

Mind the methods of determining minimal important differences: three critical issues to consider

Tahira Devji ¹, Alonso Carrasco-Labra,² Gordon Guyatt¹

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/ebmental-2020-300164>).

¹Health Research Methods, Evidence & Impact, McMaster University, Hamilton, Canada
²Oral and Craniofacial Health Science, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

Correspondence to

Dr Tahira Devji, Health Research Methods, Evidence & Impact, McMaster University, Hamilton, Canada; devjits@mcmaster.ca

Received 30 May 2020
Revised 9 July 2020
Accepted 13 July 2020
Published Online First
24 August 2020

ABSTRACT

Objective Clinical trialists, meta-analysts and clinical guideline developers are increasingly using minimal important differences (MIDs) to enhance the interpretability of patient-reported outcome measures (PROMs). Here, we elucidate three critical issues of which MID users should be aware. Improved understanding of MID concepts and awareness of common pitfalls in methodology and reporting will better inform the application of MIDs in clinical research and decision-making.

Methods We conducted a systematic review to inform the development of an inventory of anchor-based MID estimates for PROMs. We searched four electronic databases to identify primary studies empirically calculating an anchor-based MID estimate for any PROM in adolescent or adult populations across all clinical areas. Our findings are based on information from 338 studies reporting 3389 MIDs for 358 PROMs published between 1989 and 2015.

Results We identified three key issues in the MID literature that demand attention. (1) The profusion of terms representing the MID concept adds unnecessary complexity to users' task in identifying relevant MIDs, requiring meticulous inspection of methodology to ensure estimates offered truly reflect the MID. (2) A multitude of diverse methods for MID estimation that will yield different estimates exist, and whether there are superior options remains unresolved. (3) There are serious issues of incomplete presentation and reporting of key aspects of the design, methodology and results of studies providing anchor-based MIDs, which threatens the optimal use of these estimates for interpretation of intervention effects on PROMs.

Conclusions Although the MID represents a powerful tool for enhancing the interpretability of PROMs, realising its full value will require improved understanding and reporting of its measurement fundamentals.

INTRODUCTION

To judge patients' response to therapy and inform clinical decision-making, clinicians often rely on patients' self-assessment of change in health status. A typical question clinicians ask their patients is 'Since last week when we started the new treatment, are you feeling better or worse—and if so, to what extent?'. Such 'transition questions' are single items that are short, simple to administer, easy to interpret and are therefore very appealing to clinicians, patients and other healthcare stakeholders.

Although transition questions represent intuitively valuable patient-reported outcome measures (PROMs), classic measurement theory holds that

such single item measures have important limitations relative to multi-item measures: they are less stable, reliable and precise. Well-constructed multi-item instruments are also more sensitive to changes in patients' health status over time (ie, responsive), are able to measure multidimensional phenomena and provide information regarding individual domains. Thus, in clinical research, questionnaires that include multiple items addressing one or more underlying constructs have proved the most trustworthy approach to measuring aspects of health status, including symptoms, functional status and quality of life.¹

With the growing emphasis on patient-centred care, major international health policy and regulatory authorities have recognised the importance of evaluating PROMs in clinical research and, to an increasing extent, in clinical practice.^{2–5} Despite the proliferation of PROMs used in research and practice, there remain substantial challenges interpreting their results. Users of PROM results—including clinicians, guideline developers and patients—often have no intuitive notion whether an apparent treatment effect is trivial in magnitude, small but important, moderate or large. For instance, if a new pharmacological therapy to treat major depression in adults improves patients' score on the Beck Depression Inventory by three points relative to control, what are we to conclude about the effectiveness of the new treatment? Is the treatment effect large or is it trivial? To address this problem, researchers developed the concept of the minimal important difference (MID): the smallest change—either positive or negative—in an outcome that patients perceive as important.⁶

There are two commonly used approaches for determining the MID: distribution-based and anchor-based methods. Distribution-based approaches to determining MIDs are based on statistical characteristics of the study sample. There are three categories of distribution-based measures. Those that involve evaluating change in a PROM in relation to either the probability that this change occurred by chance (paired *t*-statistic,⁷ growth curve analysis⁸), sample variation (effect size—with suggestions that treatment effects including 0.2 and 0.5 are minimally important,⁹ standardised response mean,¹⁰ responsiveness statistic¹¹) or measurement precision (SE of measurement, an SD of 0.5,¹² the reliable change index¹³; online supplementary appendix 1).¹⁴

Distribution-based approaches rely solely on the variability in PROM scores and do not reflect the patients' perspective, severely limiting their usefulness in interpreting results.



© Author(s) (or their employer(s)) 2021. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Devji T, Carrasco-Labra A, Guyatt G. *Evid Based Ment Health* 2021;**24**:77–81.

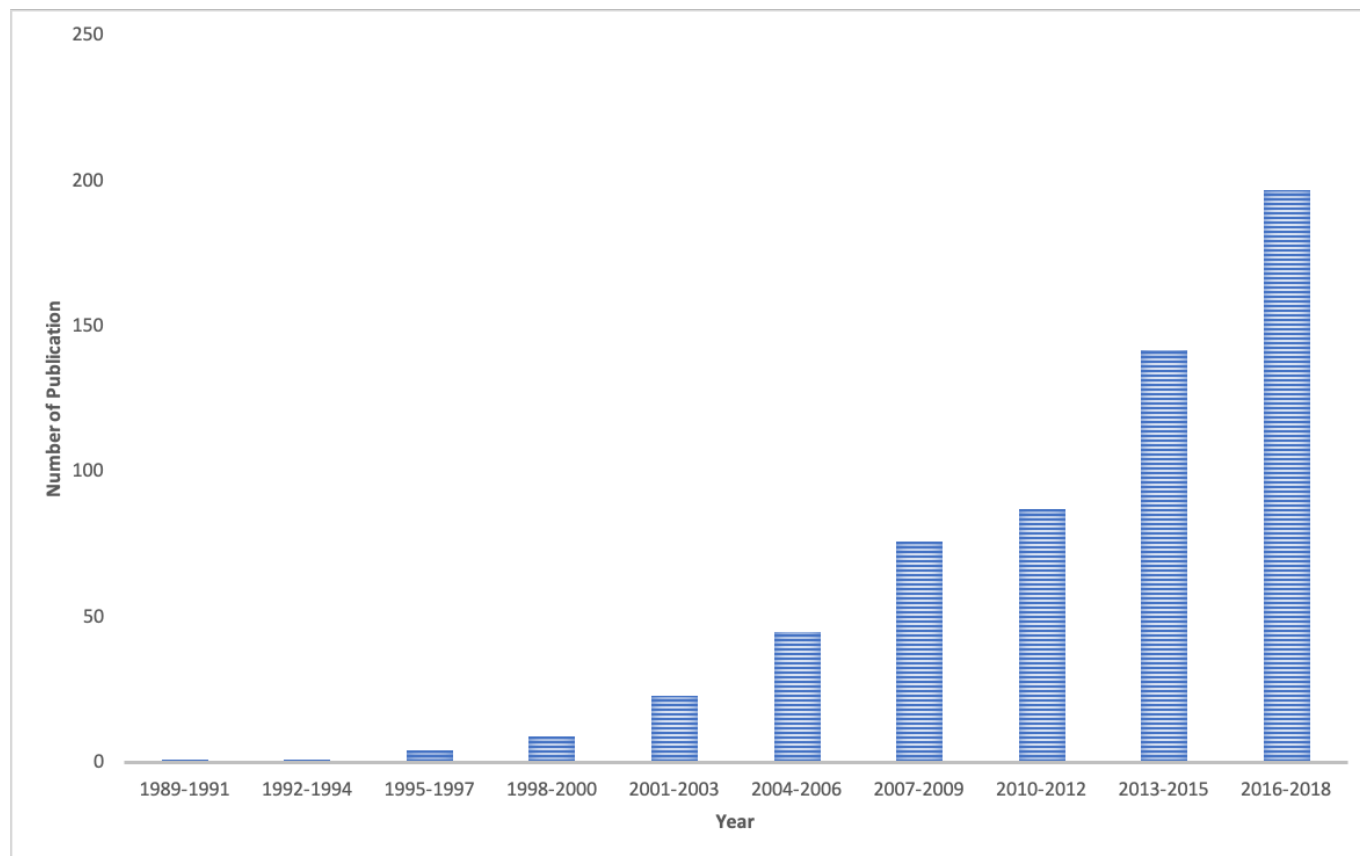


Figure 1 Number of published anchor-based minimal important difference estimation studies from 1989 up to October 2018 during each 3-year stratum.

In the anchor-based approach, investigators establish an MID by relating a difference in PROM scores to an improvement or deterioration captured by an independent measure (ie, the anchor) that is itself interpretable.¹⁵ One common anchor-based approach to estimate the MID is the use of a transition question. For example, Parker *et al* calculated the MID for Zung Self-rating Depression Rating Scale (ZDS) as the mean change in ZDS score in patients who reported themselves as ‘slightly better’ to the Health Transition Index (comparison of current health with prior to receiving treatment) of the Short-Form-36 health survey, offering response options of markedly better, slightly better, unchanged, slightly worse and markedly worse. Other independent standards include preference ratings (eg, health states), disease-related outcomes (eg, severity of disease, presence of symptoms), non-disease-related outcomes (eg, job loss, healthcare utilisation), normative population(s) (eg, functional or dysfunctional populations), or prognosis of future events (eg, mortality, hospitalisation).¹⁴

In the last three decades, the number of published studies providing anchor-based MID estimates for PROMs has grown rapidly (figure 1), and MIDs are increasingly being used in trials, meta-analysis and guidelines to enhance the interpretability of PROMs. According to the Web of Science, the paper that described the first empirically developed MID⁶ has been cited over 2500 times; in 2019, 30 years after its publication, it was cited 223 times.

In response to this expanding field, our research group has developed resources facilitating the identification and appraisal of anchor-based MIDs. We have developed an inventory of all published studies empirically estimating MIDs for PROMs, and

created a novel instrument to evaluate the credibility—the extent to which the design and conduct of studies measuring MIDs are likely to have protected against misleading estimates—of MIDs.

When summarising MIDs and evaluating their credibility, we encountered challenges and identified opportunities to improve current methodological standards. Given the widespread recognition of the usefulness of the MID, elucidating the most critical issues of which MID users should be aware may be of considerable value. Key stakeholders who may find these insights useful include committees that develop clinical practice guidelines and formularies, set market access and reimbursement policies and make regulatory decisions; as well as trialists, systematic reviewers, clinicians and patients. Improved understanding of MID concepts and awareness of common pitfalls in methodology and reporting will better inform the application of MIDs in clinical research, clinical practice and regulatory policy.

METHODS

The development of our inventory involved a systematic survey to summarise primary studies estimating an anchor-based MID for one or more PROMs in adolescent (≥ 13 to 17) or adult (≥ 18) populations across all clinical areas. PROMs of interest included self-reported patient-important outcomes of health-related quality of life, function, symptom severity and psychological distress, and well-being.¹⁶ We included any reported MID estimate irrespective of the participants’ condition or disease, type of intervention used in the study, nature of the anchor and its interpretability, or study design. We systematically searched Medline, Embase, Cumulative Index to Nursing and

Allied Health Literature (CINAHL) and PsycINFO for studies published between 1989 and April 2015 (the first empirically derived MID appeared in the medical literature in 1989).⁶ We also searched the Patient-Reported Outcome and Quality of Life Instruments Database (PROQOLID) internal library and retrieved additional relevant citations and reviewed reference lists from relevant reviews and eligible studies.

For each estimate in our MID inventory, we have abstracted information pertaining to: the study setting; patient demographics; PROM characteristics; interventions administered in the context of the MID estimation; anchor details (ie, type, construct(s), range of options/categories/values, specific anchor-based method); MID estimate, its associated measure of variability and direction; details regarding MID determination (eg, number of patients informing the MID estimate, duration of follow (if applicable), threshold on the anchor selected to represent a 'small but important difference', analytical (or estimation) approach, correlations between the PROM and anchor).

A previously published protocol¹⁶ and subsequent articles documenting the development of our MID inventory (A Carrasco-Labra, personal communication, 2020) and credibility instrument¹⁷ provide full details of the project's methods. The MID inventory and credibility instrument can be found here: www.promid.org. At the time of writing, we are in the process of updating our MID inventory to include articles published between April 2015 and October 2018; however, our findings presented herein are based on the first 338 studies reporting 3389 anchor-based MID estimates for 358 PROMs published between 1989 and 2015.

RESULTS

Lack of a consistent nomenclature

In 1987, Guyatt *et al* introduced the MID concept, labelling the concept the minimal *clinically* important difference.¹⁸ Because this terminology focused attention on the clinical arena as opposed to patients' experience, the same group of researchers later suggested dropping the word 'clinical', and relabelled the concept as the MID.^{19,20} The authors subsequently asserted that the patients who are providing information on aspects of their health status are in the best position to ultimately judge whether a difference in a PROM is or is not important.²¹ Many might consider this semantic distinction important, and the universal adoption might add clarity to the discussion.

In the 338 anchor-based MID studies currently included in our inventory, we identified 86 unique terms referring to the MID concept (online supplementary appendix 2). Most deviations from the original and revised terminology were trivial (eg, *minimum clinically* important difference), but others were more problematic (eg, clinically relevant change, minimal patient perceivable deterioration, responder definition improvement).

This profusion of terms is often a semantic matter, but sometimes represents a different concept. For instance, MIDs for improvement may sometimes differ from those of deterioration, and thus being explicit about the direction of the difference (or *change*) may be warranted (eg, minimal important *improvement*, minimal important *deterioration*).^{22–24} Further, some researchers suggest the usefulness of distinguishing between methods that rely on within-person changes and those that quantify differences between groups.^{22,25}

Such inconsistencies in terminology add unnecessary complexity to reviewers' task in comprehensively identifying relevant MIDs, requiring meticulous inspection of methodology in individual studies to ensure estimates offered truly reflect the

MID. Thus, to avoid confusion and minimise the risk of MIDs going undetected when searching the literature, effort should be made to standardise terminology, and where appropriate reflect important conceptual distinctions.

Multitude of diverse methods for MID estimation—determining whether the MID actually reflects a small but important difference

Generally speaking, the methodology underpinning an anchor-based MID relies on two key components: (1) the anchor and (2) the analytical (or estimation) approach. The appropriate use of an anchor for MID estimation requires knowledge regarding the magnitude of difference on the anchor that is small but important to patients. Two-thirds of the MIDs in our inventory were estimated using a transition rating—perhaps not surprising given that transition ratings are easily understandable for both clinicians and patients, and can easily be framed to relate closely to the construct that the PROM is measuring. Investigators must, however, choose the right response option to correspond to the MID (eg, 'a little better' would be a choice much superior to 'much better'). Quantifying a change that is small but important to patients on other anchor types such as haemoglobin levels, incontinence episodes or Crohn's Disease Activity Index are likely to be much more challenging.

Although anchors with a very limited relation to patient function or experience (such as haemoglobin) are very likely poor choices, there is no consensus on the type of anchor that is best suited for ascertaining MIDs. Moreover, for the same anchor, the threshold defining the MID often differs across studies. Applying our credibility instrument to the studies in our inventory revealed that, in our judgement, anchors reflected a small but important difference in only 43% of putative MIDs.

Further complicating matters, even after the threshold is set, there are a multitude of statistical approaches to compute the MID, each with its own merits and limitations.²⁶ The different analytical methods will yield different estimates,^{24,27} and whether one or more represent better choices remains unresolved. **Box 1** presents the different analytical methods identified in the anchor-based MID estimation studies included in our inventory.

The recent Consolidated Standards of Reporting Trials (CONSORT) Patient-Reported Outcome (PRO) Extension addresses the need for enhanced interpretation of PRO results, and encourages authors to include discussion of an MID or a responder definition in clinical trial reports.²⁸ This demand for increased MID reporting in trials will require clinical trialists and users of trial data to better understand MID methodology and distinguish between more and less trustworthy MID estimates. Failure to use credible MIDs to provide interpretable estimates of treatment effects measured by PROMs may lead to serious misinterpretations of findings from otherwise well-designed clinical trials and meta-analyses. Future research should aim to better understand the strengths and limitations of the various analytical methods used by researchers estimating anchor-based MIDs, and strive towards harmonisation of MID methods that will yield robust estimates.

Inadequate reporting and the need for the development of a reporting standard for better transparency

In addition to appropriate design, conduct and analysis, investigators must also report clear, transparent research methods and findings. In the development of our inventory of MIDs, we found major deficiencies in reporting. The usefulness of the anchor-based approach is critically dependent on the extent

Box 1 Analytical methods reported in the anchor-based minimal important difference (MID) estimation studies included in the MID inventory

1. Changes *within* patients over time

The MID is the mean change in patient-reported outcome measure (PROM) scores over time within the subgroup of participants who reported a small but important change (ie, improvement and/or worsening).

The MID is the median change in PROM scores over time within the subgroup of participants who reported a small but important change.

The MID is defined as the 75th percentile of the distribution of change in PROM scores within the subgroup of participants who reported a small but important change.

The MID is the lower or upper limit of the 95% CI of the mean change in PROM scores over time within the subgroup of participants who reported a small but important change.

The MID is estimated using a regression model (either logistic or linear), in which the dependent variable is the change in PROM score and the independent variable is the value, rating or category on the anchor that reflects a small but important change. Alternatively, the PROM score at follow-up may be the dependent variable, while the independent variables in this model are the value, rating or category on the anchor and the baseline PROM score. A second approach, although less common, involves using the anchor as a dependent variable and the PROM as the independent variable.

The MID is estimated using an analysis of variance model, in which the dependent variable is the change in PROM score and the independent variable is the value, rating or category on the anchor.

The MID is estimated using discriminant function analysis.

The MID is estimated using linkage (or scale alignment) to estimate the MID.

2. Differences *between groups* capturing changes *within* patients over time.

The MID is the mean change in PROM scores over time in the participants with a small but important change minus the mean change in PROM scores over time in the participants with no change. This method attempts to correct for the change in PROM scores in the no change group.

The MID is the change in PROM scores over time in the participants in one group minus the mean change in PROM scores over time in the participants in another group. The participants in these groups have a different status on the same condition or disease-related outcome.

3. Differences *between* patients' PROM scores at one timepoint.

The MID is the difference in PROM scores between participants who rated themselves compared with another participant, as a little bit better (or a little bit worse) versus participants who rated themselves, compared with another participant, as about the same.

The MID is the difference in PROM scores between participants in groups with a different status on the same condition or disease-related outcome.

The MID is estimated using a regression model (either logistic or linear), where the dependent variable is the PROM score and

Continued

Box 1 Continued

the independent variable is the value, rating or category on the anchor.

4. Receiver operating characteristic curve analysis

Methods selecting optimal cut-point based on lowest overall misclassifications and giving equal weight to sensitivity and specificity

Youden method: the point on the receiver operating characteristic (ROC) curve that maximises the distance to the identity (or chance) line is selected as the optimal MID.

Closest-to-(0,1) criterion: the point on the ROC curve closest to (0,1) (upper left corner of the graph) is selected as the optimal MID.

45° tangent line: A -45° tangent line is drawn from (0,1) to (1,0) intersecting the ROC curve (ie, from the top left corner to the bottom right corner of the graph). The MID is the point on the ROC curve that is closest to the -45° tangent line. In other words, this is the cut-point with equal (or almost equal) sensitivity and specificity.

Others

80% specificity rule: the MID is the cut-point that provides the best sensitivity for response while still achieving at least 80% specificity.

The MID is the cut-point associated with an optimal likelihood ratio.

to which the PROM and anchor measure the same, similar or related constructs. Thus, perhaps arguably, the single most important aspect of credibility of the MID is the correlation between the PROM and anchor. In our inventory, we found that, for 71% of MIDs, authors did not report an associated correlation coefficient.

Further, our confidence in an MID estimate will be lower if the CI around the point estimate is insufficiently narrow; yet, for 56% of MIDs, authors did not report a measure of precision. Even simpler issues, such as the number of patients informing the MID, proved unclear for 22% of MIDs; a similar number failed to report the range of the patient-reported measurement scale for which the MID was determined. Because different scales often exist for a single PROM, this is problematic. For example, the Western Ontario and McMaster Universities Osteoarthritis Index pain instrument may be rated on a 5-point likert-type scale with items summing to give a possible range of 0–20 or as a 0–100 visual analogue scale. Moreover, in some studies, investigators will transform scores to a different scale (eg, 0–100) than reported by the authors who developed the instrument.

Judging the credibility and applicability of MIDs requires complete and accurate reporting: inadequate reporting in MID determination studies will threaten the optimal use of these estimates for interpretation of intervention effects on PROMs in clinical trials, systematic reviews and clinical practice guidelines. The use of reporting guidelines for other types of research, such as the CONSORT statement for randomised trials, has resulted in improved reporting.^{29 30} In similar fashion, the development of a reporting guideline for anchor-based MID estimation studies will likely improve the completeness and transparency of MID reports and promote higher methodological standards for robust MID estimation. At the time of writing, our research group is in the initial stages of developing such a standard, and we invite

global experts on MID and PRO methods to collaborate with this initiative.

CONCLUSIONS

Although the MID represents a powerful tool for enhancing the interpretability of PROMs, realising its full value will require improved understanding and reporting of its measurement fundamentals. Some of the issues we have labelled—in particular, terminology and completeness of reporting, are easily addressed. Others, such as choice of optimal anchors and response options representing a small but important threshold difference, and optimal statistical approaches, are likely to prove more challenging. Empirical investigations in the exploration of factors explaining variability in MIDs may aid in informing the desperately needed harmonisation of methods.

Twitter Tahira Devji @TahiraDevji

Acknowledgements The authors thank Anila Qasim, Mark Phillips, Bradley C Johnston, Niveditha Devasenapathy, Dena Zeraatkar, Meha Bhatt, Xuejing Jin, Romina Brignardello-Petersen, Olivia Urquhart, Farid Foroutan, Stefan Schandellaier, Hector Pardo-Hernandez, Robin WM Vernooij, Hsiaomin Huang, Yamna Rizwan, Reed Siemieniuk, Lyubov Lytvyn, Donald L Patrick, Shanil Ebrahim, Toshi Furukawa, Gihad Nesrallah, Holger J Schunemann, Mohit Bhandari, and Lehana Thabane for their contributions on the MID inventory and credibility instrument projects.

Contributors The authors are the lead investigators of a programme of research focused on advancing PRO and MID methods. GG is Distinguished Professor in the Departments of Health, Research Methods, Evidence and Impact (HEI), and Medicine at McMaster University, is the Co-Chair of the GRADE Working Group, Co-Convenor of the Cochrane PRO methods group, and an advisor to clinical practice guideline panels worldwide. He developed the initial concept of the MID, first published in 1987, and developed the transition anchor strategy for establishing MIDs, first published in 1989. He has since been active in the health status measurement field, and particularly in interpretation of PROMs. As the senior mentor on our programme of research focused on improving MID and PRO methods, he oversees the conduct of all related projects, provides methodological guidance and facilitates the dissemination of our work through key stakeholder groups such as GRADE, MAGIC, the WHO, the National Institute for Health and Care Excellence, AHRQ, Canadian Agency for Drugs and Technologies in Health (CADTH), and Cochrane. TD and AC-L co-led the systematic review that informed the development of the MID inventory of all published anchor-based MIDs and developed the MID credibility instrument. It is through this prior work that the critical insights presented in this article have emerged. TD wrote the first draft of the manuscript. All authors interpreted the data analysis and critically revised the manuscript. TD is the guarantor.

Funding The projects informing the submitted work were funded by the Canadian Institutes of Health Research, Knowledge Synthesis grant number KRS138214.

Competing interests No, there are no competing interests.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

ORCID iD

Tahira Devji <http://orcid.org/0000-0001-8414-7410>

REFERENCES

- Bowling A. *Just one question: if one question works, why ask several?* BMJ Publishing Group Ltd, 2005.
- Allik H, Larsson J-O, Smedje H. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes* 2006;4:1–20.
- European Medicines Agency Committee for Medicinal Products for Human Use. *Appendix 2 to the Guideline on the evaluation of anticancer medicinal products in man: the use of patient-reported outcome (pro) measures in oncology studies EMA/CHMP/292464/2014*. London, England: European Medicines Agency, 2016.
- Doward LC, Gnanasakthy A, Baker MG. Patient reported outcomes: looking beyond the label claim. *Health Qual Life Outcomes* 2010;8:89.
- Valderas JM, Kotzeva A, Espallargues M, 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.
- Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407–15.
- Husted JA, Cook RJ, Farewell VT, et al. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000;53:459–68.
- Speer DC, Greenbaum PE. Five methods for computing significant individual client change and improvement rates: support for an individual growth curve approach. *J Consult Clin Psychol* 1995;63:1044–8.
- Cohen J. *Statistical power analysis for the behavioral sciences*. Academic press, 2013.
- Stucki G, Liang MH, Fossel AH, et al. Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. *J Clin Epidemiol* 1995;48:1369–78.
- Guyatt GH, Bombardier C, Tugwell PX. Measuring disease-specific quality of life in clinical trials. *CMAJ* 1986;134:889.
- Norman GR, Sloan JA, Wyvrich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003;41:582–92.
- Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research 1992.
- Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003;56:395–407.
- Guyatt GH, Osoba D, Wu AW, et al. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 2002;77:371–83.
- Johnston BC, Ebrahim S, Carrasco-Labra A, et al. Minimally important difference estimates and methods: a protocol. *BMJ Open* 2015;5:e007953.
- Devji T, Carrasco-Labra A, Qasim A, et al. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. *BMJ* 2020;369:m1714.
- Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987;40:171–8.
- 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.
- Schünemann HJ, Puhan M, Goldstein R, et al. Measurement properties and interpretability of the chronic respiratory disease questionnaire (CRQ). *COPD* 2005;2:81–9.
- Guyatt Get al. Patients at the centre: in our practice, and in our use of language. *Evid Based Med* 2004;9:6–7.
- Copay AG, Subach BR, Glassman SD, et al. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J* 2007;7:541–6.
- King MT. A point of minimal important difference (mid): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res* 2011;11:171–84.
- Mills KAG, Naylor JM, Eyles JP, et al. Examining the minimal important difference of patient-reported outcome measures for individuals with knee osteoarthritis: a model using the knee injury and osteoarthritis outcome score. *J Rheumatol* 2016;43:395–404.
- King MT, Dueck AC, Revicki DA. Can methods developed for interpreting group-level patient-reported outcome data be applied to individual patient management? *Med Care* 2019;57:S38–45.
- King MT. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. *Qual Life Res* 1996;5:555–67.
- Terwee CB, Roorda LD, Dekker J, et al. Mind the MIC: large variation among populations and methods. *J Clin Epidemiol* 2010;63:524–34.
- Calvert M, Blazeby J, Altman DG, et al. Reporting of patient-reported outcomes in randomised trials: the CONSORT pro extension. *JAMA* 2013;309:814–22.
- Cobo E, Cortés J, Ribera JM, et al. Effect of using reporting guidelines during peer review on quality of final manuscripts submitted to a biomedical Journal: masked randomised trial. *BMJ* 2011;343:d6783.
- Turner L, Shamseer L, Altman DG, et al. Consolidated standards of reporting trials (consort) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database Syst Rev* 2012;11:Mr000030.