

BMJ Open Minimal important differences for improvement in shoulder condition patient-reported outcomes: a systematic review to inform a *BMJ* Rapid Recommendation

Qiukui Hao,^{1,2} Tahira Devji,² Dena Zeraatkar,² Yuting Wang,^{1,2} Anila Qasim,² Reed A C Siemieniuk,² Per Olav Vandvik,^{3,4} Tuomas Lähdeoja,^{5,6} Alonso Carrasco-Labra,² Thomas Agoritsas,^{2,7} Gordon Guyatt^{2,8}

To cite: Hao Q, Devji T, Zeraatkar D, *et al*. Minimal important differences for improvement in shoulder condition patient-reported outcomes: a systematic review to inform a *BMJ* Rapid Recommendation. *BMJ Open* 2019;**9**:e028777. doi:10.1136/bmjopen-2018-028777

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-028777>).

Received 26 December 2018
Revised 7 January 2019
Accepted 18 January 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Qiukui Hao;
haoqiukui@gmail.com

ABSTRACT

Objectives To identify credible anchor-based minimal important differences (MIDs) for patient-reported outcome measures (PROMs) relevant to a *BMJ* Rapid Recommendations addressing subacromial decompression surgery for shoulder pain.

Design Systematic review.

Outcome measures Estimates of anchor-based MIDs, and their credibility, for PROMs judged by the parallel *BMJ* Rapid Recommendations panel as important for informing their recommendation (pain, function and health-related quality of life (HRQoL)).

Data sources MEDLINE, EMBASE and PsycINFO up to August 2018.

Study selection and review methods We included original studies of any intervention for shoulder conditions reporting estimates of anchor-based MIDs for relevant PROMs. Two reviewers independently evaluated potentially eligible studies according to predefined selection criteria. Six reviewers, working in pairs, independently extracted data from eligible studies using a predesigned, standardised, pilot-tested extraction form and independently assessed the credibility of included studies using an MID credibility tool.

Results We identified 22 studies involving 5562 patients that reported 74 empirically estimated anchor-based MIDs for 10 candidate instruments to assess shoulder pain, function and HRQoL. We identified MIDs of high credibility for pain and function outcomes and of low credibility for HRQoL. We offered median estimates for the systematic review team who applied these MIDs in Grading of Recommendations Assessment, Development and Evaluation (GRADE) evidence summaries and in their interpretations of results in the linked systematic review addressing the effectiveness of surgery for shoulder pain.

Conclusions Our review provides anchor-based MID estimates, as well as a rating of their credibility, for PROMs for patients with shoulder conditions. The MID estimates inform the interpretation for a linked systematic review and guideline addressing subacromial decompression surgery for shoulder pain, and could also prove useful

Strengths and limitations of this study

- Our review includes a comprehensive search for anchor-based minimal important differences (MIDs) for instruments commonly used in Randomized Controlled Trials (RCTs) of shoulder conditions conducted without restrictions of study design or language of publication.
- We undertook judgements of MID credibility using a formal instrument with demonstrated reliability and most studies provided highly credible estimates.
- The range of reported MIDs was wide for some of the patient-reported outcome measures.
- Although participants' disease/conditions, sample size, anchors and analytical methods varied among included studies, we cannot convincingly relate these characteristics to variability in estimates.
- For some instruments used in RCTs of surgery for shoulder, we did not find any study estimating MIDs in our target patient population.

for authors addressing other interventions for shoulder problems.

PROSPERO registration number CRD42018106531.

BACKGROUND

The shoulder is the body's most mobile joint, allowing movement in many directions. Shoulder conditions, including arthritis, adhesive capsulitis, rotator cuff conditions, dislocations, fractures, shoulder instability, and shoulder separation, are common problems that cause pain and disability.¹ Up to 26% of adults have recently experienced shoulder pain.² In the USA, the evaluation and management of one shoulder condition—rotator cuff tears—costs US\$3 billion each year.^{3,4}

The relationship between shoulder pain in an individual and the physical cause is often not clear: anatomical abnormalities are frequently not the cause of an individual patient's shoulder pain. Subacromial pain syndrome, also known as shoulder impingement syndrome or rotator cuff disease, is a broad diagnosis that includes several specific conditions and is one of most common diagnoses for patients with shoulder or upper extremity pain or disability.^{5 6} Subacromial pain syndrome encompasses all non-traumatic shoulder conditions including partial tear of the rotator cuff, tendon cuff degeneration, bursitis, tendinosis, supraspinatus tendinopathy or biceps tendinitis.⁶ It is most often unilateral.

Investigating interventions to address shoulder conditions such as shoulder pain requires measurement of patients' pain and function, best undertaken using patient-reported outcome measures (PROMs). PROMs are reported directly by the patient and address aspects of the patient's experience and perspective without interpretation by the clinician or caregiver.⁷ Investigators of interventions for shoulder conditions often include PROMs addressing shoulder pain, function and health-related quality of life (HRQoL) as their primary outcomes.^{18–14} Interpreting PROMs can, however, be challenging. In particular, interpretation requires knowing if an apparent treatment effect is trivial in magnitude, small but important, moderate or large. Statistical significance provides no insight into this issue.¹⁵

To aid interpretation of PROM findings, researchers developed the concept of the minimal important difference (MID): the smallest change—either positive or negative—that patients perceive as important.^{16 17} The MID can help clinicians, patients, and clinical practice guideline developers interpret the magnitude of effects of interventions on PROMs.^{15 18 19}

There are two common approaches for determining the MID: anchor-based and distribution-based methods.²⁰ Distribution-based methods rely solely on the statistical characteristics of PROMs (eg, mean and SD of PROM scores). These statistical characteristics do not reflect the patient's perspective, severely limiting the distribution-based approach in aiding interpretation of results.^{18 21}

Investigators using the anchor-based approach choose an independent interpretable measure as an external criterion or anchor and then examine the relation between the target PROM instrument and that anchor.¹⁸ Although there is no 'gold standard' anchor-based methodology, our group has used the existing literature and expert input to develop an instrument that measures the credibility of anchor-based MIDs. Among desirable criteria to establish a trustworthy MID is a requirement for at least a moderate correlation between change in the target PROM instrument and the change on the anchor.^{20 22}

Although systematic reviews addressing MIDs in shoulder PROMs are available,^{23–27} they are dated and have not applied an assessment of credibility. Therefore, we set out to identify the most credible anchor-based MID

estimates to inform a systematic review addressing the effectiveness of subacromial decompression surgery for shoulder pain. Our review informed an associated *BMJ* Rapid Recommendations and facilitated interpretation of critical outcomes of interest, including shoulder pain, function and HRQoL. The *BMJ* Rapid Recommendations project is a collaboration between the MAGIC foundation (www.magicproject.org) and the *BMJ*, with the goal of providing timely, trustworthy practice guidelines.²⁸

A variety of study designs could inform MIDs for PROMs chosen by investigators for the Randomized Controlled Trials (RCTs). Therefore, in this systematic review, we (1) summarise MID estimate that comes largely from observational studies for the PROMs chosen by the trialists in RCTs that investigated the effect of surgery on shoulder pain and (2) assessed the credibility of these MID estimates.

METHODS

Guideline panel and patient involvement

The *BMJ* Rapid Recommendations guideline panel provided critical oversight to this systematic review. The panel included academic and community-based practitioners (orthopaedic surgeons, general internists, physiotherapists, a rheumatologist, a general practitioner and a geriatrician), methodologists and patients with lived experience of shoulder pain. The panel members also provided input into the methodology of our review. Patients helped, in particular, to identify the outcomes of interest for which we identified MID estimates.²⁸ This study builds on methods used in a similar *BMJ* Rapid Recommendation on arthroscopic knee surgery.^{29 30}

Instruments under consideration

The *BMJ* Rapid Recommendations panel, informed by the Outcome Measures in Rheumatology shoulder core outcomes set,³¹ nominated shoulder pain, function and HRQoL as critical patient-important outcomes of interest in the management of shoulder conditions. Following guidance from the panel, the systematic review team addressing the effectiveness of surgery for subacromial pain syndrome sought evidence for each of these outcomes in the eligible RCTs. We worked closely with that review team and addressed each of the PROMs corresponding to these constructs included as outcomes in the RCTs that proved eligible for the systematic review addressing the impact of shoulder surgery (the subacromial decompression surgery) (table 1).

Literature search and study identification

This project used a database that includes all articles reporting anchor-based MID from 1989 to April 2015 (the MID concept was first described in the medical literature in 1989).^{16 32} We obtained full access to the database of these MIDs—the leaders (ACL, TD and GG) of that project are participants in the current review.

Table 1 Patient-reported outcome measure instruments considered in this review

| Instrument with full name and abbreviation | General score range | Higher scores are better or worse | Construct(s) measured |
|---|-------------------------------|-----------------------------------|---|
| Pain Numeric Rating Scale | 0–10/0–100 | Worse | Pain |
| Pain Visual Analogue Scale | 0–10/0–100 | Worse | Pain |
| PainDETECT Numerical Rating Scale | 0–10/0–100 | Worse | Pain |
| Disability of the Arm, Shoulder and Hand (DASH) | 0–100 | Worse | Symptom and function |
| Quick DASH | 0–100 | Worse | Symptom and function |
| Shoulder Disability Questionnaire | 0–100 | Worse | Pain-related function of the shoulder |
| Simple Shoulder Test | 0–12 | Better | Shoulder comfort and function |
| Oxford Shoulder Score | 0–48 | Better | Shoulder function and pain |
| Project on Research and Intervention in Monotonous work score | 0–36 for each region | Worse | Pain or other complains |
| Neer score | 0–100 | Worse | Function |
| Constant (Murley) Score | 0–100 | Better | Shoulder function, pain, Activities of Daily Living (ADL) function, the range of motion, strength |
| Watson-Sonnabend score | Pain: 0–10; Function: 0–42 | Pain: worse; Function: better | Satisfaction, pain and 0–3 discrete for 14 function items |
| Short Form Health Survey 36 (SF-36) | 0–100 | Better | Health-related quality of life |
| SF-12 | 0–100 | Better | Health-related quality of life |
| EuroQol 5-dimension 3-level index | –0.59–1 | Better | Health-related quality of life |
| 15 D | 0–1 | Better | Health-related quality of life |
| Hospital Anxiety and Depression Score | 0–42 | Worse | Anxiety and depression |

We conducted a comprehensive search for relevant studies addressing MIDs from February 2015 to August 2018 using the MEDLINE, EMBASE and PsycINFO databases. For outcomes that did not fully meet the definition of patient-reported outcomes (such as Constant score^{33 34}) or were not identified in the systematic review informing the database of MIDs, we conducted a comprehensive search for relevant studies from January 1989 to August 2018. We used the MID search strategy filter from the previous MID database development project including a shoulder filter for the relevant PROMs. We also hand searched references from related reviews. There were no language restrictions. Online supplementary appendix 1 presents the full search strategy.

Study selection

We included studies with any intervention, including expectant management. We included original reports of all studies that estimated MID(s) using anchor-based methods for any candidate PROM (table 1). If, for a particular PROM, MID(s) were available for a shoulder condition, we restricted ourselves to those MIDs. If no study estimated MIDs in patients with shoulder conditions, we used the results from studies focusing on upper extremity musculoskeletal conditions. We did not consider studies that estimated MIDs in patients with lower extremity or other conditions. Because RCTs evaluated the effects of an intervention on pain, function and HRQoL that would

require MIDs for improvement, we did not include MIDs for deterioration.

Eligible studies used any design including retrospective and prospective observational studies or clinical trials that compared the results of a target PROM instrument to an anchor, regardless of the credibility of the design, conduct or results of the study. Two reviewers independently performed title and abstract screening and, subsequently, full-text screening of studies included by either reviewer. At full-text screening, reviewers resolved the disagreement by discussion or, if needed, by consultation with a third reviewer.

Data abstraction

Six reviewers, working in three pairs, independently extracted the following data from eligible studies using a pre-designed, standardised, pilot-tested extraction table: first author name; publication year; country(ies); demographic characteristics of participants (eg, sample size, age, sex, condition or disease); intervention; characteristics of the PROM (eg, construct(s), domain(s) and range); anchor details (eg, construct(s), threshold, range of options, categories or values); details in MID determination methods (eg, number of participants used to estimate the MID, duration of follow-up from baseline, analysis methods and correlation between the anchor and PROM). Reviewers resolved disagreements by discussion.

Table 2 The criteria for credibility assessment

| Item | Assessment aspects | Results |
|------|---|---|
| 1 | Whether the anchor instrument directly addressed the patient's perspective. | 0=No 1=Yes 2=Impossible to tell |
| 2 | Whether patients could easily understand the anchor instrument. | 0=Definitely no 1=Not so much 2=To a great extent 3=Definitely yes 4=Impossible to tell |
| 3 | The correlation between the anchor and the PROM.* | 0=Definitely no 1=Not so much 2=To a great extent 3=Definitely yes NR=Not reported |
| 4 | The precision of the MID estimation. | 0=Definitely no 1=Not so much 2=To a great extent 3=Definitely yes NR=Not reported |
| 5 | Whether the threshold or difference between groups on the anchor used to estimate the MID represented a small but important change. | 0=Definitely no 1=Not so much 2=To a great extent 3=Definitely yes NR=Not reported |

*For anchors with categorical scales the Spearman rather than the Pearson's correlation, is appropriate.
MID, minimal important difference; PROM, patient-reported outcome measure.

Credibility assessment

The MID database project included the development of an instrument to assess the credibility of anchor-based MID estimates and tested its reliability (it proved reliable—manuscript in preparation, data available on request). We defined the credibility of studies estimating the MIDs as the extent to which the methodology and performance of studies are likely to have protected against misleading estimates.³² We used an abridged version of the MID credibility tool developed by our group to measure the credibility of MIDs. The tool needs to assess many aspects of the MIDs (table 2) and has proved reliable (manuscript in preparation). Six reviewers, working in three pairs, independently assessed the credibility of included studies. Reviewers resolved disagreements by discussion. We deemed that the MID estimate had high credibility if three or more of the five criteria were met (either 'definitely yes' or 'to a great extent' for each item); otherwise, we deemed that the MID had low credibility. We regard the credibility as a dichotomous variable (high and low) and do not quantify the credibility.

Synthesis of results

We described the characteristics of eligible studies including MID estimates, demographic characteristics of participants, intervention and characteristics of the instrument and anchor. We identified the median, minimum and maximum values across the range of high credibility trustworthy MID estimates generated from the eligible

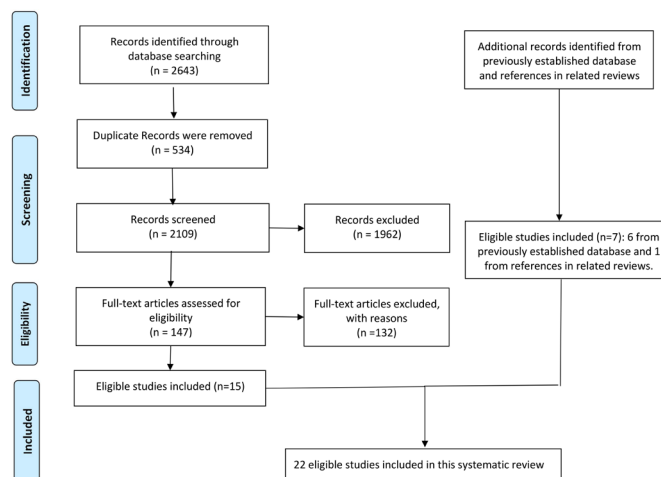


Figure 1 Flowchart for eligible studies identification according to PRISMA guidelines. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

studies for the PROMs of interest. If all MIDs estimates were of low credibility, we presented these estimates.

For each MID with multiple estimates of the MID, we considered variables that may influence the MID. These included: the intervention type (surgical or non-surgical) and, for transition anchors, the period from first to second instrument administration (<3 months vs 3 months or more). We tested the subgroup effect by examining the interaction between each variable and the MID ($p < 0.05$ was deemed statistically significant).

RESULTS

We found six eligible studies from the existing database of anchor-based MIDs and one study from the references in related reviews. We identified 2643 records through our search of electronic databases, of which 534 were duplicates, leaving 2109 records for the title and abstract screening. We excluded 1962 records based on our title and abstract screening and assessed 147 full-text articles, of which 15 were eligible. Therefore, 22 studies were eligible for this review. Figure 1 summarises the study identification process.

Table 3 presents the characteristics of the 22 eligible studies.^{24 35–55} Sample sizes ranged from 20⁴⁹ to 1856,⁴⁶ with a total of 5562 participants providing MID estimates for two relevant instruments assessing shoulder pain, one assessing function, five assessing shoulder symptoms and function and two assessing HRQoL (table 3). The 22 studies reported 74 anchor-based MIDs estimates. Twenty-one of 22 studies employed a variety of transition ratings as the anchor to determine the MIDs, of which five had a follow-up period of less than 3 months.^{38 43 44 48 49} One study used the Pen shoulder score (cut-off point: 8.6) as the anchor to determine the MIDs for pain measurement (PNRS).⁴² Of the 22 studies, 19 reported the absolute estimates for the MIDs and 3—addressing the Constant score, quick Disability of the Arm, Shoulder and Hand (DASH) and Oxford Shoulder Score (OSS)—relative

Table 3 Characteristics of eligible studies

| Author (year) | Disease/conditions | Participants in baseline | Intervention | Instrument/scale | Anchor | Follow-up period |
|--|---|--------------------------|---|---|--|------------------|
| Simovitch <i>et al</i> ⁴⁶ 2018 | Cuff tear arthropathy, a combination of osteoarthritis and rotator cuff insufficiency | 1885 | Total shoulder arthroplasty | Constant score; SST; Pain VAS; | Global rating question | 40.2–49.7 months |
| Negahban <i>et al</i> ⁴⁴ 2015 | Shoulder disorders including impingement syndrome/tendonitis, frozen shoulder, shoulder instability | 200 | Physiotherapy | DASH (Persian version) | Global rating of shoulder function | 1 month |
| Holmgren <i>et al</i> ³⁹ 2014 | Subacromial impingement syndrome | 93 | Physiotherapy | Constant (Murley) shoulder assessment score | Patient's global impression of change | 3 months |
| Rysstad <i>et al</i> ⁴⁵ 2017 | Subacromial pain syndrome | 50 | Physiotherapy | DASH (Norwegian version) | Patient's perceived recovery | 3–4 months |
| van de Water <i>et al</i> ⁴⁹ 2014 | Isolated proximal humeral fracture | 20 | Active rehabilitation | Constant score; OSS; DASH | Patient perception of change | 1.5 months |
| Christiansen <i>et al</i> ³⁵ 2015 | 8–12 weeks after arthroscopic decompression surgery for subacromial impingement syndrome | 112 | Physiotherapy | OSS ; Modified Constant score ; | Patient Global Impression of Change | 3 months |
| Kukkonen <i>et al</i> ⁴⁰ 2013 | Rotator cuff tears (both partial and full thickness) | 781 | Arthroscopy | Constant score; | The two-stage question of the patient satisfaction | 3 months |
| Michener <i>et al</i> ⁴² 2011 | Shoulder pain with or without surgery | 136 | Rehabilitation | PNRS | Pen shoulder score | 3–4 weeks |
| Christie <i>et al</i> ³⁶ 2011 | Rheumatic disease (inflammatory or degenerative disease) undergoing elective shoulder surgery | 100 | Arthroplasty or other surgery (not specified) | DASH; OSS; Pain VAS at activity; Pain VAS at rest; Constant score | Shoulder symptoms question 'At 1-year follow-up, the patients were also asked to rate their shoulder symptoms at present compared with baseline' | 12 months |
| Ekeberg <i>et al</i> ³⁸ 2010 | Rotator cuff disease | 121 | Local ultrasound-guided injections of triamcinolone and Xylocaine | OSS | Main complaint score (–9 (worst) to 9 (best)) | 2 to 6 weeks |
| Mintken <i>et al</i> ⁴³ 2009 | Shoulder pain | 101 | Physical therapy | PNRS; QuickDASH; | Global rating of change | 2–4 weeks |
| Tubach <i>et al</i> ⁴⁸ 2006 | Acute rotator cuff syndrome | 252 | NSAID therapy or placebo | PNRS; Neer score; | Response to NSAID treatment question | 7 days |
| Mahabier <i>et al</i> ⁴¹ 2017 | Humeral shaft fracture | 140 | Operative and non-operative treatment of humeral shaft fracture | DASH; Constant (Murley) score; | Transition item: perception of change in the general condition of the affected upper limb | 1.5–12 months |

Continued

Table 3 Continued

| Author (year) | Disease/conditions | Participants in baseline | Intervention | Instrument/scale | Anchor | Follow-up period |
|--|---|--------------------------|--|---|---|------------------|
| Tashjian <i>et al</i> ⁴⁷ 2017 | Osteoarthritis, rheumatoid arthritis, rotator cuff arthropathy, advanced rotator cuff disease | 326 | Total shoulder arthroplasty (primary anatomic or reverse) | SST Pain VAS | Improvement after treatment | 3.5 years |
| Dritsaki <i>et al</i> ³⁷ 2017 | Rheumatoid arthritis with pain and dysfunction of the hands and/or wrists | 488 | Tailored exercise programme | EQ-5D-3L; EQ-5D-3L VAS; SF-12-physical; SF-12-mental | Participant self-rated improvement in their hands and wrist | 4–8 months |
| Tashjian <i>et al</i> ²⁴ 2009 | Rotator cuff tendinitis, rotator cuff tear (partial or full thickness) | 81 | Non-surgical management | Pain VAS | Four-item anchor instrument: response to treatment | 3.6 months |
| Schmitt and Di Fabio ⁵² 2004 | Musculoskeletal proximal upper extremity problem | 211 | Occupational or physical therapy | DASH | Global disability rating | 3 months |
| van Kampen <i>et al</i> ⁶⁴ 2013 | Shoulder problems | 128 | Operative or non-operative treatment | DASH; QuickDASH | Global rating scale for function | 6 months |
| Lundquist <i>et al</i> ⁶⁰ 2014 | Shoulder conditions (rotator cuff/impingement, adhesive capsulitis, humeroscapular instability, humeroscapular arthrosis, humeral fracture, other or unspecified shoulder disorder) | 81 | NR | DASH (Danish version) | Global impression of change | 2.3 months |
| Tashjian <i>et al</i> ⁶³ 2010 | Rotator cuff tendinitis, rotator cuff tear (partial or full thickness) | 81 | Nonsurgical management | SST | 15-item function question; 4-item improvement question | 3.6 months |
| Marks <i>et al</i> ⁶¹ 2014 | Trapeziometacarpal joint osteoarthritis | 177 | Conservative treatment or surgery (resection/suspension/interposition arthroplasty or arthrodesis) | SF-12-physical; SF-12-mental | Patient-perceived change in thumb condition | 12 months |
| Castricini <i>et al</i> ⁶⁵ 2014 | Irreparable rotator cuff tears | 27 | Shoulder arthroplasty surgery and rehabilitation | Constant and Murley score | The three-stage question of the patient satisfaction. | 27 months |

DASH, Disabilities of the Arm, Shoulder and Hand; EQ-5D-3L, Euro-Quality of life 5-dimensions 3-level index; NR, Not reported; NSAID, Nonsteroidal anti-inflammatory drug; OSS, Oxford shoulder score; PNRS, Pain Numerical Rating Scale; SF-12, Short Form Health Survey 12; SST, Simple shoulder test; VAS, Visual analogue scale.

estimates.^{35 39 43} Patients underwent surgical interventions in 4 studies^{36 40 46 47}; 4 studies used both surgical and non-surgical interventions^{41 51 54 55}; 13 used non-surgical interventions^{24 35 37–39 42–45 48 49 52 53} and 1 did not report the type of intervention.⁵⁰

The analysis methods for estimating the MID included mean change in patients who had experienced a small but minimally important difference over time^{35–40 48 49 52 54 55}; mean difference in groups perceived to have changed versus not changed^{24 40 46 47 53} and Receiver Operating Characteristic (ROC) curves.^{35 38–45 50 51} Fourteen studies provided highly credible estimates and eight studies provided low credibility estimates.^{37 39 42 43 47 48 54 55} Studies with high credibility reported MID estimates for Constant score, Simple Shoulder Test (SST), Pain Visual Analogue Scale (VAS), DASH, OSS and Short Form Health Survey 12 (SF-12) (table 1). Studies provided low credibility MID estimates for the Pain Numeric Rating Scale, Quick DASH, Neer score and EuroQol 5 dimensions 3 levels (EQ-5D-3L) (table 1). No studies estimate MIDs for the following instruments in shoulder or upper extremity conditions: PainDETECT Numerical Rating Scale (0–10), Shoulder Disability Questionnaire (SDQ), Project on Research and Intervention in Monotonous work score, Watson-Sonnabend score, 15D, SF-36 and Hospital Anxiety and Depression Score (HADS).

Table 4 presents median, maximum and minimum estimates of MIDs according to credibility, with the best estimates suggested to the systematic review team shaded. For the MID estimates with high credibility, MIDs for the SST (1.5–2.1) and overall pain VAS (1.4–1.6) were consistent across the two available estimates. The MIDs for the Constant score (3–16.6), DASH (4.4–25.4) and OSS (4.0–14.7) were, however, inconsistent among 6–10 estimates provided.

Available evidence permitted subgroup analyses exploring potential sources of heterogeneity only for surgical versus non-surgical interventions for the Constant score and SST and follow-up time (less than 3 months or ≥ 3 months) for the OSS. In no case did these differences explain the variation in the MID. Online supplementary appendix 2 provides details of the MID estimates and the results of subgroup analysis.

DISCUSSION

We identified 22 studies involving 5562 patients that reported 74 empirically estimated anchor-based MIDs for 10 candidate instruments to assess shoulder pain, function and HRQoL. The majority of studies used a global rating of change (transition rating) as the anchor and had a follow-up period of over 3 months. We identified MIDs of high credibility for pain and function outcomes and of low credibility for HRQoL. MIDs estimates often varied widely; we offered median estimates for the systematic review team and guideline panel. We also provided the systematic review team with the median, minimum and maximum values across the range of high credibility

trustworthy MID estimates generated from the eligible studies for the PROMs of interest. The only instance in which the variability in scores was sufficiently great that choice of one of the extremes rather than the median could substantially influence conclusions was for the Constant score.

Authors of the linked review used these MIDs (Pain VAS 0–10 1.5 units, the Constant score 0–100 scale 8.3 units and EQ-5D, 0.07 units) to gauge the importance of possible difference patients in Grading of Recommendations Assessment, Development and Evaluation (GRADE) evidence summaries and to dichotomise the improvements (proportions of patients achieving MID or more); the *BMJ* Rapid Recommendations guideline panel used them to inform their judgements of magnitude of effect in formulating their recommendations. The systematic review informed the *BMJ* Rapid Recommendations panel in their development of the guideline.

Strengths of our review include a comprehensive search for anchor-based MIDs for instruments commonly used in RCTs of shoulder conditions conducted without restrictions of study design or language of publication. We undertook judgements of MID credibility using a formal instrument with demonstrated reliability. Most studies (n=14) provided highly credible estimates. These MIDs not only can help clinicians, patients and clinical practice guideline developers interpret the magnitude of effects of interventions on PROMs, they also can be used in power calculations in future trials on shoulder conditions.

For the credibility assessment, we found that the anchor instrument directly addressed the patient's perspective, and judged the understanding the anchor instrument for patients as 'definitely yes' or 'to a great extent', for all the MID estimates. Approximately half of the estimates did not report the correlation between the anchor and the PROM. We judged the precision of the MID estimation and the threshold or difference between groups on the anchor used to estimate the MID as 'definitely no' or 'not so much' for most MID estimates.

The results of our systematic review have limitations. The range of reported MIDs was wide for some of the PROMs (eg, 0.3–30 for Constant score; 4.4–25.41 for DASH). Baseline characteristics (participants' disease/conditions, sample size, PROMs or instruments), anchors and analytic methods varied among included studies; though others have detected associations between methodological approaches and MIDs,⁵⁶ our attempts to establish a clear relation between these variables and the MIDs were not successful. For some instruments used in RCTs of surgery for shoulder pain—the SDQ, SF-36 and 15D—we did not find any study estimating MIDs in our target patient population. For others, MIDs for shoulder conditions closely related to subacromial syndrome, or for shoulder conditions at all, were not available, and we, therefore, relied on estimates from any upper extremity problem population. With respect to the assessment of credibility, a formal assessment of the validity of the instrument has not been undertaken. Moreover, one might

Table 4 Summary of MIDs for improvement for interested instruments according to the credibility

| Instrument/domain (score range) | No of estimates | Median estimate | Minimum estimate | Maximum estimate |
|--|-----------------|------------------|------------------|------------------|
| High credibility | | | | |
| Absolute MIDs | | | | |
| Constant score (0–100)* | 10 | 8.3 | 3 | 16.6 |
| SST (0–12) | 2 | 1.8 | 1.5 | 2.1 |
| Pain VAS (overall) (0–10) | 2 | 1.5 | 1.4 | 1.6 |
| Pain VAS (activity) (transfer to 0–10) | 1 | 2.1 | | |
| Pain VAS (at rest) (transfer to 0–10) | 1 | 3.0 | | |
| DASH (0–100) | 6 | 10.2 | 4.4 | 25.4 |
| OSS: (0–48)† | 8 | 5.3 | 4.0 | 14.7 |
| SF-12 (0–100) | 1 | Physical: 1 | | |
| | 1 | Mental: 4 | | |
| Relative MIDs (relative to baseline) | | | | |
| Constant score (0–100) | 1 | 15% | | |
| OSS (0–48) | 1 | 11% | | |
| Low credibility | | | | |
| Absolute MIDs | | | | |
| Constant score (0–100) | 9 | 19.0 | 0.3 | 36.0 |
| SST (0–100) | 6 | 2.1 | 1.4 | 2.9 |
| Pain VAS (overall) (0–10) | 5 | 1.4 | 0.5 | 2.7 |
| DASH (0–100) | 1 | 12.4 | | |
| PNRS (0–10) | 5 | 3.4 | 1.1 | 6.3 |
| Quick DASH (0–100) | 1 | 13.4 | | |
| Neer score (0–100) | 3 | 2.0 | 1.5 | 3.7 |
| EQ-5D-3L (–0.59 to 1) | 2; | Raw index: 0.07; | 0.02; | 0.11 |
| | 2 | VAS: 7.18 | 6.86 | 7.50 |
| SF-12 (0–100) | 2; | Physical: 2.2 | 2.0; | 2.4; |
| | 2 | Mental: 0.9 | 0.9 | 1.0 |
| Relative MIDs (relative to baseline) | | | | |
| Constant score (0–100) | 1 | 22% | | |
| Quick DASH (0–100) | 1 | 8% | | |

*The range of the Constant score is 2–100 in van de Water *et al.*⁴⁹

†The range of the OSS is 12–60 in Christie *et al.*³⁶

DASH, Disabilities of the Arm, Shoulder and Hand; EQ-5D-3L, Euro-Quality of life 5-dimension 3-level index; MID, minimal important difference; NR, not reported; OSS, Oxford Shoulder Score; PNRS, Pain Numerical Rating Scale; SF-12, Short Form Health Survey 12; SST, simple shoulder test; VAS, Visual Analogue Scale.

challenge our judgement in inferring high credibility if three or more criteria were met. Finally, investigators used different methods to relate the anchor to a transition rating; the optimal approach remains uncertain.^{56 57}

Our results are consistent with previous studies.^{23–25} A previous review of MIDs of upper extremity instruments that appeared in selected orthopaedic journals from 2014 to 2016 found a wide range of MIDs for the Constant score (8–36) and reported a pain VAS MID of 1.4 on 10-point scale.²⁶ Reviews of pain VAS MIDs in shoulder injuries found a range of 0.5–3.0.^{24 36 46 47} A review of pain ratings in a wide variety of conditions reported VAS MIDs

of 0.1–8.2 and noted that absolute MIDs are higher in patients with more pain at baselines.²⁷ Only one study included in our review reported MID estimates separately according to the baseline severity⁴⁸ but these estimates had low credibility due to problems in the anchor selected and failure to report the correlation between the anchor and the instrument. Two other reviews of shoulder instrument MIDs, primarily from rotator cuff injuries reported MID values of 10.2–20 for DASH, and 4.0–13.4 for OSS.^{23 25} Participants' disease/conditions, baseline scale score and inappropriate analytic methods can cause serious bias in determining MIDs^{56 58}; researchers should

pay more attention to these factors during the MID estimation studies.

CONCLUSION

Our review provides anchor-based MID estimates, as well as a rating of their credibility, for PROMs for measurement instruments addressing patients with shoulder conditions. The review identified methodological limitations of the primary studies, future studies should strive for high precision of MID estimation, seek to identify difference between groups and reasons for those differences and report correlations between the anchor and the PROM.^{56 58}

The MID estimates inform the interpretation for a linked systematic review and guideline on arthroscopy for shoulder pain. Researchers addressing a wide variety of shoulder conditions can in future make use of our summary MIDs to inform sample size and aid in interpretation of results.

Author affiliations

¹The Center of Gerontology and Geriatrics/ National Clinical Research Center for Geriatrics, West China Hospital, Sichuan University, Chengdu, China

²Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada

³Department of Medicine, Innlandet Hospital Trust-division, Gjøvik, Norway

⁴Institute of Health and Society, Faculty of Medicine, University of Oslo, Oslo, Norway

⁵Finnish Center of Evidence Based Orthopaedics (FICEBO), University of Helsinki, Helsinki, Finland

⁶Department of Orthopaedics and Traumatology, HUS Helsinki University Hospital, Helsinki, Finland

⁷Division of General Internal Medicine and Division of Epidemiology, University Hospitals of Geneva, Geneva, Switzerland

⁸Department of Medicine, McMaster University, Hamilton, Ontario, Canada

Acknowledgements The authors thank Rachel Couban, librarian at McMaster university for testing the search strategy and members of the Rapid Recommendations panel (especially, Clare Ardern, physiotherapist, Teemu Karjalainen, orthopedic surgeon, Lyubov Lytvyn, patient partnership liaison and Rudolf Poolman, orthopedic surgeon) for critical feedback on the inclusion of relevant PROMs and for their review of this manuscript. The abridged version of the MID credibility tool used in this study is derived from a MID credibility tool that was made available to the authors under license from McMaster University.

Contributors GG and RACS conceived the study idea; QH designed the search strategy; QH, YW and DZ screened studies for eligibility; QH, TD, YW, DZ, RACS and AQ extracted data and assessed the credibility; QH wrote the first draft of the manuscript; GG, TD, POV, TL, TA, ACL and RACS interpreted the data analysis and critically revised the manuscript. QH is the guarantor.

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

Competing interests None declared.

Patient consent for publication Not required.

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, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- Codsi M, Howe CR. Shoulder conditions: diagnosis and treatment guideline. *Phys Med Rehabil Clin N Am* 2015;26:467–89.
- Walker-Bone KE, Palmer KT, Reading I, et al. Soft-tissue rheumatic disorders of the neck and upper limb: prevalence and risk factors. *Semin Arthritis Rheum* 2003;33:185–203.
- Aurora A, McCarron J, Iannotti JP, et al. Commercially available extracellular matrix materials for rotator cuff repairs: state of the art and future trends. *J Shoulder Elbow Surg* 2007;16:S171–8.
- Campbell M. Problems with large joints: shoulder conditions. *FP Essent* 2016;446:25–30.
- Mitchell C, Adebajo A, Hay E, et al. Shoulder pain: diagnosis and management in primary care. *BMJ* 2005;331:1124–8.
- Dierckx R, Bron C, Dorrestijn O, et al. Guideline for diagnosis and treatment of subacromial pain syndrome: a multidisciplinary review by the dutch orthopaedic association. *Acta Orthop* 2014;85:314–22.
- Anker SD, Agewall S, Borggreve M, et al. The importance of patient-reported outcomes: a call for their comprehensive integration in cardiovascular clinical trials. *Eur Heart J* 2014;35:2001–9.
- Payne C, Michener LA. Physiotherapists use of and perspectives on the importance of patient-reported outcome measures for shoulder dysfunction. *Shoulder Elbow* 2014;6:204–14.
- Tibaek S, Gadsboell J. Scapula alata: description of a physical therapy program and its effectiveness measured by a shoulder-specific quality-of-life measurement. *J Shoulder Elbow Surg* 2015;24:482–90.
- Paavola M, Malmivaara A, Taimela S, et al. Subacromial decompression versus diagnostic arthroscopy for shoulder impingement: randomised, placebo surgery controlled clinical trial. *BMJ* 2018;362:k2860.
- Rueda Garrido JC, Vas J, Lopez DR. Acupuncture treatment of shoulder impingement syndrome: a randomized controlled trial. *Complement Ther Med* 2016;25:92–7.
- Galace de Freitas D, Marcondes FB, Monteiro RL, et al. Pulsed electromagnetic field and exercises in patients with shoulder impingement syndrome: a randomized, double-blind, placebo-controlled clinical trial. *Arch Phys Med Rehabil* 2014;95:345–52.
- Moezy A, Sepehrifar S, Solaymani Dodaran M. The effects of scapular stabilization based exercise therapy on pain, posture, flexibility and shoulder mobility in patients with shoulder impingement syndrome: a controlled randomized clinical trial. *Med J Islam Repub Iran* 2014;28:87.
- Kinsella R, Cowan SM, Watson L, et al. A comparison of isometric, isotonic concentric and isotonic eccentric exercises in the physiotherapy management of subacromial pain syndrome/rotator cuff tendinopathy: study protocol for a pilot randomised controlled trial. *Pilot Feasibility Stud* 2017;3:45.
- Guyatt GH, Juniper EF, Walter SD, et al. Interpreting treatment effects in randomised trials. *BMJ* 1998;316:690–3.
- Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407–15.
- Schünemann HJ, Guyatt GH. Commentary-goodbye M(C)ID! Hello MID, where do you come from? *Health Serv Res* 2005;40:593–7.
- Brozek JL, Guyatt GH, Schünemann HJ. How a well-grounded minimal important difference can enhance transparency of labelling claims and improve interpretation of a patient reported outcome measure. *Health Qual Life Outcomes* 2006;4:69.
- Guyatt G, Schunemann H. How can quality of life researchers make their work more useful to health workers and their patients? *Qual Life Res* 2007;16:1097–105.
- King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res* 2011;11:171–84.
- McGlothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients. *JAMA* 2014;312:1342–3.
- Guyatt GH, Osoba D, Wu AW, et al. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 2002;77:371–83.
- Roy JS, MacDermid JC, Woodhouse LJ. Measuring shoulder function: a systematic review of four questionnaires. *Arthritis Rheum* 2009;61:623–32.
- Tashjian RZ, Deloach J, Porucznik CA, et al. Minimal clinically important differences (MCID) and patient acceptable symptomatic state (PASS) for visual analog scales (VAS) measuring pain in patients treated for rotator cuff disease. *J Shoulder Elbow Surg* 2009;18:927–32.
- St-Pierre C, Desmeules F, Dionne CE, et al. Psychometric properties of self-reported questionnaires for the evaluation of symptoms and functional limitations in individuals with rotator cuff disorders: a systematic review. *Disabil Rehabil* 2016;38:103–22.

26. Copay AG, Chung AS, Eyberg B, *et al.* Minimum clinically important difference: current trends in the orthopaedic literature, Part I: upper extremity: a systematic review. *JBJS Rev* 2018;6:e1.
27. Olsen MF, Bjerre E, Hansen MD, *et al.* Minimum clinically important differences in chronic pain vary considerably by baseline pain and methodological factors: systematic review of empirical studies. *J Clin Epidemiol* 2018;101:87–106.
28. Siemieniuk RA, Agoritsas T, Macdonald H, *et al.* Introduction to BMJ rapid recommendations. *BMJ* 2016;354:i5191.
29. 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.
30. Siemieniuk RAC, Harris IA, Agoritsas T, *et al.* Arthroscopic surgery for degenerative knee arthritis and meniscal tears: a clinical practice guideline. *BMJ* 2017;357:j1982.
31. Buchbinder R, Page MJ, Huang H, *et al.* A preliminary core domain set for clinical trials of shoulder disorders: a report from the OMERACT 2016 shoulder core outcome set special interest group. *J Rheumatol* 2017;44:1880–3.
32. Johnston BC, Ebrahim S, Carrasco-Labra A, *et al.* Minimally important difference estimates and methods: a protocol. *BMJ Open* 2015;5:e007953.
33. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. *Clin Orthop Relat Res* 1987;160–4.
34. Constant CR, Gerber C, Emery RJ, *et al.* A review of the constant score: modifications and guidelines for its use. *J Shoulder Elbow Surg* 2008;17:355–61.
35. Christiansen DH, Frost P, Falla D, *et al.* Responsiveness and minimal clinically important change: a comparison between 2 shoulder outcome measures. *J Orthop Sports Phys Ther* 2015;45:620–5.
36. Christie A, Dagfinrud H, Garratt AM, *et al.* Identification of shoulder-specific patient acceptable symptom state in patients with rheumatic diseases undergoing shoulder surgery. *J Hand Ther* 2011;24:53–61.
37. Dritsaki M, Petrou S, Williams M, *et al.* An empirical evaluation of the SF-12, SF-6D, EQ-5D and michigan hand outcome questionnaire in patients with rheumatoid arthritis of the hand. *Health Qual Life Outcomes* 2017;15:20.
38. Ekeberg OM, Bautz-Holter E, Keller A, *et al.* 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.
39. Holmgren T, Oberg B, Adolfsson L, *et al.* Minimal important changes in the Constant-Murley score in patients with subacromial pain. *J Shoulder Elbow Surg* 2014;23:1083–90.
40. Kukkonen J, Kauko T, Vahlberg T, *et al.* Investigating minimal clinically important difference for constant score in patients undergoing rotator cuff surgery. *J Shoulder Elbow Surg* 2013;22:1650–5.
41. Mahabier KC, Den Hartog D, Theyskens N, *et al.* Reliability, validity, responsiveness, and minimal important change of the disabilities of the arm, shoulder and hand and constant-murley scores in patients with a humeral shaft fracture. *J Shoulder Elbow Surg* 2017;26:e1–12.
42. Michener LA, Snyder AR, Leggin BG. Responsiveness of the numeric pain rating scale in patients with shoulder pain and the effect of surgical status. *J Sport Rehabil* 2011;20:115–28.
43. Mintken PE, Glynn P, Cleland JA. Psychometric properties of the shortened disabilities of the Arm, Shoulder, and Hand Questionnaire (QuickDASH) and numeric pain rating scale in patients with shoulder pain. *J Shoulder Elbow Surg* 2009;18:920–6.
44. Negahban H, Behtash Z, Sohani SM, *et al.* Responsiveness of two Persian-versions of shoulder outcome measures following physiotherapy intervention in patients with shoulder disorders. *Disabil Rehabil* 2015;37:2300–4.
45. Rysstad T, Røe Y, Haldorsen B, *et al.* Responsiveness and minimal important change of the Norwegian version of the Disabilities of the Arm, Shoulder and Hand questionnaire (DASH) in patients with subacromial pain syndrome. *BMC Musculoskelet Disord* 2017;18:248.
46. Simovitch R, Flurin PH, Wright T, *et al.* Quantifying success after total shoulder arthroplasty: the minimal clinically important difference. *J Shoulder Elbow Surg* 2018;27:298–305.
47. Tashjian RZ, Hung M, Keener JD, *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.
48. Tubach F, Dougados M, Falissard B, *et al.* Feeling good rather than feeling better matters more to patients. *Arthritis Rheum* 2006;55:526–30.
49. van de Water AT, Shields N, Davidson M, *et al.* Reliability and validity of shoulder function outcome measures in people with a proximal humeral fracture. *Disabil Rehabil* 2014;36:1072–9.
50. Lundquist CB, Døssing K, Christiansen DH. Responsiveness of a Danish version of the Disabilities of the Arm, Shoulder and Hand (DASH) questionnaire. *Dan Med J* 2014;61:A4813.
51. Marks M, Audigé L, Herren DB, *et al.* Measurement properties of the german michigan hand outcomes questionnaire in patients with trapeziometacarpal osteoarthritis. *Arthritis Care Res* 2014;66:245–52.
52. Schmitt JS, Di Fabio RP. Reliable change and minimum important difference (MID) proportions facilitated group responsiveness comparisons using individual threshold criteria. *J Clin Epidemiol* 2004;57:1008–18.
53. Tashjian RZ, Deloach J, Green A, *et al.* 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.
54. van Kampen DA, Willems WJ, van Beers LW, *et al.* 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.
55. Castricini R, Longo UG, De Benedetto M, *et al.* Arthroscopic-assisted latissimus dorsi transfer for the management of irreparable rotator cuff tears: short-term results. *J Bone Joint Surg Am* 2014;96:e119.
56. Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *J Clin Epidemiol* 2017;82:128–36.
57. 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.
58. Angst F, Benz T, Lehmann S, *et al.* Multidimensional minimal clinically important differences in knee osteoarthritis after comprehensive rehabilitation: a prospective evaluation from the bad zurzach osteoarthritis study. *RMD Open* 2018;4:e000685.