Minimal clinically important difference

Shengping Yang PhD, Gilbert Berdine MD

Minimal Clinically Important Difference (MCID) scores are commonly used for evaluating patient responses to treatment by clinicians. Many important clinical decisions are often made with the guidance of MCID. Since the MCID is also important in clinical research, could you give a brief introduction on that?

The concept of Minimal Clinically Important Difference (MCID) was first introduced by Jaeschke et al. in 1989, and since then it has gained increasing popularity. It is defined as the smallest change in a treatment outcome that is considered important for a patient and which would indicate a change in the patient's management. In general, there are two main approaches for estimating MCID.

1. ANCHOR-BASED METHODS

The anchor-based methods use external criteria (the anchor), which are often subjective, to quantify differences measured by an outcome instrument, e.g., comparing patients' reported outcome scores to the patients' answers to prior subjective assessments. Note that objective criteria, such as comparing patients' ratings on the pain scales to the ingested amount of pain medication, are rarely used. There are four main variations among the anchor-based methods:

(A) THE 'WITHIN-PATIENTS' SCORE CHANGE

The MCID is defined as the change in patient-reported outcome scores of a group of patients selected according to their answers to a global assessment scale. 'Within-patients' differences are important in studies in which patients serve as their own controls.

Corresponding author: **Shengping Yang**
Contact Information: Shengping.Yang@pbrc.edu
DOI: 10.12746/swrccc.v9i39.863

Examples include determinations of a self-rated outcome score, such as a standard dyspnea score, before and after treatment. The main concern with this method is to avoid score ranks that are either too coarse or too fine a distinction between adjacent scores. If too coarse, then important clinical distinctions may be missed. Too fine leads to distinctions without a meaningful clinical difference.

(B) THE 'BETWEEN-PATIENTS' SCORE CHANGE

The MCID is defined as the difference in patient-reported outcome (change) scores between two adjacent levels on a global assessment scale. 'Between-patients' differences are important when groups of patients are compared to each other. Examples include studies in which quality of life scores are compared between a treatment group and a control group. The issue with this approach is the subjective decision in choosing the two adjacent levels.

(C) THE SENSITIVITY AND SPECIFICITY BASED APPROACH

The MCID is defined as the score that best separates patients who reported an improvement and those who did not. Although the determination of the best cutoff value from a Receiver Operating Characteristic curve, via sensitivity and specificity, could be objective, the definition of improvement vs. non-improvement is arbitrary. This approach can be applied to any continuous measured outcome variable in which the result is not the value of the variable but whether the variable is above or below some arbitrarily designated threshold. Examples include positive and negative polymerase chain reaction tests.

(D) THE SOCIAL COMPARISON APPROACH

Patients rate themselves as compared to those they were paired with, and the differences are used for

estimating MCID. The challenge with this approach is to appropriately pair the patients with each other. An example would be the difference in the patient self-rated score compared to those of patients they spoke to, perhaps comparing patients who attended pulmonary rehabilitation with those who did not.

2. DISTRIBUTION-BASED METHODS

The distribution-based methods define MCID based on statistical characteristics of the obtained samples, and a number of these methods have been developed, including;

(A) THE MCID BASED ON STANDARD ERROR OF MEASUREMENT (SEM)

The SEM is a measure of how much measured test scores are spread around a "true" score. Often $1 \times$ SEM is used as a benchmark for a "true" change. Examples include a normal range for serum sodium (Na) levels.

(B) MCID BASED ON STANDARD DEVIATION (SD)

The SD is a measure of the amount of variation or dispersion of a set of values. A $0.5 \times$ SD is often used to define MCID for patient-reported outcomes. Examples include the 12% number for trial-to-trial comparisons of forced expiratory volume in one second (FEV1) in the same patient pre-and post-inhaled bronchodilator.

(C) MCID BASED ON EFFECT SIZE (ES)

The ES is the ratio of change from baseline and the SD of the baseline values, and thus it is a standardized measure of change. Examples include the 20% threshold for decrease in FEV1 during a bronchoprovocation test using a challenge with increasing doses of inhaled methacholine.

Both anchor-based and distribution-based approaches have advantages and disadvantages. For example, compared to anchor-based methods, MCIDs defined by distribution-based methods are considered more objective. However, the distribution-based methods have few common agreed benchmarks for establishing clinically significant improvement and do not accommodate the patient's perspective of a clinically important change. On the other hand, the anchor-based methods generally rely on the use of a subjective assessment, which is often arbitrary, and thus can cause difficulties in assessment standardization, interpretation, and comparison. In addition, there are large variabilities among all these methods.

While MCID has become a critical tool in clinical decision making and management, it is also widely used in clinical research. Next, we discuss its relevance in clinical research, specifically, its application in power/sample size calculation.

A large majority of clinical studies are hypothesis driven, for example, to test if there is a difference between two hypothetical treatment groups. After data collection has been completed, statistical analysis will be performed; a comparison will be made between groups by using a parametric/non-parametric test, and a conclusion will be made based on the p value of the test. However, it is important to know that the p value of a statistical test is partially affected by the sample size. In fact, with a sufficiently large sample size, it is possible to obtain a very small p value regardless of how small the difference between the two hypothetical groups is, unless the difference is exactly 0. To avoid the situation in which a trivial difference is detected due to large sample size, or a sample size that is too small to answer a research question, a power/sample size calculation is routinely performed in the planning phase of a clinical study.

3. POWER/SAMPLE SIZE CALCULATION AND MCID

Often, the goal of a clinical study is to detect the difference in a clinical outcome of interest. An investigator is expected to decide on the study design, e.g., parallel vs. crossover, to be used, consider ethical and scientific factors, and assess study validity and feasibility. Among them, it is critical to determine in advance the number of subjects needed because it

is directly related to study recruitment, duration, and costs, etc. In addition, an appropriate sample size ensures that the difference to be detected is clinically meaningful, meaning that the difference is not so trivial that it has little clinical significance.

Statistical power can be described as the probability of rejecting a null hypothesis, given there is a true difference between the groups to be compared. When planning a study, the statistical power is often pre-specified, e.g., 80% or 90%, and then a sample size calculation is conducted. The elements used for the calculation often include pre-specified type I error rate (often set at 0.05) and the effect size that a study is designed to detect. It is often convenient to use the distribution-based MCID as the effect size that would be clinically worth detection. The interpretation of a calculated sample size is that, with a type I error rate of 0.05, there is an 80% probability to detect the pre-specified difference of interest, given that the difference on average is equal to MCID and the estimates on data variation are correct. Note that should the true difference be less than MCID, then the probability of detecting a difference would be less, depending on how small the true difference is.

In summary, MCID is critical in both making and managing clinical decisions and conducting clinical research. The MCID can be estimated by using either the anchor-based or the distribution-based approach; both approaches have advantages and disadvantages. Although it could be potentially subjective, MCID estimation is expected to be transparent and provide a meaningful assessment on a patient outcome. A low MCID may result in overestimating a positive effect, while a high MCID may result in failing to declare a beneficial effect when it does exist. The MCID is also used in power/sample size calculation to ensure that the conclusion made from a clinical study is meaningful.

REFERENCES

1. 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):160–166.
2. Copay AG, Subach BR, Glassman SD, Polly Jr DW, Schuler TC. (2007) Understanding the minimum clinically important difference: a review of concepts and methods. Spine J 2007; 7:541–546.
3. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 Quality of Life measurement instruments in patients with osteoarthritis of the lower extremities. Arthritis care & research 2001;45:384–391.
4. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407–415.
5. Rai SK, Yazdany J, Fortin PR, Aviña-Zubieta JA. Approaches for estimating minimal clinically important differences in systemic lupus erythematosus. Arthritis research & therapy 2015; 17:143.