Journal of Clinical Epidemiology 133 (2021) 61−71

Journal of Clinical Epidemiology

Minimal important difference estimates for patient-reported outcomes: A systematic survey

Alonso Carrasco-Labra[a,b,#,*], Tahira Devji[a,#], Anila Qasim[a], Mark R. Phillips[a], Yuting Wang[a], Bradley C. Johnston[a,c], Niveditha Devasenapathy[d], Dena Zeraatkar[a], Meha Bhatt[a], Xuejing Jin[e,f], Romina Brignardello-Petersen[a], Olivia Urquhart[g], Farid Foroutan[h], Stefan Schandelmaier[a,i], Hector Pardo-Hernandez[j], Qiukui Hao[a,k], Vanessa Wong[a], Zhikang Ye[a], Liam Yao[a], Robin W.M. Vernooij[l,m], Hsiaomin Huang[n], Linan Zeng[a,o], Yamna Rizwan[p], Reed Siemieniuk[a], Lyubov Lytvyn[a], Donald L. Patrick[q], Shanil Ebrahim[a], Toshi A. Furukawa[r], Gihad Nesrallah[s,t], Holger J. Schünemann[a,u], Mohit Bhandari[a,v], Lehana Thabane[a], Gordon H. Guyatt[a,u]

[a]*Department of Health Research Methods, Evidence and Impact, McMaster University, 1280 Main St East, Hamilton, Ontario L8S 4L8, Canada*
[b]*Department of Oral and Craniofacial Health Science, School of Dentistry, University of North Carolina at Chapel Hill, 385 S Columbia St, Chapel Hill, NC 27599, USA*
[c]*Department of Nutrition, Texas A&M University, College of Agriculture and Life Sciences, College Station, TX B3H 1V7, USA*
[d]*Indian Institute of Public Health-Delhi, Public Health Foundation of India, Plot 47, Sector 44, Institutional Area, Gurgaon, Haryana 122002, India*
[e]*Centre for Evidence-Based Chinese Medicine, Beijing University of Chinese Medicine, Beijing, China*
[f]*School of Public Health, University of Alberta, 85 Avenue, Edmonton, Alberta AB T6G 2R3, Canada*
[g]*Department of Evidence Synthesis & Translation Research, ADA Science & Research Institute, American Dental Association, 211 E Chicago Avenue, Chicago, Illinois, USA*
[h]*Ted Rogers Centre for Heart Research, University Health Network, Toronto, Ontario, Canada*
[i]*Institute for Clinical Epidemiology and Biostatistics, Department of Clinical Research, University of Basel, Basel, Switzerland*
[j]*Iberoamerican Cochrane Centre, Biomedical Research Institute (IIB Sant Pau), CIBERESP, Hospital de la Santa Creu i Sant Pau, C/ Sant Antoni M. Claret 167 Pavelló 18 planta 0, 08025 Barcelona, Spain*
[k]*The Center of Gerontology and Geriatrics/National Clinical Research Center for Geriatrics, West China Hospital, Sichuan University, Chengdu, China*
[l]*Department of Nephrology & Hypertension, University Medical Center Utrecht, Utrecht, The Netherlands*
[m]*Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands*
[n]*Department of Orthopaedic Surgery, University of Michigan, 1500 E Medical Center Dr, Ann Arbor, MI 48109, USA*
[o]*Pharmacy Department/Evidence-based Pharmacy Center, West China Second University Hospital, Sichuan University, Chengdu, China*
[p]*Department of Psychology, University of Guelph, 50 Stone Road E, Guelph, Ontario N1G 2W1, Canada*

Author's contributions: A.C.L., T.D., B.C.J., G.N., S.E., and G.H.G. conceived the study idea; A.C.L., T.D., B.C.J., A.Q., M.P., and G.G. created the data extraction form for the MID inventory and led the development of the credibility instrument; A.C.L., T.D., A.Q., M.P., Y.W., N.D., D.Z., M.B., X.J., R.B.P., O.U., F.F., S.S., H.P.H., Q.H, V.W., Z.Y., L.Y., R.W.M.V., H.H., L.Z., Y.R., R.S., and L.L. extracted data and assessed the credibility of MIDs in our inventory; A.C.L. and T.D. wrote the first draft of the manuscript; A.C.L., T.D., A.Q., M.P., Y.W., N.D., B.C.J., D.Z., M.B., X.J., R.B.P., O.U., F.F., S.S., H.P.H., Q.H., V.W., ZY., L.Y., R.W.M.V., H.H., L.Z., Y.R., R.S., L.L., D.L.P., S.E., T.A.F., G.N., H.J.S., M.B., L.T., and G.H.G. interpreted the data analysis, critically revised and approved this manuscript. A.C.L. and T.D. are the guarantors.

Conflict of interest statement: ACL, TD, and GHG hold the copyright of the credibility tool to evaluate minimal important difference estimates. (Devji T, Carrasco-Labra A, Qasim A, Phillips M, Johnston BC, Devasenapathy *N*, Zeraatkar D, Bhatt M, Jin X, Brignardello−Petersen R, et al. 2020. Evaluating the credibility of anchor-based estimates of minimal important differences for patient-reported outcomes: Instrument development and reliability study. BMJ (Clinical research ed). 369:m1714.)

Funding/support: This research was funded in part by the Canadian Institutes of Health Research (CIHR), Knowledge Synthesis grant number DC0190SR.

Role of the sponsor: The funding organization did not influence the design or conduct of the study; collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; nor the decision to submit this manuscript for publication.

Ethical approval statement: Not required. The Minimal Important Difference Inventory, authored by Dr Alonso Carrasco−Labra et al., is the copyright of McMaster University (Copyright © 2018, McMaster University, Hamilton, Ontario, Canada). The Minimal Important Difference Inventory has been provided under license from McMaster University and must not be copied, distributed, or used in any way without the prior written consent of McMaster University. Contact the McMaster Industry Liaison Office at McMaster University, email: milo@mcmaster.ca for licensing details.

* Corresponding author: Alonso Carrasco-Labra, DDS, MSc, Ph.D. Department of Health Research Methods, Evidence and Impact, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4L8, Canada. Tel.: 905-525-9140 x 22338.
E-mail address: alonsocarrascolabra@gmail.com (A. Carrasco-Labra).

https://doi.org/10.1016/j.jclinepi.2020.11.024
0895-4356/© 2020 Elsevier Inc. All rights reserved.

[q]Department of Health Services, University of Washington, 1705 NE Pacific St, Seattle, WA 98195, USA
[r]Department of Health Promotion and Human Behavior, Department of Clinical Epidemiology, Kyoto University Graduate School of Medicine / School of Public Health, Yoshidakonoecho, Sakyo Ward, Kyoto 606-8501, Japan
[s]Nephrology Program, Humber River Regional Hospital, 1235 Wilson Avenue, Toronto, Ontario M3M 0B2, Canada
[t]Faculty of Medicine, University of Toronto, 1 King's College Cir, Toronto, Ontario M5S 1A8, Canada
[u]Department of Medicine, McMaster University, 1280 Main St East, Hamilton, Ontario L8S 4L8, Canada
[v]Department of Surgery, McMaster University, 1280 Main St East, Hamilton, Ontario L8S 4L8, Canada

Accepted 29 November 2020; Published online 13 December 2020

Abstract

Objectives: The objective of the study was to develop an inventory summarizing all anchor-based minimal important difference (MID) estimates for patient-reported outcome measures (PROMs) available in the medical literature.

Study Design and Setting: We searched MEDLINE, EMBASE, CINAHL, PsycINFO, and the Patient-Reported Outcome and Quality of Life Instruments Database internal library (January 1989−October 2018). We included primary studies empirically calculating an anchor-based MID estimate for any PROM in adults and adolescents. Pairs of reviewers independently screened and selected studies, extracted data, and evaluated the credibility of the MIDs.

Results: We identified 585 eligible studies, the majority conducted in Europe (n = 211) and North America (n = 179), reporting 5,324 MID estimates for 526 distinct PROMs. Investigators conducted their studies in the context of patients receiving surgical (n = 105, 18%), pharmacological (n = 85, 15%), rehabilitation (n = 65, 11%), or a combination of interventions (n = 194, 33%). Of all MID estimates, 59% (n = 3,131) used a global rating of change anchor. Major credibility limitations included weak correlation (n = 1,246, 23%) or no information regarding the correlation (n = 3,498, 66%) between the PROM and anchor and imprecision in the MID estimate (n = 2,513, 47%).

Conclusion: A large number of MIDs for assisting in the interpretation of PROMs exist. The MID inventory will facilitate the use of MID estimates to inform the interpretation of the magnitude of treatment effects in clinical research and guideline development. © 2020 Elsevier Inc. All rights reserved.

Keywords: Patient-reported outcome measure; Minimal important difference

1. Introduction

Outcomes that matter to patients have become a key focus in studies evaluating the effects of health care interventions. Patient-reported outcome measures (PROMs) provide information regarding a patient's health condition directly from the patient without interpretation by a clinician or anyone else [1]. Investigators have developed PROM measuring constructs such as function, pain, dyspnea, and fatigue. Many instruments measure a number of domains that bear on broader constructs, including functional status, emotional function, and health-related quality of life.

The evaluation and results of PROMs as outcomes in clinical trials, systematic reviews, clinical practice guidelines, although undeniably important, suffer from difficulties with intuitive understanding regarding the magnitude of change that patients have experienced [2]. The minimal important difference (MID), initially defined as ''the smallest difference that patients perceive as beneficial and that would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management'' [3] is the most widely used approach to facilitating the interpretation of PROMs. An update of this definition includes the patient's perception not only of the benefits but also of harms and the possibility of an ''informed proxy'' as a valid informant when the patient is incapable of providing the information [4,5].

Investigators use two primary strategies to determine an MID: distribution and anchor-based. Distribution-based approaches rely on the statistical characteristics of the sample that fail to incorporate the patient perspective and yield MIDs that will vary widely depending on sample characteristics [6,7]. Anchor-based approaches relate a change in a PROM to an external criterion (i.e., the anchor) that is itself interpretable and provide meaning to the change experienced in the PROM [8]. Empirical evidence suggests that estimates from distribution-based approaches differ from one another and from anchor-based approaches and thus are of limited use [9,10].

Although widely accepted, the use of anchor-based MID estimates also presents challenges. Clinical trialists, systematic review authors, and guideline developers wishing to use MIDs to enhance PROM interpretability must conduct systematic searches to identify primary studies ascertaining MIDs. They will often find multiple MIDs and will often lack training and skills to choose the most credible and applicable to their context [11−13]. Therefore, to facilitate the interpretation of PROMs and to increase our understanding of and access to MIDs, we summarized all anchor-based MID estimates for PROMs available in the medical literature and evaluated their credibility.

What is new?

Key findings

- Authors frequently incorporate anchors that appropriately rely on patient reports and are relevant to patients but seldom report the correlation between the PROM and the anchor, a key credibility item, and often fail to enroll sufficient participants to ensure the precision of MID estimates.

What this add to what is known?

- This is the first systematic survey evaluating the completeness of reporting among primary studies empirically ascertaining anchor-based MIDs for PROMs and the impact of reporting on MID credibility assessment.

What is the implication, what should change now?

- A large number of MID estimates are available to inform the interpretation of a great many PROMs across a wide variety of clinical areas, but their credibility is often limited. Clinical trialists, systematic review authors, and guideline panels can use the MID inventory in interpreting the results of PROMs.

2. Methods

Readers can find a detailed report of the methods of our review in a previously published protocol [14]. This report adheres to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses criteria that are relevant for this systematic survey [15].

2.1. Eligibility criteria

We included primary studies empirically estimating an anchor-based MID for one or more PROMs (in our terminology, the target instruments) in adolescent (≥12 to 17 years) or adult (≥18 years) populations. PROMs of interest measured health-related quality of life, functional ability, symptom severity, and psychological distress and well-being [14].

We included any reported MID irrespective of the participants' condition or disease, type of intervention used in the study, or nature of the anchor. We included reports using any MID-related terminology (e.g., minimally clinically important difference, subjective significant difference, clinically important difference, and minimally detectable change) and any anchor to which results on the target instrument were related, irrespective of the interpretability of that anchor [14]. Eligible studies with longitudinal (e.g., global rating of change, prognosis of future events, and change in disease-related outcomes) or cross-sectional (e.g., comparison with another group with a different status on the same condition or domain and preference rating) designs were considered [6].

We excluded systematic reviews of anchor-based MID estimation studies, abstracts from conferences, studies in which authors explicitly targeted a moderate or large important difference as opposed to an MID, and MIDs estimated using a combined anchor and distribution-based approach.

2.2. Literature search

We searched Medline, EMBASE, CINAHL, and PsycINFO for studies published between 1989 and October 2018 (an approach to estimate an MID was first described in the medical literature in 1989 [3]). The search strategy, adapted to each database, included terms representing the MID concept along with terms addressing PROMs (eTable 1). To complement this search, we also accessed the Patient-Reported Outcome and Quality of Life Instruments Database [17] internal library and reviewed reference lists from relevant reviews and eligible studies.

2.3. Study selection, data collection, and analysis

Reviewers, working in pairs, independently screened titles and abstracts for potentially eligible studies. Studies identified as potentially relevant were retrieved for full-text evaluation, again conducted in duplicate. Reviewers resolved disagreement by discussion or, if needed, by consultation with a third reviewer (ACL, TD).

Before commencing data extraction, all reviewers received extensive training and participated in calibration exercises in which reviewers abstracted and thoroughly discussed data from up to seven studies. The unit of data extraction was the MID estimate. For each MID, reviewers abstracted information about the country of the study; patient demographics; PROM characteristics (i.e., construct(s), domain (if applicable), and scale); interventions administered in the context of MID estimation; anchor details (i.e., type, construct(s), range of options/categories/values, threshold selected to represent a ''small but important difference,'' and specific anchor-based method); MID estimate, its associated measure of precision; and details regarding MID determination (e.g., number of patients included in MID estimation, in longitudinal studies duration of follow-up, analytical approach, and correlations between the PROM and anchor). Using a previously published taxonomy, [16] we classified PROMs in two main categories with two and four subcategories: 1) generic (health profiles and utility measures) and 2) specific (disease/condition, symptom, function, and population-specific). Each pair of reviewers resolved disagreements by discussion with input from a third reviewer (ACL, TD, AQ, and MP).

We used descriptive statistics such as frequencies and percentages to summarize the data.

2.4. Credibility assessment

We defined credibility as "the extent to which the design and conduct of studies measuring MIDs is likely to have protected against misleading estimates" [14]. We assessed the credibility of MID estimates using an instrument developed in the context of this project; we elsewhere report the development of the instrument, its characteristics, and its high reliability [18]. (Table 1 summarizes the instrument items and provide examples of high credibility from studies empirically ascertaining minimal important difference estimates). The instrument is designed for assessment of an individual MID estimate; thus, each MID estimate from a single study providing multiple estimates warrants its own credibility evaluation. The instrument includes two components: 1) core items with five criteria applicable to any anchor-based MID estimation and 2) four additional items addressing global ratings of change—also referred to as a transition rating—anchors. With the exception of item number one in the core criteria, which has a yes/no response, each item in the instrument provides a five-point adjectival scale with response options definitely yes (high credibility), to a great extent, not so much, definitely no, or impossible to tell (low credibility). Reviewers, working in pairs, independently conducted the credibility evaluation, resolving disagreements by discussion with input from a third reviewer for quality assurance (ACL, TD, AQ, and MP).

3. Results

3.1. Search results

Of 14,840 citations identified from our search, reviewers screened 10,469 titles and abstracts, of which 2,161 studies proved eligible for full-text evaluation. Of these, reviewers ultimately deemed eligible 585 studies reporting on 5,324 MID estimates for 526 distinct PROMs (Fig. 1).

3.2. Study-level characteristics

Most of the studies were conducted in Europe and North America. Many investigators conducted their studies in the context of patients receiving surgical, pharmacologic, rehabilitation, or a combination of interventions (Table 2). Investigators most commonly enrolled adults under age 65 years (45%) or adults of all ages (45%); 4% enrolled only those over 65 years. Most of the studies (n = 366, 63%) reported estimates for one PROM, while 162 (28%) included two or three PROMs (Fig. 2A). The median number of MIDs reported per study was 4 (interquartile range, 2−10).

3.3. PROM-level characteristics

Of the 526 PROMs for which MID estimates were available, 67% were specific for a disease/condition, 21% were symptom-specific, and 6% were function-specific; whereas only 5% were classified as generic health profiles or utility indices (Table 2). Disease/condition-specific PROMs most commonly addressed musculoskeletal disorders, cancer, and urologic/gynecologic conditions. Symptom-specific PROMs most frequently evaluated pain, fatigue, and dyspnea, whereas function-specific PROMs frequently assessed physical function and activities of daily living. Most PROMs have more than one MID available and often multiple MIDs for a given PROM are estimated within a single study. Five PROMs (i.e., the EORTC Core Quality Of Life Questionnaire (EORTC QLQ-C30), the 36-Item Short Form Survey version 1 and 2, the Knee injury and Osteoarthritis Outcome Score, and the Western Ontario and McMaster Universities Osteoarthritis Index) had more than 100 MID estimates available (Fig. 2B), and for 42 (7%) PROMs, investigators reported between 25 and more than 100 estimates within a single study (Fig. 2C).

3.4. MID-level characteristics

In total, the included studies reported 5,324 individual MID estimates. Most studies addressed the MID related to participants' improvement (66%), whereas 31% addressed worsening or conducted analyses assuming that MIDs were similar for improvement and worsening (Table 2). Most MID estimates (n = 4,707; 88%) were generated from studies using longitudinal designs. In such studies, patients responded to the target instrument on two occasions and at follow-up a global rating of change, a measure of satisfaction, ratings of another PROM, or report of a clinical end point. In studies using cross-sectional study designs, investigators either asked participants to compare their status on the target domain with others or compared target instrument scores from groups that differed on the anchor (Table 2).

3.5. Nature and source of information of the anchor

The nature of the anchor (e.g., global rating of change, disease-related outcome, and comparison with another group) and the source of the information (e.g., self-report or proxy report and performance-based measure) varied considerably (Table 2). Investigators typically used anchors in which patients reported their status (83%), most commonly a global rating of change or transition rating (59%) and less frequently change in a disease-related outcome (23%) or comparison with another group (11%). Investigators used a proxy as the source of information for 470 MID estimates (9%), which was almost always a clinician's impression of change in health status (452 MIDs, 96%). Investigators less frequently used other anchors such as clinical or laboratory data (e.g., hemoglobin

Table 1. Examples of high credibility from studies empirically ascertaining minimal important difference estimates[a]

Credibility item	Rationale for the item	Example of high credibility	Study reference
Is the patient or necessary proxy responding directly to both the PROM and the anchor?	As anchor-based MIDs aim to assist in the interpretation of the magnitude of a difference in a PROM, it is desirable that the anchor instrument used to ascertaining the MID is also informed by patients, as opposed to clinicians or a distant proxy. Thus, the MID used to interpret the PROM truly reflect the patients' perspective.	''At each visit, patients completed the HAQ-DI (0–3) and VAS for pain, fatigue, sleep, and global status, which ranged from 0 (none) to 100 mm (very severe). Patients also completed a 5-point Likert scale of change that asked, ''How would you describe your overall status since the last visit?''	Sekhon S, Pope J; Canadian Scleroderma Research Group, Baron M. The minimally important difference in clinical practice for patient-centered outcomes including health assessment questionnaire, fatigue, pain, sleep, global visual analog scale, and SF-36 in scleroderma. J Rheumatol. 2010 Mar; 37(3):591-8.
Is the anchor easily understandable and relevant for patients or necessary proxy?	Anchors that are easily understandable and relevant to patients directly measure change in health status (function, symptoms, treatment success, disease severity, or prognosis of future events such as death, job loss, disability, etc.)	''Compared with how your urinary incontinence was before treatment, do you feel you are (very much better, much better, better, about the same, worse, much worse, very much worse?).''	Barber MD, Spino C, Janz NK, Brubaker L, Nygaard I, Nager CW, Wheeler TL; Pelvic Floor Disorders Network. The minimum important differences for the urinary scales of the Pelvic Floor Distress Inventory and Pelvic Floor Impact Questionnaire. Am J Obstet Gynecol. 2009 May; 200(5):580.e1-7.
Has the anchor shown good correlation with the PROM?	The anchor and PROM should be measuring the same or similar underlying construct, and thus should be moderate to highly correlated. A poorly correlated anchor with the PROM will likely provide misleading MID estimates.	''The correlations between the GRCS and the score changes of both scales were r = 0.72 (95% CI: 0.66, 0.78) for the DASH and r = 0.71 (95% CI: 0.64, 0.76) for the QuickDASH ($P < .001$ for both).''	Franchignoni F, Vercelli S, Giordano A, Sartorio F, Bravini E, Ferriero G. Minimal clinically important difference of the disabilities of the arm, shoulder, and hand outcome measure (DASH) and its shortened version (QuickDASH). J Orthop Sports Phys Ther. 2014 Jan; 44(1):30-9.
Is the MID precise?	As with any estimation method, MID point estimates should be accompanied by a measure of variability, for example, 95% confidence intervals. To judge the extent to which an MID estimate is precise, the confidence interval can be used to define the likelihood that inferences about the magnitude of a treatment effect would differ at the lower and upper boundaries of the interval.	''In a pooled analysis the estimated MCID [for the Fibromyalgia Impact Questionnaire total score] (95% confidence interval) was 14% (13; 15) and for FIQ stiffness it was 13% (12; 14)''	Bennett RM, Bushmakin AG, Cappelleri JC, Zlateva G, Sadosky AB. Minimal clinically important difference in the fibromyalgia impact questionnaire. J Rheumatol. 2009 Jun; 36(6):1304-11.
Does the threshold or difference between groups on the anchor used to estimate the MID reflect a small but important difference?	Even if the authors of the study claim to estimate an MID, judgment is needed to determine whether the selected groups or thresholds compared on the anchor instrument actually reflect a small but important difference, as supposed to a moderate to large one.	''We calculated the changes in RQLQ score between consecutive visits for each patient and related these to the individual patient's global rating of change score.'' ''They respond on a 15-point scale from −7 (a very great deal worse) to 0 (no change) to +7 (a very great deal better). Previous studies have shown that global rating scores of −1, 0, and +I are trivial; but changes of +2 or greater and −2 or less are important to patients and can be considered	Juniper EF, Guyatt GH, Griffith LE, Ferrie PJ. Interpretation of rhinoconjunctivitis quality of life questionnaire data. J Allergy Clin Immunol. 1996 Oct; 98(4):843-5.

(*Continued*)

Table 1. Continued

Credibility item	Rationale for the item	Example of high credibility	Study reference
		clinically important. Changes of +3, +2, −2, and −3 can be considered minimally important changes (i.e., the MID) …''	

Abbreviations: MID, minimal important difference; PROM, patient-reported outcome measure; HAQ-DI, Health Assessment Questionnaire Disability Index; VAS, visual analogu scale; DASH, Disabilities of the Arm; Shoulder and Hand; GRCS, global rating of change scale; MCID, minimal clinically important difference; FIQ, Fibromyalgia Impact Questionnaire; RQLQ, Rhinoconjunctivitis Quality of Life Questionnaire.

[a] Credibility instrument from Devji T, Carrasco–Labra A, Qasim A, Phillips M, Johnston BC, Devasenapathy *N*, Zeraatkar D, Bhatt M, Jin X, Brignardello–Petersen R, Urquhart O, Foroutan F, Schandelmaier S, Pardo-Hernandez H, Vernooij RW, Huang H, Rizwan Y, Siemieniuk R, Lytvyn L, Patrick DL, Ebrahim S, Furukawa T, Nesrallah G, Schünemann HJ, Bhandari M, Thabane L, Guyatt GH. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. BMJ. 2020 Jun 4; 369:m1714.

level, number of metastatic sites, and forced vital capacity), performance-based measures (e.g., accelerometry data and best-corrected visual acuity), and administrative data (e.g., occurrence of death and rehospitalization) (Table 2).

3.6. Analytical approach for MID estimation

Investigators used a variety of analytical approaches to compute the MID estimate (Table 3). In MIDs from longitudinal study design (n = 4,707), investigators most frequently measured the change in target instrument score in those who reported a small but important change on the anchor (49%) or compared the change in those reporting a small but important difference with another group (e.g., patients reporting no change) (18%). Less frequently, authors used a receiver operating characteristic curve analysis and seldom other approaches (e.g., regression and ANOVA modeling, discriminant function analysis, and linkage or scale-alignment approaches). In cross-sectional studies (n = 616), investigators most frequently compared scores on the target instrument in groups that differed on the anchor (72%), with one-third using regression modeling.

3.7. Credibility assessment of available MID estimates

In most cases, MID estimates met the first criterion (n = 4,456; 84%)—patients or proxies usually responded to both the target instrument and the anchor. Investigators also usually chose easily understandable anchors (second criterion) (n = 4,713; 89%). Unfortunately, these easily understandable anchors frequently represented a threshold or difference between groups that failed to reflect a small but important difference (n = 1,267; 24%) and sometimes were so poorly presented that judgment was not possible (fifth criterion) (n = 834; 16%). Investigators typically failed to meet the third and fourth criteria, usually neglecting to report the correlation between the target instrument and the anchor (n = 3,498; 66%), and not enrolling sufficient patients to ensure a precise estimate of the MID (n = 2,513; 47%). For the 64% (n = 3,409) that used a global rating of change as the anchor, very few satisfied the four additional criteria in the extension of the credibility tool. The duration of time between the first and second administration of the target PROM was excessively long—more than 3 months—in 50% (n = 1,709) of the MIDs, whereas only 8% (n = 272) reported correlations between the transition

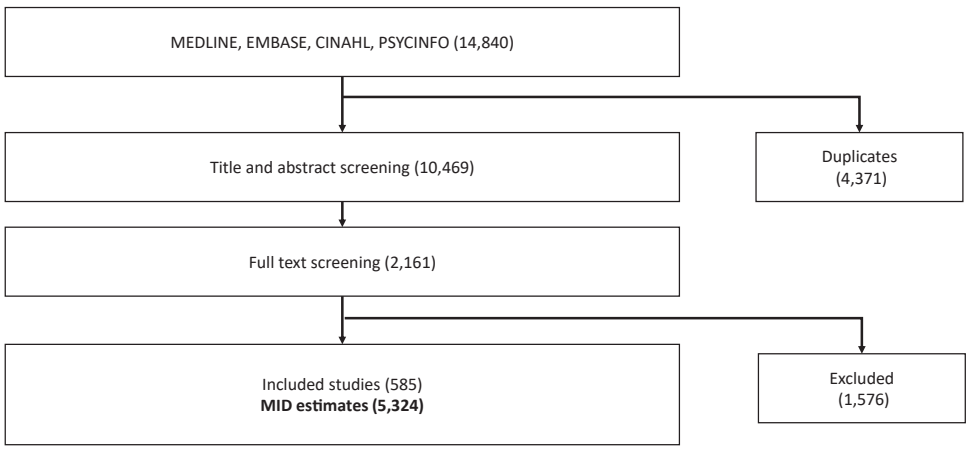

Fig. 1. PRISMA flowchart study selection process.

Table 2. Characteristics of the included studies, patient-reported outcome measures, and reported minimal important difference estimates

Study-level data (n = 585)	
Regions (%)	
Europe	211 (36)
North America	179 (31)
Asia	46 (8)
Oceania	19 (3)
South America	1 (0)
Africa	1 (0)
Multiple continents	36 (6)
Not reported	92 (16)
Interventions (%)	
Surgical/invasive	105 (18)
Pharmacological	85 (15)
Rehabilitation	65 (11)
No intervention	13 (2)
Alternative medicine	9 (2)
Behavioral	3 (1)
Other	43 (7)
Combination of interventions	194 (33)
Not reported	68 (12)
Design (%)	
Longitudinal	539 (92)
Cross-sectional	25 (4)
Both	21 (4)
PROM-level data (n = 526)	
Type of PROM and specific area addressed (%)	
Disease/condition specific	350 (67)
MSK disorders	89 (26)
Cancer	49 (14)
Urologic/gynecologic	29 (8)
Neurologic	36 (10)
Pain	21 (6)
Other	126 (36)
Symptom specific	109 (21)
Pain	48 (44)
Fatigue	14 (13)
Dyspnea	12 (11)
Gastrointestinal symptoms	6 (5)
Other	29 (27)
Function specific	33 (6)
Physical function	18 (55)
Activities of daily living	5 (15)
Sleep	4 (12)
Social function	3 (9)
Sexual function	2 (6)
Work limitations	1 (3)
Utility index	17 (3)
Generic health profile	13 (2)
Other	4 (1)
MID-level data (n = 5,324)	
MID direction (%)	
Improvement	3,514 (66)
Worsening	956 (18)
Improvement/worsening	695 (13)
Unclear	159 (3)
Nature of the anchor (%)	
Global rating of change	3,131 (59)
Change in disease-related outcome	1,242 (23)
Comparison with another group	574 (11)
Satisfaction scale	328 (6)
Prognosis of future events	12 (0)
Comparison with known population(s)	8 (0)
Combination of methods	29 (1)
Source of anchor information (%)	
Self-reported	4,399 (83)
Proxy-reported	470 (9)
Laboratory data	147 (3)
Performance-based measure	115 (2)
Combination of types	56 (1)
Self- and proxy-reported	30 (0)
Administrative data	13 (0)
Unclear	94 (2)

Abbreviations: PROM, patient-reported outcome measure, MSK, musculoskeletal, MID, minimal important difference.

score and the prescore and postscore on the target instrument (Table 4).

On the basis of the results of this systematic survey, we developed an inventory of anchor-based MID estimates that will allow users to search for all available MIDs for PROMs across all clinical disciplines. For each MID, we have summarized information pertaining to the study design, PROM characteristics, patient demographics, intervention details, MID methodology, anchor details, and credibility assessment. Individuals interested in accessing the inventory can do so here: www.promid.org.

4. Discussion

4.1. Main findings

This first systematic summary of all available anchor-based MID estimates for PROMs in the medical literature identified 585 primary studies reporting on more than 5,300 anchor-based MID estimates applicable to 526 distinct PROMs. Studies representing a wide variety of clinical disciplines, most frequently addressed disease/condition-specific PROMs and used longitudinal designs with self-reported global ratings of change. The credibility of the MID estimates varied substantially, and reporting issues often limited the credibility evaluation.

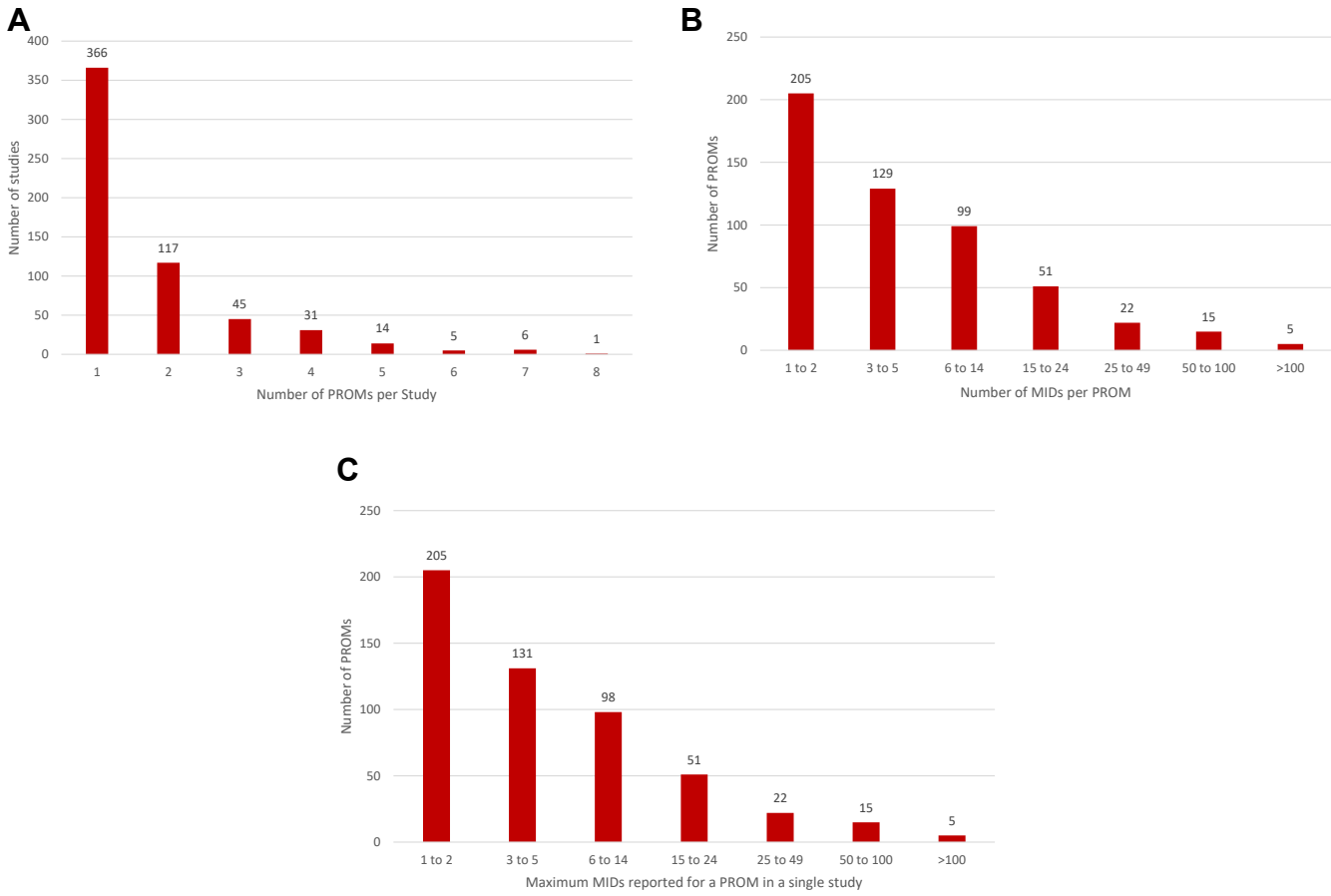

Fig. 2. (A) Frequency of PROMs reported in individual studies; (B) Frequency of MIDs available for PROMs in the inventory; (C) Maximum number of MIDs reported for a PROM in a single study.

4.2. Applications of MIDs for PROM interpretation in clinical research

Several reports have provided guidance for the use of MID estimates to facilitate PROM interpretation, an application that the inventory will greatly facilitate [19]. Investigators have proposed examining the magnitude of treatment effects in relation to the MID and also examining the proportion of patients in intervention and control groups who have experienced improvement or worsening greater than the MID—a so-called "responder analysis." [20] This approach allows the presentation of effect estimates using relative (risk ratio and odds ratio) and absolute measures (risk difference and number needed to treat for benefit or harm) [21].

When conducting a meta-analysis in which studies use different PROMs measuring the same construct, authors can report mean difference in MID units, as an alternative to the standardized mean difference (a measure associated with considerable challenges in interpretability) [22]. Another approach suggests the use of MIDs for the calculation of the probability for trial participants to experience a treatment effect that is greater than or at least equal to the MID [23,24]. Authors have also suggested a role for MID estimates for determining sample size calculation [8,24,25].

4.3. Strengths and limitations

We conducted extensive screening using broad inclusion criteria at a title and abstract level, minimizing the risk of missing MID estimates due to inconsistencies in terminology. Thus, it is likely that our inventory includes a near-complete collection of anchor-based MIDs in adolescents and adults reported in the peer-reviewed medical literature, with a description of salient characteristics, including the credibility of MID estimates. To ensure documentation of relevant characteristics and methodological aspects of MID estimation studies we: i) used a piloted form that underwent iterative testing, ii) conducted extensive calibration processes and selecting and extracting data in duplicate, iii) implemented a quality assurance process in which a third researcher participated as arbiter when clarifying data discrepancies, and iv) created and applied a novel instrument that proved highly reliable to assess MID estimate credibility [18].

This study also has limitations. The lack of standardized reporting and nomenclature for MID estimation studies presented challenges when building search strategies and conducting screening at title and abstract and full-text level, leaving the possibility that our search missed some MIDs. Although we likely achieved a near-complete accounting

Table 3. Analytical approach as per study design and operational definition (n = 5,324)

Design	Analytical approach	*n* (%)	Operational definition
Longitudinal design (n = 4,707)	Mean change	2,288 (49)	The MID was the mean change in PROM scores over time within the subgroup of participants who reported a small but important improvement (or worsening).
	Mean difference	848 (18)	The MID was the difference in PROM scores over time in the participants in one group minus the mean change in PROM scores over time in the participants in another group. The participants in the defined groups typically have a different status on the same condition or disease-related outcome. When a global rating of change anchor is used, often the participants who reported a small but important improvement (or worsening) are compared with those in the no change group.
	Receiver operating characteristic curve	1,054 (22)	The MID was the cutoff point that is defined by determining the lowest overall misclassifications (e.g., point closest to 0, 1 criterion, closest to the −45° tangent line, maximizing the distance to the identity line). Other approaches to ROC analysis include but are not limited to an 80% specificity rule and the use of an optimal likelihood ratio.
	Regression and ANOVA modeling	458 (10)	The MID was estimated using regression modelling (either logistic or linear), where the dependent variable was the change in the PROM score and the independent variable was the value, rating, or category on the anchor (e.g., ratings on a GROC or score on an anchor instrument). Alternatively, PROM score at follow-up was used as the dependent variable, whereas the independent variables were the value, rating, or category on the anchor and the baseline PROM score. A second approach, although less common, involves using the anchor as a dependent variable and the PROM as the independent variable. When using ANOVA modelling, the MID was estimated using as the dependent variable the change in PROM score and the independent variable the value, rating, or category on the anchor (e.g., ratings on a GROC or score on the anchor instrument).
	Other	10 (0)	The MID was estimated using discriminant function analysis, linkage or scale-alignment, or a combination of different analytical approaches.
	Unclear	49 (1)	Although a longitudinal design was used, there was insufficient information to determine the MID analytical method.
Cross-sectional design (n = 616)	Mean difference	444 (72)	The MID is the difference in PROM scores between participants who rated themselves as a little bit better (or a little bit worse) than another participant and participants who rated themselves as about the same as compared with another participant or the difference in PROM scores between participants in groups with a different status on the same condition or disease-related outcome.
	Regression modelling	164 (27)	The MID is estimated using regression modelling (either logistic or linear), where the dependent variable is the PROM score and the independent variable is the value, rating, or category on the anchor (e.g., score on the anchor instrument).
	Unclear	8 (1)	Although a cross-sectional design was used, there was insufficient information to determine the MID analytical method.
Both designs combined (n = 1)			The MID was estimated using a combination of both longitudinal and cross-sectional designs.

Abbreviations: PROM, patient-reported outcome measure, MID, minimal important difference, GROC, global rating of change, ROC, receiver operating characteristic curve.

of MID studies published in the peer-reviewed medical literature, ensuring completeness in future updates will require including the gray literature (e.g., conference abstracts and other PROM databases). Finally, our study is currently comprehensive only to October 2018. This inventory is, however, part of a continuous effort to secure access to the most updated MID estimates; the process of retrieving studies published since 2018 is currently underway and is part of a plan to continuously add new MIDs to our living web-based MID inventory (PROMID—www.promid.org).

Finally, how to proceed when there are numerous MIDs available remains a challenge. As we note in the following sections, the credibility assessment is certain to provide useful guidance. However, our preliminary work suggests that for a given instrument, there may be multiple alternative credible MIDs from which to choose. Addressing this situation is an important part of our ongoing work with the inventory.

4.4. Insights and implications for the use of MID estimates

A number of insights emerged from this study. First, there are a large number of MID estimates available that investigators can use to inform the interpretation, in randomized trials, systematic reviews, and clinical practice guidelines, of a great many PROMs across a wide variety

Table 4. Credibility assessment of MID estimates [a]

Core credibility items (n = 5,324): Count (%)	Definitely no	Not so much	To a great extent	Definitely yes	Impossible to tell
1. Is the patient or necessary proxy responding directly to both the PROM and the anchor?	856 (16)	-	-	4,456 (84)	12 (0)
2. Is the anchor easily understandable and relevant for patients or necessary proxy?	152 (3)	345 (6)	950 (18)	3,763 (71)	114 (2)
3. Has the anchor shown good correlation with the PROM?	333 (6)	913 (17)	466 (9)	114 (2)	3,498 (66)
4. Is the MID estimate precise?	2,513 (47)	808 (15)	487 (9)	802 (15)	714 (13)
5. Does the threshold or difference between groups on the anchor used to estimate the MID reflect a small but important difference?	1,267 (24)	1,275 (24)	1,628 (31)	320 (6)	834 (16)
Extension credibility items (n = 3,409): Count (%)					
1. Is the amount of elapsed time between baseline and follow-up measurement for MID estimation optimal?	1,709 (50)	698 (20)	307 (9)	576 (17)	119 (3)
2. Does the transition item have a substantial positive correlation with the PROM score at follow-up?	10 (0)	8 (0)	13 (0)	82 (2)	3,296 (97)
3. Does the transition item correlate negatively or very weakly positively with the PROM score at baseline?	13 (0)	12 (0)	23 (1)	25 (1)	3,336 (98)
4. Is the correlation of the transition item with the PROM change score appreciably greater than the correlation of the transition item with the PROM score at follow-up?	49 (1)	18 (1)	10 (0)	9 (0)	3,323 (97)

Abbreviations: PROM, patient-reported outcome measure; MID, Minimal important difference.

[a] Devji T, Carrasco-Labra A, Qasim A, et al. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. BMJ (Clinical research ed). 2020; 369:m1714.

of clinical areas. Second, individual studies often report several MIDs, usually for only one or two PROMs; for individual PROMs, there are often between one and five available MID estimates but occasionally far more. Third, investigators make use of a variety of methodologies to generate anchor-based MIDs, the relative merits of which remain to be established. Fourth, although easily understandable and relevant anchors to which patients or proxies responded directly informed the majority of MIDs, most studies failed to report the correlation between the PROM and the anchor and failed to enroll sufficient patients to ensure precise estimates. The substantial deficiencies observed in many studies highlight the need for improvements in the methodology of developing MID estimates.

All alternatives for using MIDs rely on the availability of credible and applicable MID estimates for the context of interest. Currently, choosing an optimal MID estimate for a given PROM presents two important challenges: 1) users of MIDs need to, ideally, conduct comprehensive systematic reviews to identify primary studies reporting MID estimates for the PROM of interest and 2) as our study showed, more than one estimate would likely be available, requiring decisions of which estimate(s) to use. The credibility assessment of the MID constitutes a key, if not a preeminent criterion, for this choice.

Recent publications provide examples of practical applications of MID estimates for improving the interpretation of PROMs in the context of primary studies, systematic reviews, and clinical practice guidelines [13,23,26–28]. By providing easy access to available MIDs, including assessments of their credibility, this inventory will reduce the time, effort, and likelihood of error in MID estimate selection and, in doing so, close the gap between MID estimation studies and their subsequent application in clinical research and practice.

Acknowledgments

The authors would like to thank Tamsin Adams–Webber at the Hospital for Sick Children and Paul Alexander for their assistance with developing the initial literature search. We would also like to thank Shahrzad Motaghi Pisheh, Brittany Dennis, Marc Jacobs, Yuqing Zhang, Kevin Quach, Nigar Sekercioglu, Sean Kennedy, William Zhang, Samantha Craigie, Iván Flórez, Yutong Fei, Brian Younho Hong, Aran Tajika, Nozomi Takeshima, Naotsugu Iwakami, Yu Hayasaka, Angela Kaminski, Barbara Nussbaumer, and Luis Colunga for their contribution on an early stage of this project.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2020.11.024.

References

[1] FDA-NIH Biomarker Working Group. In: BEST (Biomarkers, EndpointS, and other Tools) Resource. Silver Spring, MD: Food and

Drug Administration (US); 2016. Available from: https://www.ncbi.nlm.nih.gov/books/NBK326791/. Accessed May 1, 2020. Co-published by National Institutes of Health (US), Bethesda (MD).
[2] Glassman SD, Carreon LY. Thresholds for health-related quality of life measures: reality testing. Spine J 2010;10(4):328−9.
[3] Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin trials 1989;10:407−15.
[4] Schunemann HJ, Puhan M, Goldstein R, Jaeschke R, Guyatt GH. Measurement properties and interpretability of the Chronic respiratory disease questionnaire (CRQ). COPD 2005;2(1):81−9.
[5] Schunemann HJ, Guyatt GH. Commentary−goodbye M(C)ID! Hello MID, where do you come from? Health Serv Res 2005;40:593−7.
[6] Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol 2003;56:395−407.
[7] 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(3):160−6.
[8] Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. Mayo Clin Proc 2002;77(4):371−83.
[9] Puhan MA, Mador MJ, Held U, Goldstein R, Guyatt GH, Schunemann HJ. Interpretation of treatment changes in 6-minute walk distance in patients with COPD. Eur Respir J 2008;32(3):637−43.
[10] Turner D, Schunemann 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.
[11] Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et al. Mind the MIC: large variation among populations and methods. J Clin Epidemiol 2010;63:524−34.
[12] de Vet HC, Ostelo RW, Terwee CB, van der Roer N, Beckerman H, Boers M, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. Qual Life Res 2007;16(1):131−42.
[13] 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(5):e015587.
[14] 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(10):e007953.
[15] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009;339:b2535.
[16] Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med 1993;118:622−9.
[17] Pinotti R. PROQOLID. J Med Libr Assoc 2016;104(1):91−2.
[18] Devji T, Carrasco-Labra A, Qasim A, Phillips M, Johnston BC, Devasenapathy N, et al. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. BMJ 2020;369:m1714.
[19] Koynova D, Luhmann R, Fischer R. A Framework for managing the minimal clinically important difference in clinical trials. Ther Innov Regul Sci 2013;47(4):447−54.
[20] Schunemann HJ, Akl EA, Guyatt GH. Interpreting the results of patient reported outcome measures in clinical trials: the clinician's perspective. Health Qual Life Outcomes 2006;4:62.
[21] Thorlund K, Walter SD, Johnston BC, Furukawa TA, Guyatt GH. Pooling health-related quality of life outcomes in meta-analysis-a tutorial and review of methods for enhancing interpretability. Res Synth Methods 2011;2(3):188−203.
[22] Johnston BC, Thorlund K, Schunemann HJ, Xie F, Murad MH, Montori VH, 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.
[23] McGlothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients. JAMA 2014;312:1342−3.
[24] Neely JG, Karni RJ, Engel SH, Fraley PL, Nussenbaum B, Paniello RC. Practical guides to understanding sample size and minimal clinically important difference (MCID). Otolaryngol Head Neck Surg 2007;136(1):14−8.
[25] 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.
[26] Brignardello-Petersen R, Guyatt GH, Buchbinder R, Poolman RW, Schandelmaier S, Chang Y, et al. Knee arthroscopy versus conservative management in patients with degenerative knee disease: a systematic review. BMJ Open 2017;7(5):e016114.
[27] Siemieniuk RAC, Harris IA, Agoritsas T, Poolman RW, Brignardello-Petersen R, Van de Velde S, et al. Arthroscopic surgery for degenerative knee arthritis and meniscal tears: a clinical practice guideline. BMJ 2017;357:j1982.
[28] 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.