

Minimal Clinically Important Difference as Applied in Rheumatology: An OMERACT Rasch Working Group Systematic Review and Critique

Beyza Doganay Erdogan, Ying Ying Leung, Christoph Pohl, Alan Tennant, and Philip G. Conaghan

ABSTRACT. *Objective.* We aimed to evaluate how minimal (clinically) important differences (MCID/MID) were calculated in rheumatology in the past 2 decades and demonstrate how the calculation is compromised by the lack of interval scaling.

Methods. We conducted a systematic literature review on articles reporting MCID calculation in osteoarthritis (OA) and rheumatoid arthritis (RA) from January 1, 1989, to May 9, 2014. We evaluated the methods of MCID calculation and recorded the ranges of MCID for common patient-reported outcome measures (PROM). Taking data from the Health Assessment Questionnaire (HAQ), we showed the effects of performing mathematical calculations on ordinal data.

Results. A total of 330 abstracts were reviewed and 123 articles chosen for full text review. Thirty-six (19 OA, 16 RA and 1 OA-RA) articles were included in the final evaluation. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) was the most frequently reported PROM with relevant calculations in OA, and the HAQ in RA. Sixteen articles used anchor-based methods alone for calculation of MCID, and 1 article used distribution-based methods alone. Nineteen articles used both anchor and distribution-based methods. Only 1 article calculated MCID using an interval scale. Wide ranges in MCID for the WOMAC in OA and HAQ in RA were noted. Ordinal-based derivations of MCID are shown to understate true change at the margins, and overstate change in the mid-range of a scale.

Conclusion. The anchor-based method is commonly used in the calculation of MCID. However, the lack of interval scaling is shown to compromise validity of MCID calculation. (First Release June 1 2015; J Rheumatol 2016;43:194–202; doi:10.3899/jrheum.141150)

Key Indexing Terms:

MINIMAL CLINICALLY IMPORTANT DIFFERENCE ANCHOR-BASED APPROACHES
DISTRIBUTION-BASED APPROACHES PATIENT-REPORTED OUTCOME MEASURES
RHEUMATOID ARTHRITIS OSTEOARTHRITIS

Over the last 25 years the concept of a minimal clinically important difference (MCID) has emerged in the outcomes

From the Ankara University Faculty of Medicine, Department of Biostatistics, Ankara, Turkey; Department of Rheumatology and Immunology, Singapore General Hospital, Singapore; Duke-NUS Graduate Medical School, Singapore; Department of Internal Medicine II, Rheumatology, Clinical Immunology, Osteology, Physical Therapy, and Sports Medicine, Schlosspark-Klinik, Teaching Hospital of the Charité, University Medicine Berlin, Berlin, Germany; Swiss Paraplegic Research, Nottwil, Switzerland; Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds and UK National Institute for Health Research Leeds Musculoskeletal Biomedical Research Unit, Leeds, UK. PGC is funded in part through a grant from Arthritis Research UK.

B. Doganay Erdogan, PhD, Assistant Professor, Ankara University Faculty of Medicine, Department of Biostatistics; Y.Y. Leung, MD, Assistant Professor, Duke-NUS Graduate Medical School, Singapore; C. Pohl, MD, Department of Internal Medicine II, Rheumatology, Clinical Immunology, Osteology, Physical Therapy and Sports Medicine, Schlosspark-Klinik, Teaching Hospital of the Charité, University Medicine Berlin; A. Tennant, PhD, MD, Scientific Adviser, Swiss Paraplegic Research; P.G. Conaghan, MBBS, PhD, FRACP, FRCP, Professor of Musculoskeletal Medicine, Leeds Institute of Rheumatic and Musculoskeletal Medicine.

Address correspondence to B. Doganay Erdogan, Ankara University Faculty of Medicine, Department of Biostatistics, 06100, Sıhhiye, Ankara, Turkey. E-mail: beyzadoganay@gmail.com

literature. A clinically important difference is defined as a change or difference in the outcome measure that would be perceived as important and beneficial by the clinician or the patient, assuming the absence of serious adverse effects and excessive costs^{1,2,3}. A MCID is therefore a threshold value for such change. A number of terms have emerged in the literature that differ slightly in definition and may be confusing. The most common are the minimally important difference (MID), MCID, minimal clinically significant difference (MCSD), and minimal clinically important improvement (MCII). A review by King details the definitions and methods for determining the MCID⁴.

In determining MCID, both distribution-based and anchor-based approaches have been described. Distribution-based or data driven approaches depend on the statistical characteristics of the data. From a statistical perspective, a significant difference means the difference that is unlikely to occur by chance, and is a decision based on probabilistic calculations. Such probabilistic calculations are often affected by sample size; thus a small difference may be regarded as significant owing to large sample size, but at the same time

may mean little to the patients or their clinician. Thus clinicians needed to ascertain the importance of statistically significant results for their own patients. A range of indicators such as the standard error of measurement (SEM), minimum detectable change (MDC), effect size (ES), standardized response mean (SRM), or Guyatt's responsiveness index (GRI) may be used to define variability in the data. In contrast, anchor-based approaches link the change in the outcome measure to a meaningful external anchor that accounts for the patient's perspective. For example, patients rate themselves according to their last condition state from "much worse," "slightly worse," "same," "slightly better" to "much better." The MCID for improvement can be defined as the mean difference of the outcome measures of patients who rated themselves "same" and "slightly better." The range of categorical rating scale can be varied from study to study and the selection of the score group(s) to calculate MCID is arbitrary. In Outcome Measures in Rheumatology (OMERACT) meetings 5 to 7, the anchor-based method was recommended as the method of choice^{1,5,6}; while in OMERACT 8, reporting the proportion of patients achieving anchor-based acceptable status was recommended as important and complementary information in clinical trials⁷. The US Federal Drug Administration (FDA) guidelines for patient-reported outcome measures (PROM) affirmed that anchor-based methodology was required in reporting the proportion responding to treatment in the evaluation of all medical devices and drugs⁸. While there may be some debate over appropriateness of the external anchor, in terms of wording, category levels, and time frame, there is a weakness that is more fundamental. This is connected to the nonlinearity of estimates obtained from PROM. PROM are ordinal scales, in which the difference between 2 levels of a response cannot be assumed to be the same as the difference between 2 other levels. However, the calculations of MCID rely on data that meet criteria for interval scaling⁹. To inform the debate about the validity of MCID/MID on PROM, we conducted a systematic review on how MCID has been calculated in 2 common rheumatologic conditions, osteoarthritis (OA) and rheumatoid arthritis (RA), in the past 2 decades, in particular on calculation methods and scaling. In addition, we show how the approach lacks validity when used on ordinal data.

MATERIALS AND METHODS

Search strategies. A systematic review of the SCOPUS and MEDLINE databases from January 1, 1989, to May 9, 2014 was performed to identify English-language original research reports related to all the MCID-related terms: ("minimally clinically important difference" OR "minimal clinically important difference" OR "minimum clinically important difference" OR "minimally clinical important difference" OR "minimum clinical important difference" OR "minimally important difference" OR "minimal important difference" OR "minimum important difference" OR "minimal clinically important improvement" OR "subjectively significant difference" OR "clinically important difference" OR "clinically significant change" OR "minimally important change" OR "minimal clinically important improvement") AND ("Osteoarthritis" OR

"Rheumatoid Arthritis") in the title or abstract. The references of all retrieved articles were also screened for potentially relevant publications.

Selection of articles. Two reviewers (BDE, YYL, or CP; 2 working on each article) independently assessed inclusion or exclusion of articles, with disputes resolved by another reviewer (AT). We included publications that calculated MCID in the evaluation of any PROM in patients with OA or RA. We excluded articles that used published results of MCID without calculation of MCID and articles that did not report MCID calculation on OA or RA separately.

Data collection. One reviewer (BDE, YYL, or CP) extracted MCID data using a standardized checklist, and results were double-checked by another reviewer. Information on whether the MCID calculated was based on distribution or anchor-based methods, anchor characteristics, and whether the calculation was based on ordinal or interval scales was gathered. MCID values of commonly used PROM were pooled. From the selected articles, we extracted anchor-based MCID improvement and deterioration along with distribution-based indicators including ES, SRM, SEM, MDC 90%, MDC 95%, and GRI.

RESULTS

A total of 330 abstracts were identified from the literature search and 123 articles were chosen for full text review (Figure 1). We excluded 188 articles in the abstract review that were commentaries, reviews, metaanalyses or systematic reviews, articles that included patients other than those with OA or RA, articles on outcome measures other than PROM, articles on single-item outcome measures, and articles that used only published MCID for PROM without calculation of MCID. Thirty-six (19 OA, 16 RA, 1 RA-OA) articles were included in the final evaluation^{10-15,16-20,21-25,26-30,31-36,37-40,41-46}.

The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and the generic Medical Outcomes Study Short Form-36 (SF-36) were the most frequently reported PROM with relevant calculations in OA. The Health Assessment Questionnaire (HAQ), SF-36, and Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F) scales were the most frequent reported PROM with relevant calculations in RA (Table 1). In determining the MCID values, anchor-based approaches (44.4%) were used frequently; 52.8% of articles used more than 1 approach (Table 1). There was only 1 article that used only distributional methods in calculation of MCID¹⁰.

There was heterogeneity in the anchor used in different studies. The majority of articles (n = 27, 75%) used a question asking patients' health status compared with an earlier timepoint. Two articles used a patient conversation method, in which patients were required to rate their health status in relation to another patient. There was diversity in the category level and time frame (period of time to compare current health status with the prior health status), and 5 articles had multiple time frames (Table 1). There were 8 articles (22.2%) that used anchors that were part of the clinical assessment and not patient-derived, such as Simplified Disease Activity Index (SDAI), physician's global assessment, pain visual analog scale (VAS), swollen/tender joint counts, and the 28-joint Disease Activity Score.

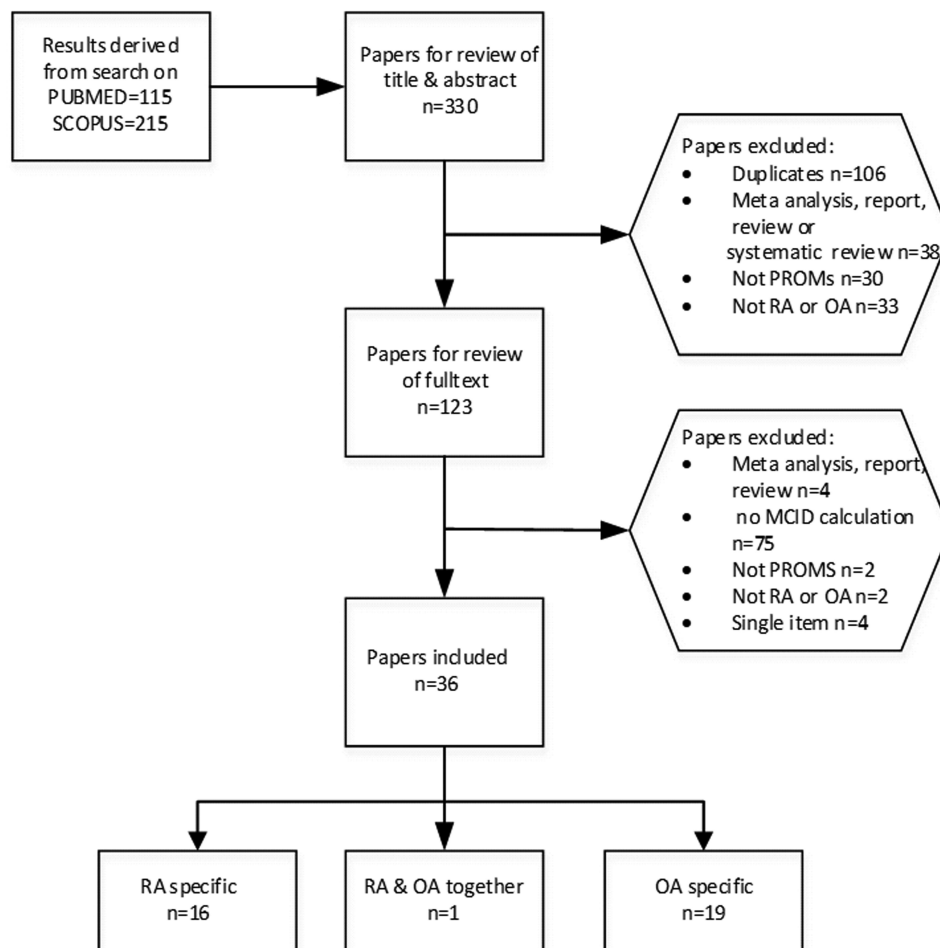


Figure 1. Selection of articles. PROM: patient-reported outcome measures; RA: rheumatoid arthritis; OA: osteoarthritis; MCID: minimal clinically important differences.

The range of MCID for WOMAC and SF-36 for knee or hip OA are shown in Appendix 1, and the range of MCID for HAQ and FACIT-F for RA are shown in Appendix 2. Data are presented separately for anchor or distribution-based methods. Regardless of calculation methods, there was a wide range of MCID values.

The majority of articles calculated MCID for PROM using the raw scores, which were all ordinal scales. There was 1 exception that used the Rasch model to convert the raw scores of ABILHAND (a Rasch-built measure of manual ability of the hand) to interval measures¹⁰.

Calculating MCID on ordinal data. Table 2 illustrates the actual problem when looking at the HAQ. This is typically scored as 8 items with a 0–3 range, giving a raw score of 0–24, usually divided by 8 to give a 0–3 range. Each increment of raw score gives a 0.125 increase in score on the 0–3 range (columns 1–3). Columns 4–6 provide the information of the metric equivalent of HAQ, which is derived from fit of HAQ data to the Rasch model (mean item residual 0.0758; SD 1.1822; person residual –0.287; SD 1.0693;

chi-square 32.3¹⁶; $p = 0.120$; alpha 0.90). Location (column 4) is the estimation of the probability of a person to achieve a change in the corresponding raw scores. This “location” is then linear transformed into Metric-8 to mimic the 0–3 scale from HAQ-8. Metric-8 is the equivalent of the HAQ represented on a metric or interval scale. The first point to note is that the increment of Metric-8 in column 6 does not match the HAQ-8 (ordinal) increment shown in column 3. In practice, each distance from one raw point to the next has a different magnitude of increment for the Metric-8. The magnitude of difference is most at the margins, and least in the center of the scale.

Consider the implications of this for an example of a 0.5-point change on the classical HAQ-8, where a shift of 4 raw score points is required. Columns 7 and 8 show the results for the change in raw score of HAQ-8 needed to achieve the 0.5-point change in HAQ, for improvement (column 7) and deterioration (column 8), respectively. On the classical HAQ-8 such an improvement will be obtained by a 4-point raw score (0.5) shift in either direction. On the Metric-8 (when units are

Table 1. Methods of calculation, anchor characteristics, and scaling of MCID calculation in included articles (n = 36).

	Total, n = 36	RA, n = 17	OA Knee and Hip, n = 4	OA Knee, n = 12	OA Hip, n = 5	OA Hand, n = 1
Calculation method						
Distribution only	1 (2.8%)	1 (6%)				
Anchor only	16 (44.4%)	7 (41%)	3 (75%)	3 (25%)	3 (60%)	1 (100%)
Distribution and anchor	19 (52.8%)	9 (53%)	1 (25%)	9 (75%)	2 (40%)	
Articles with anchor						
External patient anchor	27 (75%)	8 (47%) [‡]	4 (100%)	12 (100%)	5 (100%)	1 (100%)
Other internal anchors (non-patient derived)	8 (22.2%)	8 (47%)				
External patient anchor properties						
No. of level of response (range)	3 to 15	3 to 7	5 to 15	3 to 15	3 to 5	8
Time frame of anchor (range)	2 wks to 5 yrs	2 wks to 1 yr	2 wks to 3 mos	4 wks to 5 yrs	4 wks to 1 yr	2 wks to 4 wks
Multiple recall points	5 (13.9%)	1 (6%)	—	3 (25%)	—	1 (100%)
Scaling						
Ordinal	35 (97.2%)	16 (94%)	4 (100%)	12 (100%)	5 (100%)	1 (100%)
Interval	1 (2.8%)	1 (6%)	—	—	—	—
PROM						
WOMAC	11 (31%)		2 (50%)	7 (58%)	4 (80%)	—
SF36	10 (28%)	5 (29%)	1 (25%)	4 (33%)	1 (20%)	—
FACIT	3 (8%)	3 (18%)	—	—	—	—
HAQ	8 (22%)	8 (47%)	—	—	—	—
Others	20 (56%)	12 (71%)	2 (50%)	7 (58%)	2 (40%)	1 (100%)

[‡] 2 of 8 were group conversations. n: number of articles; PROM: patient-reported outcome measures; WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index; SF-36: Medical Outcomes Study Short Form-36; FACIT-F: Functional Assessment of Chronic Illness Therapy–Fatigue; HAQ: Health Assessment Questionnaire.

of equal size), the 0.5-point change can be obtained by just a 2- to 3-point raw score improvement at the margins, whereas it requires 6 raw score points in the center of the scale. The reason there are differences for the 0.5-point improvement and deterioration is explained in the item threshold distribution histogram as shown in Figure 2. The items of HAQ-8 are now calibrated on a metric scale. An item's location in logits is the relative difficulty respondents describe regarding that item on a scale. The score points can be considered as a ruler of physical function, from most disabled at the right end to least disabled at the left end. When moving from right to left on the metric ruler, raw score points are lost in a nonlinear fashion (i.e., the patient is improving). For example, a patient starting at 4 logits would quickly lose 3 score points as he improves (1 logit down to 3 logits); but when he continues to improve, he needs to move 2 more logits before he loses another 3 raw score points (with the caveat that this is set within a probabilistic framework). This is typical of an ordinal scale where the distances between each raw score point are not equal. Further, when a patient is becoming more disabled (moving from left to right on the ruler), score points are picked up in a different fashion, thus explaining the differences in improvement versus deterioration on the ordinal scale. This also makes a difference for MCID calculated by distribution-based approaches, where the same unit and associated measurement error is assumed for all parts of the instrument.

DISCUSSION

At OMERACT 11, a special interest group (SIG) on Rasch model analysis was formed and reported on the acceptance and increasing use of Rasch model analysis in evaluation and development of PROM⁴⁷. However, it also highlighted the lack of availability and reporting of transformation tables, which limited the application of transformed interval measurement in clinical practice. Meaningful measurement is based on the arithmetical property of interval scales⁴⁸, and this applies particularly to the responsiveness of PROM and “discrimination” in the OMERACT filter⁴⁹. During OMERACT 12 (May 2014), we evaluated the application of interval scaling in the domain of responsiveness and MCID calculation. We noted a significant deficiency in the use of interval scaling, which resulted in potential misinference of reported MCID. During the SIG in OMERACT 12, there was unanimous agreement from 39 participants that it was crucial to promote the use of the Rasch interval scale in measurement. Thus an international collaboration has been established to provide data for the establishment of Rasch transformed scales for commonly used PROM in rheumatology, which should be made available in the public domain.

The use of MCID or similar calculations in musculoskeletal disorders is growing rapidly. As shown in our systematic review for PROM in OA and RA, the majority of cases have failed to recognize that ordinal scales do not support the mathematical calculations required for MCID

Table 2. Effects of ordinal scores on calculations of MCID.

1 Raw Score	2 HAQ-8	3 Increment on HAQ-8 (ordinal scale)	4 Location	5 Metric-8	6 Increment on Metric-8 (metric scale)	7 ↑	8 ↓
0	0.000		-4.777	0.000			2
1	0.125	0.125	-3.911	0.260	0.260	1	4
2	0.250	0.125	-3.280	0.449	0.189	2	5
3	0.375	0.125	-2.821	0.587	0.138	3	5
4	0.500	0.125	-2.442	0.701	0.114	4	6
5	0.625	0.125	-2.108	0.801	0.100	4	6
6	0.750	0.125	-1.799	0.894	0.093	5	6
7	0.875	0.125	-1.505	0.982	0.088	5	6
8	1.000	0.125	-1.220	1.068	0.086	6	6
9	1.125	0.125	-0.939	1.152	0.084	6	6
10	1.250	0.125	-0.662	1.236	0.083	6	6
11	1.375	0.125	-0.386	1.318	0.083	6	6
12	1.500	0.125	-0.111	1.401	0.083	6	6
13	1.625	0.125	0.166	1.484	0.083	6	5
14	1.750	0.125	0.447	1.569	0.084	6	5
15	1.875	0.125	0.738	1.656	0.087	6	5
16	2.000	0.125	1.043	1.747	0.092	6	4
17	2.125	0.125	1.369	1.845	0.098	6	4
18	2.250	0.125	1.721	1.951	0.106	5	4
19	2.375	0.125	2.105	2.066	0.115	5	3
20	2.500	0.125	2.526	2.193	0.126	4	3
21	2.625	0.125	2.999	2.335	0.142	4	2
22	2.750	0.125	3.553	2.501	0.166	3	2
23	2.875	0.125	4.276	2.718	0.217	3	1
24	3.000	0.125	5.216	3.000	0.282	2	

Columns 7 and 8 show the change in raw score of HAQ-8 needed to achieve a 0.5-point change in HAQ-8, for improvement (↑) and deterioration (↓), respectively. HAQ: Health Assessment Questionnaire; location: location on the metric scale (in logits); Metric-8: HAQ-8 metric equivalent; MCID: minimal clinically important differences.

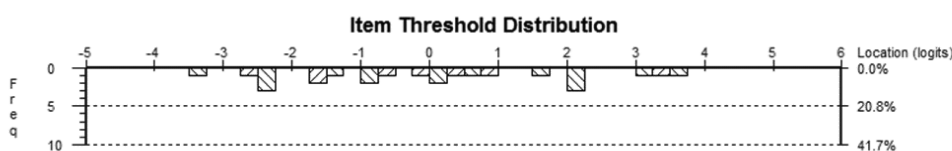


Figure 2. Item threshold distribution histogram showing ordinal points of the Health Assessment Questionnaire on a metric scale.

calculation. Only in 1 case, with the ABILHAND, was the MCID calculated using an appropriate metric⁴⁹. This situation is not unique to rheumatology, and only rarely can a metric-based MCID be found elsewhere⁵⁰. In practice, it generally means that where an original MCID calculation has been undertaken on a raw ordinal score, then subsequently patients will be less likely to achieve an MCID when they are moving across the outer 50% of the scale range, and more likely to achieve an MCID across the middle range of the scale. Thus, as we have demonstrated using the HAQ, at the margins they will improve a great deal for a 4-point change, and much less so as they pass across the middle of the score range.

The problems associated with MCID/MID calculations have been noted for some time. There is the debate over the appropriate methodology to ascertain patient perceived change^{7,51}. It has been recognized that MCID/MID appear to differ according to the initial health status of study subjects^{52,53}, and are different for improvement and deterioration. Both issues can be solved when MCID/MID are calculated on the metric or interval scale. Rasch analysis contributes to a valid calculation of MCID/MID by providing an interval scale metric for this purpose, and revealing how the ordinal-based calculations are invalid and biased. On the other hand, there are issues, particularly concerning the

external anchor used in the calculation of MCID on metric scales, that cannot be resolved, including the difference in MCID for improvement/deterioration as determined by the anchor method; this is because patients' perception of "important" changes for improvement or deterioration may be different. However, the question as to whether the patient weights his/her judgment differentially is a separate matter. Because the Rasch metric shows a difference in the unit change (in relation to the raw score) depending on whether the patient is improving or deteriorating, does it matter if the subjective anchor for that change is also different? We must accept the subjective anchor, and link the Rasch metric accordingly. From a methodological standpoint, the issue is whether the subjective judgment of change (i.e., the anchor) is consistent within and between patients and across samples. Given this, the rheumatology community should now state that MCID and their equivalent should always be calculated on a metric transformation. This may also help to facilitate development of a gold standard for calculating MCID/MID, which has yet to be established and has resulted in problems of interpretation due to varied results.

Our study has some limitations. The systematic review was constrained to English language publications, and information presented in other languages may have been missed. We found large ranges of MCID values calculated for either OA or RA for commonly used instruments, regardless of calculation method. However, we cannot differentiate how much is attributed by difference in anchor or anchor category levels, and how much is scaling. Because the terms used to indicate MCID have been numerous, evolving, and confusing, we may have missed some terms. But we think we have included most terms that were commonly used. We used only 1 scale (HAQ) to demonstrate the lack of interval scaling and its effect upon calculating the MCID. However, the nonlinearity of ordinal raw scores in PROM is now well established in the literature, including other widely used scales such as the WOMAC⁵⁴.

The existing calculations of MCID/MID based on an ordinal scale are invalid. Making MCID/MID dependent on the initial health status of study subjects and direction of change leads to potential mis-inference of it. This has most likely contributed to the wide range of reported MCID/MID in commonly used PROM. The availability of interval scaling based on Rasch model-transformed ordinal scales in the calculation of MCID/MID is urgently required.

REFERENCES

- Wells G, Beaton D, Shea B, Boers M, Simon L, Strand V, et al. Minimal clinically important differences: review of methods. *J Rheumatol* 2001;28:406-12.
- Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J* 2007;75:541-6.
- Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407-15.
- King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res* 2011;11:171-84.
- Wells G, Anderson J, Beaton D, Bellamy N, Boers M, Bombardier C, et al. Minimal clinically important difference module: summary, recommendations, and research agenda. *J Rheumatol* 2001;28:452-4.
- Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003;56:395-407.
- Tubach F, Ravaud P, Beaton D, Boers M, Bombardier C, Felson DT, et al. Minimal clinically important improvement and patient acceptable symptom state for subjective outcome measures in rheumatic disorders. *J Rheumatol* 2007;34:1188-93.
- U.S. Department of Health and Human Services Food and Drug Administration, Center of Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH). Clinical/Medical. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. [Internet. Accessed April 14, 2015.] Available from: www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf
- Merbitz C, Morris J, Grip JC. Ordinal scales and foundations of misinference. *Arch Phys Med Rehabil* 1989;70:308-12.
- Marra CA, Woolcott JC, Kopec JA, Shonjania K, Offer R, Brazier JE, et al. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. *Soc Sci Med* 2005;60:1571-82.
- Batcho CS, Durez P, Thonnard JL. Responsiveness of the ABILHAND questionnaire in measuring changes in rheumatoid arthritis patients. *Arthritis Care Res* 2011;63:135-41.
- Kersten P, White PJ, Tennant A. The visual analogue WOMAC 3.0 scale – internal validity and responsiveness of the VAS version. *BMC Musculoskelet Disord* 2010;11:80.
- Wells GA, Tugwell P, Kraag GR, Baker PR, Groh J, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. *J Rheumatol* 1993;20:557-60.
- Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE Jr. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. *Arthritis Rheum* 2000;43:1478-87.
- 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.
- de Boer YA, Hazes JM, Winia PC, Brand R, Rozing PM. Comparative responsiveness of four elbow scoring instruments in patients with rheumatoid arthritis. *J Rheumatol* 2001;28:2616-23.
- Angst F, Aeschlimann A, Michel BA, Stucki G. Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. *J Rheumatol* 2002;29:131-8.
- Walters SJ, Brazier JE. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. *Health Qual Life Outcomes* 2003;1:4.
- Cella D, Yount S, Sorensen M, Chartash E, Sengupta N, Grober J. Validation of the Functional Assessment of Chronic Illness Therapy Fatigue Scale relative to other instrumentation in patients with rheumatoid arthritis. *J Rheumatol* 2005;32:811-9.
- Quintana JM, Escobar A, Bilbao A, Arostegui I, Lafuente I,

- Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after hip joint replacement. *Osteoarthritis Cartilage* 2005;13:1076-83.
21. 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.
22. Chiou CF, Sherbourne CD, Cornelio I, Lubeck DP, Paulus HE, Dylan M, et al. Development and validation of the revised Cedars-Sinai health-related quality of life for rheumatoid arthritis instrument. *Arthritis Rheum* 2006;55:856-63.
23. Tubach F, Dougados M, Falissard B, Baron G, Logeart I, Ravaud P. Feeling good rather than feeling better matters more to patients. *Arthritis Rheum* 2006;55:526-30.
24. Dawson J, Doll H, Coffey J, Jenkinson C. Oxford and Birmingham Foot and Ankle Clinical Research Group. Responsiveness and minimally important change for the Manchester-Oxford foot questionnaire (MOXFQ) compared with AOFAS and SF-36 assessments following surgery for hallux valgus. *Osteoarthritis Cartilage* 2007;15:918-31.
25. Escobar A, Quintana JM, Bilbao A, Aróstegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. *Osteoarthritis Cartilage* 2007;15:273-80.
26. Wells G, Li T, Maxwell L, MacLean R, Tugwell P. Determining the minimal clinically important differences in activity, fatigue, and sleep quality in patients with rheumatoid arthritis. *J Rheumatol* 2007;34:280-9.
27. Pouchot J, Kherani RB, Brant R, Lacaille D, Lehman AJ, Ensworth S, et al. Determination of the minimal clinically important difference for seven fatigue measures in rheumatoid arthritis. *J Clin Epidemiol* 2008;61:705-13.
28. Pope JE, Khanna D, Norrie D, Ouimet JM. The minimally important difference for the health assessment questionnaire in rheumatoid arthritis clinical practice is smaller than in randomized controlled trials. *J Rheumatol* 2009;36:254-9.
29. Shauver MJ, Chung KC. The minimal clinically important difference of the Michigan hand outcomes questionnaire. *J Hand Surg Am* 2009;34:509-14.
30. Greco NJ, Anderson AF, Mann BJ, Cole BJ, Farr J, Nissen CW, et al. Responsiveness of the International Knee Documentation Committee Subjective Knee Form in comparison to the Western Ontario and McMaster Universities Osteoarthritis Index, modified Cincinnati Knee Rating System, and Short Form 36 in patients with focal articular cartilage defects. *Am J Sports Med* 2010;38:891-902.
31. Kvamme MK, Kristiansen IS, Lie E, Kvien TK. Identification of cutpoints for acceptable health status and important improvement in patient-reported outcomes, in rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis. *J Rheumatol* 2010;37:26-31.
32. Akinpelu AO, Odole AC, Raheem S. Minimal clinically important difference of Ibadan knee hip osteoarthritis outcome measure (IKHOAM) a cross-sectional study. *J Phys Ther* 2011;3:52-60.
33. Ornetti P, Dougados M, Paternotte S, Logeart I, Gossec L. Validation of a numerical rating scale to assess functional impairment in hip and knee osteoarthritis: comparison with the WOMAC function scale. *Ann Rheum Dis* 2011;70:740-6.
34. Quintana JM, Aguirre U, Barrio I, Orive M, García S, Escobar A. Outcomes after total hip replacement based on patients' baseline status: what results can be expected? *Arthritis Care Res* 2012;64:563-72.
35. Williams VJ, Piva SR, Irrgang JJ, Crossley C, Fitzgerald GK. Comparison of reliability and responsiveness of patient-reported clinical outcome measures in knee osteoarthritis rehabilitation. *J Orthop Sports Phys Ther* 2012;42:716-23.
36. Clement ND, Macdonald D, Simpson AH. The minimal clinically important difference in the Oxford knee score and Short Form 12 score after total knee arthroplasty. *Knee Surg Sports Traumatol Arthrosc* 2013;22:1933-9.
37. Dures EK, Hewlett SE, Cramp FA, Greenwood R, Nicklin JK, Urban M, et al. Reliability and sensitivity to change of the Bristol Rheumatoid Arthritis Fatigue scales. *Rheumatology* 2013;52:1832-9.
38. Ebert JR, Smith A, Wood DJ, Ackland TR. A comparison of the responsiveness of 4 commonly used patient-reported outcome instruments at 5 years after matrix-induced autologous chondrocyte implantation. *Am J Sports Med* 2013;41:2791-9.
39. Escobar A, García Pérez L, Herrera-Españeira C, Aizpuru F, Sarasqueta C, Gonzalez Sáenz de Tejada M, et al. Total knee replacement; minimal clinically important differences and responders. *Osteoarthritis Cartilage* 2013;21:2006-12.
40. Harris KK, Dawson J, Jones LD, Beard DJ, Price AJ. Extending the use of PROMs in the NHS—using the Oxford Knee Score in patients undergoing non-operative management for knee osteoarthritis: A validation study. *BMJ Open* 2013;3:e003365.
41. Hays RD, Spritzer KL, Fries JF, Krishnan E. Responsiveness and minimally important difference for the Patient-Reported Outcomes Measurement Information System (PROMIS) 20-item physical functioning short form in a prospective observational study of rheumatoid arthritis. *Ann Rheum Dis* 2015;74:104-7.
42. Sorensen AA, Howard D, Tan WH, Ketchersid J, Calfee RP. Minimal clinically important differences of 3 patient-rated outcomes instruments. *J Hand Surg Am* 2013;38:641-9.
43. Paulsen A, Roos EM, Pedersen AB, Overgaard S. Minimal clinically important improvement (MCII) and patient-acceptable symptom state (PASS) in total hip arthroplasty (THA) patients 1 year postoperatively. *Acta Orthop* 2014;85:39-48.
44. Singh JA, Luo R, Landon GC, Suarez-Almazor M. Reliability and clinically important improvement thresholds for osteoarthritis pain and function scales: a multicenter study. *J Rheumatol* 2014;41:509-15.
45. Strand V, Kosinski M, Gnanasakthy A, Mallya U, Mpofu S. Secukinumab treatment in rheumatoid arthritis is associated with incremental benefit in the clinical outcomes and HRQoL improvements that exceed minimally important thresholds. *Health Qual Life Outcomes* 2014;12:31.
46. Ward MM, Guthrie LC, Alba MI. Clinically important changes in individual and composite measures of rheumatoid arthritis activity: thresholds applicable in clinical trials. *Ann Rheum Dis* 2014 May 1 (E-pub ahead of print).
47. Leung YY, Png ME, Conaghan P, Tennant A. A systemic literature review on the application of Rasch analysis in musculoskeletal disease - a special interest group report of OMERACT 11. *J Rheumatol* 2014;41:159-64.
48. Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil* 1989;70:857-60.
49. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino MA, et al. Developing core outcome measurement sets for clinical trials: OMERACT Filter 2.0. *J Clin Epidemiol* 2014;67:745-53.
50. Mills RJ, Calabresi M, Tennant A, Young CA. Perceived changes and minimum clinically important difference of the Neurological Fatigue Index for multiple sclerosis (NFI-MS). *Mult Scler* 2013;19:502-5.
51. Gatchel RJ, Mayer TG. Testing minimal clinically important difference: consensus or conundrum? *Spine J* 2010;10:321-7.
52. Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research. How meaningful is it? *Pharmacoeconomics* 2000;18:419-23.
53. Rouquette A, Blanchin M, Sébille V, Guillemin F, Côté SM,

APPENDIX 1. Anchor-based and distribution-based MCID values for scales in osteoarthritis (OA).

Scale	MCID Improvement	MCID Deterioration	MCID Determined by Anchor Methods			MCID Determined by Distributional Methods		
			ES	SRM	SEM	MDC90	MDC95	GRI
WOMAC pain*								
Knee and hip	-8.3 to -7.5 (n = 2)	6.4 to 11 (n = 2)	—	—	—	—	—	—
Knee	-29.9 to -7.5 (n = 3)	27.1 to 29.9 (n = 1)	0.98 to 1.86 (n = 2)	0.91 to 1.55 (n = 2)	8.08 to 8.6 (n = 2)	—	22.39 (n = 1)	1.2 to 1.74 (n = 1)
Hip	-29.26 to 35.8 (n = 2)	-2.5 to 20 (n = 2)	2.1 to 2.24 (n = 1)	1.86 to 1.98 (n = 1)	7.71 (n = 1)	—	21.98 (n = 1)	1.10 to 2.18 (n = 1)
WOMAC stiffness*								
Knee and hip	-10.1 to -7.2 (n = 2)	2.9 to 5.1 (n = 2)	—	—	—	—	—	—
Knee	-20.84 to -6.3 (n = 2)	—	0.51 to 1.69 (n = 2)	0.40 to 1.32 (n = 2)	4.73 (n = 1)	—	13.11 (n = 1)	1.16 to 1.31 (n = 1)
Hip	-25.91 (n = 1)	1.79 (n = 1)	1.61 to 1.81 (n = 1)	1.39 to 1.53 (n = 1)	10.1 (n = 1)	—	27.98 (n = 1)	0.81 to 1.12 (n = 1)
WOMAC function*								
Knee and hip	-8.0 to 6.7 (n = 2)	10.3 to 13.3 (n = 2)	—	—	—	—	—	—
Knee	-33.5 to -5.3 (n = 6)	32.8 to 33.9 (n = 1)	0.66 to 1.32 (n = 3)	0.64 to 1.13 (n = 3)	4.6 to 10.5 (n = 2)	—	29.12 (n = 1)	0.56 to 0.62 (n = 1)
Hip	-26.54 to -9.42(n = 4)	4.0 to 10.29 (n = 2)	0.45 to 2.58 (n = 2)	0.49 to 1.97 (n = 2)	4.3 (n = 1)	—	11.93 (n = 1)	1.45 to 1.79 (n = 1)
WOMAC total*								
Knee and hip	-8.2 to 6.7 (n = 2)	9.6 to 12.9 (n = 2)	—	—	—	—	—	—
Knee	-11.5 to 11.5 (n = 2)	—	0.96 to 1.19 (n = 1)	0.91 to 1.13 (n = 1)	4.8 to 7.2 (n = 1)	11.2 to 16.7 (n = 1)	13.4 to 20 (n = 1)	0.85 to 1.00 (n = 1)
Hip	-14.7 (n = 1)	—	—	—	—	—	—	—
SF-36 PF								
Knee and hip	3.3 (n = 1)	-5.3 (n = 1)	—	—	—	—	—	—
Knee	0.75 to 17.5 (n = 3)	—	0.1 to 1.10 (n = 3)	0.33 to 0.83 (n = 2)	7.03 to 7.74 (n = 2)	14.52 (n = 1)	19.5 (n = 1)	0.38 to 0.66 (n = 1)
Hip	20.40 (n = 1)	-8.75 (n = 1)	1.54 to 1.60 (n = 1)	1.10 (n = 1)	6.85 (n = 1)	—	18.99 (n = 1)	1.50 to 1.54 (n = 1)
SF-36 RP								
Knee	3.77 to 13.16 (n = 2)	—	0.21 to 1.15 (n = 3)	0.2 to 0.69 (n = 2)	6.08 to 9.73 (n = 2)	14.38 (n = 1)	26.97 (n = 1)	0.28 (n = 1)
Hip	10.78 (n = 1)	-20.0 (n = 1)	1.23 to 1.72 (n = 1)	0.74 to 0.99 (n = 1)	8.19 (n = 1)	—	22.71 (n = 1)	0.26 (n = 1)
SF-36 BP								
Knee and hip	7.8 (n = 1)	-7.2 (n = 1)	—	—	—	—	—	—
Knee	6.69 to 20.35 (n = 2)	—	0.56 to 0.66 (n = 3)	0.45 to 0.57 (n = 2)	10.28 to 13.86 (n = 2)	24.93 (n = 1)	37.91 (n = 1)	0.21 to 0.65 (n = 1)
Hip	14.67 (n = 1)	-15.71 (n = 1)	1.15 to 1.26 (n = 1)	0.91 to 0.96 (n = 1)	13.74 (n = 1)	—	38.09 (n = 1)	0.04 to 0.63 (n = 1)
SF-36 GH								
Knee	-7.3 to 9.58 (n = 2)	—	-0.07 to 0.32 (n = 2)	-0.07 to 0.16 (n = 2)	8.43 to 9.88 (n = 2)	15.15 (n = 1)	27.4 (n = 1)	-0.31 to 0.04 (n = 1)
Hip	0.40 (n = 1)	-1.71 (n = 1)	0.09 to 0.25 (n = 1)	0.09 to 0.27 (n = 1)	10 (n = 1)	—	27.73 (n = 1)	0.04 to 0.63 (n = 1)
SF-36 VT								
Knee	2.60 to 3.86 (n = 3)	—	0.14 to 0.46 (n = 3)	0.01 to 0.46 (n = 2)	7.72 to 10.76 (n = 2)	16.21 (n = 1)	29.84 (n = 1)	0.12 to 0.21 (n = 1)
Hip	10.14 (n = 1)	-3.33 (n = 1)	0.79 to 0.80 (n = 1)	0.73 to 0.78 (n = 1)	11.31 (n = 1)	—	31.35 (n = 1)	0.03 to 0.35 (n = 1)
SF-36 SF								
Knee	6.15 to 11.66 (n = 2)	—	0.04 to 0.57 (n = 3)	0.29 to 0.58 (n = 2)	8.7 to 14.87 (n = 2)	24.33 (n = 1)	41.23 (n = 1)	-0.31 to 0.04 (n = 1)
Hip	8.63 (n = 1)	-27.08 (n = 1)	0.72 to 0.74 (n = 1)	0.69 to 0.71 (n = 1)	15.17 (n = 1)	—	42.05 (n = 1)	0.55 to 0.83 (n = 1)
SF-36 RE								
Knee	2.42 to 7.65 (n = 2)	—	0.15 to 0.48 (n = 3)	0.16 to 0.49 (n = 2)	5.86 to 10.29 (n = 2)	16.3 (n = 1)	28.52 (n = 1)	0.04 to 0.17 (n = 1)
Hip	-6.45 (n = 1)	-25.0 (n = 1)	0.29 to 0.31 (n = 1)	0.26 to 0.30 (n = 1)	10.94 (n = 1)	—	30.33 (n = 1)	0.37 to 0.44 (n = 1)
SF-36 MH								
Knee	-0.32 to 4.48 (n = 2)	—	0.13 to 0.32 (n = 3)	0.12 to 0.34 (n = 2)	7.34 to 8.73 (n = 2)	14.14 (n = 1)	24.19 (n = 1)	-0.02 to 0.16 (n = 1)
Hip	8.99 (n = 1)	-12.0 (n = 1)	0.54 to 0.56 (n = 1)	0.54 to 0.60 (n = 1)	8.42 (n = 1)	—	23.33 (n = 1)	-0.13 (n = 1)
SF-36 nPCS								
Knee	4.6 to 4.8 (n = 1)	—	0.44 to 0.98 (n = 2)	0.41 to 1.00 (n = 2)	—	—	—	—
SF-36 nMCS								
Knee	0.3 to 2.7 (n = 1)	—	0.27 to 0.38 (n = 2)	0.29 to 0.40 (n = 2)	—	—	—	—

*Scores standardized to 0–100 scale. n: number of articles; MCID: minimal clinically important difference; ES: effect size; SRM: standardized response mean; SEM: standard error of measurement; MDC90: minimum detectable change 90%; MDC95: minimum detectable change 95%; GRI: Guyatt's responsiveness index; WOMAC: The Western Ontario and McMaster Universities Osteoarthritis Index; SF-36: Medical Outcomes Study Short Form-36; SF-36 PF: SF-36 physical functioning; SF-36 RP: SF-36 role physical; SF-36 BP: SF-36 bodily pain; SF-36 GH: SF-36 general health; SF-36 VT: SF-36 vitality; SF-36 SF: SF-36 social functioning; SF-36 RE: role emotional; SF-36 MH: SF-36 mental health; SF-36 nPCS: normalized physical component summary; SF-36 nMCS: normalized mental component summary.

APPENDIX 2. Anchor-based and distribution-based MCID values for scales in rheumatoid arthritis.

Scale	MCID Determined by Anchor Methods			MCID Determined by Distributional Methods	
	MCID Improvement	MCID Deterioration	ES	SRM	SEM
HAQ	−0.04 to −0.71 (n = 7)	0.01 to 0.495 (n = 4)	0.12 to 0.15 (n = 2)	−0.65 (n = 1)	—
FACIT-F	2.5 to 14.2 (n = 3)	−6.8 to 13.7 (n = 3)	—	—	4.06 to 4.24 (n = 1)

n: number of articles; MCID: minimal clinically important difference; ES: effect size; SRM: standardized response mean; SEM: standard error of measurement; HAQ: Health Assessment Questionnaire; FACIT-F: Functional Assessment of Chronic Illness Therapy–Fatigue.