

BMJ Open Minimally important difference estimates and methods: a protocol

Bradley C Johnston,^{1,2,3,4} Shanil Ebrahim,^{1,4,5,6,7} Alonso Carrasco-Labra,^{4,8} Toshi A Furukawa,⁹ Donald L Patrick,¹⁰ Mark W Crawford,¹ Brenda R Hemmelgarn,¹¹ Holger J Schunemann,^{4,12} Gordon H Guyatt,^{4,12} Gihad Nesrallah^{13,14,15}

To cite: Johnston BC, Ebrahim S, Carrasco-Labra A, *et al.* Minimally important difference estimates and methods: a protocol. *BMJ Open* 2015;**5**:e007953. doi:10.1136/bmjopen-2015-007953

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2015-007953>).

Received 10 April 2015

Revised 21 July 2015

Accepted 31 July 2015



CrossMark

For numbered affiliations see end of article.

Correspondence to Dr Gihad Nesrallah; gnesrallah@hrh.ca

ABSTRACT

Introduction: Patient-reported outcomes (PROs) are often the outcomes of greatest importance to patients. The minimally important difference (MID) provides a measure of the smallest change in the PRO that patients perceive as important. An anchor-based approach is the most appropriate method for MID determination. No study or database currently exists that provides all anchor-based MIDs associated with PRO instruments; nor are there any accepted standards for appraising the credibility of MID estimates. Our objectives are to complete a systematic survey of the literature to collect and characterise published anchor-based MIDs associated with PRO instruments used in evaluating the effects of interventions on chronic medical and psychiatric conditions and to assess their credibility.

Methods and analysis: We will search MEDLINE, EMBASE and PsycINFO (1989 to present) to identify studies addressing methods to estimate anchor-based MIDs of target PRO instruments or reporting empirical ascertainment of anchor-based MIDs. Teams of two reviewers will screen titles and abstracts, review full texts of citations, and extract relevant data. On the basis of findings from studies addressing methods to estimate anchor-based MIDs, we will summarise the available methods and develop an instrument addressing the credibility of empirically ascertained MIDs. We will evaluate the credibility of all studies reporting on the empirical ascertainment of anchor-based MIDs using the credibility instrument, and assess the instrument's inter-rater reliability. We will separately present reports for adult and paediatric populations.

Ethics and dissemination: No research ethics approval was required as we will be using aggregate data from published studies. Our work will summarise anchor-based methods available to establish MIDs, provide an instrument to assess the credibility of available MIDs, determine the reliability of that instrument, and provide a comprehensive compendium of published anchor-based MIDs associated with PRO instruments which will help improve the interpretability of outcome effects in systematic reviews and practice guidelines.

BACKGROUND

For decades, longevity and major morbid events (eg, stroke, myocardial infarction)

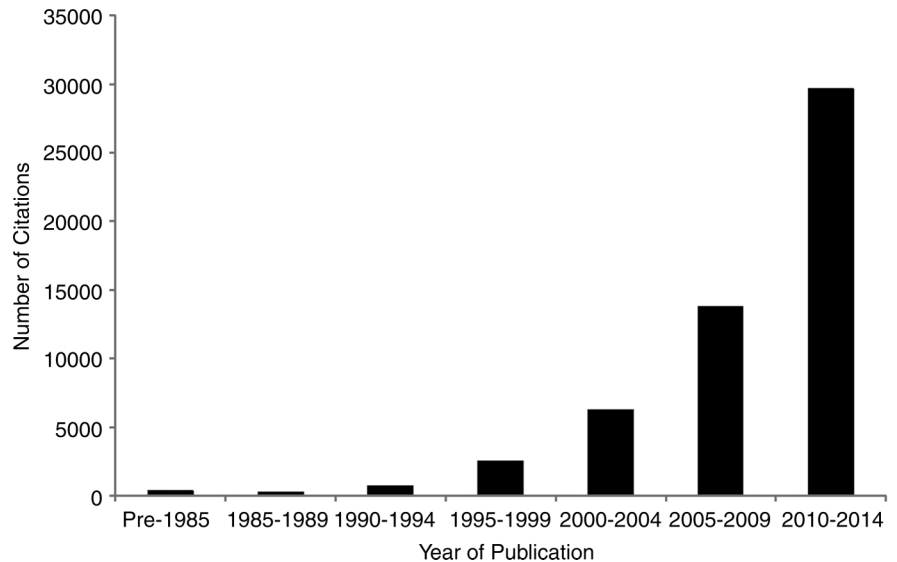
have been a primary focus in health research. Increasingly, investigators and clinicians have acknowledged the critical role of disease and treatment-related symptoms, as well as the function and perceptions of well-being for informed clinical decision-making. Typically measured by direct patient inquiry, these outcomes, previously generally referred to as 'quality of life' or 'health-related quality of life' measures, are now most commonly referred to as patient-reported outcomes (PROs). PROs provide patients' perspectives on treatment benefits and harms, and are often the outcomes of most importance to patients.

The PRO literature has grown exponentially over the past three decades ([figure 1](#)), with several instruments (eg, Short-Form-36,¹ Beck Depression Inventory²) in routine use in research and clinical practice. The number of clinical trials evaluating the impact of interventions on PROs has also steadily increased ([figure 2](#)), and PROs are increasingly considered in practice guidelines ([figure 3](#)).

Although PROs are often measured as primary outcomes in clinical trials, challenges remain in their application. In addition, although evidence supporting reliability, validity and responsiveness exists for many PRO instruments, interpretation of their results remains a challenge.

Interpretability has to do with understanding the changes in instrument scores that constitute trivial, small but important, moderate or large differences in effect. For instance, if a treatment improves a PRO score by three points relative to control, what are we to conclude? Is the treatment effect large, warranting widespread dissemination in clinical practice, or is it trivial, suggesting that the new treatment should be abandoned? Recognition of this potentially serious limitation has led to increasing interest in the interpretation of treatment effects on PROs.^{3 4}

Figure 1 Number of citations found in PubMed with search terms of patient reported outcome, by 5-year stratum.



The minimally important difference (MID) provides a measure of the smallest change in the PRO of interest that patients perceive as important, either beneficial or harmful, and that would lead the patient or clinician to consider a change in management.⁵ Knowledge of the MID allows decision-makers to better interpret the magnitude of treatment effect and assess the trade-off between beneficial and harmful outcomes. Patients, clinicians and clinical practice guideline developers require knowledge of the MID to guide their decisions. For example, a guideline developer using the GRADE approach⁶ might consider the MID as a decision threshold for determining if the quality of evidence for a given intervention should be rated down (for imprecision) if the CI surrounding a pooled effect estimate includes the MID,⁷ or if it is sufficiently precise, lying well above the MID threshold. The MID also provides a metric for clinical trialists planning sample sizes for their studies. This is accomplished by first calculating the proportion of patients achieving an MID or greater change, and subsequently determining the difference in the proportion of

responders that trialists would like to examine between the treatment and control that would constitute a clinically important difference. The widespread recognition of the usefulness of the MID is reflected in the exponential growth in the number of citations reporting MID since the concept was first introduced into the medical literature in 1989 (figure 4).⁴

There are two primary approaches for estimating an MID: distribution-based and anchor-based methods. Distribution-based methods rely on the distribution around the mean scores of the measure of interest (eg, SD).⁸ In the anchor-based approach, investigators examine the relation between the target PRO instrument and an independent measure that is itself interpretable—the anchor. An appropriate anchor will be relevant to patients (eg, measures of symptoms, disease severity or response to treatment).⁹ Investigators often use global ratings of change (patients classifying themselves as unchanged, or experiencing small, moderate and large improvement or deterioration) as an anchor. It is generally agreed that the patient-reported anchor-based

Figure 2 Number of citations found in PubMed with the search terms of patient reported outcome limited to clinical trials, by 5-year stratum.

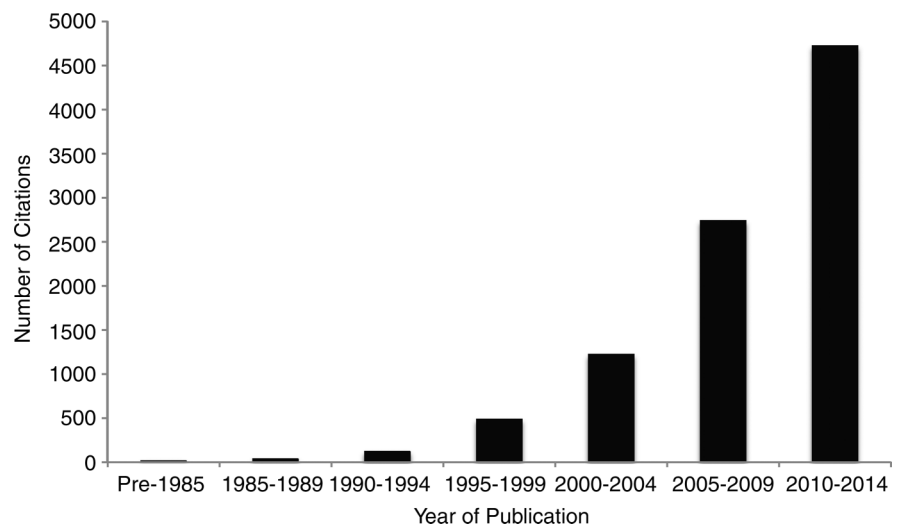
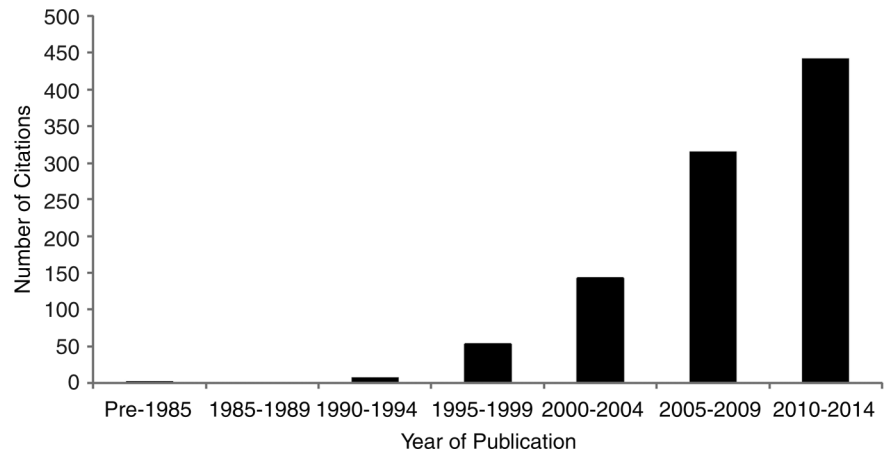


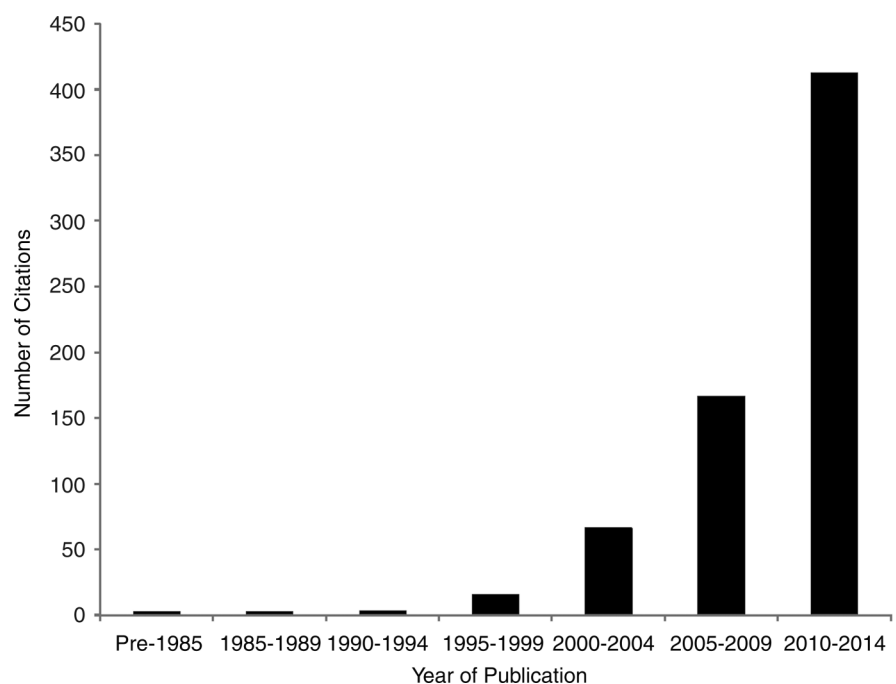
Figure 3 Number of citations found in PubMed with search terms of patient reported outcome and practice guidelines, by 5-year strata.



approach is the optimal way to determine the MID because it directly captures the patients' preferences and values,^{8 10} although it can still be problematic if the credibility of the anchor is in question (eg, is the anchor itself interpretable and are responses on the anchor independent of responses on the PRO?).

Currently, there is no study or database that systematically documents all available anchor-based MIDs associated with PRO instruments. In addition, there are currently no accepted standards for appraising the credibility of an MID determination; incorrect methods, assumptions or interpretation of MID could be detrimental to otherwise well-designed clinical trials, systematic reviews and guidelines. For example, systematic review authors and guideline developers might interpret trials using PROs incorrectly, and provide misleading guidance to patients and clinicians.^{11 12} In addition, erroneous MIDs may lead to inappropriate sample size calculations.

Figure 4 Number of citations found in PubMed with search terms of minimal (clinically) important difference, by 5-year stratum.



Our objectives are therefore to:

1. Summarise the anchor-based methods that investigators have used to estimate MIDs and the criteria thus far suggested to conduct studies optimally (henceforth referred to as a 'systematic survey addressing the anchor-based methods used to estimate MIDs').
2. Develop an instrument for evaluating the credibility of MIDs that emerge from identified PRO instruments.
3. Document published anchor-based MIDs associated with PRO instruments used in evaluating the effects of interventions on chronic medical and psychiatric conditions in adult and paediatric populations (henceforth referred to as a 'systematic survey of the inventory of published anchor-based MIDs').
4. Apply the credibility criteria (developed in objective 2) to each of the MID estimates that emerge from our synthesis in objective 3.
5. Determine the reliability of our credibility instrument (developed in objective 2).

To ensure the feasibility and a manageable scope of the project and given that the primary use of PROs is in the management of chronic medical and psychiatric conditions,¹³ we have restricted our focus to these clinical areas.

METHODS

Eligibility criteria

We define an anchor-based approach as any independent assessment to which the PRO instrument is compared, irrespective of the interpretability or the quality of the anchor. We will include two types of publications: (1) Methods articles addressing MID estimation using an anchor-based approach (theoretical descriptions, summaries, commentaries, critiques). We will include only studies that dedicate a minimum of two paragraphs to discuss methodological issues when estimating MIDs from both of these types of studies. (2) Original reports of studies that document the empirical development of an anchor-based MID for a particular instrument in a wide range of chronic medical and psychiatric conditions. That is, studies will compare the results of a PRO instrument to an independent standard (the 'anchor'), irrespective of the interpretability or the quality of the anchor. We will include adult (≥ 18 years of age) and paediatric populations (< 18 years of age). PROs of interest will include self-reported patient-important outcomes of health-related quality of life, functional ability, symptom severity and measures of psychological distress and well-being.

In our definition of a PRO, if the patient is incapable of responding to the instrument and a proxy is used, then the study is still eligible. If the clinician completes only the anchor, then it is still eligible. For anchor-based MIDs identified in the paediatric population for children under the age of 13, we will include both patient-reported and caregiver-reported instruments.

We will exclude studies when only the clinician completes the PRO instrument (ineligible proxy), as well as studies reporting only distribution-based MIDs without an accompanying anchor-based MID.

Information sources and search

We will search MEDLINE, EMBASE and PsycINFO for studies published from 1989 to the present (the MID concept was first introduced into the medical literature in 1989).³ Search terms will include database subject headings and text words for the concepts: 'minimal important difference', 'minimal clinical important difference', 'clinically important difference', 'minimal important change', alone and in combination, and adapted for each of the chosen databases. Table 1 presents the MEDLINE search strategy. One of the co-authors (DLP) works closely with MAPI Trust, which maintains PROQOLID—a large repository of commonly used, well-validated PRO instruments.¹³ To supplement our search, the investigator has granted us access to the PROQOLID

internal library, which houses approximately 180 citations related to MID estimates. We will, in addition, search the citation lists of included studies and collect narrative and systematic reviews for original studies that report an MID for a given instrument.

Study selection

Teams of two reviewers will independently screen titles and abstracts to identify potentially eligible citations. To determine eligibility, the same reviewers will review the full texts of citations flagged as potentially eligible.

Data collection, items and extraction

Teams of data extractors will, independently and in duplicate, extract data using two pilot-tested data collection forms: one for the systematic survey addressing the anchor-based methods used to estimate MIDs, and the second for the systematic survey of the inventory of published anchor-based MIDs. On study initiation, we will conduct calibration exercises until sufficient agreement is achieved. Our data collection forms will include the following items: study design, description of population, interventions, outcomes, characteristics of candidate instruments (eg, generic or disease specific), characteristics of independent anchor measures, critiques/commentary on methods to estimate MID(s), and credibility criteria for method(s) to estimate MIDs.

For the extraction of anchor-based methods studies used to estimate MIDs, a team of methodologists familiar with MID methods will use standard thematic analysis techniques¹⁴ to abstract concepts related to the methodological quality of MID determinations until reaching saturation. We will review coding and revise the taxonomy of methodological factors iteratively until informational redundancy is achieved. Appendix A presents the data extraction items for the systematic survey of the inventory of published anchor-based MIDs. Reviewers will resolve disagreements by discussion and, if needed, a third team member will serve as an adjudicator.

Subsequently, we will synthesise and complete each objective as follows:

OBJECTIVE 1: METHODS TO DEVELOP ANCHOR-BASED METHODS USED TO ESTIMATE MIDS

We will classify the anchor-based approaches used to determine MIDs into separate categories, describe their methods, and summarise their advantages, disadvantages and important factors that constitute a high-quality anchor.

OBJECTIVE 2: DEVELOPMENT OF A CREDIBILITY INSTRUMENT FOR STUDIES DETERMINING MIDS

We define credibility as the extent to which the design and conduct of studies measuring MIDs are likely to have protected against misleading estimates. Similar definitions have been used for instruments measuring the credibility of other types of study designs.^{15 16}

Table 1 Search strategies for MEDLINE, January 1989 to present

1	(clinical* important difference? or clinical* meaningful difference? or clinical* meaningful improvement? or clinical* relevant mean difference? or clinical* significant change? or clinical* significant difference? or clinical* important improvement? or clinical* meaningful change? or mcid or minim* clinical* important or minim* clinical* detectable or minim* clinical* significant or minim* detectable difference? or minim* important change? or minim* important difference? or smallest real difference? or subjectively significant difference?).tw.
2	"Quality of Life"/
3	"outcome assessment(health care)"/or treatment outcome/or treatment failure/
4	exp pain/
5	exp disease attributes/or exp "signs and symptoms"/
6	or/2–5
7	1 and 6
8	health status indicators/or "severity of illness index"/or sickness impact profile/or interviews as topic/or questionnaires/ or self report/
9	Pain Measurement/
10	patient satisfaction/or patient preference/
11	or/8–10
12	7 and 11
13	limit 12 to yr="1989 -Current"
14	(quality of life or life qualit??? or hrqol or hrql).mp.
15	(assessment? outcome? or measure? outcome? or outcome? studies or outcome? study or outcome? assessment? or outcome? management or outcome? measure* or outcome? research or patient? outcome? or research outcome? or studies outcome? or study outcome? or therap* outcome? or treatment outcome? or treatment failure?).mp.
16	pain?????.mp.
17	((activity or sever* or course) adj3 (disease or disabilit* or symptom*)).mp.
18	or/14–17
19	1 and 18
20	(questionnaire? or instrument? or interview? or inventor* or test??? or scale? or subscale? or survey? or index?? or indices or form? or score? or measurement?).mp.
21	(patient? rating? or subject* report? or subject* rating? or self report* or self evaluation? or self appraisal? or self assess* or self rating? or self rated).mp.
22	(patient? report* or patient? observ* or patient? satisf*).mp.
23	anchor base???.mp.
24	or/20–23
25	19 and 24
26	limit 25 to yr="1989 -Current"
27	13 or 26

The systematic survey addressing methods to estimate MIDs (objective 1) will identify all the available methodologies and concepts along with their strengths and limitations, and will thus inform the item generation stage of instrument development. On the basis of the survey of the methods literature and our group's experience with methods of ascertaining MIDs,^{3 4 17–22} we will develop initial criteria for evaluating the credibility of anchor-based MID determinations. Our group has used these methods successfully for developing methodological quality appraisal standards across a wide range of topics.^{16 23–26}

Using a sample of eligible studies, we will pilot the draft instrument with four target users, specifically researchers interested in the credibility of MID estimates, who will be identified within our international network of knowledge users (*please see 'Knowledge Translation' section below*). The data collected at this stage will inform item modification and reduction. This iterative process will be conducted until we achieve consensus for the final version of the instrument.

OBJECTIVE 3: SYSTEMATIC SURVEY OF STUDIES GENERATING AN ANCHOR-BASED MID OF TARGET PRO INSTRUMENTS

We will summarise MID estimates separately for paediatric and adult populations, along with study design, intervention, population characteristics, characteristics of the PRO, characteristics of the anchor and credibility ratings. If multiple MID estimates are captured for the same PRO instrument across similar clinical conditions, we will summarise all the estimates.

OBJECTIVE 4: MEASURING CREDIBILITY OF COMPILED MIDS

Using the instrument (the development of which we have described in objective 2), teams of two reviewers will undertake the credibility assessment for all eligible studies identified in the review. The appraisal will be performed in duplicate, using prepiloted forms. Disagreements will be resolved by discussion between the reviewers and, if needed, with a third team member.

Knowledge users (eg, future systematic review authors or guideline developers) could then use their judgement to consider the credibility of the MID on a continuum from 'highly credible' to 'very unlikely to be credible'.

OBJECTIVE 5: RELIABILITY STUDY OF THE CREDIBILITY INSTRUMENT

We will conduct a reliability study of our instrument to measure the credibility of MIDs by calculating the inter-rater reliability and associated 95% CI as measured by weighted κ with quadratic weights. We will complete reliability analyses using classical test theory. We will consider a reliability coefficient of at least 0.7 to represent 'good' inter-rater reliability.^{27–29} According to Walter *et al*,³⁰ considering three replicates per study (three raters), a minimally acceptable level of reliability of 0.6 and an expected reliability of 0.7, an α of 0.05 and a β of 0.2, we would require a minimum of 133 observations/study assessed per rater. We will use all the studies identified in the systematic survey of estimated MIDs to calculate the reliability estimate. On the basis of initial pilot screening, we estimate that we will have approximately 400 eligible studies.

KNOWLEDGE TRANSLATION

Our multidisciplinary study team is composed of knowledge users and researchers with a broad range of content and methodological expertise. We represent and provide direct links to international networks of key knowledge users including the GRADE Working Group, the Cochrane Collaboration, the WHO and the Canadian Agency for Drugs and Technologies in Health. Collectively, these organisations provide opportunities for disseminating our findings to an international network of clinical trialists, systematic review authors, guideline developers and researchers involved in the development of PROs.

Our dissemination strategies include the incorporation of our knowledge products into Cochrane reviews via the GRADEprofiler (<http://www.guidelinedevelopment.org>—a globally-adopted platform that is widely used by our target audience for conducting systematic reviews and developing practice guidelines) and MAGICapp (a guideline and evidence summary creation tool that uses the GRADE framework). We will offer our work for relevant portions of the printed and online versions of the Cochrane Handbook, and we will develop interactive education sessions (workshops) and research briefs to inform knowledge-users from various health disciplines about our findings and their implications for clinical trial, systematic review and guideline development.

DISCUSSION

Our systematic survey will represent the first overview of methods to develop anchor-based MIDs, and the first comprehensive compendium of published anchor-based

MIDs associated with PRO instruments used in evaluating the effects of interventions on chronic medical and psychiatric conditions. Our systematic survey on methods to estimate MIDs will draw attention to the methodological issues and challenges involved in MID determinations. In doing so, we will deepen the understanding and improve the quality of reporting and use of MIDs by our target knowledge users. The methods review will inform the development of an instrument to determine the credibility of anchor-based MIDs that will allow us to address existing MIDs and can subsequently be used to evaluate new studies offering anchor-based MIDs. This work will also help knowledge users identify anchor-based MIDs that may be less credible, misleading or inappropriate with respect to the average magnitude of change that is important to patients.

We recognise that some variability in MIDs will be attributable to context and patient characteristics. For example, anchor-based MIDs are established using average magnitudes of change that are considered important to patients. Without individual patient data, we will not be able to explore, for example, subgroups of patients with mild or severe disease, gender differences and the relative contributions of these factors. We will alert our knowledge users to this potential limitation.

Collectively, these efforts will promote better-informed decision-making by clinical trialists, systematic review authors, guideline developers and clinicians interpreting treatment effects on PROs.

Author affiliations

¹Department of Anaesthesia and Pain Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada

²Child Health Evaluative Sciences, The Research Institute, The Hospital For Sick Children, Hospital for Sick Children, Toronto, Ontario, Canada

³Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

⁴Department of Clinical Epidemiology & Biostatistics, McMaster University, Hamilton, Ontario, Canada

⁵Department of Anesthesia, McMaster University, Hamilton, Ontario, Canada

⁶Department of Medicine, Stanford Prevention Research Center, Stanford University, Stanford, California, USA

⁷Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, California, USA

⁸Evidence-Based Dentistry Unit, Faculty of Dentistry, Universidad de Chile, Independencia, Santiago, Chile

⁹Department of Health Promotion and Human Behavior, School of Public Health, Kyoto University Graduate School of Medicine, Kyoto, Japan

¹⁰Department of Health Services, University of Washington, Seattle, Washington, USA

¹¹Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada

¹²Department of Medicine, McMaster University, Hamilton, Ontario, Canada

¹³Nephrology Program, Humber River Regional Hospital, Toronto, Ontario, Canada

¹⁴Division of Nephrology, University of Western Ontario, London, Ontario, Canada

¹⁵Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Ontario, Canada

Acknowledgements The authors would like to thank Tamsin Adams-Webber at the Hospital for Sick Children and Mr. Paul Alexander for their assistance with developing the initial literature search.

Contributors BCJ, SE, GHG and GN conceived the study design, and AC-L, TAF, DLP, BRH and HJS contributed to the conception of the design. BCJ and SE drafted the manuscript, and all authors reviewed several drafts of the manuscript. All authors approved the final manuscript to be published.

Funding This project is funded by the Canadian Institutes of Health Research, Knowledge Synthesis grant number DC0190SR. SE is supported by an MITACS Elevate and SickKids Restrcomp Postdoctoral Fellowship Awards.

Competing interests TAF has no COI with respect to the present study but has received lecture fees from Eli Lilly, Meiji, Mochida, MSD, Otsuka, Pfizer and Tanabe-Mitsubishi, and consultancy fees from Sekisui Chemicals and Takeda Science Foundation. He has received royalties from Igaku-Shoin, Seiwa-Shoten and Nihon Bunka Kagaku-sha publishers. He has received grant or research support from the Japanese Ministry of Education, Science, and Technology, the Japanese Ministry of Health, Labour and Welfare, the Japan Society for the Promotion of Science, the Japan Foundation for Neuroscience and Mental Health, Mochida and Tanabe-Mitsubishi. He is a diplomate of the Academy of Cognitive Therapy. All other authors declare no conflicts of interest.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

- Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473–83.
- Beck AT, Ward CH, Mendelson M, et al. An inventory for measuring depression. *Arch Gen Psychiatry* 1961;4:561–71.
- Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407–15.
- Johnston BC, Thorlund K, Schunemann HJ, 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.
- Schunemann HJ, Guyatt GH. Commentary—goodbye M(C)ID! Hello MID, where do you come from? *Health Serv Res* 2005;40:593–7.
- Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–6.
- Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines 6. Rating the quality of evidence—imprecision. *J Clin Epidemiol* 2011;64:1283–93.
- Guyatt GH, Osoba D, Wu AW, et al. Clinical Significance Consensus Meeting G. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 2002;77:371–83.
- Guyatt GH, Juniper EF, Walter SD, et al. Interpreting treatment effects in randomised trials. *BMJ* 1998;316:690–3.
- King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res* 2011;11:171–84.
- Johnston BC, Patrick DL, Thorlund K, et al. Patient-reported outcomes in meta-analyses—part 2: methods for improving interpretability for decision-makers. *Health Qual Life Outcomes* 2013;11:211.
- Johnston BC, Bandayrel K, Friedrich JO, et al. Presentation of continuous outcomes in meta-analysis: a survey of clinicians' understanding and preferences. *Cochrane Database Syst Rev* 2013; Suppl 1(212).
- MAPI Research Trust. Patient-Reported Outcome and Quality of Life Database (PROQOLID). Secondary Patient-Reported Outcome and Quality of Life Database (PROQOLID). 2013. <http://www.proqolid.org>
- Morse J. Designing funded qualitative research. In: Denzin NK, Lincoln YS, eds. *Handbook for qualitative research*. Thousand Oaks, CA: Sage, 1994:220–35.
- Murad MH, Montori VM, Ioannidis JP, et al. How to read a systematic review and meta-analysis and apply the results to patient care: users' guides to the medical literature. *JAMA* 2014; 312:171–9.
- Sun X, Briel M, Walter SD, et al. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ* 2010;340:c117.
- Fallah A, Akl EA, Ebrahim S, et al. Anterior cervical discectomy with arthroplasty versus arthrodesis for single-level cervical spondylosis: a systematic review and meta-analysis. *PLoS ONE* 2012;7:e43407.
- Turner D, Schunemann HJ, Griffith LE, et al. The minimal detectable change cannot reliably replace the minimal important difference. *J Clin Epidemiol* 2010;63:28–36.
- Turner D, Schunemann HJ, Griffith LE, et al. Using the entire cohort in the receiver operating characteristic analysis maximizes precision of the minimal important difference. *J Clin Epidemiol* 2009;62:374–9.
- Juniper EF, Guyatt GH, Willan A, et al. Determining a minimal important change in a disease-specific Quality of Life Questionnaire. *J Clin Epidemiol* 1994;47:81–7.
- Johnston BC, Thorlund K, da Costa BR, et al. New methods can extend the use of minimal important difference units in meta-analyses of continuous outcome measures. *J Clin Epidemiol* 2012;65:817–26.
- Yalcin I, Patrick DL, Summers K, et al. Minimal clinically important differences in Incontinence Quality-of-Life scores in stress urinary incontinence. *Urology* 2006;67:1304–8.
- Levine M, Ioannidis J, Haines T, Harm (observational studies). In: Guyatt G, Rennie D, Meade MO, et al., eds. *Users' guides to the medical literature: a manual for evidence-based clinical practice*. McGraw-Hill, 2008:363–82.
- Randolph A, Cook DJ, Guyatt G. Prognosis. In: Guyatt G, Rennie D, Meade MO, et al., eds. *Users' Guides to the Medical Literature: a manual for Evidence-Based Clinical Practice*: McGraw-Hill, 2008:509–22.
- Furukawa TA, Jaeschke R, Cook D, et al. Measuring of patients' experience. In: Guyatt G, Drummond R, Meade MO, et al., eds. *Users' guides to the medical literature: a manual for evidence-based clinical practice*. 2nd edn. New York: McGraw-Hill, Inc, 2008:249–72.
- Akl E, Sun X, Busse JW, et al. Specific instructions for estimating unclearly reported blinding status in randomized trials were reliable and valid. *J Clin Epidemiol* 2012;65:262–7.
- Heppner PP, Kivlighan DM, Wampold BE. *Research design in counseling*. Pacific Grove, CA: Brooks/Cole, 1992.
- Kaplan RM, Sacuzzo DP. *Psychological testing: principles, applications, and issues*. 4th edn. Pacific Grove, CA: Brooks/Cole, 1997.
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing interrater reliability. *Psychol Bull* 1979;86:420–8.
- Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med* 1998;17:101–10.

Appendix A. Data Abstraction Form

1	RefID number	
2	Write your initials	
3	FULL Journal Name	
4	Last Name of First Author or the first meaningful word in the bi-line	
5	Publication year	
6	Country(ies) where the study was conducted (Check all that apply)	<input type="checkbox"/> United States <input type="checkbox"/> Canada <input type="checkbox"/> United Kingdom <input type="checkbox"/> Australia <input type="checkbox"/> New Zealand <input type="checkbox"/> Japan or other Asia, specify: _____ <input type="checkbox"/> Continental Europe, specify: _____ <input type="checkbox"/> Other (specify): _____ _____
7	Study Design (Check the correct option)	<input type="checkbox"/> Clinical/experimental trial <input type="checkbox"/> Observational study <input type="checkbox"/> Secondary research (e.g. meta-analyses, systematic review) <input type="checkbox"/> Other (specify): _____ _____
8	If this is an experimental trial, please state the study interventions	_____ _____
9	Medical condition under investigation	_____
10a	Total number of participants at baseline	_____
10b	Total number of participants used for the timepoint where the MID was calculated	_____
10c	Total number of participants that was specially used to calculate the MID estimate	_____
11	% of female participants at baseline	_____

12	Mean/median age of participants	_____
13	Standard deviation/interquartile range of the age of participants	_____
14a	Is the PRO instrument disease-specific, generic or unclear? (Check the correct option)	<input type="checkbox"/> Disease-specific <input type="checkbox"/> Generic <input type="checkbox"/> Unclear, comment: _____ -
14b	Please specify the name of the PRO instrument	_____
15	Number of items in the PRO instrument	_____
16a	What is the corresponding range of scores of the PRO? (Please pay especial attention to the minimum score. The same instrument can sometimes have 0 or 30 (for example) as the minimum depending on whether each item is rated 0-6 or 1-7)?	_____
16b	Are higher scores on the PRO attributed to a better or worse outcome?	<input type="checkbox"/> Worse <input type="checkbox"/> Better
16c	What are the type of response options for the questions/items in the instrument?	<input type="checkbox"/> Nominal/ordinal (i.e., categorical - e.g., likert scale) <input type="checkbox"/> Ratio/interval (i.e., continuous - e.g, VAS scale) <input type="checkbox"/> Both
17	What is the mean (SD) baseline score of the PRO?	Mean: _____ SD: _____
18	What is the mean (SD) end score of the PRO?	Mean: _____ SD: _____
19	What is the mean (SD) change score of the PRO?	Mean: _____ SD: _____
20	Does the PRO instrument measure a single domain or multiple domains?	<input type="checkbox"/> Single domain, please specify _____ <input type="checkbox"/> Multiple domains, please specify _____ <input type="checkbox"/> Unclear
21	Indicate if the PRO is a self-report (completed by the patient) or proxy-report (completed by someone else other than the patient)	<input type="checkbox"/> Self-reported <input type="checkbox"/> Proxy-reported, please specify _____ <input type="checkbox"/> Both <input type="checkbox"/> Unclear

22	Was an anchor-based method used to determine the MID?	<input type="checkbox"/> Yes <input type="checkbox"/> No
23	Describe the methods used to determine the anchor-based MID?	<hr/> <hr/> <hr/>
24a	Do the authors report the instrument's MID value? (Check the correct option)	<input type="checkbox"/> Yes – please specify the mean MID value: _____ <input type="checkbox"/> No <input type="checkbox"/> Unclear, comment: _____
24b	Indicate whether the authors report relative or absolute measures of the MID or both?	<input type="checkbox"/> Relative <input type="checkbox"/> Absolute <input type="checkbox"/> Both
25	Please report the type of precision estimate of the MID reported	<input type="checkbox"/> SD, please specify _____ <input type="checkbox"/> SE, please specify _____ <input type="checkbox"/> 95% CI, please specify _____ <input type="checkbox"/> IQR, please specify _____ <input type="checkbox"/> Range, please specify _____ <input type="checkbox"/> Not reported
26	Do the authors report the instrument's MID value specific to males? (Check the correct option)	<input type="checkbox"/> Yes – please specify the MID value: _____ <input type="checkbox"/> No
27	Do the authors report the instrument's MID value specific to females? (Check the correct option)	<input type="checkbox"/> Yes – please specify the MID value: _____ <input type="checkbox"/> No
28	Please specify what anchor was used?	<hr/>
29	Number of items on the anchor instrument	<hr/>
30	What is the corresponding range of scores of the anchor instrument?	<hr/>
31a	What is the type of response options of the anchor?	<input type="checkbox"/> Nominal/ordinal (i.e., categorical - e.g., likert scale) <input type="checkbox"/> Ratio/interval (i.e., continuous - e.g, VAS scale) <input type="checkbox"/> Both
31b	What is the score (or range of scores) on the anchor that is classified as “important change” or “important deterioration”?	<hr/>

32	Does the anchor measure a single domain or multiple domains?	<input type="checkbox"/> Single domain, please specify _____ <input type="checkbox"/> Multiple domains, please specify _____ <input type="checkbox"/> Unclear
33	Indicate if the anchor is a self-report (completed by the patient) or proxy-report (completed by someone else other than the patient) (Check the correct option)	<input type="checkbox"/> Self-report <input type="checkbox"/> Proxy, please specify _____ <input type="checkbox"/> Both <input type="checkbox"/> Unclear
34	Were the PRO measure and anchor-based measure administered simultaneously?	<input type="checkbox"/> Yes <input type="checkbox"/> No
35	If the answer is 'no' to question 34 above, what was the time interval between administration of the PRO measure and anchor-based measure, in weeks?	_____