JAMA Guide to Statistics and Methods

Minimal Clinically Important Difference
Defining What Really Matters to Patients

Anna E. McGlothlin, PhD; Roger J. Lewis, MD, PhD

When assessing the clinical utility of therapies intended to improve subjective outcomes, the amount of improvement that is important to patients must be determined.[1] The smallest benefit of value to patients is called the minimal clinically important difference (MCID). The MCID is a patient-centered concept, capturing both the magnitude of the improvement and also the value patients place on the change. Using patient-centered MCIDs is important for studies involving patient-reported outcomes,[2] for which the clinical importance of a given change may not be obvious to clinicians selecting treatments. The MCID defines the smallest amount an outcome must change to be meaningful to patients.[1]

Related article page 1313

In this issue of *JAMA*, Hinman et al[3] report findings from a clinical trial evaluating whether acupuncture (needle, laser, and sham laser) improved pain or overall functional outcomes compared with no acupuncture among patients with chronic knee pain. Pain was measured on a numerical rating scale and functional status by the Western and McMaster Universities Osteoarthritis Index (WOMAC) score. The MCIDs for both end points were based on prior experience with these scoring systems. The MCID for pain was determined using an expert consensus, or Delphi approach,[4] while the MCID for function was determined using an "anchor" approach, based on patients' qualitative assessments of their own responses to treatment.[5]

Use of the Method

Why Is the MCID Used?

The appropriate clinical interpretation of changes on a numerical scale must consider not only statistical significance, but also whether the observed change is meaningful to patients. Identical changes on a numerical scale may have different clinical importance in different patient populations (eg, different ages, disease severity, injury type). Furthermore, statistical significance is linked to the sample size. Given a large enough sample, statistical significance between groups may occur with very small differences that are clinically meaningless.[6]

When determining how many patients to enroll in a study, the calculation usually reflects the intention to reliably find a clinically important effect of a treatment, such as the MCID. The smaller the treatment effect sought, the larger the required number of study participants.[7]

The MCID can be calculated using consensus, anchor, and distribution-based methods. Consensus (also known as Delphi) methods convene an expert panel to provide independent assessments of what constitutes a clinically relevant change. The assessments are then revised after the panel members review each other's assessments. This process is repeated until a consensus is reached regarding a numerical value for the MCID. The MCID for the pain assessment scale used in Hinman et al[3] was determined by the Delphi method using a panel of 6 rheumatology experts.[4]

Anchor-based methods determine the MCID by associating the change in the numerical scale for an outcome to some other subjective and independent assessment of improvement. For example, patients may be asked if they felt "about the same," "a little bit better," or "quite a bit better" after receiving treatment. These categorical responses are then related to the numerical measurement scale used in the study, "anchoring" the numerical outcome scale to the qualitative, categorical assessment that is presumably more meaningful to patients. The MCID for the WOMAC measure of functional status in the study by Hinman et al[3] was based on the 75th percentile of the WOMAC score; 75% of patients categorizing themselves as having experienced benefit (the anchor) had an improvement equal to or larger than the derived MCID using this definition.[5]

Distribution-based methods for defining the MCID involve neither expert opinion nor patient assessments. These methods rely on the statistical properties of the distribution of outcome scores, particularly how widely the scores vary between patients. These methods determine what magnitude of change is required to show that the change in an outcome measure in response to an intervention is more than would be expected from chance alone. Because distribution-based methods are not derived from individual patients, they probably should not be used to determine an MCID.[6]

What Are the Limitations of MCID Derivation Methods?

Consensus methods use clinical and domain experts, rather than patients, to define the MCID. In many settings, expert opinion may not be a valid and reliable way to determine what is important to patients.

Anchor-based methods are limited by the choice of anchor, which is a subjective assessment. For example, when an anchor is based on asking a patient whether he or she improved after receiving treatment, the response may be susceptible to recall bias. A patient's current status tends to influence recollection of the past. The anchor's validity and reliability are crucial for determination of a valid MCID.

Anchor-based methods may be influenced by the statistical distribution of scores within each category of the anchor. If the data are highly skewed, such as occurs with length-of-stay information because of the occasional outlying patient with a complicated clinical course, the derivation of the MCID may be affected by the outliers. Furthermore, anchor methods often rely on an MCID estimate derived from only a subset of patients (those within a particular category of the anchor). Not accounting for information from patients outside of this group may result in erroneous MCID

estimates if the characteristics of the excluded patients differ from those who were included.

Because distribution-based methods are based on purely statistical reasoning, they can only identify a minimal detectable effect: that is, an effect that is unlikely to be attributable to random measurement error. The lack of an anchor that links the numeric scores to an assessment of what is important to patients causes distribution-based methods to fall short of identifying important, clinically meaningful outcomes for patients. In fact, the term MCID is sometimes replaced by "minimal detectable change" when the difference is calculated by distribution-based methods.[6] Distribution-based methods are not recommended as a first-line means for determining MCID.

Ideally, determination of the MCID should consider different thresholds in different subsets of the population. For example, patients with substantial amounts of pain at baseline might require greater pain reduction to perceive treatment benefit compared with those patients who have little baseline pain.

Why Did the Authors Use MCID in This Particular Study?

Hinman et al[3] specified an MCID for each end point to establish an appropriate sample size for their study and to facilitate clinically meaningful interpretation of the final outcome data. The number of patients enrolled was selected to provide sufficient power (ie, probability) for detecting a change in outcomes resulting from the intervention that was at least as large as the MCID for each end point.

How Should MCID Findings Be Interpreted in This Particular Study?

The actual treatment effects observed in the study of Hinman et al[3] were quite modest, ranging from an improvement of 0.9 to 1.2 units in pain, relative to an MCID of 1.8 units, and an improvement of 4.4 to 5.1 units in function, relative to an MCID of 6 WOMAC units. Although there were statistically significant differences between groups, the clinical importance of these differences is uncertain.[3]

Caveats to Consider When Looking at Results Based on MCIDs

In the study by Hinman et al[3] the observed effect is smaller than the predefined MCID, yet the differences between groups still achieved statistical significance. This phenomenon is not uncommon and occurred in another recently published study in *JAMA* on the effect of vagal nerve stimulation for obesity treatment.[8] This occurs because the study sample size is selected to achieve a high probability of detecting a benefit equal to the MCID, resulting in a substantial chance of demonstrating statistical significance even when the effect of an intervention is smaller than the MCID.

In the study by Hinman et al,[3] acupuncture therapies were compared with control groups by measuring difference in mean changes in pain and function scores. An alternative experimental design would be based on a "responder analysis," namely, comparing the proportion of patients within each therapy who experienced a change greater than the MCID. This type of data presentation could be more informative because it focuses on patients who experience an improvement at least as large as the MCID.[2] This approach is useful when the data are highly skewed by outliers in such a way that the calculated mean value may be above the MCID even when most patients do not have an effect greater than the MCID.

A fundamental aspect of MCIDs that is often ignored is the need to consider potential improvements from an intervention in relation to costs and complications. When selecting an MCID for a clinical trial, defining a meaningful improvement from the patient's perspective ideally involves considering all aspects of clinical care, both favorable and unfavorable.

ARTICLE INFORMATION

Author Affiliations: Berry Consultants, Austin, Texas (McGlothlin, Lewis); Department of Emergency Medicine, Harbor-UCLA Medical Center, Torrance, California (Lewis); Los Angeles Biomedical Research Institute, Torrance, California (Lewis); David Geffen School of Medicine at UCLA, Los Angeles, California (Lewis).

Corresponding Author: Roger J. Lewis, MD, PhD, Department of Emergency Medicine, Harbor-UCLA Medical Center, Bldg D9, 1000 W Carson St, Torrance, CA 90509 (roger@emedharbor.edu).

Conflict of Interest Disclosures: Both authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

REFERENCES

1. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407-415.

2. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. Food and Drug Administration. http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf. Accessed August 27, 2014.

3. Hinman RS, McCrory P, Pirotta M, et al. Acupuncture for chronic knee pain: a randomized clinical trial. *JAMA*. doi:10.1001/jama.2014.12660.

4. Bellamy N, Carette S, Ford PM, et al. Osteoarthritis antirheumatic drug trials: III, setting the delta for clinical trials: results of a consensus development (Delphi) exercise. *J Rheumatol*. 1992;19(3):451-457.

5. Tubach F, Ravaud P, Baron G, et al. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. *Ann Rheum Dis*. 2005;64(1):29-33.

6. Turner D, Schünemann HJ, Griffith LE, et al. The minimal detectable change cannot reliably replace the minimal important difference. *J Clin Epidemiol*. 2010;63(1):28-36.

7. Livingston EH, Elliot A, Hynan L, Cao J. Effect size estimation: a necessary component of statistical analysis. *Arch Surg*. 2009;144(8):706-712.

8. Ikramuddin S, Blackstone RP, Brancatisano A, et al. Effect of reversible intermittent intra-abdominal vagal nerve blockade on morbid obesity: the ReCharge randomized clinical trial. *JAMA*. 2014;312(9):915-922.