Understanding the Minimal Clinically Important Difference (MCID) of Patient-Reported Outcome Measures

Ahmad R. Sedaghat, MD, PhD

Abstract

Objective. The minimal clinically important difference (MCID) of a patient-reported outcome measure (PROM) represents a threshold value of change in PROM score deemed to have an implication in clinical management. The MCID is frequently used to interpret the significance of results from clinical studies that use PROMs. However, an understanding of the many caveats of the MCID, as well as its strengths and limitations, is necessary. The objective of this article is to provide a review of the calculation, interpretation, and caveats of MCID.

Data Sources. MEDLINE and PubMed Central.

Review Methods. Literature search—including primary studies, review articles, and consensus statements—pertinent to the objectives of this review using PubMed.

Conclusions. The MCID of a PROM may vary depending on the patients and clinical context in which the PROM is given. The primary approaches for calculating MCID are distribution-based and anchor-based methods. Each methodology has strengths and limitations, and the ideal determination of a PROM MCID includes synthesis of results from both approaches. The MCID of a PROM is also not perfect in detecting patients experiencing a clinically important improvement, and this is reflected in its accuracy (eg, sensitivity and specificity).

Implications for Practice. Interpretation or application of MCID requires consideration of all caveats underlying the MCID, including the patients in whom it was derived, the limitations of the methodologies used to calculate it, and its accuracy for identifying patients who have experienced clinically significant improvement.

Keywords

minimal clinically significant difference, MCID, outcomes, quality of life

Received January 4, 2019; accepted May 3, 2019.
substantiate that any reported improvement in an outcome measure is not only noticeable to the patient but also surpasses a threshold of benefit or worsening that would mandate action (e.g., changing medical practice patterns). Prospective clinical studies must also be powered to detect changes of at least the minimal change in the primary outcome measure that constitutes a clinically meaningful difference. It is important to note that the definition of clinically meaningful may be different based on the stakeholder’s perspective. For example, what may be deemed a meaningful change by the patient may be different from what a physician may deem to be meaningful or what an insurance company/third-party payor may deem to be meaningful.6,7 In addition, what constitutes a meaningful change in a PROM does not account for cost or economic considerations needed to achieve that change either. Although the following discussion is applicable to many varieties of clinical outcomes and stakeholders, this review will focus on patients as the primary stakeholders and PROMs—reflecting patients’ assessment of self.

Patient-reported outcomes generally consist of elements that are reflective of patient perception of disease and may be influenced by patients’ perspectives, feelings, moods, or attitude.8,9 The concept of a minimal clinically important difference (MCID) for a PROM was introduced in 1989 by Jaeschke et al,10 who defined the MCID as

the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient’s management.

This definition of MCID has also been paraphrased by others such as Wyrwich and Tardino,11 who described the MCID as “a difference [in] score that is large enough to have an implication for the patient’s treatment or care.” Similar concepts to the MCID, including the minimal important difference (MID)12 and minimally important change (MIC),13 have also been described but are more patient centered by being formally defined based on what patients deem to be minimally important.

The MCID has subsequently been used in clinical studies as a benchmark for the magnitude of improvement in a PROM that reflects clinical significance. However, the interpretation of clinical study results through application of an MCID is dependent on the many caveats and assumptions underlying the MCID itself, many of which are frequently unclear or misunderstood. In this review, the various methods for calculating the MCID of a PROM will be described and their underlying assumptions discussed. Finally, these and other considerations for the interpretation of a reported MCID will be discussed.

**Methods**

The purpose of this article was to provide a state-of-the-art review of the calculation, interpretation, and caveats of MCID. With that purpose in mind, the MEDLINE and PubMed Central databases were queried using PubMed for studies, review articles, and consensus statements that addressed the objectives of this review. Searches were performed using primary search terms, including minimal clinically significant difference, minimal important difference, subjectively significant difference, minimally important change, and minimal detectable change. The timeframe of literature searched is from 1989 (when the MCID was first described) to November 2018, and the author (A.R.S.) was the sole reviewer of articles. The references of identified articles were also searched for pertinent articles. Given the objective of this review, the focus of the search was primary literature and studies that directly described MCID methodology and caveats. The final articles included were subjectively chosen as those best addressing the objectives of this review.

**Discussion**

**Calculation of the MCID**

The MCID of an instrument can be calculated in many different ways, and there is no consensus for the best methodology to calculate an MCID.14 Methods for calculating MCID can broadly be categorized as distribution-based and anchor-based methods.15,16 Distribution-based methods rely on the statistical characteristics of a group’s baseline PROM scores to determine—given the spread of a group’s baseline PROM scores—how much of a change may be clinically important. In comparison, anchor-based methods compare the change in patients’ PROM scores to how patients score on a second, explicit metric of patient improvement. These 2 categories of methods, and various well-described statistical approaches comprising each, are discussed further below.

**Choice of Patient Population**

Previous work has shown that the calculated value of the MCID for a PROM is dependent on the context.17 Context includes characteristics of the patient population being studied to derive the MCID—characteristics that include socio-economic status, disease severity, or patient expectations.18 Context can also include the clinical scenario, such as the specific disease as well as the type of intervention that is being provided.19 For this reason, calculation of an MCID first begins by choosing an appropriate patient population—with characteristics and in a clinical context—as close as possible to the patient population to which the PROM and its MCID will be applied.

**Distribution-Based Methods for Calculating MCID**

**Conceptual framework.** Distribution-based methods for determining MCID are defined based on statistical parameters reflecting statistical spread/variation and measurement accuracy of the PROM. The rationale for distribution-based methods is related to the inherent variability in scoring on PROMs, which is observed both at the individual and group level. A patient’s score on a PROM may vary from one
time point to another even when the patient does not have any important change in that corresponding outcome. This is also reflected at the group level as variability in the group’s mean PROM score over time without an overall change in the corresponding outcome for the group. Conceptually, an MCID should be related to the baseline random variability in PROM score.

**Calculation methods and corresponding derivations.** The 2 primary statistical parameters reflecting spread most commonly used for distribution-based calculation of MCID are the standard deviation (SD) and standard error of measurement (SEM). SD and SEM, although both descriptors of statistical spread, represent different aspects of variability, and this will be reflected in the rationale for using one over the other.20

The SD of PROM scores at baseline can be used in several ways to calculate a MCID. The most common and well-described distribution-based formula for calculation of the MCID of a PROM using SD is

$$\text{MCID} = 0.5 \times \text{SD},$$

where SD represents the SD of baseline or pretreatment PROM scores of the patient population being studied.21

Equation (1) was initially proposed after a systematic review of studies that calculated the MCID of PROMs found that the MCIDs of these PROMs, regardless of methodology used to calculate them, converged to roughly half of the SD of the baseline PROM scores.21 The authors hypothesized that the remarkable consistency of equation (1) was related to the previously reported finding that the upper limit of discrimination by individuals appears to be a 7-point scale22: assuming patients’ scores on a 7-point scale having a rectangular distribution that is 7 points wide (and therefore an SD of 2.16), if the MCID is a 1-unit change on the 7-point scale, then the MCID would be equal to 1/2.16 or roughly 0.5 SD. A subsequent study and systematic review23 has suggested that MCID = 0.3 SD, but repeat analysis of that study’s data set could also be interpreted as consistent with MCID = 0.5 SD.24

Defining MCID in terms of SD can also be thought of with respect to an effect size for the change in the PROM. As its name implies, the effect size refers to the expected magnitude of change in a group’s mean PROM score from one time point to another, for example, after a response to a treatment.25 The effect size, represented with Cohen’s d, is calculated as the difference in a group’s mean PROM score from pretreatment to posttreatment divided by the standard deviation of the pretreatment scores26:

$$d = \frac{\text{Score}_{\text{Mean Posttreatment}} - \text{Score}_{\text{Mean Pretreatment}}}{\text{SD}_{\text{Pretreatment}}}. \quad (2)$$

Values of Cohen’s d corresponding to small, medium, and large effect sizes have been previously recommended as 0.2, 0.5, and 0.8.26 If the minimal change in a cohort’s mean PROM score representing a clinically important difference reflects the MCID (ie, MCID = \(\text{Score}_{\text{Mean Posttreatment}} - \text{Score}_{\text{Mean Pretreatment}}\), then the MCID can be defined as MCID = \(d \times \text{SD}_{\text{Pretreatment}}\). The remaining consideration in this calculation of the MCID is what value of d (reflecting which effect size) to assume. The choice for the value of d naturally may depend on the clinical situation and the outcome itself.25 Without an a priori rationally chosen value of d, a clinically significant difference is usually considered at least a medium effect size (ie, MCID ≥ 0.5 SD), based on Cohen’s description that “medium [effect sizes] represent an effect likely to be visible to the naked eye of a careful observer.”27 However, previous calculations of MCID have also assumed a small effect size,28 which although intended to be noticeably smaller than a medium effect size, is described by Cohen to be “not so small as to be trivial.”27

Distribution-based calculation of MCID using SD has therefore been proposed as 0.2, 0.3, or 0.5 times the pretreatment SD of PROM scores, depending on the study and underlying assumptions of effect size. However, although the use of SD for distribution-based calculation of MCID is well established,12,14 a disadvantage is that SD is a property of the cohort being studied so the resultant MCID calculation may be sample dependent.20 This may make the MCID calculation less generalizable.

In contrast to SD, according to classical test theory the SEM of PROM scores is independent of the patient cohort being studied and an intrinsic property of the PROM assessment tool. From the standpoint of broad generalizability, this characteristic of the SEM thus makes it a more desirable metric upon which to calculate the MCID as an unbiased characteristic of the PROM. Several different methods of calculating MCID using SEM have been described. Because in all cases MCID is calculated as a function of SEM, it is first necessary to calculate the SEM associated with the PROM. In classical test theory,

$$\text{SEM} = \text{SD} \times \sqrt{1 - r}, \quad (3)$$

where r is the PROM’s reliability. Previous studies of MCID have used Cronbach’s α as a metric for reliability to calculate SEM for MCID.29,30 Although it is conveniently calculated, Cronbach’s α is a suboptimal metric of reliability for determining SEM in the calculation of MCID.31 The conceptual rationale for using distribution-based methods is that MCID should be greater than any random variation that is observed in a group’s mean PROM score without a corresponding clinical change. Cronbach’s α, while a measure of reliability, is a reflection of internal consistency—how well the individual items of a PROM are reflective of the outcome they purportedly measure.32 This is not, however, the type of reliability that is the conceptual basis for using SEM to measure MCID. Instead, the most appropriate metric of reliability that should ideally be used to determine SEM for calculation of MCID is test-retest reliability.33,34 Test-retest reliability is calculated by evaluating the correlation between scores on the PROM for a group of individuals at 2
time points separated by a period of time that is too short for any change to occur in the outcome measured by the PROM but too long for each individual to recall how he or she completed the PROM at the first time point. Test-retest reliability therefore represents the metric of reliability that underlies how distribution-based methods were derived to determine clinical importance (ie, change in a PROM, which would exceed random day-to-day variation).

MCID has been independently proposed to be equal to 1 SEM, 1.96 SEM, and 2.77 SEM. The derivation of

\[
\text{MCID} = 1 \text{ SEM}
\]

was based on secondary analyses of validated PROMs used in randomized controlled trials.\(^2^9\)\(^3^0\) Equation (4) has been independently confirmed in other studies as well.\(^3^6\)\(^3^7\) More stringent definitions of MCID have been proposed as follows:

\[
\text{MCID} = 1.96 \text{ SEM}
\]

and

\[
\text{MCID} = 1.96 \sqrt{2} \text{ SEM} = 2.77 \text{ SEM}.
\]

Equation (5) has been proposed to represent MCID as a change in a cohort’s mean pretreatment PROM score that is greater than random variation in PROM score in a statistically significant manner (ie, the MCID is beyond the 95% confidence interval of the expected random variation in PROM scores—there is less than a 5% chance that the proposed MCID value is within the expected random variability of PROM scores).\(^5\) Equation (6) was derived from equation (5) to include a multiplicative term of \(\sqrt{2}\) since MCID reflects the difference in a cohort’s mean PROM scores at 2 time points.\(^5\) In contrast to the underlying assumptions of equations (5) and (6), equation (4) suggests that an MCID may be smaller than the expected random variation in the corresponding PROM.\(^3^4\)

Although SEM has the benefit of being sample independent, there are nevertheless limitations to the use of SEM for calculating MCID. While equations (4) to (6) have been derived using specific rationales, studies have suggested that the relationship between SEM and MCID is not universally true or consistent in all circumstances, with MCID reportedly equal to larger multiples of SEM for cohorts of patients with higher baseline severity of disease\(^3^6\)\(^3^9\) (ie, patients with higher baseline disease severity require greater improvement in the measured outcome for clinically importance\(^3^4\)). Moreover, like the use of SD for MCID calculation, SEM-based methods also suffer from the same limitation of all distribution-based methods, which de Vet and colleagues\(^3^1\) described as “the major disadvantage of all distribution-based methods is that they do not, in themselves, provide a good indication of the importance of the observed change.”

---

**Anchor-Based Methods**

**Conceptual framework.** Anchor-based methods of calculating a PROM’s MCID use an explicit scale for patients to rate the change in the outcome that is measured by the PROM.\(^1^0\)\(^4^0\) This explicit scale is referred to as the “anchor”; it is applied at the posttreatment time point to patients and serves as an external standard against which changes in the PROM score can be compared. For example, an anchor question related to change in general health status after an intervention might be, “Since [intervention], how has your general health changed?” An anchor question can be designed to give patients any number of discrete answer options (referred to as “transition ratings”).\(^3^1\) The number of transition ratings provided with an anchor question is tailored to the outcome or assessment tool being studied or the disease/condition that it is being applied to. Response to the anchor question “Since [intervention], how has your general health changed?” may have only 3 simple choices: “worse,” “no change,” or “better.” By convention, anchor questions have an odd number of possible answer options (centered on no change in the outcome), most often ranging from 3 to 7 answer options (Figure 1).\(^1^4\) However, anchor questions with 15 transition ratings have also been used, although the transition ratings are often grouped for analysis to effectively reduce the number of transition ratings from 15 to 7.\(^1^0\) To calculate an MCID in a group of patients using an anchor-based method, patients’ scores on the PROM are compared to how the patients responded to the anchor.
Calculation methods and corresponding derivations. Anchor-based methods can be performed cross-sectionally or longitudinally and, in each case, have been implemented in multiple ways. Cross-sectional anchor-based methods, for example, compare the mean scores on a PROM for a cohort reported for each transition rating on an anchor question to determine the relationship between PROM score and reporting of one transition rating over another. However, because the concept of the MCID is most frequently applied in the context of assessing improvement (eg, after an intervention) and is based on change in a PROM, this review will focus on longitudinal anchor-based methods.

Longitudinal anchor-based methods generally calculate MCID by comparing the changes in PROM scores for patients in a cohort after an intervention to responses on an anchor question that is assessed after the intervention. Although a number of longitudinal anchor-based methods for calculating MCID have been described, this calculation is performed in 2 primary ways: comparison of mean changes in PROM scores for patients reporting each transition rating and through the direct study of change in PROM score as a tool to detect individual patients who have experienced clinically important improvement on the anchor question out of all patients in the cohort. In all cases, an MCID must first be assigned to the anchor question (ie, which transition rating on the anchor question represents a clinically important improvement). Frequently, the transition rating corresponding to “a little” improvement is considered the MCID of the anchor question, but this depends on the clinical scenario and the outcome being studied.

Comparison of changes in PROM scores in relation to transition ratings can be performed in a number of ways (Figure 2). The most direct method is to simply calculate the mean PROM score change for patients reporting the transition rating corresponding to the anchor question MCID and equate that value to the PROM’s MCID (Figure 2, solid double arrow). Some investigators subtract the mean PROM score change for patients reporting the immediately adjacent and lower transition rating or for patients reporting no change (Figure 2, dotted double arrow).

An alternative anchor-based method for deriving an MCID is by conceptualizing change in the PROM score as a diagnostic test for detecting patients in a cohort who have experienced clinically important improvement in the outcome (ie, reporting at least the MCID transition rating, or greater improvement, on the anchor question) (Figure 3). In this approach, the anchor question serves as the “gold standard” for whether the patient has experienced clinical improvement. Most commonly with the use of receiver operating characteristic (ROC) analysis, the MCID of the PROM is calculated as the threshold change in

![Figure 2](image-url)  
**Figure 2.** Mean change in score on the 22-item Sinonasal Outcome Test (SNOT-22)—a patient-reported outcome measure (PROM) reflecting disease-specific quality of life—for patients with chronic rhinosinusitis receiving medical management is plotted against transition rating. Data and graphics adapted from Figure 3 of Phillips et al. The minimal clinically important difference (MCID) calculation using just the mean change in SNOT-22 scores for patients reporting the MCID of the anchor and subtracting the mean change in SNOT-22 scores for the immediately adjacent (and lower) transition rating is shown with the dotted double arrow. The MCID calculation using the mean change in SNOT-22 scores for patients reporting the MCID of the anchor is shown with the solid double arrow. The MCID calculation using the mean change in SNOT-22 scores for the immediately adjacent (and lower) transition rating is shown with the dotted double arrow.

![Figure 3](image-url)  
**Figure 3.** Change in score on the 22-item Sinonasal Outcome Test (SNOT-22)—a patient-reported outcome measure (PROM) reflecting disease-specific quality of life—for patients with chronic rhinosinusitis is plotted against transition rating. Data and graphics adapted from Figure 3 of Phillips et al. The data are separated by a horizontal, dotted green line that represents the minimal clinically important difference (MCID) of the SNOT-22 while the vertical solid green line represents the MCID of the anchor question. The MCID of the PROM (SNOT-22) is derived as the value that maximizes the sum of sensitivity (true positives/true positives + false negatives) and specificity (true negatives/true negatives + false positives).
Detectable vs Clinically Important Differences

In the discussion of MCID, it is important to touch upon the concept of the minimally detectable difference (MDD), which is sometimes used in conjunction with the MCID. Statistically, the MDD is the smallest change in the PROM that is detectable by the patient or the smallest change in the PROM that exceeds the normal random variation in PROM scores that occur in stable patients. The MCID differs from the MDD by reflecting a judgment on the change in a PROM score as minimally important. Intuitively, the MCID may be expected to be greater than the MDD, but this is not always the case, as on a population level, a clinically important difference does not necessarily equate to a change in PROM score greater than its random variation in a statistically significant manner. While the MDD is an informative metric and sometimes a close approximate for the MCID, it is not a substitution for the MCID, which is calculated with the judgment of clinical importance in mind.

Implications for Practice

The MCID—while conceptually straightforward—is multifaceted in practice. The MCID not only varies based on the patients and clinical context being studied but also is derived through varied methods each with specific underlying assumptions that affect the value and accuracy of the final result. The concept of MCID is useful and necessary in setting numerical thresholds for clinically meaningful improvement for PROMs that are increasingly used in clinical studies. The MCID, however, should not be blindly applied or accepted as universal fact. Instead, application of MCID to the interpretation of clinical data must be done judiciously, keeping in mind the subtleties of its derivation as well as its limitations, which are summarized in Table 1. These considerations are used as motivation for 3 recommendations below for understanding, applying, and interpreting MCID with respect to clinical data.

Recommendation 1: The interpretation and application of a PROM’s MCID should include consideration for the patient population in which it was calculated. The MCID of a PROM may be different depending on the disease and the patient populations to which it is given. When interpreting the relevance of an MCID, one must carefully consider whether the patient population in which it was calculated is reflective of the patient population to which it is being applied (eg, similar clinical context, disease severity, and treatment given). Most important, compared to the patient population to which the MCID is being applied, did the patient population in which the MCID was derived have the same disease and reasonably similar expectations for how much improvement in the PROM might be noticeable or significant?

Finally, it is important to understand that application of the MCID can be to a group of patients or to individual patients. For example, a treatment may be applied to a
Recommendation 2: Determination of a PROM’s MCID should be through the synthesis of calculations using multiple approaches, including distribution-based and anchor-based methods. Although there is no universally accepted methodology for calculating MCID, it is generally accepted that the preferred approach is to use and report the results from both distribution-based and anchor-based methods, which are complementary to each other. A recent systematic review of studies reporting MCIDs showed that almost half of all studies used both distribution-based and anchor-based methods. Among studies using only 1 approach, anchor-based MCID calculations were used approximately 7 times more frequently than distribution-based calculations. Although MCID should ideally be derived through the synthesis of both distribution-based and anchor-based calculations, if faced with the choice of using 1 of the 2 approaches, anchor-based methods are preferred since they are based on an explicit definition of clinical significance in the form of the anchor. Distribution-based methods, which incorporate the precision of the PROM, nevertheless serve as important and independent means for confirming the numerical range of MCID results established by anchor-based methods. There is no consensus for how to reconcile differing values of MCID that result from different calculation methods. Proposal of an MCID value from different possible choices should include consideration for accuracy (sensitivity and specificity of the proposed MCID to identify patients with clinically important change), methodology used (ie, anchor-based methods preferable over distribution-based), and the general consensus (or lack thereof) by at least some of the methods used (eg, do all anchor-based or all distribution-based methods suggest the approximate same value or do even methods within the same category suggest wildly different values?). Moreover, these considerations should ideally be transparently and explicitly reported in studies proposing MCIDs so that potential future users of the MCID can also make their own judgments.

Recommendation 3: Interpretation and application of a PROM’s MCID should include consideration for its accuracy in identifying patients experiencing clinical improvement. Previous study has shown that a significant number of patients not experiencing an MCID of improvement on a PROM may nevertheless report having experienced a clinically important improvement. For this reason, reported PROM MCIDs would ideally be reported with their accuracy in distinguishing patients who have experienced clinically significant improvement from those who have not. In fact, the US Food and Drug Administration (FDA) has placed an emphasis on the accurate identification of individual treatment responders rather than generalizations about efficacy by broad application of an MCID to changes in group-level mean PROM scores.

When reported, analyses of PROM MCIDs’ accuracies to detect patients experiencing clinically important improvement have shown MCIDs can be more specific than sensitive and, in some cases, poorly sensitive. Therefore, the application of an MCID should be done with its accuracy (and its limitations in that respect) in mind. For example, as MCID has generally been shown to be more specific than sensitive, and it may be well used in a scenario where strict criteria are needed for establishing clinically important improvement. On the other hand, MCID may be ill-suited for situations where detecting patients who have experienced clinically significant improvement is a higher priority than avoiding false-positive detection of patients not experiencing clinically significant improvement.
Conclusion

The MCID is a useful and necessary tool in the interpretation of clinical improvements assessed using PROMs. However, the determination of a PROM’s MCID is fraught with subtleties and assumptions, all of which are different depending on the methodology used and all of which can affect the final result. Application and interpretation of an MCID to clinical outcomes data require knowledge of the methodologies used to calculate the MCID, as well as their strengths and limitations.

Author Contributions

Ahmad R. Sedaghat, performed the literature review, synthesized the literature review, wrote/revised the manuscript.

Disclosures

Competing interests: None.

Sponsorships: None.

Funding source: None.

References

22. Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. 1956. Psychol Rev. 1994;101:343-352.
30. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful
60. de Vet HC, Terwee CB. The minimal detectable change should not replace the minimal important difference. J Clin Epidemiol. 2010;63:804-805, author reply 806.


