drawn. Content experts confirmed that the range of patients and treatments included in the final selection of MIDs was satisfactory to inform MIDs for the population, intervention and comparator included addressed by the recommendation.

More than one study provided estimates for six of the PROs, and all studies derived MIDs for more than one PRO or PRO domain. Two studies^{14 26} used more than one anchor to estimate MIDs for the same PRO. Two

studies^{20 29} estimated MIDs for multiple cohorts of patients and reported the estimates separately. Follow-up duration ranged from 20 days to 24 months. Three studies^{14 25 29} estimated MIDs for more than one length of follow-up. Investigators used ROC curves to calculate the MID in one study,²⁷ mean change methods in nine studies^{21–26 28 30 31} and both approaches in three studies.^{14 20 29} Altogether, 13 unique studies included in our review reported a total of 95 empirically estimated anchor-based MIDs.

In 20 instances, the correlation between the anchor and the PRO for which the MID was estimated was <0.4 . Nine studies^{21–23 25–28 30 31} providing 21 MIDs did not provide correlation coefficients. We deemed these 41 estimates not trustworthy and thus did not include them in the plausible range of MIDs. For these reasons, we were unable to present credible MIDs for the VAS pain and 36-item Short Form Survey (SF-36) bodily pain and physical function domains.

Table 2 presents the median absolute MID estimate for the Western Ontario and McMaster University Osteoarthritis Index (WOMAC) and Knee injury and Osteoarthritis Outcome Score (KOOS) pain and function domains, and EuroQol five dimensions Questionnaire (EQ-5D), along with the minimum and

Figure 1 Study flow diagram. MID, minimally important difference; PRO, patient-reported outcome.

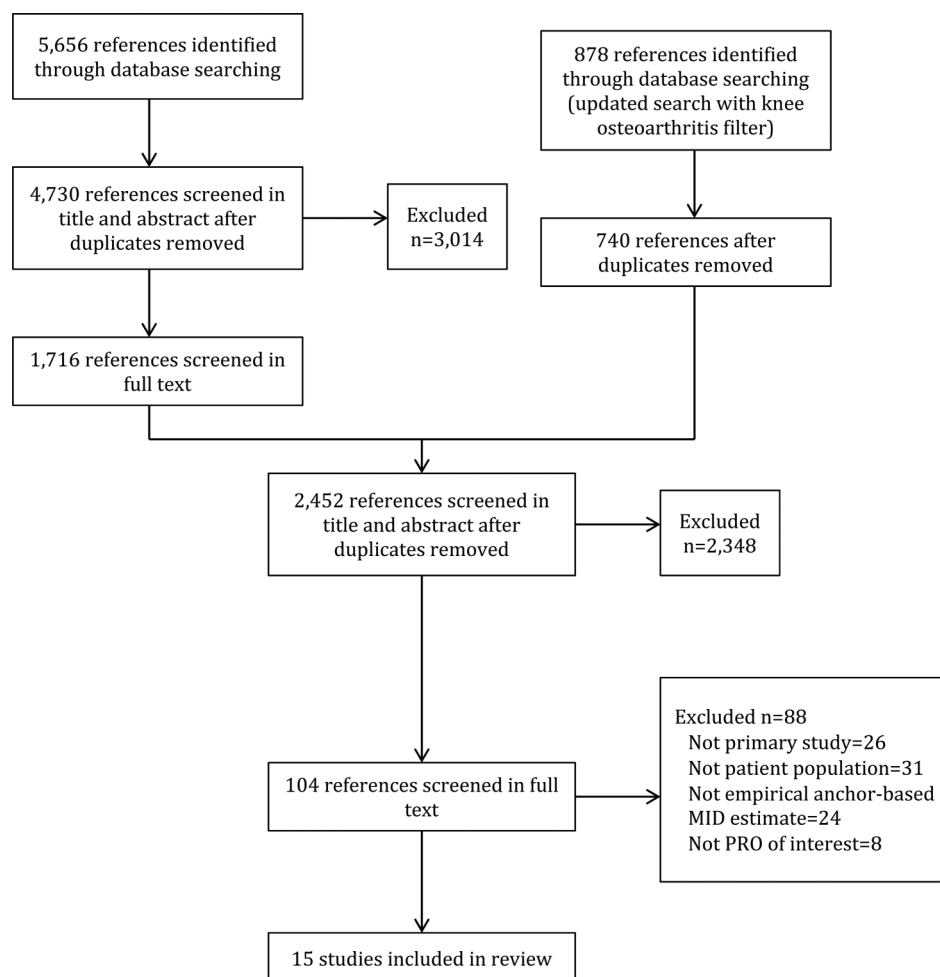


Table 1 Characteristics of included studies reporting MIDs for knee symptoms and health-related quality of life

Study	Disease/condition	Intervention	Instrument/scale (abbreviated name)	Score range	Construct(s) measured	Anchor
Angst <i>et al</i> 2002 ²²	Knee/hip OA	Comprehensive rehabilitation intervention	WOMAC	0 to 100*	Pain, function, stiffness	GROC (5-point)
Bellamy <i>et al</i> 2015 ²³	Knee/hip OA	NSAIDs	WOMAC	0 to 100*	Pain, function, stiffness	GROC 2-step approach (4-point)
Browne <i>et al</i> 2010²⁴	NR	TKA	EQ-5D	-0.4 to 1	HRQoL	GROC (5-point)
Escobar <i>et al</i> 2007 ²⁵	Knee OA	TKA	WOMAC	0 to 100*	Pain, function, stiffness	GROC (5-point)
Escobar <i>et al</i> 2013/2014^{18 20}	Knee OA	TKA	SF-36 WOMAC†	0 to 100 0 to 100	HRQoL Pain, Function	Transition rating for pain and function (5-point)
Ornetti <i>et al</i> 2011²⁶	Knee OA	NSAIDs	WOMAC	0 to 100*	Function	GROC (4-point)
Mills <i>et al</i> 2016¹⁴	Knee OA	Osteoarthritis chronic care programme‡	KOOS	0 to 100	Pain, ADL, HRQoL	2 anchors: walking and knee health (7-point)
Monticone <i>et al</i> 2013 ²⁷	Knee OA	Rehabilitation following TKA	KOOS§	0 to 100	Pain, symptoms, ADL, sport/rec, HRQoL	GROC (5-point)
Terwee <i>et al</i> 2009 ²⁸	Knee/hip OA	TKA	WOMAC¶	0 to 20*0 to 68*	Pain Function	Transition rating for pain and function (15-point)
Terwee <i>et al</i> 2010²⁹	^a Patients who reported knee pain in the past 12 months	^a Usual care	WOMAC	0 to 100	Pain, function	Transition rating for pain (6-point)
	^b Patients who visited their GP with a new episode of non-traumatic knee symptoms	^b Usual care	WOMAC	0 to 100	Pain, function	Transition rating for knee symptoms (6-point)
	^c Patients who visited their GP with a new episode of knee symptoms	^c Usual care	WOMAC	0 to 100	Pain, function	Transition rating for knee symptoms (5-point)
	^d Knee OA	^d Behavioural-graded activity or usual care	WOMAC	0 to 100	Pain, function	Transition rating for knee symptoms (8-point)
	^e Knee OA	^e TKA	WOMAC	0 to 100	Pain, function	Transition rating for pain and function (15-point)
Tubach <i>et al</i> 2005/2006 ^{19 21}	Knee OA	NSAIDs	VAS pain	0 to 100*	Pain	Response to treatment (5-point)
Tubach <i>et al</i> 2012 ³⁰	Knee and/or hip OA	NSAIDs	WOMAC	0 to 100*	Function	GROC 2-step approach (4-point)
Walters <i>et al</i> 2005 ³¹	Patients with knee OA recruited from rheumatology clinics and those assessed preoperatively for TKA	NR	Pain NRS EQ-5D	0 to 10* -0.59 to 1	Pain HRQoL	GROC (5-point)

Bold text were those that provided credible MIDs (those with a correlation of 0.4 or greater between change in the index instrument and the global rating of change).

a–eTerwee *et al*²⁹ reported on five cohorts of patients from different studies. a=cohort 1; b=cohort 2; c=cohort 3; d=cohort 4; e=cohort 5.

*Higher scores on the PRO scale represent a worse outcome.

‡Surgical therapy: debridement, shaving, drilling, autologous chondrocyte implantation, abrasion arthroplasty, microfracture and cell therapy.

‡Osteoarthritis chronic care programme: multidisciplinary non-surgical management strategy.

¶Dutch.

§Italian.

†Spanish.

ADL, activities of daily living; EQ-5D, EuroQol five dimensions Questionnaire; GP, general practitioner; GROC, global rating of change; HRQoL, health-related quality of life; KOOS, Knee injury and Osteoarthritis Outcome Score; MID, minimal important difference; NR, not reported; NRS, Numeric Rating Scale; NSAID, non-steroidal anti-inflammatory drugs; OA, osteoarthritis; OKS, Oxford Knee Score; PRO, patient-reported outcome; SF-36, 36-Item Short Form Survey; TKA, total knee arthroplasty; VAS, visual analogue scale; WOMAC, Western Ontario and McMaster University Osteoarthritis Index.

maximum values across the range of plausible trustworthy estimates (ie, those in which correlations were 0.4 or greater). Among PROs with more than one estimated MID, even among those with correlations of 0.4 or greater, we found wide variation in MID values. Online supplementary appendix 2 presents the MID estimates, as well as details regarding MID estimation for each PRO measure. The content experts confirmed that the MID thresholds generated were consistent with their impressions from use of the instruments in clinical practice.

We only performed subgroup analyses exploring potential sources of heterogeneity for the WOMAC pain and function domains, as estimates for the KOOS pain and activities of daily living, and EQ-5D came from a single study. Type of intervention (ie, total knee arthroplasty (TKA) vs conservative management) was significantly associated with magnitude of the MID for WOMAC pain ($p < 0.00001$; figure 2) and function ($p < 0.00001$; figure 3). For pain, the weighted pooled MID for TKA was 25 (95% CI 24 to 27) in TKA and for conservative management 8 (95% CI 3 to 13). For function, the weighted pooled MID for TKA was 28 (95% CI 27 to 29), and for conservative management 19 (95% CI 3 to 17). We found no association between the hiatus between initial and follow-up visits, nor between the analytic method (ROC or mean change) and the MID.

Incorporation into the systematic review informing *BMJ* Rapid Recommendations

The results of this study informed both the systematic review of treatment effects and the Rapid Recommendations panel in their development of recommendations for arthroscopic surgery versus conservative management in patients with degenerative knee disease.^{4 12} The panel members reviewed the evidence summary (GRADE Summary of Findings table—table 3) from the systematic review with data addressing pain, function, HRQoL and adverse events; they discussed recommendations through teleconferences.

The Summary of Findings for short-term and long-term outcome of pain (table 3) exemplifies how the MID for the KOOS pain domain informed this PRO assessment. Although results from the systematic review favoured arthroscopic surgery in the short term, the estimate of this difference (5.4 points) and its CI (1.9 to 8.8) show magnitudes of effects less than the MID of 12 points established for the index (KOOS) instrument. The systematic review found—by dichotomising outcomes—that 12.4% (95% CI 4.4% to 20.4%) more patients receiving arthroscopy reported a small but important benefit in pain or function at 3 months, which was no longer apparent at 1 year. Sensitivity analyses using the upper and lower estimates across the range of credible MIDs for each instrument, and based on the standardised mean difference (SMD), revealed similar results. The risk difference when using the lowest value of the range was 10.5% (95% CI 4.3% to 16.7%)

Table 2 Summary of the range of plausible credible MIDs for improvement for PRO measures used to inform the systematic review of treatment effects

PRO instrument/ domain (score range)	Absolute MID*	Minimum	Maximum
WOMAC			
Pain (0 to 100)	12	2	30
Function (0 to 100)	13	3	34
KOOS			
Pain (0 to 100)	12	4	20
ADL (0 to 100)	8	3	9
EQ-5D (−0.59 to 1)	0.15	NR	NR

MID estimates are presented as positive values, regardless of the direction of change.

*Median MID estimate.

ADL, activities of daily living; EQ-5D, EuroQol five dimensions Questionnaire; KOOS, Knee injury and Osteoarthritis Outcome Score; MID, minimal important difference; NR, not reported; PRO, patient-reported outcome; WOMAC, Western Ontario and McMaster University Osteoarthritis Index.

and when using the highest value of the range it was 11.3% (95% CI 2.9% to 19.7%). The risk difference based on the SMD was 9% (95% CI 1.7% to 15.7%).

The panel was confident in concluding that any benefit from arthroscopic knee surgery is small or very small, and is less important than the burden and transient pain and limitation associated with the arthroscopy procedure itself. The information provided by the MID informed these judgements, which motivated the panel's decision to make a strong recommendation against arthroscopy in patients with degenerative knee disease.

DISCUSSION

Principal findings

In this review, we identified 13 studies reporting MIDs for eight PROs and/or domains measuring knee pain, function and HRQoL in patients with degenerative knee disease, yielding 95 empirically estimated anchor-based MIDs. Investigators used the same anchor-based approach, relying on transition ratings (global ratings of change).⁸ For the majority of the PROs, more than one study provided MID estimates, and did so at more than one duration of follow-up, using different anchors, and using various analytic methods, resulting in multiple estimates for the same PRO.

MID estimates for the same instrument varied widely across all estimates, as well as when restricted to studies meeting our credibility criterion of a correlation of 0.4 or greater between change in the index instrument and the transition rating. Including only MIDs generated from data meeting this criterion, we were able to provide a range of plausible trustworthy estimates for PROs identified as critical outcomes to inform the systematic review of treatment effects and rapid recommendation (table 2). The systematic

Figure 2 Subgroup analysis for WOMAC pain by intervention type. MID, minimally important difference; WOMAC, Western Ontario and McMaster University Osteoarthritis Index.

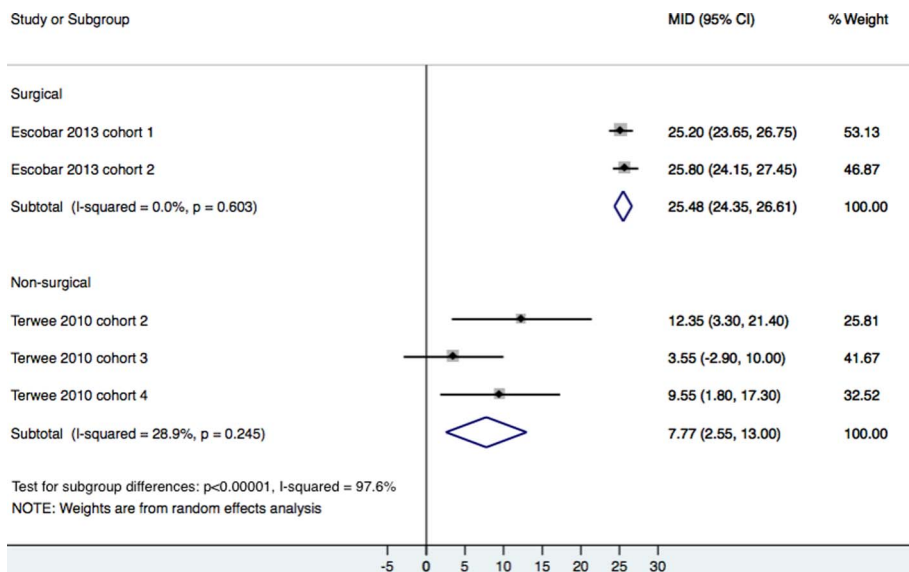
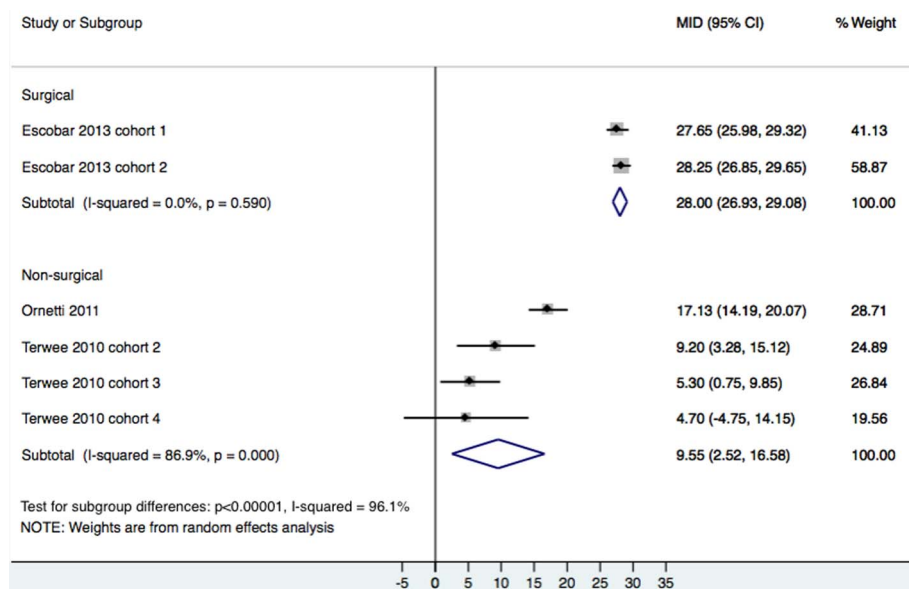


Figure 3 Subgroup analysis for WOMAC function by intervention type. MID, minimally important difference; WOMAC, Western Ontario and McMaster University Osteoarthritis Index.



review team used the most credible MIDs identified in our review to contextualise mean differences and calculate risk differences, and to conduct appropriate sensitivity analyses, expressing the proportion of patients achieving improvements greater than the MID (table 3).

Strengths and limitations

Our study represents the first comprehensive synthesis and evaluation of anchor-based MIDs for self-reported patient-important outcomes commonly assessed in RCTs of degenerative knee disease. We undertook a transparent approach to appraising the credibility of MIDs that allowed identification of the highest credibility MIDs for each instrument. Both the systematic review team and the Rapid Recommendation panel found the MIDs useful in understanding and presentation of the evidence; in particular, the recommendation panel, to a

considerable extent, based recommendations on our findings.

This review also has limitations. There are no established credibility criteria for MIDs with measurement properties, particularly reliability, that have been formally tested. We have therefore focused on a single criterion that is indisputably important, and can be ascertained without judgement, and thus without error: correlations between change in the index instrument and the global rating of change of 0.4 or greater.

Even applying our credibility criterion of a sufficiently high correlation, the range of MIDs reported was very wide (table 2). This raises questions regarding whether the criterion is sufficient—that is, credibility of the estimates may still be limited. The results may, however, reasonably represent a range in which the MID actually lies. We have dealt with this issue by recommending a sensitivity analysis including the full

Table 3 Summary of findings for the outcome of short-term pain presented to *BMJ* Rapid Recommendations panel

Outcome time frame	Study results and measurements	Absolute effect estimates		Certainty in effect estimates (quality of evidence)	Summary
		Conservative management	Arthroscopy		
Short term					
Pain (difference in change from baseline) 3 months	Measured by: different instruments converted to scale of index instrument (KOOS pain subscale) Scale: 0–100 high better, MID 12 Data from 1231 patients in 10 studies Follow-up 3 months	15.0 points (mean) Difference: mean difference 5.4 more (CI 95% 1.9 more to 8.8 more)	20.0 points (mean)	High	On average, knee arthroscopy results in very small extra reduction in pain scores when compared to control
Pain (difference in patients who achieve a change higher than the MID) 3 months	Data from 1102 patients in 9 studies Follow-up 3 months	669 per 1000 Difference: 124 more per 1000 (CI 95% 44 more to 204 more)	793 per 1000	High	Knee arthroscopy increases the number of patients with an important reduction in short-term pain by ~12 in 100
Long term					
Pain (difference in change from baseline) 1–2 years	Measured by: different instruments converted to scale of index instrument (KOOS pain subscale—MID 12) Scale: 0–100 high better Based on data from 1097 patients in 8 studies Follow-up 2 years	19.0 points (mean) Difference: mean difference 3.13 more (CI 95% 0.17 fewer to 6.43 more)	22.0 points (mean)	High	On average, knee arthroscopy results in no difference, or a very small reduction, in pain

KOOS, Knee injury and Osteoarthritis Outcome Score; MID, minimally important difference.

range of plausible MIDs, an approach that the linked systematic review has followed. In future, development and testing of other credibility criteria, and their application in establishing trustworthy MIDs, will strengthen the field.

No MIDs were assessed in the patients of direct interest for the associated systematic review and guideline: patients undergoing knee arthroscopic interventions. Patients included in the eligible studies either underwent major surgery (knee arthroplasty) or non-surgical interventions. Patients did, however, suffer from degenerative knee disease, the condition in which knee arthroscopy is of putative benefit.

For the WOMAC pain and function domains, MIDs estimated in patients undergoing TKA were, on average, appreciably higher than the median MIDs we used as the best estimates. These results suggest that patients undergoing knee arthroplasty, versus those undergoing non-surgical interventions, require a greater degree of change on the PRO measure to consider themselves having an important improvement. In other words, differences in the magnitude of the MID may be related to patient expectations with regard to surgical interventions, as compared with non-surgical or less invasive interventions. The intervention of interest for this Rapid Recommendation—arthroscopy—is, in its invasiveness and immediate consequences, intermediate between non-surgical interventions and total joint arthroplasty. To the extent that, as a result, our best estimate of the MID underestimates the true MID for arthroscopy, the conclusion in the linked systematic review that the effects of arthroscopy are small or very small is actually strengthened.

One of our PROs, the EQ-5D, had only one MID estimate with a correlation of over 0.4. Moreover, this estimate, 0.15, is inconsistent with evidence from other studies that suggest that 0.15 approximates the entire burden of moderate osteoarthritis, and that the MID for the EQ-5D is appreciably <0.15 .³² For the purpose of the review, our work is informing, however, this issue was of minimal concern: the benefits of arthroscopic surgery on quality of life in the short term and long term were not statistically significantly different from patients receiving conservative management, and thus the MID was not needed to further contextualise these results.

Another issue concerns the possible influence of baseline score on the MID.³³ Three of the included studies^{18 21 22} in our review reported MIDs for patients stratified according to baseline severity status. Given that MIDs were consistently higher in magnitude with increasing baseline severity, expression of the MID as a relative change may in instances be superior to an absolute difference. A recent report examining the merits of expressing MIDs as relative or absolute estimates in a number of studies suggested, however, that absolute changes generally correlate higher with global change ratings and are simpler to use and interpret.³⁴

The following considerations mitigate the concerns regarding the credibility of the MID estimates that guided the panel's recommendation. First, our best estimates of the MID approximate 10% of the instruments' total range, a value that is both intuitive and consistent with MID estimates for other instruments. Second, our best estimates of the MID are consistent with the experience of clinicians who have used the instruments as part of their clinical practice. Third, estimates for the risk difference in proportion improved with arthroscopy from the sensitivity analyses in the linked systematic review show that using the upper and lower boundaries of the MID that we have suggested, and a value based on the SMD, approximate those using our best estimate of the MID.^{4 12}

Implications of the findings for future directions

Our review focused on studies using an anchor-based approach, relying on transition ratings as the anchor, to estimate MIDs; we have highlighted shortcomings in their application. We have focused on a single criterion, correlations of 0.4 and greater, to define credible MIDs. The variability in MIDs generated when this criterion was met suggests residual variation in credibility that warrants further investigation.

Authors have suggested—either explicitly or implicitly, when commenting on strengths and limitations of their studies—criteria for judging credibility of MID estimates emerging from empirical studies. Our group has conducted a systematic survey of such commentaries, and on that basis has developed credibility criteria for studies that define MIDs (manuscript in preparation). Feedback from a wider community will be necessary to establish the robustness, appropriateness and comprehensiveness of these criteria, as well as the empirical studies necessary to establish their reliability.

Given the current uncertainty around MIDs, we recommend that triallists, systematic reviews, guideline panellists and other end users of clinical trial PRO data triangulate their interpretation of these subjective outcomes with additional strategies that complement use of the MID. These include viewing the magnitude of effect in relation to the range of the scale for specific PROs, relying on the experience of clinicians using the instruments in their practice, as well as the use of other summary effect measures (eg, SMD). If interpretations of the results are consistent across approaches, this will strengthen interpretation of the magnitude of intervention effects.

CONCLUSIONS

The MID has the potential to help interpret the magnitude of treatment effects and thus guide clinical decision-making in chronic disease management. This study provides a model for applying the MID concept to aid in the interpretation of evidence, and the formulation of recommendations for clinical practice

guidelines.^{4 12} Investigators and guideline panellists can use the approaches reported here to make their systematic reviews more informative, and their recommendations more informed, appropriate and useful.

Author affiliations

¹Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada

²Department of Medicine, McMaster University, Hamilton, Ontario, Canada

³Oslo University Hospital, Oslo, Norway

⁴Evidence-Based Dentistry, Faculty of Dentistry, Universidad de Chile, Santiago, Chile

⁵Heart Failure/Transplant Program, Toronto General Hospital, University Health Network, Toronto, Ontario, Canada

⁶HIV/STI Surveillance Research Center, and WHO Collaborating Center for HIV Surveillance, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

⁷Department of Epidemiology and Preventive Medicine, School of Public Health & Preventive Medicine, Monash University, Melbourne, Australia

⁸Monash Department of Clinical Epidemiology, Cabrini Institute; Suite 41 Cabrini Medical Centre, Malvern, Victoria, Australia

⁹Department of Orthopaedic Surgery, Joint Research, OLVG, Amsterdam, The Netherlands

¹⁰South Western Sydney Clinical School, Sydney, New South Wales, Australia

¹¹Whitlam Orthopaedic Research Centre, Level 2, Ingham Institute for Applied Medical Research, Liverpool, New South Wales, Australia

¹²Department of Medicine, University of Toronto, 1 King's College Circle, Toronto, Ontario, Canada

¹³Department of Medicine, Innlandet Hospital Trust-Gjøvik, Gjøvik, Norway

Contributors TD, GHG, AC-L, RACS and POV conceived the study idea. TD and AC-L performed the literature search. AC-L and GHG, among other colleagues, developed the credibility tool (core criteria) used in this study. TD performed the data analysis. TD, GHG, RWP, RB and RACS interpreted the data analysis. TD and GHG wrote the first draft of the manuscript. TD, LL, BS and FF acquired the data and performed credibility assessments. TD, GHG, POV, RWP and RB-P critically revised the manuscript. TD had full access to all of the data in the study, and takes responsibility for the integrity of the data and the accuracy of the data analysis. TD is the guarantor.

Funding TD, ACL, and GHG are Canadian Institutes of Health Research, Knowledge Synthesis grant recipients for projects related to MID methods (ACL and GHG grant # DC0190SR; TD, ACL, and GHG grant # DC0190SR). RB-P is funded by an Australian National Health and Medical Research Council (NHMRC) Senior Principal Research Fellowship.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

1. Lawrence RC, Felson DT, Helmick CG, *et al.* Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. *Arthritis Rheum* 2008;58:26–35.
2. Altman R, Brandt K, Hochberg M, *et al.* Design and conduct of clinical trials in patients with osteoarthritis: recommendations from a task force of the Osteoarthritis Research Society. Results from a workshop. *Osteoarthritis Cartilage* 1996;4:217–43.
3. Bellamy N, Kirwan J, Boers M, *et al.* Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. *J Rheumatol* 1997;24:799–802.
4. Brignardello-Petersen R, Guyatt G, Schandlmaier S, *et al.* Knee arthroscopy versus conservative management in patients with degenerative knee disease: a systematic review. *BMJ Open* 2017;7:e016114.
5. Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. *BMJ Qual Saf* 2014;23:508–18.
6. Schünemann HJ, Guyatt GH. Commentary—Goodbye M(C)ID! Hello MID, where do you come from? *Health Serv Res* 2005;40:593–7.
7. Schünemann HJ, Puhan M, Goldstein R, *et al.* Measurement properties and interpretability of the chronic respiratory disease questionnaire (CRQ). *COPD* 2005;2:81–9.
8. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407–15.
9. Siemieniuk RA, Agoritsas T, Macdonald H, *et al.* Introduction to BMJ Rapid Recommendations. *BMJ* 2016;354:i5191.
10. Alonso-Coeello P, Oxman AD, Moberg J, *et al.* GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines. *BMJ* 2016;353:i2089.
11. Laine C, Taichman DB, Mulrow C. Trustworthy clinical guidelines. *Ann Intern Med* 2011;154:774–5.
12. Siemieniuk RAC, Harris IA, Agoritsas T, *et al.* Arthroscopic surgery for degenerative knee arthritis and meniscal tears: a clinical practice guideline. *BMJ* 2017;257:j1982.
13. Johnston BC, Ebrahim S, Carrasco-Labra A, *et al.* Minimally important difference estimates and methods: a protocol. *BMJ Open* 2015;5:e007953.
14. Mills KA, Naylor JM, Eyles JP, *et al.* Examining the minimal important difference of patient-reported outcome measures for individuals with knee osteoarthritis: a model using the knee injury and osteoarthritis outcome score. *J Rheumatol* 2016;43:395–404.
15. Revicki D, Hays RD, Cella D, *et al.* Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008;61:102–9.
16. Turner D, Schünemann HJ, Griffith LE, *et al.* The minimal detectable change cannot reliably replace the minimal important difference. *J Clin Epidemiol* 2010;63:28–36.
17. Thorlund K, Walter SD, Johnston BC, *et al.* Pooling health-related quality of life outcomes in meta-analysis—a tutorial and review of methods for enhancing interpretability. *Res Synth Methods* 2011;2:188–203.
18. Escobar A, Riddle D. Concordance between important change and acceptable symptom state following knee arthroplasty: the role of baseline scores. *Osteoarthritis Cartilage* 2014;22:1107–10.
19. Tubach F, Dougados M, Falissard B, *et al.* Feeling good rather than feeling better matters more to patients. *Arthritis Rheum* 2006;55:526–30.
20. Escobar A, Garcia Perez L, Herrera-Espineira C, *et al.* Total knee replacement; minimal clinically important differences and responders. *Osteoarthritis Cartilage* 2013;21:2006–12.
21. Tubach F, Ravaud P, Baron G, *et al.* Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. *Ann Rheum Dis* 2005;64:29–33.
22. Angst F, Aeschlimann A, Michel BA, *et al.* Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. *J Rheumatol* 2002;29:131–8.
23. Bellamy N, Hochberg M, Tubach F, *et al.* Development of multinational definitions of minimal clinically important improvement and patient acceptable symptomatic state in osteoarthritis. *Arthritis Care Res* 2015;67:972–80.
24. Browne JP, van der Meulen JH, Lewsey JD, *et al.* Mathematical coupling may account for the association between baseline severity and minimally important difference values. *J Clin Epidemiol* 2010;63:865–74.
25. Escobar A, Quintana JM, Bilbao A, *et al.* Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. *Osteoarthritis Cartilage* 2007;15:273–80.
26. Ornetti P, Dougados M, Paternotte S, *et al.* Validation of a numerical rating scale to assess functional impairment in hip and knee osteoarthritis: comparison with the WOMAC function scale. *Ann Rheumatic Dis* 2011;70:740–6.

27. Monticone M, Ferrante S, Salvaderi S, *et al.* Responsiveness and minimal important changes for the knee injury and osteoarthritis outcome score in subjects undergoing rehabilitation after total knee arthroplasty. *Am J Phys Med Rehabil* 2013;92:864–70.
28. Terwee CB, Roorda LD, Knol DL, *et al.* Linking measurement error to minimal important change of patient-reported outcomes. *J Clin Epidemiol* 2009;62:1062–7.
29. Terwee CB, Roorda LD, Dekker J, *et al.* Mind the MIC: large variation among populations and methods. *J Clin Epidemiol* 2010;63:524–34.
30. Tubach F, Ravaud P, Martin-Mola E, *et al.* Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: results from a prospective multinational study. *Arthritis Care Res (Hoboken)* 2012;64:1699–707.
31. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res* 2005;14:1523–32.
32. Sullivan PW, Ghushchyan V. Preference-based EQ-5D index scores for chronic conditions in the United States. *Med Decis Making* 2006;26:410–20.
33. Rouquette A, Blanchin M, Sébille V, *et al.* The minimal clinically important difference determined using item response theory models: an attempt to solve the issue of the association with baseline score. *J Clin Epidemiol* 2014;67:433–40.
34. Zhang Y, Zhang S, Thabane L, *et al.* Although not consistently superior, the absolute approach to framing the minimally important difference has advantages over the relative approach. *J Clin Epidemiol* 2015;68:888–94.

Linked articles in this *BMJ* Rapid Recommendations cluster

- ▶ Siemieniuk RAC, Harris IA, Agoritsas T, *et al.* Arthroscopic surgery for degenerative knee arthritis and meniscal tears: a clinical practice guideline. *BMJ* 2017;257:j1982. doi:10.1136/bmj.j1982
Summary of the results from the Rapid Recommendation process
- ▶ Brignardello-Peterson R, Guyatt GH, Schandelmaier S, *et al.* Knee arthroscopy versus conservative management in patients with degenerative knee disease: a systematic review. *BMJ Open* 2017;7:e016114. doi:10.1136/bmjopen-2017-016114
Review of all available randomised trials that assessed the benefits of knee arthroscopy compared with non-operative care and observational studies that assessed risks
- ▶ Devji T, Guyatt GH, Lytvyn L, *et al.* Application of minimal important differences in degenerative knee disease outcomes: a systematic review and case study to inform *BMJ* Rapid Recommendations. *BMJ Open* 2017;7:e015587. doi:10.1136/bmjopen-2016-015587
Review addressing what level of individual change on a given scale is important to patients (minimally important difference). The study informed sensitivity analyses for the review on net benefit, informed discussions on patient values and preferences, and was key to interpreting the magnitude of effect sizes and the strength of the recommendation
- ▶ MAGICapp (www.magicapp.org)
Expanded version of the results with multilayered recommendations, evidence summaries, and decision aids for use on all devices