J Shoulder Elbow Surg (2020) 29, 1484–1492

ELSEVIER

Journal of Shoulder and Elbow Surgery

www.elsevier.com/locate/ymse

REVIEW ARTICLES

Minimal clinically important difference (MCID) for patient-reported shoulder outcomes

Ian A. Jones, BA[a], Ryan Togashi, BS[b], Nathanael Heckmann, MD[b], C. Thomas Vangsness Jr, MD[b,*]

[a]School of Medicine, University of California Irvine, Irvine, CA, USA
[b]Department of Orthopaedic Surgery, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

Background: The minimal clinically important difference (MCID) is used when interpreting the importance of outcome data. However, a consensus regarding the MCID for commonly used patient-reported outcomes in shoulder surgery has not been established. The purpose of this systematic review was to evaluate the available literature on shoulder MCID to improve clinical interpretation of shoulder outcome data.
Methods: A systematic review of the literature was conducted to identify studies reporting anchor-based MCID values for the patient-reported outcomes recommended by the American Shoulder and Elbow Surgeons (ASES): Veterans Rand 12 score, ASES score, Single Assessment Numeric Evaluation (SANE) score, Western Ontario Rotator Cuff (WORC) score, Western Ontario Osteoarthritis Score (WOOS), Western Ontario Shoulder Instability Index (WOSI), Pennsylvania Shoulder Score, and Oxford Shoulder Score (OSS).
Results: A total of 14 articles reporting anchor-based MCID values were included in the final analysis. No studies reporting the Western Ontario Osteoarthritis Score (WOOS) were identified. The ASES score (6 studies), OSS (4 studies), and WORC score (2 studies) were the only instruments investigated in more than 1 study. The average reported MCID values for the ASES, OSS, and WORC scores were 15.5 (15% total difference), 275.7 (13% total difference), and 6 (13% total difference), respectively. The vast majority of studies failed to report information necessary to validate the credibility of these MCID values.
Discussion and conclusion: The current utility of the MCID for patient-report shoulder outcome instruments is limited by poor study methodology, inadequate reporting, and a lack of data. Further research is needed to more clearly define the MCID values for commonly used patient-reported outcomes in shoulder surgery.
Level of evidence: Systematic Review; Basic Science

Keywords: Minimal clinically important difference; MCID; shoulder; rotator cuff; outcome measurement

A number of patient-reported shoulder outcomes are well-established in the literature; however, the clinical interpretability of these instruments remains a challenge. The minimal clinically important difference (MCID) attempts to bridge the gap between numeric results and patient experience by assigning a difference threshold for clinical importance.[25] However, many commonly used patient-reported outcomes in shoulder surgery do not have established MCID values based on patient perception.

The statistical methods used to calculate the MCID fall into 2 broad categories: (1) anchor based and (2) distribution based. Anchor-based methods link the change in outcome measure to an external criterion that accounts for the patient's perspective using a global rating-of-change scale.[14] For shoulder assessment instruments, the anchor is often a questionnaire that asks patients to rate improvement in shoulder function

No institutional review board approval was required for this systematic review.

*Reprint requests: C. Thomas Vangsness, Jr, MD, Department of Orthopaedic Surgery, Keck School of Medicine of USC, HCT 1520 San Pablo St, Ste 2000, Los Angeles, CA 90033, USA.
E-mail address: Vangsness@usc.edu (C.T. Vangsness).

1058-2746/$ - see front matter
https://doi.org/10.1016/j.jse.2019.12.033

retrospectively.[25] For example, a 4-point anchor might ask patients to rate the change in pain since last visit as "worse," "no different," "improved," or "much improved." The shoulder outcome scores of patients answering "worse" or "no change" could then be averaged and compared with the scores of patients who answered "improved." Distribution-based methods rely on the variance of a given instrument.[41] As such, distribution-based methods, which can be calculated a variety of different ways (using the standard error of measurement, standard deviation, effect size, minimal detectable change, reliable change index, standardized response mean, and so on), depend on the distribution of scores within a given cohort rather than a patient's perception of improvement.[47] Although distribution-based methods are still widely used in the orthopedic literature and can act as a proxy for the MCID in certain cases, they are generally considered less informative than anchor-based estimates because they have no external reference point and rely on the statistical properties of the distribution.[9,13,14,17,19,41,47,48,60,66,68] As such, this review excludes distribution-based estimates of the MCID.

Inconsistent nomenclature and the historical diversity of the techniques used to calculate the MCID have limited widespread application in the field of orthopedic surgery (Fig. 1). In addition to the MCID, the terms "minimal important change,"[30] "minimal clinically important improvement,"[38] and "minimal important difference"[52,53] have all been used to describe patients who reach a "responder" threshold. This article does not distinguish between different nomenclature variants, and the term "MCID" will be used nonspecifically to refer to a difference threshold for clinically meaningful change.

A number of articles have reviewed shoulder assessment instruments.[7,45,51] However, these reviews focused on the psychometric properties of the instruments and did not assess the methods by which these values were determined. The purpose of this review was to systematically aggregate the available data from published shoulder studies investigating the MCID to provide clinicians with a resource that can be used when interpreting patient-reported outcome shoulder data.

Methods

Selection of patient-reported outcomes

This review focused exclusively on the patient-reported outcomes recommended by the American Shoulder and Elbow Surgeons (ASES) Value Committee.[23] On the basis of its recommendations, the Veterans Rand 12 (VR-12),[54] ASES score,[37,42,50] Single Assessment Numeric Evaluation (SANE) score,[64] Western Ontario Rotator Cuff (WORC) score,[34] Western Ontario Osteoarthritis Score (WOOS),[40] Western Ontario Shoulder Instability Index (WOSI),[36] Pennsylvania Shoulder Score,[39] and Oxford Shoulder Score (OSS)[15] were included. A general overview of each instrument is provided in Table I.

Literature search and study identification

The overall approach used to identify and screen studies is outlined in Figure 2 and is based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) checklist items 1-13.[43] A comprehensive search of the literature was conducted in November 2018 using the electronic databases PubMed, Medline, and Cochrane Library. The following search terms were used: (1) "minimal important difference Veterans Rand 12"; (2) "minimal important difference American Shoulder and Elbow Surgeons"; (3) "minimal important difference (SANE) Single Assessment Numeric Evaluation Score"; (4) "minimal important difference Western Ontario Rotator Cuff"; (5) "minimal important difference Western Ontario Stability Index"; (6) "minimal important difference Western Ontario Osteoarthritis Score"; (7) "minimal important difference Oxford Shoulder Score"; (8) "minimal important difference Pennsylvania Shoulder Score"; and (9) "minimal important difference shoulder." An additional 5 studies reporting on the MCID were identified from article references.

Studies that derived MCID values using distribution-based approaches exclusively were excluded. Although distribution-based methods provide supportive information and may be useful temporary substitutes when anchor-based values are unavailable, they may not accurately reflect improvement from a patient perspective.[9,13,14,17,19,48,60,68] As such, only articles using anchor-based methods were included in our final analysis.

Data presentation and organization

Basic descriptive data were extracted, including participant demographic characteristics, length of follow-up, and anchor information. MCID values were organized by outcome and reported alongside information needed to validate their credibility (as discussed later). To provide additional perspective, MCID values for each outcome were averaged (if reported in multiple studies) and then divided by the maximum possible score.

Credibility assessment

An assessment was performed to identify credible MCID values based on criteria described previously.[16,27] In brief, studies that reported a correlation of $R \geq 0.4$ between the change in the patient-reported outcome under consideration and the anchor (ie, global rating of change) were considered reasonably "credible," whereas those that reported a correlation of $R < 0.4$ or failed to report the correlation entirely were considered "questionable."

Results

A total of 14 publications reporting anchor-based MCID values were included (Table II). The sample size and length of follow-up were highly variable, ranging from 44 to 1856 patients and 1 to 157 months, respectively. Most studies used a single anchor that measured overall improvement. Five-point anchors were used most commonly, although

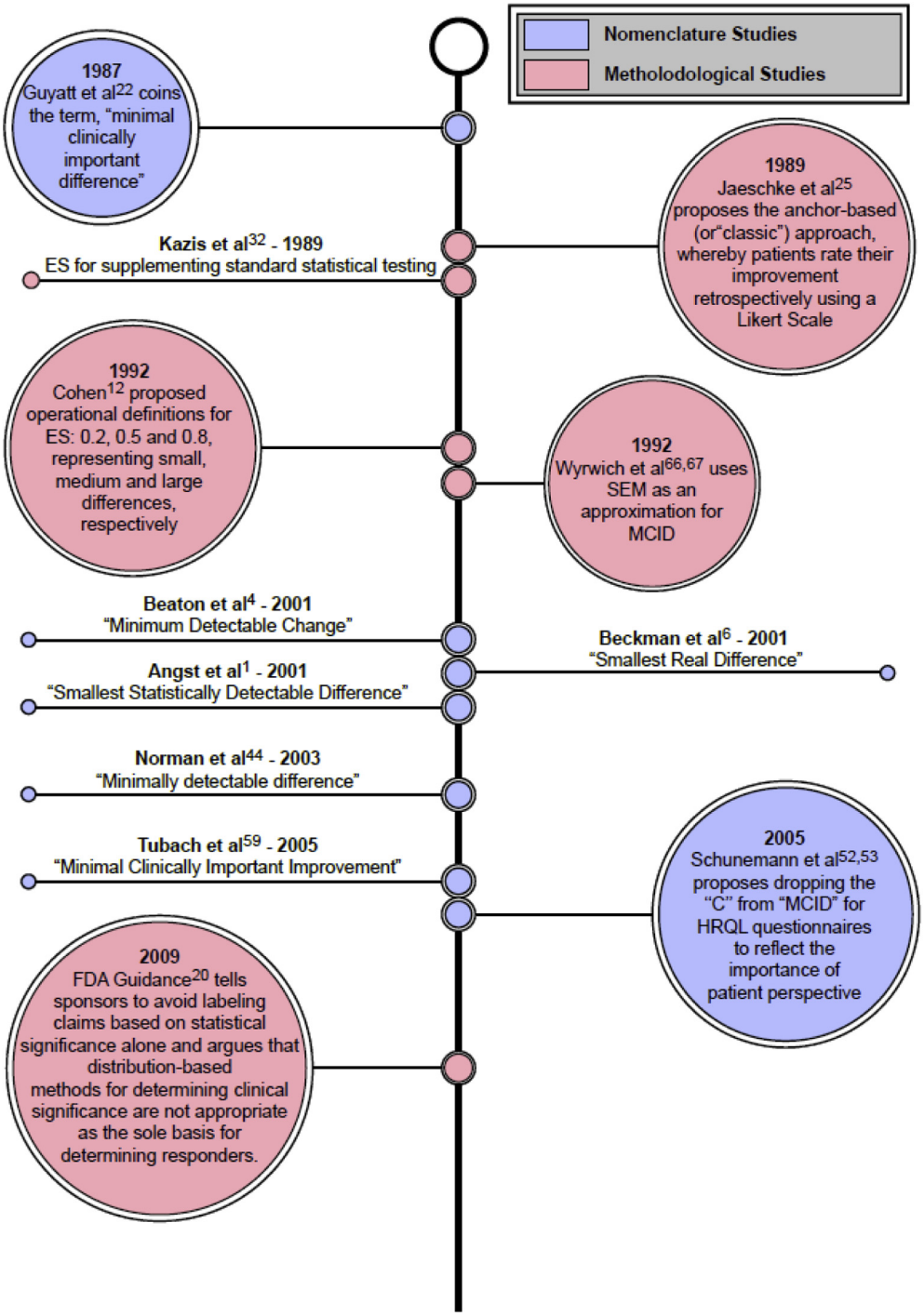

Figure 1 Historical timeline of key publications establishing nomenclature (*blue*) and methodology (*pink*). Particularly notable studies are expanded and include key findings or propositions.[1,4,6,12,20,22,32,44,59,67] *ES*, effect size; *SEM*, standard error of measurement; *MCID*, minimal clinically important difference; *HRQL*, health-related quality of life; *FDA*, US Food and Drug Administration.

anchors with as few as 2 questions or as many as 11 questions were also used.

The majority of studies failed to report the duration of follow-up, and several studies that did report follow-up used data recorded over a large follow-up window. For example, follow-up times varied by almost a year in the first study by Tashjian et al[57] and by 3 years in their second study.[58] In one study, data were recorded within a 24- to 157-month follow-up window.[55]

Several studies included in the final analysis failed to provide critical information related to the MCID values (Table III). More than one-third of studies, by failing to provide confidence intervals, standard deviation, or standard error, did not quantify the data distribution. In

Table I Overview of commonly used shoulder outcome instruments

Clinical utility	Questionnaire	Domains	No. of questions	Minimum-maximum score
General quality of life	Veterans Rand 12 score	Health perception, physical functioning, role limitations due to physical and emotional problems, bodily pain, energy-fatigue, social functioning, mental health	12	SD*
Shoulder specific	American Shoulder and Elbow Surgeons score	Pain, function	11	0-100
Ease of implementation	Single Assessment Numeric Evaluation score	Satisfaction	1	0-100
Disease specific	Western Ontario Rotator Cuff score	Physical symptoms, sport/recreation, work function, lifestyle function, emotional function	21	0-2100
	Western Ontario Osteoarthritis Score	Physical symptoms, sport/recreation/ work function, lifestyle function, emotional function	19	0-1900
	Western Ontario Shoulder Instability score	Physical symptoms, sport/recreation/ work function, lifestyle function, emotional function	21	0-2100
	Oxford Shoulder Score	Pain, activities	12	0-48
	Pennsylvania Shoulder Score	Pain, satisfaction, function	24	0-100

SD, standard deviation.
* In the United States, each increment of 10 points above or below 50 corresponds to 1 standard deviation away from the population average.

addition, the vast majority of studies failed to report the correlation between the patient-reported outcome and the anchor. Because of this, only 3 of the included estimates were considered credible.

To evaluate the relative magnitude of the reported MCID values, the average MCID (excluding Veterans Rand 12 [VR-12] score and Western Ontario Osteoarthritis Score [WOOS]) and corresponding percentages for each patient-reported outcome score are listed in Table IV. The range of MCID values proposed by different studies investigating the ASES score was large, mostly because of 1 study,[42] which reported a low MCID value. The SANE and Pennsylvania Shoulder Score MCID percentages, which were based on a single patient cohort, were higher than the percentages for the ASES score, WORC score, Western Ontario Shoulder Instability Index (WOSI), and OSS.

Discussion

This review highlights poor study methodology, inadequate reporting, and a lack of data on the MCID for ASES-recommended shoulder outcome instruments. Anchor-based MCID values for most of the outcomes included in this study have only been reported once, and even among the outcomes studied in different patient cohorts, MCID values have often been reported without key information needed to validate their credibility. The ASES score, WORC score, and OSS were the only instruments that had multiple studies reporting anchor-based MCID values. The WORC score and OSS were also the only instruments for which at least 1 credible MCID value was reported in this review.

In one of the few reviews of shoulder assessment instruments to include MCID data, the MCID reported for the ASES score was 6.4.[51] However, this value was based solely on the work of Michener et al,[42] whereas the MCID value provided in this review (15.5) was derived by averaging the results of 6 independent studies. Moreover, the MCID value for the ASES score reported by Michener et al appeared to be an outlier, suggesting that 6.4 may be an underestimation. The low MCID value obtained by Michener et al also could have been caused by the short follow-up used in their study. However, it should be noted that reporting of MCID values as a whole was poor, and none of the studies investigating the ASES score were considered credible.

The anchor used to estimate the MCID should be considered when appraising MCID values. Anchors that have too few options may induce ‘‘end-aversion bias,’’[56] which refers to an inherent tendency not to pick extreme ends of a scale. It has been suggested that anchors with a scale of 7-11 points may offer the best compromise between patient preference, discriminative ability, and test-retest reliability,[46] but most of the MCID values examined in this review were derived using anchors with less than 7 points. Anchor scales should also be balanced, with an equal number of options on both sides and a point labeled ‘‘unchanged.’’[31] However, many of the MCID values

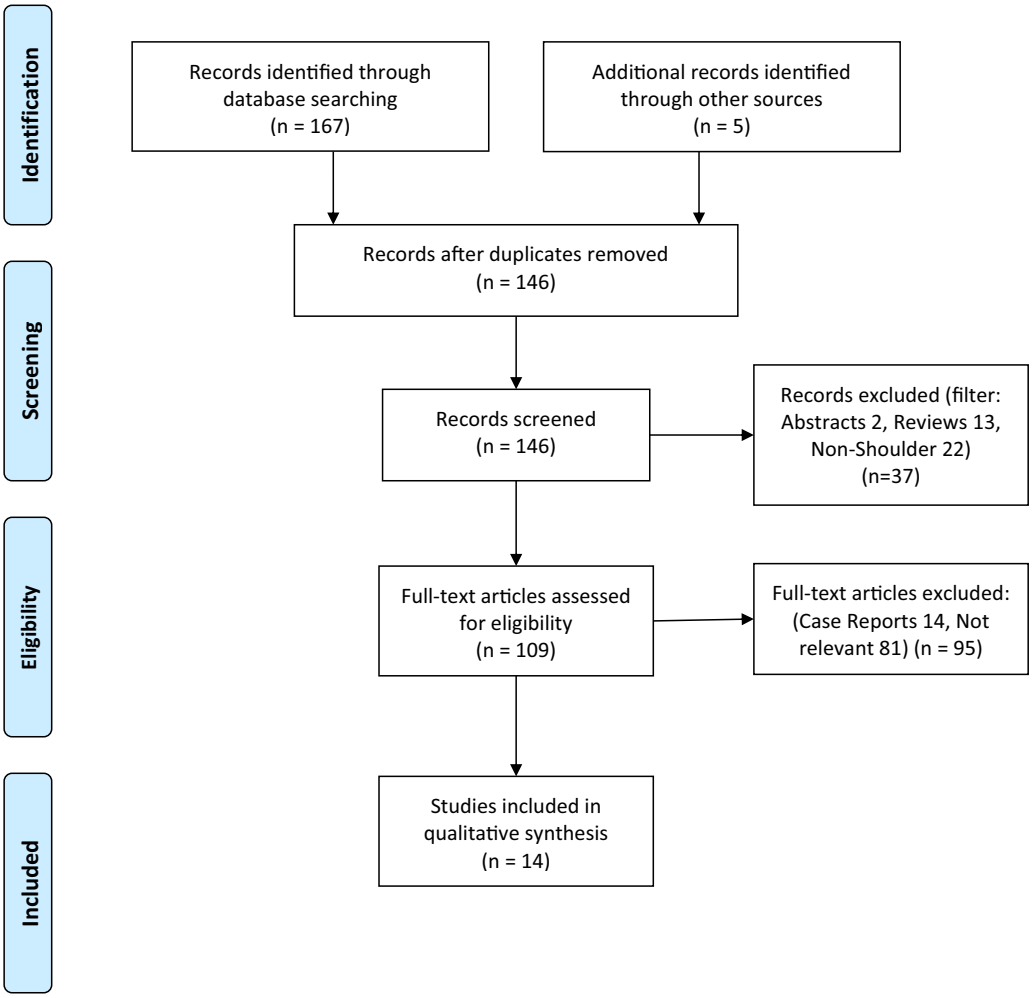

Figure 2 Methodology for systematic review.

included in this review were derived using uneven scales. Another possible anchor-related limitation is anchor domain, which should agree with the domains it is trying to evaluate.[62] For example, the ASES score has both pain and function domains, so pain and function anchors should be used when estimating the MCID.

The methodologic heterogeneity used to derive the MCID presents a major challenge, and issues related to MCID methodology have led some authors to argue that the MCID should be used only as a supplementary instrument not as a basis for clinical decision making.[2] Indeed, Beaton et al[5] showed that different approaches produce different thresholds for interpretable change in shoulder pain, raising broader concerns about the interpretability of responder-type analysis. However, this review focused exclusively on MCID values derived using anchor-based methods, and anchor-based MCID values are less susceptible to methodologic variability than distribution-based methods.[60] For example, in the study of Beaton et al, only 2 of the 13 different methodologies used to calculate the MCID values used anchor-based techniques, and the difference between the 2 anchor-based MCID values was negligible.

The MCID varies based on the population and methodology used to derive it.[65] There are several practical problems in estimating the MCID that make absolute thresholds suspect, and no one MCID value is valid for all applications.[33] As such, it is important that the MCID be considered in the broader context of the disease being treated, study-specific factors, currently available interventions, and the overall risk-benefit ratio of the treatment itself.[17] Longer follow-up times are generally thought to increase the MCID owing to recall bias,[14] and small differences are more likely to be clinically significant when symptoms are more severe.[65] For example, the MCID for the treatment of full-thickness rotator cuff tears is likely lower than the MCID for the treatment of partial-thickness tears.[24]

The 2013 American Academy of Orthopaedic Surgeons Evidence Based Guidelines adopted a 10% difference threshold as an MCID for evaluating patient-reported knee outcomes.[26] The 10% threshold, which was based on empirical derivations of the MCID (presented at a US Food and Drug Administration public meeting in 2012), was controversial[29] and prompted some researchers to argue

Table II Characteristics of studies reporting MCID values for shoulder patient-reported outcomes

Study	Instrument	Sample size	Follow-up range, mo*	Mean age, yr	Disease or condition	Treatment	Anchor domain	Anchor scale
Tashjian et al,[57] 2010	ASES	81	1.5-12.5	51	RC tear or RC tendinitis	Nonoperative	Function, pain, and overall	Mixed
Gagnier et al,[21] 2018	ASES and WORC	222	14.8-15	60.6	RC tears	Nonoperative and operative	Overall	11 points
Werner et al,[63] 2016	ASES	490	NR	68	RC tears	Operative	Work, activities, and overall	5 points
Tashjian et al,[58] 2017	ASES	326	24-60	67.4	RC tears	Operative	Overall	4 points
Christiansen et al,[10] 2015	OSS	126	2-3	48.2	Subacromial impingement syndrome	Nonoperative	Overall	7 points
Ekeberg et al,[18] 2010	OSS and WORC	122	NR	51	Nonspecific	Nonoperative	Overall	2 points
Van Kampen et al,[61] 2013	OSS	164	NR	41	Nonspecific	Nonoperative and operative	Pain and function	7 points
Michener et al,[42] 2002	ASES	63	0.75-1.25	51.7	Nonspecific	Nonoperative and operative	Function	5 points
Simovitch et al,[55] 2018	ASES	1856	24-157	69.6	Nonspecific	Operative	Overall	4 points
Christie et al,[11] 2011	OSS	100	NR	63.2	Rheumatic shoulder disease	Operative	Overall	5 points
Kirkley et al,[35] 2003	WORC	44	NR	NR	RC tear or RC tendinitis	Nonoperative	Overall	5 points
	WOSI	NR	NR	NR	NR	NR	Overall	5 points
Leggin et al,[39] 2006	PENN	109	NR	49.1	Nonspecific	Nonoperative	Overall	5 points
Braun and Handoll,[8] 2018	WORC	64	NR	50	RC tear	Nonoperative	Overall	7 points
Zhou et al,[69] 2018	VR-12 and SANE	222	NR	60.5	RC tear	Nonoperative and operative	Overall	5 points

MCID, minimal clinically important difference; *ASES*, American Shoulder and Elbow Surgeons; *RC*, rotator cuff; *WORC*, Western Ontario Rotator Cuff; *OSS*, Oxford Shoulder Score; *NR*, not reported; *WOSI*, Western Ontario Shoulder Instability; *PENN*, Pennsylvania Shoulder Score; *VR-12*, Veterans Rand 12; *SANE*, Single Assessment Numeric Evaluation.

* When follow-up was reported in weeks, 4 weeks was assumed to be 1 month.

that a 10% difference runs the risk of condemning potentially viable treatments in the context of limited available alternatives.[2] As discussed, stringent application of the MCID is indeed problematic; however, the results of this study support the American Academy of Orthopaedic Surgeons methodology, suggesting that 10% may actually be a fairly conservative estimate in most cases, at least for patient-reported shoulder outcomes. Moreover, the purpose of this study was not to advocate for (or against) any particular application of the MCID; the literature on the MCID should not be discounted entirely. Johnston et al[28] suggested that as the pooled estimate falls below 0.5 of the MCID, it becomes progressively less likely that an appreciable number of patients will achieve important benefits from the treatment. This approach may afford some practical utility to the average MCID values presented in this study, particularly when similar studies estimating MCID values are not available.

When considering the average MCID values and ranges presented in this review, it is important to understand that the MCID is not a universal threshold. Clinicians must consider multiple factors, including study size, methodology, and quality, as well as the severity of the pathology, follow-up time, and cost of treatment.[49] It is also important to understand that important change needs to be considered differently for individuals and groups,[3] and the group mean change may bear little relation to an important improvement for a particular individual.[17]

Conclusion

This review demonstrates that the literature regarding the MCID for the most popular shoulder assessment instruments is fairly limited. Most studies to date fail to report information necessary to validate the credibility of the reported MCID. Shoulder surgeons should take this into account when interpreting studies that use these instruments.

Table III Studies reporting MCID values for commonly used shoulder assessment instruments

	Average final follow-up, mo	MCID	Data distribution	Credibility
VR-12 score				
Zhou et al,[69] 2018	14.5	PCS: 2.57	PCS: 90% CI, −1.62 to 6.76 SD, ±10.84	Questionable
		MCS: 1.87	MCS: 90% CI, −2.07 to 5.80 SD, ±10.18	Questionable
ASES score				
Michener et al,[42] 2002	1	6.4	AUC, 0.82	Questionable
Tashjian et al,[57] 2010	3.6	16.72	NR	Questionable
Gagnier et al,[21] 2018	14.8	21.9	95% CI, 4.1 to 39.6	Questionable
Werner et al,[63] 2016	24	13.5	95% CI, 4.8 to 22.3	Questionable
Tashjian et al,[58] 2017	24	20.9	NR	Questionable
Simovitch et al,[55] 2018	44.9	13.6	95% CI, 13.4 to 13.8 SD, ±2.3	Questionable
SANE score				
Zhou et al, 2018	14.5	27.25	90% CI, 16.17 to 38.33 SD, ±28.65	Questionable
WORC score				
Ekeberg et al,[18] 2010	1.5	275	NR	Credible
Braun and Handoll,[8] 2018	3	300	NR	Questionable
Kirkley et al,[35] 2003	3	245.26	NR	Questionable
Gagnier et al,[21] 2018	14.8	282.6	95% CI, −39.0 to 604.3	Questionable
WOOS				
No studies identified				
WOSI				
Kirkley et al,[35] 2003	NR	220	NR	Questionable
OSS				
Ekeberg et al,[18] 2010	1.5	5	NR	Credible
Christiansen et al,[10] 2015	3	6	AUC, 0.81	Credible
Van Kampen et al,[61] 2013	6	6	SD, ±5.3	Questionable
Christie et al,[11] 2011	12	6.9	NR	Questionable
PENN				
Leggin et al,[39] 2006	1	11.4	SD, ±9.5	Questionable

MCID, minimal clinically important difference; *PCS*, physical component; *CI*, confidence interval; *SD*, standard deviation; *MCS*, mental component; *VR-12*, Veterans Rand 12; *ASES*, American Shoulder and Elbow Surgeons; *AUC*, area under curve; *NR*, not reported; *SANE*, Single Assessment Numeric Evaluation; *WORC*, Western Ontario Rotator Cuff; *WOOS*, Western Ontario Osteoarthritis Score; *WOSI*, Western Ontario Shoulder Instability; *OSS*, Oxford Shoulder Score; *PENN*, Pennsylvania Shoulder Score.

Table IV Overview of MCID values

Outcome	MCID			Approximate % difference
	Mean	Range	SD	
ASES score	15.5	6.4-21.9	5.7	15
SANE score	NA	NA	NA	27
WORC score	275.7	245.3-300	22.8	13
WOSI	NA	NA	NA	12
OSS	6	5-6.9	0.8	13
PENN	NA	NA	NA	24

MCID, minimal clinically important difference; *ASES*, American Shoulder and Elbow Surgeons; *SANE*, Single Assessment Numeric Evaluation; *NA*, not applicable (value based on single study); *WORC*, Western Ontario Rotator Cuff; *WOSI*, Western Ontario Shoulder Instability; *OSS*, Oxford Shoulder Score; *PENN*, Pennsylvania Shoulder Score; *VR-12*, Veterans Rand 12; *SD*, standard deviation.

Disclaimer

The authors, their immediate families, and any research foundations with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

References

1. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. Arthritis Rheum 2001;45:384-91.

2. Bannuru RR, Vaysbrot EE, McIntyre LF. Did the American Academy of Orthopaedic Surgeons osteoarthritis guidelines miss the mark? Arthroscopy 2014;30:86-9. https://doi.org/10.1016/j.arthro.2013.10.007
3. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. Curr Opin Rheumatol 2002;14:109.
4. Beaton DE, Bombardier C, Katz JN, Wright JG, Wells G, Boers M, et al. Looking for important change/differences in studies of responsiveness. OMERACT MCID Working Group. Outcome Measures in Rheumatology. Minimal clinically important difference. J Rheumatol 2001;28:400-5.
5. Beaton DE, van Eerd D, Smith P, van der Velde G, Cullen K, Kennedy CA, et al. Minimal change is sensitive, less specific to recovery: a diagnostic testing approach to interpretability. J Clin Epidemiol 2011;64:487-96. https://doi.org/10.1016/j.jclinepi.2010.07.012
6. Beckerman H, Roebroeck ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek AL. Smallest real difference, a link between reproducibility and responsiveness. Qual Life Res 2001;10:571-8.
7. Bot SDM, Terwee CB, van der Windt DAWM, Bouter LM, Dekker J, de Vet HCW. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. Ann Rheum Dis 2004;63:335-41.
8. Braun C, Handoll HH. Estimating the Minimal Important Difference for the Western Ontario Rotator Cuff Index (WORC) in adults with shoulder pain associated with partial-thickness rotator cuff tears. Musculoskelet Sci Pract 2018;35:30-3. https://doi.org/10.1016/j.msksp.2018.02.003
9. Cella D, Eton DT, Lai J-S, Peterman AH, Merkel DE. Combining anchor and distribution-based methods to derive minimal clinically important differences on the Functional Assessment of Cancer Therapy (FACT) anemia and fatigue scales. J Pain Symptom Manage 2002;24:547-61.
10. Christiansen DH, Frost P, Falla D, Haahr JP, Frich LH, Svendsen SW. Responsiveness and minimal clinically important change: a comparison between 2 shoulder outcome measures. J Orthop Sports Phys Ther 2015;45:620-5. https://doi.org/10.2519/jospt.2015.5760
11. Christie A, Dagfinrud H, Garratt AM, Ringen Osnes H, Hagen KB. Identification of shoulder-specific patient acceptable symptom state in patients with rheumatic diseases undergoing shoulder surgery. J Hand Ther 2011;24:53-60. https://doi.org/10.1016/j.jht.2010.10.006. quiz 61.
12. Cohen J. A power primer. Psychol Bull 1992;112:155-9.
13. Crosby RD, Kolotkin RL, Williams GR. An integrated method to determine meaningful changes in health-related quality of life. J Clin Epidemiol 2004;57:1153-60. https://doi.org/10.1016/j.jclinepi.2004.04.004
14. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol 2003;56:395-407. https://doi.org/10.1016/s0895-4356(03)00044-1
15. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. J Bone Joint Surg Br 1996;78:593-600.
16. Devji T, Guyatt GH, Lytvyn L, Brignardello-Petersen R, Foroutan F, Sadeghirad B, 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. https://doi.org/10.1136/bmjopen-2016-015587
17. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008;9:105-21. https://doi.org/10.1016/j.jpain.2007.09.005
18. Ekeberg OM, Bautz-Holter E, Keller A, Tveitå EK, Juel NG, Brox JI. A questionnaire found disease-specific WORC index is not more responsive than SPADI and OSS in rotator cuff disease. J Clin Epidemiol 2010;63:575-84. https://doi.org/10.1016/j.jclinepi.2009.07.012
19. Eton DT, Cella D, Yost KJ, Yount SE, Peterman AH, Neuberg DS, et al. A combination of distribution- and anchor-based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale. J Clin Epidemiol 2004;57:898-910. https://doi.org/10.1016/j.jclinepi.2004.01.012
20. Food and Drug Administration. Patient-reported outcome measures: use in medical product development to support labeling claims. Silver Spring, MD: Food and Drug Administration; 2009.
21. Gagnier JJ, Robbins C, Bedi A, Carpenter JE, Miller BS. Establishing minimally important differences for the American Shoulder and Elbow Surgeons score and the Western Ontario Rotator Cuff Index in patients with full-thickness rotator cuff tears. J Shoulder Elbow Surg 2018;27:e160-6. https://doi.org/10.1016/j.jse.2017.10.042
22. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis 1987;40:171-8.
23. Hawkins RJ, Thigpen CA. Selection, implementation, and interpretation of patient-centered shoulder and elbow outcomes. J Shoulder Elbow Surg 2018;27:357-62. https://doi.org/10.1016/j.jse.2017.09.022
24. Holmgren T, Oberg B, Adolfsson L, Björnsson Hallgren H, Johansson K. Minimal important changes in the Constant-Murley score in patients with subacromial pain. J Shoulder Elbow Surg 2014;23:1083-90. https://doi.org/10.1016/j.jse.2014.01.014
25. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407-15.
26. Jevsevar DS. Treatment of osteoarthritis of the knee: evidence-based guideline, 2nd edition. J Am Acad Orthop Surg 2013;21:571-6. https://doi.org/10.5435/JAAOS-21-09-571
27. Johnston BC, Ebrahim S, Carrasco-Labra A, Furukawa TA, Patrick DL, Crawford MW, et al. Minimally important difference estimates and methods: a protocol. BMJ Open 2015;5:e007953. https://doi.org/10.1136/bmjopen-2015-007953
28. Johnston BC, Thorlund K, Schünemann HJ, Xie F, Murad MH, Montori VM, et al. Improving the interpretation of quality of life evidence in meta-analyses: the application of minimal important difference units. Health Qual Life Outcomes 2010;8:116. https://doi.org/10.1186/1477-7525-8-116
29. Jones IA, Togashi R, Wilson ML, Heckmann N, Vangsness CT. Intra-articular treatment options for knee osteoarthritis. Nat Rev Rheumatol 2018;85(Suppl 3):49. https://doi.org/10.1038/s41584-018-0123-4
30. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. J Clin Epidemiol 1994;47:81-7.
31. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. J Man Manip Ther 2009;17:163-70. https://doi.org/10.1179/jmt.2009.17.3.163
32. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27(Suppl):S178-89.
33. King MT. A point of minimal important difference (MID): a critique of terminology and methods. Expert Rev Pharmacoecon Outcomes Res 2011;11:171-84. https://doi.org/10.1586/erp.11.9
34. Kirkley A, Alvarez C, Griffin S. The development and evaluation of a disease-specific quality-of-life questionnaire for disorders of the rotator cuff: the Western Ontario Rotator Cuff Index. Clin J Sport Med 2003;13:84-92. https://doi.org/10.1097/00042752-200303000-00004
35. Kirkley A, Griffin S, Dainty K. Scoring systems for the functional assessment of the shoulder. Arthroscopy 2003;19:1109-20. https://doi.org/10.1016/j.arthro.2003.10.030
36. Kirkley A, Griffin S, McLintock H, Ng L. The development and evaluation of a disease-specific quality of life measurement tool for shoulder instability. The Western Ontario Shoulder Instability Index (WOSI). Am J Sports Med 1998;26:764-72.
37. Kocher MS, Horan MP, Briggs KK, Richardson TR, O'Holleran J, Hawkins RJ. Reliability, validity, and responsiveness of the American Shoulder and Elbow Surgeons subjective shoulder scale in patients

with shoulder instability, rotator cuff disease, and glenohumeral arthritis. J Bone Joint Surg Am 2005;87:2006-11. https://doi.org/10.2106/JBJS.C.01624
38. Kvien TK, Heiberg T, Hagen KB. Minimal clinically important improvement/difference (MCII/MCID) and patient acceptable symptom state (PASS): what do these concepts mean? Ann Rheum Dis 2007;66(Suppl 3):iii40-1. https://doi.org/10.1136/ard.2007.079798
39. Leggin BG, Michener LA, Shaffer MA, Brenneman SK, Iannotti JP, Williams GR. The Penn shoulder score: reliability and validity. J Orthop Sports Phys Ther 2006;36:138-51. https://doi.org/10.2519/jospt.2006.36.3.138
40. Lo IK, Griffin S, Kirkley A. The development of a disease-specific quality of life measurement tool for osteoarthritis of the shoulder: the Western Ontario Osteoarthritis of the Shoulder (WOOS) index. Osteoarthritis Cartilage 2001;9:771-8.
41. McGlothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients. JAMA 2014;312:1342-3. https://doi.org/10.1001/jama.2014.13128
42. Michener LA, McClure PW, Sennett BJ. American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, patient self-report section: reliability, validity, and responsiveness. J Shoulder Elbow Surg 2002;11:587-94. https://doi.org/10.1067/mse.2002.127096
43. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred Reporting Items for Systematic Review and Meta-analysis Protocols (PRISMA-P) 2015 statement. Syst Rev 2015;4:1. https://doi.org/10.1186/2046-4053-4-1
44. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care 2003;41:582-92. https://doi.org/10.1097/01.MLR.0000062554.74615.4C
45. Placzek JD, Lukens SC, Badalanmenti S, Roubal PJ, Freeman DC, Walleman KM, et al. Shoulder outcome measures: a comparison of 6 functional tests. Am J Sports Med 2004;32:1270-7. https://doi.org/10.1177/0363546503262193
46. Preston CC, Colman AM. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. Acta Psychol (Amst) 2000;104:1-15.
47. Rai SK, Yazdany J, Fortin PR, Aviña-Zubieta JA. Approaches for estimating minimal clinically important differences in systemic lupus erythematosus. Arthritis Res Ther 2015;17:143. https://doi.org/10.1186/s13075-015-0658-6
48. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008;61:102-9. https://doi.org/10.1016/j.jclinepi.2007.03.012
49. Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. Health Qual Life Outcomes 2006;4:70. https://doi.org/10.1186/1477-7525-4-70
50. Richards RR, An KN, Bigliani LU, Friedman RJ, Gartsman GM, Gristina AG, et al. A standardized method for the assessment of shoulder function. J Shoulder Elbow Surg 1994;3:347-52.
51. Roy J-S, MacDermid JC, Woodhouse LJ. Measuring shoulder function: a systematic review of four questionnaires. Arthritis Rheum 2009;61:623-32. https://doi.org/10.1002/art.24396
52. Schünemann HJ, Guyatt GH. Commentary—goodbye M(C)ID! Hello MID, where do you come from? Health Serv Res 2005;40:593-7. https://doi.org/10.1111/j.1475-6773.2005.00374.x
53. Schünemann HJ, Puhan M, Goldstein R, Jaeschke R, Guyatt GH. Measurement properties and interpretability of the Chronic Respiratory Disease Questionnaire (CRQ). COPD 2005;2:81-9. https://doi.org/10.1081/COPD-200050651
54. Selim AJ, Rogers W, Fleishman JA, Qian SX, Fincke BG, Rothendler JA, et al. Updated U.S. population standard for the Veterans RAND 12-item Health Survey (VR-12). Qual Life Res 2009;18:43-52. https://doi.org/10.1007/s11136-008-9418-2
55. Simovitch R, Flurin P-H, Wright T, Zuckerman JD, Roche CP. Quantifying success after total shoulder arthroplasty: the minimal clinically important difference. J Shoulder Elbow Surg 2018;27:298-305. https://doi.org/10.1016/j.jse.2017.09.013
56. Streiner DL, Norman GR, Cairney J. Health measurement scales. New York: Oxford University Press, USA; 2014.
57. Tashjian RZ, Deloach J, Green A, Porucznik CA, Powell AP. Minimal clinically important differences in ASES and simple shoulder test scores after nonoperative treatment of rotator cuff disease. J Bone Joint Surg Am 2010;92:296-303. https://doi.org/10.2106/JBJS.H.01296
58. Tashjian RZ, Hung M, Keener JD, Bowen RC, McAllister J, Chen W, et al. Determining the minimal clinically important difference for the American Shoulder and Elbow Surgeons score, Simple Shoulder Test, and visual analog scale (VAS) measuring pain after shoulder arthroplasty. J Shoulder Elbow Surg 2017;26:144-8. https://doi.org/10.1016/j.jse.2016.06.007
59. 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;64:29-33. https://doi.org/10.1136/ard.2004.022905
60. Turner D, Schünemann HJ, Griffith LE, Beaton DE, Griffiths AM, Critch JN, et al. The minimal detectable change cannot reliably replace the minimal important difference. J Clin Epidemiol 2010;63:28-36. https://doi.org/10.1016/j.jclinepi.2009.01.024
61. van Kampen DA, Willems WJ, van Beers LWAH, Castelein RM, Scholtes VAB, Terwee CB. Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). J Orthop Surg Res 2013;8:40. https://doi.org/10.1186/1749-799X-8-40
62. Ward MM, Guthrie LC, Alba M. Domain-specific transition questions demonstrated higher validity than global transition questions as anchors for clinically important improvement. J Clin Epidemiol 2015;68:655-61. https://doi.org/10.1016/j.jclinepi.2015.01.028
63. Werner BC, Chang B, Nguyen JT, Dines DM, Gulotta LV. What change in American Shoulder and Elbow Surgeons score represents a clinically important change after shoulder arthroplasty? Clin Orthop Relat Res 2016;474:2672-81. https://doi.org/10.1007/s11999-016-4968-z
64. Williams GN, Gangel TJ, Arciero RA, Uhorchak JM, Taylor DC. Comparison of the Single Assessment Numeric Evaluation method and two shoulder rating scales. Outcomes measures after shoulder surgery. Am J Sports Med 1999;27:214-21.
65. Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). J Man Manip Ther 2012;20:160-6. https://doi.org/10.1179/2042618612Y.0000000001
66. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. Med Care 1999;37:469-78.
67. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. J Clin Epidemiol 1999;52:861-73.
68. Yost KJ, Cella D, Chawla A, Holmgren E, Eton DT, Ayanian JZ, et al. Minimally important differences were estimated for the Functional Assessment of Cancer Therapy-Colorectal (FACT-C) instrument using a combination of distribution- and anchor-based approaches. J Clin Epidemiol 2005;58:1241-51. https://doi.org/10.1016/j.jclinepi.2005.07.008
69. Zhou L, Natarajan M, Miller BS, Gagnier JJ. Establishing minimal important differences for the VR-12 and SANE scores in patients following treatment of rotator cuff tears. Orthop J Sports Med 2018;6. https://doi.org/10.1177/2325967118782159. 2325967118782159.