Journal of Clinical Epidemiology 89 (2017) 188–198

Journal of Clinical Epidemiology

REVIEW

Minimal important difference to infer changes in health-related quality of life—a systematic review

Ravishankar Jayadevappa[a,b,c,d,*], Ratna Cook[a], Sumedha Chhatre[e]

[a]Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[b]Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[c]Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA 19104, USA
[d]Corporal Michael J. Crescenz VAMC, Philadelphia, PA 19104, USA
[e]Department of Psychiatry, School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

Accepted 3 June 2017; Published online 1 July 2017

Abstract

Objectives: The objective of the study was to assess the usability of minimal important difference (MID) and minimal clinically important difference (MCID) for measuring meaningful changes in disease-specific and generic health-related quality-of-life (HRQoL) outcomes in patient-centered care.

Study Design and Setting: We adopted a two-step literature review process. First, we used PubMed and Google scholar to identify a broad range of search terms. Next, we searched OVID Medline, JSTOR, and PubMed for terms ''MID,'' and ''MCID.'' We excluded non-English language studies, articles older than 1995, those not related to generic- and disease-specific HRQoL measures, and protocols of future studies. Studies were grouped according to generic- and disease-specific measures. We assessed MID or MCID calculation methods, effect sizes, estimated values, and significance.

Results: Eighty articles satisfied the inclusion criteria. Our synthesis provides a comprehensive assessment of MID or MCID for 10 generic-specific and 80 disease-specific instruments. We observed a lack of consistency in the application of methods for computing MID or MCID for generic and disease-specific HRQoL measures. Only 43 (54%) studies used both anchor and distribution methods to elicit MID or MCID. Thirty-four articles estimated MID values only, whereas 47 articles estimated MCID.

Conclusion: The anchor-based method yields conservative estimates of MID or MCID, compared to the distribution-based method. The distribution method does not take into account patient perspectives and should be accompanied by anchor method while computing MID. The MID should be interpreted with caution, and available estimates for a particular instrument must be used. This will help in integrating the MID estimates into the overall research or clinical plan for a specific context.

Keywords: Prostate cancer; Health-related quality of life; Minimal important difference; Minimal clinically important difference; Anchor based; Distribution based

1. Introduction

As patient-centered care increases in prominence, understanding how best to alleviate patients' symptom burden has important implications for patients, providers, researchers, and payers. Interpretation of changes in scores of patient-reported outcomes (PROs), such as health-related quality of life (HRQoL), is a challenge to the meaningful integration of PRO measures in patient-centered care and policy [1–6]. A PRO is defined as ''any report coming directly from patients about how they function or feel in relation to a health condition and its therapy'' [7–9]. The Patient-Centered Outcomes Research Institute methodology core is currently investigating strategies for integrating PROs into electronic health records. The PROs represent patients' perspectives on treatment benefits and outcomes beyond survival, disease, and physiological markers. These are often outcomes of great importance to patients and are elicited from interviews, self-completed questionnaires, and diaries, preferably via methods that are rigorous, scientific, and validated [7–9]. Just as laboratory values are routinely examined in clinical care, PROs should be assessed in clinically/minimally meaningful contexts. Thus, individual

Conflict of interest: None.

Funding: This study was supported by the Patient Centered Outcomes Research Institute, CE-12-11-4973.

* Corresponding author. 224, Ralston-Penn Center, 3615 Chestnut Street, Philadelphia, PA 19104-2676, USA. Tel.: +1-215-898-3798; fax: +1-215-573-8684.

E-mail address: jravi@mail.med.upenn.edu (R. Jayadevappa).

http://dx.doi.org/10.1016/j.jclinepi.2017.06.009
0895-4356/

What is new?

- Our synthesis provides a comprehensive assessment of minimal important difference (MID) or minimal clinically important difference (MCID) for 10 generic-specific and 80 disease-specific instruments.

Key findings

- Variation exists in the MID estimates derived from two different methods (distribution-based method and anchor-based method).
- Compared to the distribution-based method, anchor-based MID estimates are conservative.

What this adds to what was known?

- The concept of MID or MCID was developed to aid providers and patients in interpreting change in health status and can play an important role in patient-centered care and comparative effectiveness assessment.

What is the implication and what should change now?

- There is an urgent need to develop MIDs for generic- and disease-specific instruments in different settings for effective patient-centered care.

change standards are necessary to provide meaningful interpretation of effects of an intervention on HRQoL and classify changes in patient health. Change can be meaningful from the societal perspective that takes into account population level with small differences or from institutional perspective that focuses on the degree of change required to influence health care policies [6,10–15]. With these differences in individual perspectives, the definition of meaningful change is discordant [16].

Minimal important difference (MID) or minimal clinically important difference (MCID) has become a standard approach in the interpretation of clinical relevance of changes in PROs. The MID is a truly patient-centered concept that detects not only the magnitude of improvement that is meaningful to patients but also the value that patients place on the change [17]. The MCID is defined as "the smallest difference which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management" [3]. On the other hand, MID is defined as the "the smallest difference in score in the domain of interest that patients perceive as important, either beneficial or harmful, and which would lead the clinician to consider a change in the patient's management" [18]. The two broad methods of estimating MID or MCID, the distribution method and the anchor-based method, are conceptually different [1–6]. The anchor-based method is more patient-centered and uses direct response from the patient using another interpretive method of assessing change in outcome to elucidate the meaning of a particular degree of change [17]. The anchor-based method may use individual-focused strategy that uses a single anchor or population-focused strategy that requires multiple anchors. The distribution method on the other hand uses statistical properties of the distribution of outcome scores, particularly how the scores differ between patients. The distribution methods may use approaches based on standard error of measurement (SEM), standard deviation (SD), effect size, minimal detectable change, reliable change index, or the standardized response mean [16]. The SEM is considered to be a characteristic of the measure, not the sample [19]. Another measure of variability is the SD, defined as the variation among a group of scores, for which 0.5 SD has been suggested to correspond to the MCID [20]. Finally, effect size represents the standardized change in scores [20]. Effect sizes of 0.20, 0.50, and 0.80 indicate small, moderate, and large effects, respectively [21], and would be smaller in patients reporting no change compared with those reporting a great improvement [16]. The distribution approach to MID considers the effect size of the difference between treatment and control groups to measure variability, standardized response mean, standard error, and the responsiveness statistics [22,23]. As this method does not consider patient perspectives, it is sometimes argued that this method alone should not be used to derive final MID or MCID values. By contrast, most anchor-based approaches are presumed to be sample independent as they use external criterion like retrospective judgment of change [1–6,10,11,13,14]. However, anchor-based method can be at risk for recall bias, especially for long-term treatment effectiveness, and may be affected by a patient's current state and events pertaining to the disease in question [2]. Individuals have unique perspectives about the needs and goals of treatment, and these perspectives may vary by outcomes. Thus, the consensus is to use a combination of anchor-based and distribution-based methods to derive MID or MCID. The objective of this study was to conduct a comprehensive systematic review to assess the use of MID and MCID for measuring meaningful changes in disease-specific and generic HRQoL outcomes in patient-centered care.

2. Methods

2.1. Search strategy

We conducted a systematic literature review by searching OVID Medline, JSTOR, Google scholar, and PubMed for word "minimal important difference," "minimal clinically important difference," "MID," or "MCID." Studies were

limited to English language and for the years between 1995 and 2016. After an initial check for duplicated articles, the abstracts of remaining articles were screened to rule out literature reviews, meta-analyses, basic science/nonhuman studies, study protocols, editorials, those not reporting patient-centered outcomes, and those without an abstract. We included randomized control trials and prospective and retrospective observational (cohort and case control) studies in human subjects that used standardized HRQoL instruments (generic and disease specific). The eligible studies determined MID using anchor-based and/or distribution method.

2.2. Data extraction and validity assessment

Two reviewers independently screened the study titles, abstracts, selected full texts, and reference lists of the studies retrieved by the literature search. The methodological quality of the included studies was assessed using the Consensus-based Standards for the selection of health Measurement Instruments checklist [24,25]. Two reviewers assessed methodological quality independently and resolved disagreements by discussion. Studies that met eligibility criteria were grouped according to generic and disease-specific measures and different clinical treatment areas. We then assessed the MID or MCID calculation methods and developed tables to display instrument names, MID/MCID values, calculation methods, effect sizes, population size, and mean/median age of population for selected studies.

3. Results

Fig. 1 presents Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram for study selection [26]. We conducted the literature search between January 2016 and May 2016 and identified 2,207 studies that met our criteria. After excluding those not meeting inclusion criteria by abstract and title screening, we reviewed 132 articles in full text for determining eligibility. We included 80 studies which investigated the MID or MCID measurement properties of generic and disease-specific HRQoL instruments. As shown in Fig. 2, 33 of these studies used only anchor-based estimates for MID/MCID calculation, four used distribution-based methods, and 43 used both. More than half of studies using anchor-based calculations alone or combined with distribution-based estimates calculated MCID values. None of the studies reported MID values using only the distribution-based calculation method. One study separately calculated estimates for both MID

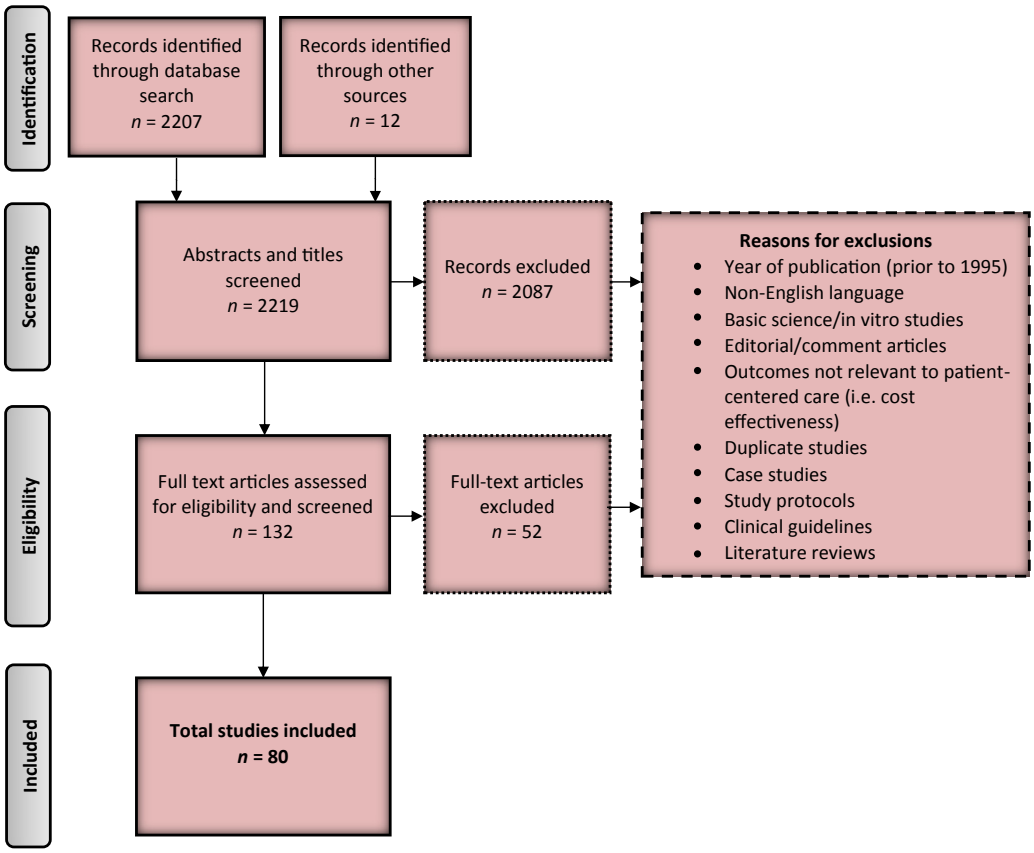

Fig. 1. Literature search flow diagram (October 1995 to April 2016).

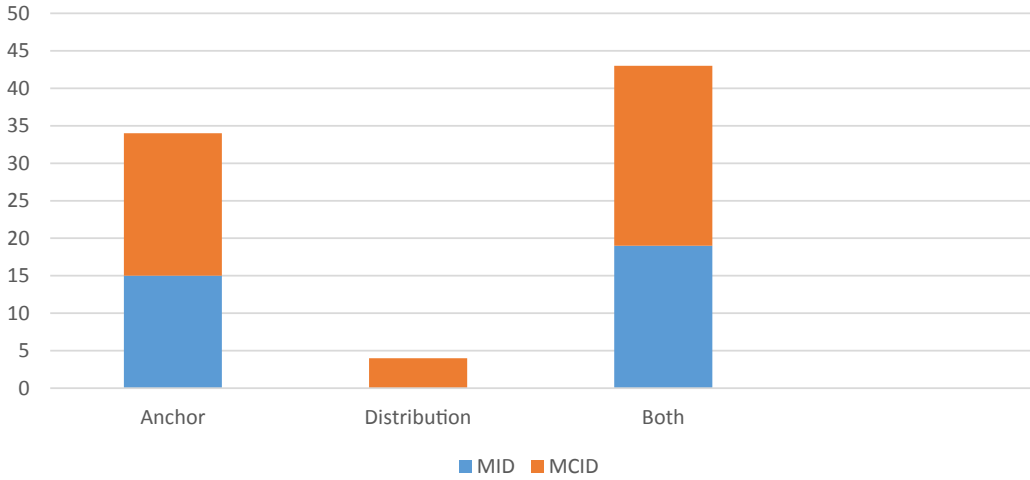

Fig. 2. Studies classified by MID/MCID calculation method. MCID, minimal clinically important difference; MID, minimal important difference.

and MCID and reported a combined value using an anchor. Table 1 presents a summary of disease condition, instruments used, and method of MID calculation. Similar information for MCID is presented in Table 2. Please see Appendix−eTable 1 at www.jclinepi.com for additional details.

3.1. Disease-specific health-related quality of life

eTable 2 (Appendix) at www.jclinepi.com summarizes the MID/MCID for disease-specific HRQoL instruments stratified by disease and references (see Appendix for abbreviations at www.jclinepi.com).

3.2. Oncology/Hematology

Among cancer instruments, the ESAS questionnaire for advanced stage patients in hospice care showed similar values for both anchor and multiple distribution-based estimates of MCID between 1.1 and 1.8, even within different subscales. The EORTC-QLQ-C30 is a cancer-specific quality-of-life questionnaire used in several studies. The MCID for esophageal cancer using EQoL was determined to be 0.5 across all domains. One study used two separate clinical anchors against which changes in selected scales of the EORTC-QLQ-C30 and QLQ-BN20 were calibrated. The MCID values for subscales were as follows: physical function = 6-9; role function = 14-12; cognitive functioning = 8; global health status = 7-4; fatigue = 12-9; motor dysfunction = 4-5; and communication deficit = 9-7. Using the World Health Organization Performance Status (WHO PS) as the anchor, MID value was found to be 8 for improvement and deterioration; however, Mini−Mental State Examination as an anchor showed corresponding values of 11 and 2, respectively. Using an anchor-based approach in breast cancer patients, the range for MCID was identified as 6.9−10.7. The MCID values for EORTC-QLQ-C30 and QLQ-BN20 ranged from 12.4 to 30.5. The subscale values were as follows: improvements: pain = 30.5, painful site = 20.1, painful characteristic = 30.5, functional interference = 19.6; decreases: emotional functioning = 12.4, global health status = 22.4, and financial issues = 13.5. In a study of advanced cancer, the MID for EORTC-QLQ-C30 ranged from 7.2 to 23.5. Specifically when assessing subgroups of the instrument, pain was observed to have the largest point value in most studies over physical or cognitive decline. Analyses of MCID using anchor and distribution methods for functional assessment cancer therapy−lung cancer (FACT-L) for advanced non−small-cell lung cancer patients demonstrated a 2- to 3-point difference on the lung cancer subscale and a 5 to 7-point difference on the trial outcomes index as clinically relevant change. Although these changes do not necessarily represent MID, they correspond to a moderate effect. Within the breast cancer population, four different studies using four different HRQoL instruments (EORTC-QLQ-BN20, functional assessment of breast cancer therapy [FACT-B], FACT-Cog, and WHOQoL-100) were identified. The MCID values for FACT-Cog were 6.9 using distribution-based method and 9.6 using anchor-based method. One study used a combination of anchor and distribution methods to estimate MID for FACT-B instrument. The MID estimate for overall FACT-B was 7−8 points, breast cancer subscales was 2−3 points, trial outcome index was 5−6 points, and FACT-G scale was 5−6 points, with a low−moderate effect size. The results reported convergence between distribution- and anchor-based estimates of MID. The EORTC-QLQ-BN20 among patients with advanced cancer brain metastasis had MCID values of deterioration of seizures = 6.1, weakness of legs = 13.8 (both using anchor methods), and deterioration of seizures = 6.2−7.7 and weakness of legs = 12.4−15.8 (using distribution methods). Similarly, results of a study for functional assessment of cancer therapy-colorectal (FACT-C) indicated an MID ranging from 1 to 2 points for the colorectal cancer subscale, 4 to 6 points

Table 1. Disease-specific and generic-specific minimal important difference

Disease condition	Instruments	%9. Anchor	Distribution
Oncology/hematology			
Breast cancer; laryngeal cancer	EORTC-QLQ-C30; FACT H&N; FACT-G	X	
Breast cancer; colorectal cancer; prostate cancer; cancer (advanced)	FACT-B; FACT-C; UCLA PCI; EORTC-QLQ-C30	X	X
Cardiovascular/neurovascular			
Heart disease	MacNew	X	
Flushing	FAST	X	X
Respiratory/ENT			
Chronic airflow limitation; dyspnea; COPD; chronic cough	SGRQ; TDI; VSRQ; LCQ	X	
Upper resp. infection; idiopathic pulmonary fibrosis; asthma; rhinitis; interstitial lung disease	WURSS-44, WURSS-21; SGRQ; ASUI; RCAT; K-BILD	X	X
Genitourinary			
Urinary disorders in MS	Qualiveen	X	
Overactive bladder	King's Health Questionnaire	X	X
Musculoskeletal/Rheumatology			
Upper-limb musculoskeletal disorders	QuickDASH	X	
Upper extremity problems for at least 3 months; low back pain; idiopathic scoliosis	DASH, SPADI, PRWE; RDMQ; SRS-22	X	X
Pediatric			
Allergic disease; acute otitis media	PADQLQ; AOM-SOS	X	
Child hydrocephalus	HOQ	X	X
Miscellaneous (Psychiatry/Endocrinology/Dermatology/Gastroenterology/Ophthalmology)			
GERD, dyspepsia, gastroparesis	PAGI-QOL	X	
Chronic idiopathic urticaria; eating disorders	DLQI; HeRQoLEDv2	X	X
Generic—HRQoL measures			
Spinal cord injury; systemic sclerosis; stroke	SF-6D; SF-6D and EQ-5D; EQ-5D and SF-6D	X	
IBS/leg ulcer/knee osteoarthritis/limb reconstruction/early rheumatoid arthritis/COPD; musculoskeletal upper extremity problems; leg ulcer/back pain/early rheumatoid arthritis/limb reconstruction/irritable bowel syndrome/acute myocardial infarction/osteoarthritis/COPD; COPD; idiopathic pulmonary fibrosis; prostate cancer	SF-6D; SF-12-PCS; SF-6D and EQ-5D; HADS; SF-36, SF-36	X	X

Abbreviations: COPD, chronic obstructive pulmonary disease; HRQoL, health-related quality of life; MS, multiple sclerosis; GERD, gastroesophageal reflux disease; IBS, irritable bowel syndrome.

See Appendix for abbreviations at www.jclinepi.com.

for the trial outcome index, and 5 to 8 points for total FACT-C score. This study strongly encouraged the use of multiple anchors to derive a range of MID. Similarly for prostate cancer, MID estimates derived from combining distribution- and anchor-based methods for the prostate cancer–specific HRQoL measures varied from 8 to 11 for subscale. The MID for Functional Assessment Therapy–Head and Neck scale indicated a score of 4–9% of the instrument range. This study used an anchor-based approach to derive one rule of thumb benchmark of 5–10% of the instrument range as a rough guide to derive MID. The MID for FACT H&N was 6.22 (+ve), and 12.4 (−ve). The MID for FACT-G was 4.37 (+ve) and 8.0 (−ve). Another study that used both anchor and distribution methods for an Immune Thrombocytopenic Purpura Patient Assessment Questionnaire showed moderate effect sizes. The MCID value ranged between 8 and 15. From analyses of FACT-Fatigue and FACT-Anemia data, MCID for five commonly aggregated summary scores was developed. The MCID for FACT-Fatigue was 3 and FACT-Anemia was 7.

3.3. Cardiovascular/Neurovascular

Among stroke population, five instruments were assessed (STREAM, BI, SIS-16, ABILHAND, and SIS) and all reported vastly different values from anchor- and distribution-based approaches. For patients with stroke, MCID values using STREAM scale ranged from 2.2 to 4.8. In another study of stroke patients, MCID score for BI was 1.85 using anchor-based method and 1.45 using distribution method. For stroke patients, MCID ranged from 9.4 to 14.1 for SIS-16 scale. MCID values for ABILHAND using distribution method was 0.26 and using anchor method was 0.35 in patients with stroke. Another study that used SIS scale in patients with stroke reported MCID values of 17.8. Two studies assessed MID values for the MacNew heart disease questionnaire.

Table 2. Disease-specific and generic-specific minimal important difference

Disease condition	Instrument	%9. Anchor	Distribution
Oncology/hematology			
Esophageal cancer	EQOL	X	
Cancer anemia and fatigue; lung cancer; immune thrombocytopenic purpura; brain cancer; cancer with bone metastasis; cancer advanced; cancer cognitive decline; cancer; advanced cancer with brain metastasis	FACT-An, FACT-G, FS, TOI-F, TOI-An; FACT-L, LCS, TOI; ITP-PAQ; EORTC-QLQ-C30, EORTC-QLQ-BN20; EORTC-QLQ-BM22, QLQ-C30; ESAS; FACT-Cog; ESAS; EORTC-QLQ-BN20	X	X
Cardiovascular/neurovascular			
Stroke; heart disease; stroke; stroke	STREAM; MacNew, EQ-5D; SIS-16, SIS	X	
Stroke; stroke	BI; ABILHAND	X	X
Respiratory/ENT			
Chronic rhinosinusitis	SNOT-22	X	
Chronic lung disease; COPD; asthma and allergic rhinitis	SOBQ; CCQ; CARAT	X	X
Genitourinary			
Incontinence	I-QOL	X	
Musculoskeletal/Rheumatology			
Tendinitis, arthritis, or nerve compression syndromes from forearm to hand; upper-limb musculoskeletal disorders; fatigue in systemic lupus erythematosus (SLE); low back pain; RA/fatigue; fibromyalgia; rheumatoid arthritis, CTS, DRS	DASH, QuickDASH, PRWE; DASH, QuickDASH; NRS, FSS, VT, MAF, CFS, FACIT-F, MFI-20; OD, NPRS; FSS, VT, MAF, MFI, FACIT-F, CFS, global RS; FIQ; MHQ	X	
Lower extremity musculoskeletal dysfunction; suspected internal derangement of knee	LEFS; KQoL-26	X	X
Carpal tunnel syndrome	SSS-CTS		X
Neurological			
Walking balance	FGA	X	
Vestibular schwannoma; Parkinson's disease; Parkinson's disease; dementia	PANQOL, SF-36; MDS-UPDRS; PDSS-2; NPI-Q	X	X
Pediatric			
Juvenile rheumatoid arthritis	CHAQ	X	X
Obesity/weight; diabetes (types 1 and 2); epilepsy (pediatric, chronic)	IWQOL-Kids; PedsQL; PESQ		X
Miscellaneous (Psychiatry/Endocrinology/Dermatology/Gastroenterology/Ophthalmology)			
Graves' ophthalmopathy; gastrointestinal QoL post-cholecystectomy; schizophrenia	GO-QOL; GIQLI; QWB, PANSS, QOLS Lenert PANSS	X	
Dyspepsia; schizophrenia	NDI; Heinrichs-Carpenter QoL	X	X
Generic—HRQoL measures			
Schizophrenia; fatigue in systemic lupus erythematosus	QWB; NRS, VT	X	
Early-stage breast cancer; cancer (advanced); vestibular schwannoma	WHOQOL-100; EORTC-QLQ-C30; SF-36	X	X

Abbreviations: COPD, chronic obstructive pulmonary disease; HRQoL, health-related quality of life.
See Appendix for abbreviations at www.jclinepi.com.

The first older study used a distribution approach to report a value of seven with a small to moderate effect size. The newer study had smaller effect size and reported MID value of 0.5 using the EQ-5D instrument as an anchor. Another study estimated MID for the flushing assessment tool for patients receiving increasing dosages of niacin therapy for hyperlipidemia. The range of mean MID values was 0.41–0.82 for high dose and 0.29–0.58 for low dose using combined anchor and distribution methods; the overall range was from 0.29 to 0.82.

3.4. Respiratory/ENT

Twelve separate studies estimated MID values for various respiratory conditions. The St. Georges respiratory questionnaire had MID of 3.05 using an anchor-based approach in chronic airflow obstruction, and a range from 5 to 8 points in idiopathic pulmonary fibrosis. Using the anchor-based approach for the physician's global evaluation scale, a one-unit change in the transition dyspnea index focal score was shown to represent MID. The shortness of breath questionnaire, however, had a larger MID value of 5 using a combination of methods. Of the two studies that looked at instruments for chronic obstructive pulmonary disease (COPD), the clinical COPD questionnaire (CCQ) and visual simplified respiratory questionnaire, both yielded very different MID values of 0.44 and 3.4, respectively. A study estimated MID for the two versions of WURSS questionnaire for upper respiratory function. The longer WURSS-44 had a larger value of 18.5 compared to 10.3 for WURSS-21. The only study in our review that

used the consensus method, in addition to an anchor-based approach, was for the Leicester cough questionnaire that found a combined MID value of 1.3. The K-BILD questionnaire for interstitial lung disease estimated lower MID value of 5.8 using distribution method, compared with anchor method of 8.2. Three studies that looked at questionnaires for population suffering from rhinitis (RCAT, SNOT-22, and CARAT) produced different MID values, using a combination of methods. The MID value for RCAT was 3, MCID for SNOT-22 was 8.9, and MCID for CARAT was 3.5. MID value for the asthma symptom utility index was estimated to be 0.09 using a combination of methods; however, this was only based on a one-point scale.

3.5. Genitourinary

Results from studies using King's Health Questionnaire, an overactive bladder disease-specific questionnaire, yielded similar values of MID, a score of five or higher. The MCID estimate for urinary incontinence quality of life (I-QoL) was be almost four points different when patients were receiving treatment (2.5), compared to between treatments (6.5). Using an anchor-based approach, MID for Qualiveen, a survey for urinary disorders in neurological conditions, was found to be 0.5.

3.6. Musculoskeletal/Rheumatology

In one study, the global MCID for subscale measures of fatigue for rheumatoid arthritis patients was reported as Fatigue Severity Scale (FSS), 20.2; Multidimensional Assessment of Fatigue (MAF), 18.7; Multidimensional Fatigue Inventory (MFI), 16.6; Functional Assessment of Chronic Illness Therapy Fatigue, 15.9; Chalder Fatigue Scale (CFS), 9.9; and Global RS, 17.7. However, the MCID of the same instrument in population of systemic lupus erythematosus showed significantly lower values: FSS was 2−7, MAF was 3.2, CFS was 1.4, Functional Assessment of Chronic Illness Therapy was 3.4, and MFI-20 was 2.9. One study reported two separate MCID scores based on if patients underwent surgical treatment for their knee condition or conservative medical management. Using KQoL-26, the MCID ranged between 3 and 24. Using a widely used anchor of SF-36 quality-of-life instrument, the MCID of the lower extremity functional scale was found to be nine and that of the Roland Morris disability questionnaire for low back pain was equivalent to a 30% reduction of the scores. For patients with upper extremity problems for at least 3 months, MID values using anchor method were 12.6 for DASH, 13.2 for SPADI, and 24 for PRWE. The values using distribution method were 5.35 for DASH, 7.75 for SPADI, and 5.22 for PRWE. Similarly, a percentage change from baseline score (of 14%) for the fibromyalgia impact scale instead of a numerical value was observed. The magnitude of effect size of the QuickDASH questionnaire showed that it was a responsive questionnaire, and the MID value was 19. Between three separate studies that estimated the MID of QuickDASH, the range was 14−19 points. In comparison, the longer and original questionnaire, DASH, showed a range of MID values of 10−12.6. MCID value among lower back pain patients for ODI was 10. In a study that estimated MCID for the Michigan Hand Outcomes Questionnaire for three separate conditions, pain and function of the hand were greatly improved after surgery for carpal tunnel syndrome and rheumatoid arthritis. However, high postoperative satisfaction a few months after surgery for patients with distal radius fracture prevented any domains from showing discriminative ability. The MCIDs were 23 for pain, 13 for function, 8 for work, 11 for pain, and 3 for daily living. Of the 13 studies of musculoskeletal or rheumatologic conditions, 9 used both anchor and distribution methods to estimate MID. The MID value was 0.6 for SRS-22 in patients with idiopathic scoliosis. The MCID value for SSS-CTS was 1.04 among patients with carpal tunnel syndrome.

3.7. Neurological

Five studies looked at populations with various neurological conditions. Two reported separate values for MCID and found that the distribution method yielded a smaller estimate compared to the anchor method. For patients with dementia, MCID values for NPI-Q was 3.18 for severity and 3.95 for distress (using anchor-based method), 2.77 for severity, and 3.10 for distress (using distribution method). MCID values for PANQoL was 11 using anchor-based method and 9 using distribution-based method. The MCID values for improvement in both Parkinson's disease−specific instruments, PDSS-2 and MDS-UPDRS, were similar at about 3.44 and 3.25, respectively. Using an anchor-based approach, MID of the functional gait assessment was found to be four.

3.8. Pediatric

Seven different studies estimated MID values for seven different conditions in pediatric population. The CHAQ score was relatively insensitive to important short-term changes in children with juvenile rheumatoid arthritis, as the MID values, 0.188 for improvement and 0.125 for worsening, were close to the smallest potential difference of 0.125. When analyzing the MCID scores for the PedsQL inventory for youth with type 1 and 2 diabetes, it was found that clinically meaningful (greater than or equal to 1 MCID) improvements in total score for at least one module of the questionnaire could be predicted by private insurance, lower BMI, and lower A1C at baseline. For type 1, MCID values for generic core were 4.88 (parent) and 4.72 (youth); MCID values for diabetes module were 4.54 (parent) and 5.27 (youth). For type 2, MCID values for generic core were 6.27 (parent) and 5.41 (youth);

MCID values for diabetes module were 6.06 (parent) and 5.96 (youth). The overall MID score for the PESQ survey for chronic pediatric epilepsy was found to be 3.25 using an SEM distribution method; however, there was variability in the MID values across the questionnaire subscales. An SEM distribution method estimated the MID of the IWQoL-kids questionnaire for childhood obesity to be 8.8 for physical comfort, 7.7 for body esteem, 8.1 for social life, 6.2 for family relations, and 4.8 for overall quality of life.

Using an anchor, MID for the pediatric allergic disease quality-of-life questionnaire was found to be 0.3 out of a seven-point scale and that of the acute otitis media severity of symptoms scale was estimated to be 3.8 of 100 points. The MID for HOQ, a questionnaire for children already treated for hydrocephalus, was found to be 0.10−0.12 using an anchor approach and 0.05 on a one-point scale using the SEM distribution method. This study found that the predictors associated with a worse overall quality of life appeared to be related to shunt complications, which might be modifiable.

3.9. Miscellaneous (Psychiatry/Endocrinology/Dermatology/Gastroenterology/Ophthalmology)

One study found the MCID range for the GO-QoL index for Graves' ophthalmopathy to be between 6 and 10, with increasing values for more invasive therapies. Three questionnaires that assessed gastrointestinal conditions, PAGI-QOL, NDI, GIQLI, reported very different MID values using a combination of methods. For the PAGI-QOL instrument, the MID value was 0.36 and MCID value was 0.47. The MCID value for NDI was 10 and that for GIQLI ranged from 6.42 to 7.64. The MID and MCID values for the HeRQoLEDv2 questionnaire for eating disorders had MID range of 0.7 in the physical role to 14.5 in the emotional role, using combined anchor and distribution method. The relationship between minimal detectable change (MDC) and MID for this questionnaire indicated that MDC was larger than MID. One year after initial assessment, patients with an eating disorder reported significant improvement and effect size of above 0.30. The MID for dermatology life quality index for patients with chronic idiopathic urticaria was 2.24 to 3.10 using distribution method and 2.97 to 3.21 using anchor-based method. Two studies assessed MCID for patients with schizophrenia. The MCID values for the Heinrichs-Carpenter QoL questionnaire were 5.30 using anchor-based approach and ranged from 3.37 to 6.61 using distribution-based method. The MCID value for Lenert PANSS was 0.15, for QOLS was 1.13, and for PANSS was 2.2.

3.10. Generic health−related quality of life

eTable 3 (Appendix) at www.jclinepi.com summarizes the MID/MCID estimates for generic HRQoL for various diseases and references. MID values for SF-6D were 0.051 using distribution method and 0.01 to 0.48 using anchor-based method. For SF-12, the MID values were 6.8 using anchor-based method, and 4.22 using distribution method. Comparison of MID for SF-6D and EQ-5D using distribution- and anchor-based methods revealed that MID for EQ-5D (mean = 0.074) had approximately twice the value for SF-6D (mean = 0.041) scale. This was evident in both systemic sclerosis and stroke populations that estimated comparable MID values for both these instruments. Among stroke patients, the MID values for EQ-5D ranged from 0.08 to 0.12, and for SF-6D, from 0.04 to 0.14. The MID values for SF-6D were 0.05 for improvement and −0.04 for deterioration. MID for EQ-5D was 0.08 for improvement and −0.13 for deterioration. The MCID value for QWB among patients with schizophrenia was 0.17. Thus, the MID estimates for each scale appear to be proportionally equivalent in the context of range of utility score for each scale. One study used anchor-based approach to derive MID for the SF-6D health utility scale for patients with spinal cord injuries. Among those who reported being somewhat worse or better at follow-up, MID was 0.03 (SD = 0.17) and it was 0.10 (SD = 0.14) for those who were only somewhat worse. The vitality subscale for rheumatoid arthritis had a higher MCID value of 14.8 compared to systemic lupus erythematosus with a value of 3.1. The World Health Organization quality-of-life questionnaire (WHOQOL-100) was assessed in early-stage breast cancer population. A combined overall score of one point was calculated using both anchor and distribution method. Using anchor-based method, MCID for decline ranged from −1.56 to −0.71; and using distribution method, it ranged from 0.64 to 0.94. The MID of the Hospital Anxiety and Depression Scale was 1.5 for patients with chronic obstructive pulmonary disease, corresponding to a 20% change from baseline values. Estimates of MID for SF-36 in pulmonary fibrosis patients using anchor and distribution methods ranged from 2 to 4 points. Similarly, for prostate cancer, MID estimates were derived from combining distribution- and anchor-based methods for the generic HRQoL instrument of SF-36 and ranged from 6 for mental health to 14 for role physical. In patients diagnosed with vestibular schwannoma, as anchor-based approach produced an MID value of seven and distribution method estimated it to be five.

4. Discussion

Our synthesis provides a comprehensive assessment of MID and MCID for 10 generic and 80 disease-specific instruments. We observed that for any instrument that measures patient-reported outcomes, there is no single measure of MID or MCID. The MID estimate will depend on the context of disease, disease severity, characteristics of population in the study, unit of interest (whether an individual or a group), the baseline values observed, and the

change in values. Therefore, it is important to use caution while interpreting MID and consider the available estimates for a particular instrument. This will help to integrate the MID estimates into the overall research or clinical plan that is context specific [15].

Most of the studies in our review used anchor-based methods, either alone or in combination with the distribution methods. We also noted that the distribution method yielded smaller estimates than the anchor-based method, suggesting that a smaller change is needed to provide a clinically important difference [27]. Smaller MID estimates suggest that treatment is beneficial to these patients, when it may not be so. Distribution-based approaches rely on relating the difference between treatment and control groups to some measure of variability [28]. In addition, the distribution method calculations mostly used the SEM, SD, or effect size approaches. However, as distribution methods are derived from statistical analysis of the population and not linked to actual clinical outcomes, they should only be used when anchor-based calculations are unavailable. Anchor-based approaches compare the change in a PRO to a second, external, and more clearly understood measure of change, that is, an anchor. Although it is agreed that anchors must be easily interpretable, widely used and at least moderately correlated with the instrument being explored [18,29,30], there is no agreement regarding appropriate anchors. The anchor-based approach is also vulnerable to recall bias, and as was evident in our review, different anchors may produce widely different estimates for the same instrument [18,29,31]. Sociodemographic characteristics of patients can also impact report of change [32]. Although some studies indicate convergence of MID estimates derived from anchor-based and distribution methods [33], there are no guidelines for action during nonconvergence. Thus, in the absence of established ''gold standard'' methodology for achieving meaningful estimates of MID, interpretation of PRO scores is subject to disagreement [18,19,31,33–42].

The MID is used for interpreting therapeutic changes in generic as well as disease-specific PRO. In our review, we observed that in general, the disease-specific instruments were more responsive to change than the generic ones. Generic instruments cover common domains of health so as to be relevant to the general population, whereas the focus of the condition-specific instruments is on the domains relating to that condition [43]. Moreover, as a generic instrument measures the generic health status, its MID should not vary across different populations and contexts. Therefore, disease-specific instruments are expected to be more responsive in detecting change that is directly related to the condition.

An important issue in the interpretability of PRO is whether one makes inferences with respect to individuals or populations. A societal perspective takes into account the degree of importance at a population level, where a small difference may be important because of the large number of individuals who may be affected [4,20,44,45]. The societal perspective is interested in the degree of change required between different strategies to make adjustments in health care policies [16]. On the other hand, an individual perspective takes into account the degree of importance at an individual level, where a large difference may be needed to interpret the change as clinically important.

Another issue is the degree of change required to stimulate clinicians to consider an intervention. The MID is an important patient-centered concept as it captures both the magnitude of improvement and the value patients place on the change. The degree of improvement and value may not take into account long-term outcomes such as functional and survival, or physiologic findings that may not be symptomatic to the patient. However, this improvement may not be as evident to a clinician who is selecting treatment. Thus, it is suggested that for HRQoL measures, responsiveness should be based on the patient's perception of meaningful change, whereas for the measures of disease activity, a clinician can provide appropriate judgment [16,46]. In this regards, several multidisciplinary entities have developed a core set of disease-specific outcome domains for PROs [16]. However, if these outcomes measures must be viewed as a composite index score or individually to aid in treatment decisions is unclear [16]. Determination of the MID must also consider different thresholds for different subsets of population. For example, compared with patients with little pain at baseline, those with substantial pain may need larger pain reduction to perceive the treatment as beneficial.

The MID can also be used to relate changes in health status to changes in more established clinical measures if both are available, or it can be used for comparing outcomes from different studies. In the era of comparative effectiveness, health status measures are assuming an important role in evaluation of PROs. A side effect or symptom that is most important to one patient might not be so to another. Thus, a major challenge in the PRO field is the measurement and interpretation of MID. Comparative effectiveness can facilitate choosing meaningful threshold and yield data on proportion of patients achieving a small but important benefit or those reporting poorer outcomes.

We acknowledge following limitations of this review: (1) In our review, MID and MCID were combined. Although structurally and conceptually different, both MID and MCID help us in interpreting changes resulting from interventions. Our objective was to assess important concordance and discordance between these two measures as it applies to patient-centered care. (2) Factors that may affect MCID scores are specific to the study population and are nontransferable across patient groups. Although we compared the same instruments between disease groups, we acknowledge that this limitation makes it difficult to elicit a single, standard MID/MCID value for an instrument. (3) Determining if studies were estimating MID or MCID values was a challenge, as many studies used these terms interchangeably. (4) When MID or MCID scores are reported as a single-point estimate, there is absence of confidence intervals that is

essential to represent the distribution of changes. Therefore, studies that use a single-point estimate of MID or MCID may wrongly classify scores below the mean as not improved when, in fact, they might have [47]. (5) Finally, patients with the same conditions may not have homogeneous disease states, stages, and symptoms, and thus, a single validated MID or MCID score may be inaccurate.

Patient-reported outcome measures are the cornerstone of patient-centered care. For clinicians as well as researchers, it is crucial that MID value is a valid and stable measure. A low MID value may overestimate the positive effects of the treatment, whereas a high MID value may misclassify patients as not responsive to treatment when in fact it was helpful [47]. The variation in the methods used makes it difficult to interpret the results and apply it to clinical practice. Clinicians also face barriers to use of questionnaires as a way of documenting treatment outcomes [28]. Barriers include separate questionnaires corresponding to the different conditions encountered in a clinic, and the assessment and interpretation of the scores [48]. In addition, the instruments used may be unfamiliar to clinicians and patients [28,49]. The score in an individual patient may not reflect a true change, and due to measurement error, these scores cannot be used for individual diagnosis [50,51]. When a change equal or above the estimated MID for the instrument is detected, it can signal the clinician to begin a comprehensive dialogue with the patient about disease symptoms [50,51].

In conclusion, currently there exists limited information regarding established MID across the vast number of generic and disease-specific PRO measures. Thus, there is an urgent need for assessments of MIDs across generic and disease-specific instruments for various disease and population settings to enrich the knowledgebase in the context of patient-centered care.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jclinepi.2017.06.009.

References

[1] Barrett B, Brown D, Mundt M, Brown R. Sufficiently important difference: expanding the framework of clinical significance. Med Decis Making 2005;25:250−61.

[2] Crosby RD, Kolotkin RL, Williams R. Defining clinically meaningful change in health related quality of life. J Clin Epidemiol 2003;56:395−407.

[3] Jeschke R, Singer J, Guyatt GH. Measurement of health status-ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407−15.

[4] Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in disease-specific quality of life questionnaire. J Clin Epidemiol 1994;47:81−7.

[5] Norman GR, Sridhar FG, Guyatt GH, Walter SD. Relation of distribution-and anchor based approaches in interpretation of changes in health related quality of life. Med Care 2001;39:1039−47.

[6] van Walraven C, Mahon JL, Moher D, Bohm C, Laupacis A. Surveying physicians to determine the minimal important difference: implications for sample-size calculation. J Clin Epidemiol 1999;52:717−23.

[7] Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimal important differences for patient-reported outcomes. J Clin Epidemiol 2008;68:102−9.

[8] Lipscomb J, Gotay CC, Snyder CF. Patient-reported outcomes in cancer: a review of recent research and policy initiatives. CA Cancer J Clin 2007;52:278−300.

[9] Valderas JM, Kotzeva A, Espallargus M, Guyatt G, Ferrans CE, Halyard MY, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. Qual Life Res 2008;17:179−93.

[10] Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. Curr Opin Rheumatol 2002;14:109−14.

[11] Coeytaux RR, Kaufman JS, Chao R, Mann JD, DeVellis RF. Four methods of estimating the minimal important difference score were compared to establish a clinically significant change in Headache Impact Test. J Clin Epidemiol 2006;59:374−80.

[12] Jordan K, Dunn KM, Lewis M, Croft P. A minimal clinically important difference was derived for the Roland-Morris disability questionnaire for low back pain. J Clin Epidemiol 2006;59:45−51.

[13] Redelmeier DA, Guyatt GH, Goldstein RS. Assessing the minimal important difference in symptoms: a comparison of two techniques. J Clin Epidemiol 1996;49:1215−9.

[14] Wyrwich KW, Tierney WM, Babu AN, Kroenke K, Wolinsky FD. A comparison of clinically important differences in health related quality of life for patients with chronic lung disease, asthma, or heart disease. Health Serv Res 2005;40:577−97.

[15] King MT. A point of minimal important difference (MID): a critique of terminology and methods. Expert Rev Pharmacoecon Outcomes Res 2011;11(2):171−84.

[16] Rai SK, Wazdany J, Fortin PR, Avina-Zubieta JA. Approaches for estimating minimal clinically important difference in systemic lupus erythematosus. Arthritis Res Ther 2015;17(1):143.

[17] McGlothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients. JAMA 2014;312:1342−3.

[18] Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. Mayo Clin Proc 2002;77:371−83.

[19] Wyrwich KW, Tirney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health related quality of life. J Clin Epidemiol 1999;52:861−73.

[20] 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;7(5):541−6.

[21] Cohen J. Statistical power analysis for the behavioral sciences. Hillsadale, NJ: Lawrence Erlbaum Associated; 1988.

[22] Anastasi A, Urbina S. Psychological testing. 7tn ed. Upper Saddle River, NJ: Prentice-Hall; 1997.

[23] Bonniaud V, Bryant D, Parratte B, Guyatt G. Qualiveen, a urinary-disorder specific instrument: 0.5 corresponds to the minimal important difference. J Clin Epidemiol 2008;61:505−10.

[24] Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012;21:651−7.

[25] Mokkink LB, Terwee CB, Patrick DL, Alonso G, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties

for health-related patient-reported outcomes. J Clin Epidemiol 2010; 63:737−45.
[26] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009;62:1006−12.
[27] Horvath K, Aschermann Z, Acs P, Deli G, Janszky J, Komoly S, et al. Minimal clinically important difference on the motor examination part of MDS-UPDRS. Parkinsonism Relat Disord 2015;21:1421−6.
[28] 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.
[29] Gyatt G, Schunemann H. How can quality of life researchers make their work more useful to health workers and their patients? Qual Life Res 2007;16:1097−105.
[30] Maringwa J, Quinten C, King M, Rinjash J, Osoba D, Coens C, et al. Minimal clinically meaningful differences for the EORTC QLQ-C30 and EORTC QLQ-BN20 scales in brain cancer patients. Ann Oncol 2011;22:2107−12.
[31] Osoba D, Rodrigues G, Myles J, Zee B, Poter J. Interpreting the significance of changes in health-related-quality of life score. J Clin Oncol 1998;16:139−44.
[32] Lauridsen HH, Hartvigsen J, Manniche C, Korsholm L, Nilsson G. Responsiveness and minimal clinically important difference for pain and disability instruments in low back pain patients. BMC Musculoskelet Discord 2006;7:82.
[33] Eton DT, Cella D, Yost KJ, Yount SE, Peterman AH, Neuberg DS, et al. A combination of distribution-and anchor based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale. J Clin Epidemiol 2004;57:898−910.
[34] Barrett B, Brown R, Mundt M. Comparison of anchor-based and distributional approaches in estimating important difference in common cold. Qual Life Res 2008;17:75−85.
[35] Chen RC, Clark JA, Talcott JA. Individualizing quality-of-life outcomes reporting: how localized prostate cancer treatments affect patients with different levels of baseline urinary, bowel, and sexual function. J Clin Oncol 2009;27:1−13.
[36] Jaeschke R, Guyatt GH, Keller J, Singer J. Interpreting changes in quality-of-life score in N or 1 randomized trials. Control Clin Trials 1991;12:226S−33S.
[37] Kulkarni AV. Distribution and anchor-based approaches provided different interpretability estimates for the hydrocephalus outcome questionnaire. J Clin Epidemiol 2006;59:176−84.
[38] Lemieux J, Beaton DE, Hogg-Johnson S, Bordeleau LJ, Goodwin PJ. Three methods for minimally important difference: no relationship was found with the net proportion of patients improving. J Clin Epidemiol 2007;60:448−55.
[39] Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health related quality of life-the remarkable universality of half a standard deviation. Med Care 2003;41:582−92.
[40] 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.
[41] Patrick DL, Chiang YP. Measurement of health outcomes in treatment effectiveness evaluations. Med Care 2000;38. II-14−II-25.
[42] Schmitt JS, Fabio Di RP. Reliable change minimum important difference proportions facilitated group responsiveness comparisons using individual threshold criteria. J Clin Epidemiol 2004;57: 1008−18.
[43] Chuang LH, Garratt A, Brealey S. Comparative responsiveness and minimal change of the knee quality of life 26-item (KQoL-26) Questionnaire. Qual Life Res 2013;22:2461−75.
[44] Kazis LE, Anderson SJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27:S178−89.
[45] Wyrwich KW. Minimal important difference threshold and the standard error of measurement: is there a connection? J Biopharm Stat 2004;14(1):97−110.
[46] Liang MH. Longitudinal constuct validity: establishment of clinical meaning in patient evaluative instrument. Med Care 2000;38. II-84−II-90.
[47] 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.
[48] Polson K, Reid D, McNair PJ, Larmer P. Responsiveness, minimal important difference and minimal detectable change scores of the shortened disability arm shoulder hand (QuickDASH) questionnaire. Man Ther 2013;15:404−7.
[49] Fayers PM, Machin D. Quality of life: assessment, analysis and interpretation. Chichester: Wiley; 2000.
[50] Cella D, Eton DT, Fairclough DL, Bonomi P, Heyes AE, Silberman C, et al. What is a clinically meaningful change on the Functional Assessment of Cancer Therapy-Lung (FACT-L) Questionnaire? Results from Eastern Cooperative Oncology Group (ECOG) Study 5592. J Clin Epidemiol 2002;55:285−95.
[51] Cella D, Eton DT, Lai J, Peterman AH, Merkel DE. Combining anchor and distribution-based methods to derive clinically important differences on the functional assessment of cancer therapy (FACT) anemia and fatigue scales. J Pain Symptom Manage 2002;24: 547−61.