A Standard Method for Determining the Minimal Clinically Important Difference for Rehabilitation Measures

James F. Malec, PhD, FACRM, ABPP-Cn, Rp (Corresponding author)
Senior Research Professor Emeritus,
Physical Medicine and Rehabilitation,
Indiana University School of Medicine
Emeritus Professor of Psychology, Mayo Clinic
6722 Meadow Lawn Circle, New Market, MD 21774
jfmalec@iupui.edu
Phone: 507-202-5103

Jessica M. Ketchum, PhD
Research Department, Craig Hospital, Englewood, CO
Traumatic Brain Injury Model Systems National Data and Statistical Center, Englewood, CO
3425 S. Clarkson St, Englewood CO 80113
JKetchum@craighospital.org
Phone: 303-789-8682
Fax: 303-789-8441

This work was partially supported by Grant Funding for the NIDILRR - Traumatic Brain Injury National Data and Statistical Center (90DP0084). Reprints are not available for this article. The authors have no conflicts of interest to declare.

This is the author's manuscript of the article published in final edited form as:

Malec, J. F., & Ketchum, J. M. (2020). A Standard Method for Determining the Minimal Clinically Important Difference for Rehabilitation Measures. Archives of Physical Medicine and Rehabilitation. https://doi.org/10.1016/j.apmr.2019.12.008

A Standard Method for Determining the Minimal Clinically Important Difference for Rehabilitation Measures

Abstract

The Minimal Clinically Important Difference (MCID) is receiving increasing interest and importance in medical practice and research. The MCID is the smallest improvement in scores in the domain of interest which patients perceive as beneficial. In clinical trials, comparing the proportion of individuals between treatment and control groups who obtain a MCID may be more informative than comparisons of mean change between groups since a statistically significant mean difference does not necessarily represent a difference that is perceived as meaningful by treatment recipients. The MCID may also be useful in advancing personalized medicine by characterizing those who are most likely to benefit from a treatment. In clinical practice, the MCID can be used to identify if a participant is experiencing a meaningful change in status.

A variety of methods have been used to determine the MCID with no clear agreement on the most appropriate approach. Two major sets of methods are either (1) distribution-based, i.e., referencing the MCID to a measure of variability or effect size in the measure of interest, or (2) anchor-based, i.e., referencing the MCID to an external assessment of change in the condition, ability, or activity represented by the measure of interest. In prior literature, using multiple methods to "triangulate" on the value of the MCID has been proposed. In this commentary, we describe a systematic approach to triangulate on the MCID using both distribution-based and anchor-based methods. Adaptation of a systematic approach for obtaining the MCID in rehabilitation would facilitate communication and comparison of results among rehabilitation researchers and providers.

Minimal Clinically Important Difference

Key Words: Quality of Health Care; Outcome and Process Assessment (Health Care); Patient Outcome Assessment; Minimal Clinically Important Difference

List of Abbreviations

GIC Global impression of change

IRT Item-response theory

MCID Minimal clinically important difference

MCR Mean change response

MOI Measure of interest

RCI Reliable change index

RCID Robust clinically important difference

ROC Receiver operating characteristic

SEM Standard error of measurement

The Minimal Clinically Important Difference (MCID) is gaining increasing interest and importance in medical research and practice. Jaeschke and colleagues[1] originally proposed the concept in 1989 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." As such, use of an anchor-based MCID as described below epitomizes a marked departure from traditional statistics, such as, significance testing (*p*-values) and effect sizes.

Traditionally, studies have estimated and compared the average improvement (change) in an outcome measure of interest (MOI) between treatment and control groups using statistical tests to determine if the improvement in the treatment group is "statistically significantly" greater

than that in the control group, beyond what is expected by chance (i.e., p-value < 0.05). However, statistical significance does not necessarily equate to *clinical* significance. Large sample sizes have the power to find that small differences that *are not* clinically meaningful are statistically significant, and small samples sizes lack the power to demonstrate that large differences that *are* clinically meaningful are statistically significant.

The magnitude of the within-group improvement or the between-group differences in improvement (relative to the variability at baseline) are often reported as measures of within- and between-group effect size. For example, Cohen's d-family effect sizes of 0.2, 0.5, and 0.8 are commonly interpreted as small, medium, and large effect sizes, respectively. Statistical testing can be used to answer the question: "Does treatment group improve more than control group beyond what we would expect by chance?" Effect sizes attempt to extend interpretation beyond statistical significance towards clinical significance by answering the question: "On average, is the within group improvement (or between group comparison of improvements) small or large (relative to the degree of variability across subjects)?" However, this is still an interpretation of treatment effect at the group level, that is, the average response to treatment across many individuals, and may not reflect the treatment effect for a particular individual. The degree of improvement may vary considerably across individuals and may be dependent upon subject-specific characteristics (measured or unmeasured). Furthermore, the effect size is not expressed in units of the MOI and is not interpretable at the individual level (e.g., in a clinical setting when presented with a single patient's pre- and post-treatment values). Finally, and perhaps most importantly, the effect size may not reflect what the persons served consider a meaningful difference in their quality of life.

The MCID is expressed in the same units as the MOI and can be more appropriately used in a clinical setting to identify if a specific individual has had a meaningful response to treatment when making decisions to continue or alter treatment. The MCID can also be used in a research setting and for program evaluation to better understand treatment effect, enabling researchers to quantify the proportion of people who had a meaningful response to treatment. In addition, this proportion can be interpreted at the individual level as an estimate of the probability that an individual will respond to treatment. Statistical significance testing can be used to compare the proportions of *responders*, i.e., those achieving a MCID or better, between treatment and control groups to determine if the response rate is greater in the treatment group beyond what would be expected by chance. Additional analyses can be conducted to describe and compare characteristics between responders and non-responders to identify subject-specific factors that are associated with increased or decreased likelihood (or probability) of response to treatment.

Without a good appreciation of how much improvement is actually meaningful to persons served, studies may not be appropriately powered to detect clinically meaningful differences. The MCID can be used to better design studies so that statistical and clinical significance are more aligned. Studies can be powered to have sufficient sample size to detect a meaningful change rather than a statistically significant difference based on effect sizes. For example, studies are often powered to detect a Cohen's d effect size of 0.5 (difference/SD), which could represent different magnitudes of change depending on the SD. Furthermore, studies are often underpowered to conduct analyses assessing response to treatment as comparisons of the proportion of responders between groups and the factors associated with the likelihood of response often require larger sample sizes than comparisons of mean change. Studies of treatment efficacy should be adequately powered to have sufficient sample to detect differences

in mean changes between groups as well as difference in the proportion of responders between groups in order to maximize understanding of the treatment effect being studied.

Statistical testing, effect size, and MCID each provide researchers and clinicians with unique information regarding treatment efficacy at the group level. However, the MCID can also be used specifically at the individual level. It is expressed in the same units as the MOI making it easily implemented in a clinical setting when considering continuing or altering treatment. Response to treatment analyses and advances in personalized medicine research can help further guide clinicians' treatment selection by identifying subject-specific characteristics associated with increased or decreased likelihood of treatment response. As we will describe in this paper, the value of the MCID, like other measures of treatment effect, can be obtained in a statistically reliable manner but may be substantially different in value from measures of effect size or other types of distribution-based indicators.

Despite its potential value, computation of the MCID is controversial as a recent exchange of Letters to the Editor in the *Archives* illustrates.[2] Two major methods have been proposed to derive the MCID: a distribution-based approach and an anchor-based approach.[3-9] The distribution-based approach references statistical indicators of significant change, such as, the standard error of measurement (SEM), indicators of various effect sizes, such as, a standard deviation (SD), or factors of these basic indicators. The anchor-based approach estimates the MCID in reference to another estimate of meaningful change by the person served or a service provider. Most commonly, a Global Impression of Change (GIC) rating is used as the anchor. Within the anchor-based approach, the degree of change in the MOI that indicates meaningful change is derived either by a mean change response (MCR) or a receiver operating characteristic (ROC) analysis. MCR compares the means of individuals indicating improvement on the anchor

to those who do not report improvement. ROC provides a similar comparison based on the proportion of agreement between the MOI and the anchor. ROC analysis yields sensitivity, specificity, and accuracy statistics, similar to evaluation of diagnostic procedures or of other types of classification analyses.

Early descriptions recommended using multiple methods to determine the MCID and then "triangulating" on the best value.[4-5] However, a specific or systematic method for this triangulation has not been suggested. Subsequently, methodologists have favored an anchor-based approach and emphasized the importance of representing the perspective of the person served in determining meaningful change.[3-5,9]

Studies attempting to identify the MCID for various measures have used a wide variety of methods. In their review, Engel and colleagues[9] describe the methods used and found that only about half used an anchor-based approach. In practice, a distribution-based value in the neighborhood of ½ SD has typically been identified as the MCID and has been recommended for use in the absence of an empirically established value.[5-6] We will not comprehensively review this literature; the interested reader is referred to recent reviews[8-9] and other papers cited previously for more detailed information about the methods and history of the MCID. In this commentary, we describe a method for systematically "triangulating" on the most appropriate value for the MCID using both distribution-based and anchor-based approaches. We have used a similar method previously.[10] In this paper, we present this method systematically and add additional reliability tests. We believe the method described here is appropriate for use with many standard rehabilitation measures and suggest that the use of a consistent method to derive MCIDs in rehabilitation will support communication about and comparability across studies. We have previously suggested that an indicator of a substantial improvement in status, the Robust

Clinically Important Difference (RCID), might also be determined to identify cases in which change is not only minimally meaningful but impressive.[11] The method described here systematically identifies values both for the MCID and RCID.

Method

The proposed method for systematically identifying the value of the MCID is fundamentally an anchor-based method. We agree with others cited previously that an anchor-based method is preferable to using distribution-based indicators alone. However, we also believe that distribution-based indicators provide familiar and well-accepted benchmarks for evaluating measurement error and effect size. Consequently, initial steps in the proposed method determine a range of distribution-based indicators that are then further evaluated through anchor-based procedures.

The recommended distribution-based indicators are the standard error of measurement (SEM), the baseline (or pre-treatment) standard deviation (SD), and three factors of these basic indicators: ½SD, 1.96SEM, and the Reliable Change Index (RCI). Their values range from the smallest amount of change that can be determined by the MOI (i.e., SEM) to very large change (i.e., 1 SD).

We agree with Engel and colleagues[9] that, when evaluating the proposed MCID in reference to an anchor, a ROC approach is preferable to a MCR approach. The MCR approach compares the mean change between those achieving the minimum amount of change (responders) and those who do not (non-responders), and consequently may not be sensitive to a minimally meaningful change among those responders whose change scores fall below the mean change. The method described below is a ROC approach. A ROC computation provides the *sensitivity* and *specificity* of the range of values of the MOI relative to the GIC. *Accuracy* can be

computed by taking the weighted sum of sensitivity and specificity, with weights corresponding to proportion of individuals above and below the MCID (i.e., the prevalence). We also recommend computing Youden's Index.[12] By combining sensitivity and specificity, Youden's Index provides an overall indicator of the performance of these metrics and, unlike the other more familiar indicators, is independent of the prevalence of responders and nonresponders. Youden's Index can vary between 0 and 1 with higher values indicating a smaller overall proportion of false negatives and false positives. Definitions and formulas for these metrics are provided in Text Box 1.

Since the validity of an anchor-based approach assumes that the anchor is representative of , that is, is associated with change on the MOI, the correlation between the anchor measure and change on the MOI is computed prior to any other computations. A Spearman correlation is suggested since most anchors, including the GIC recommended here, are ordinal measures. While a correlation of at least .3 to .35 has been recommended as a minimum correlation between the change score and the anchor,[5] we suggest that a stronger correlation indicating at least 50% or better shared variance (i.e., correlation of .7 or higher) provides greater confidence that the anchor is sensitive to change on the MOI and that both these measures represent the same construct.

The change in the MOI and the anchor may not be adequately correlated for a number of reasons. Most commonly, (a) the MOI change score does not have adequate reliability or precision; (b) the time between measurements on the MOI is too great, leading to recall bias or response shift; or (c) the MOI is unreliable because of the participant's impaired self-awareness. Lack of reliability or precision in the MOI and consequently MOI change can be avoided by carefully selecting statistically sound measures for evaluation. Measures with an interval level of

scaling are required since, with such measures, the change score will indicate the same degree of change regardless of the initial level on the measure. Ordinal measures can be transformed to interval scaling through Rasch or other item-response theory (IRT) procedures. Recall bias is distortion in perceived change due to difficulty in recalling the progression of one's condition over an extended period of time. An optimal period of time between initial and final measurement has not been well-defined and may vary with the MOI and the anchor.[13] Response shift refers to a change in one's perception of one's condition over time. In other words, the factors that the rater considered in making the initial rating changed over the course of time and are different at the time of the final rating. Unreliability due to recall bias or response shift can probably not be addressed retrospectively and most likely prevents a valid MCID determination. Unreliability due to impaired self-awareness is also difficult to address retrospectively in participant ratings. In such cases, ratings made by a more objective observer are preferable for determining the MCID.

If a lack of correlation between change in the MOI and the anchor prevents computation of the MCID, ½ SD may be used as the putative MCID, as recommended by others.[5-6] Alternatively, if a more conservative estimate of the clinically important difference is appropriate in the context of the research, the RCI may be used. This is the approach we used in prior work[14] in which a substantial correlation between change in the MOI and anchor was not obtained due to the extended time (5 years) between measurements of the MOI. Computing both these proxy values mirrors the derivation of a MCID and RCID. However, the use of such proxy values should only be used in specific research situations (e.g., the time between measurements is extremely long) where derivation of the MCID and RCID is not possible. Proxy values should not be substituted for systematic and precise derivation of the MCID and RCID in the long term.

Example

The basic steps for obtaining the MCID and RCID briefly are as follows: (1) obtain a representative sample (being aware that the MCID may vary among samples of varying severity of illness, chronicity, demographic, and other factors), (2) determine if the correlation between the MCID and the anchor is adequate to proceed (≥ 0.7), (3) compute the sensitivity, specificity, accuracy, and Youden's Index for the MOI relative to the GIC in identifying those who indicate that their condition is "Better" or "Much Better", and then (4) select the MCID and RCID corresponding to the highest accuracy and optimal sensitivity and specificity as indicated by Youden's Index. A more detailed, step-by-step description of the method is provided as Supplementary Material 1.

To demonstrate this method, we have constructed a mock data set (available as Supplemental Material 2) consisting of 100 cases. In this mock data set, the MOI was expressed as an integer (no decimal values) T-score between 0 and 100 with a mean of 50 and a SD of 10 at time 1. Reliability (*r*) was assumed to be 0.9. GIC values on an ordinal scale indicating much worse (-2), a little worse (-1), about the same (0), a little better (1), much better (2) were selected to generally agree with change on the MOI; however, to mirror reality, some values did not agree. The Spearman correlation between change on the MOI and the GIC was .88. With an adequate correlation between MOI change and the GIC (≥ 0.7), distribution-based indicators were calculated as follows: SEM=3.2; ½ SD=5.0; 1.96×SEM=6.2; RCI=8.8; 1 SD=10.0. Because the measure was integer-based, decimal values for the distribution-based indicators were rounded to the nearest whole integer as displayed in Tables 1-3.

In this example, inspection of Table 1 shows both 1.96×SEM and .5 have the same accuracy and acceptable sensitivity and specificity. However, Youden's Index favors .5 SD as

the MCID. Inspection of Table 2 suggests the RCI and 1 SD as possible values for the RCID; both show good accuracy, sensitivity and specificity. However, RCI has a slightly higher Youden's Index and is selected as the potential RCID. These proposed values are then evaluated for the entire sample (Table 3). Inspection of Table 3 shows that both the proposed MCID and RCID continue to perform well for the entire sample and are selected as the final MCID and RCID.

Concluding Comments

Determination of the MCID for statistically reliable measures used in rehabilitation has significant potential value as described in the introduction to this paper. In contrast to effect size, the MCID is expressed in the units of the measure itself rather than referenced to the variability of its distribution and represents the smallest change that is clinically significant and meaningful to the person served. As such, the MCID for a measure may vary across different populations (e.g., diagnostic groups) as well as with severity, chronicity, demographic and other factors within these populations. Effect size represents the magnitude of change between or within treatment groups and is not indicative of individual treatment response. Whereas, the MCID represents a degree of change that will be perceived as meaningful by most persons served and can be used to inform individual treatment decisions.

Measures developed using IRT should be used in deriving the MCID since they are reliable and are equivalent to interval measures in providing change scores of consistent value regardless of the initial level of the measure. While the impression of the treatment recipient is of paramount importance in determining whether a meaningful change has been obtained, the impression of a more objective observer is also of value, particularly in assessments in which there is substantial risk of impaired self-awareness on the part of the treatment recipient. The

method described here focused on the evaluation of positive change since this is most often of interest in evaluating rehabilitation interventions. However, a similar method might also be used to evaluate negative change or deterioration. We attempted to provide a clear and straightforward approach to determining the MCID. Nonetheless, as shown in our example, reliability indicators for potential values may be very similar and some judgement may be required in the final determination.

We proposed the determination of a RCID in addition to the MCID. However, we wish to emphasize the value of determining the minimal change that is meaningful to participants and providers, and to caution against the ascendance of the RCID as a more important indicator. The RCID is of interest only in identifying those who had an outstanding response to treatment. In some fields, the RCI is embraced as the premier measure of significant change. However, while the RCI indicates a value that is very unlikely to occur by statistical chance, it does not address the issue of meaningful change since it is derived from a distribution-based approach. As described in MCID reviews and studies cited previously, participants may *reliably* perceive a meaningful change at a level much less than the RCI.

As investigations of methods for personalized medicine expand, both the MCID and RCID should be useful in characterizing individuals who benefit from specific treatments. On the historic timeline for the development of scientific methods (which can span a century), the MCID—first proposed 30 years ago—is just reaching adolescence. Consequently, further evolution of this concept and methodologies can be expected. For example, the best method to compute the standard error used in the calculation of some distribution-based indicators is debated,[15] as can be the optimal measure for reliability. We have proposed that the correlation between the MOI and the anchor should be relatively strong, i.e, .7 or higher, while others[5] have

suggested a correlation as low as .3. Future systematic empirical investigation is required to determine the recommended correlation between the MOI and the anchor. In the interim, adaptation of a consistent approach to determining the MCID in rehabilitation will support clearer communications and comparison of results among rehabilitation providers and researchers.

References

1. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. Control Clin Trials 1989;10(4):407–15.
2. Letters to the Editor. Ovacik U, Çelik D. Minimal clnically important difference of Berg Balance Scale in People with multiple sclerosis; Cattaneo D, Gervasoni E, Montesano A, Jonsdotir J. Response to letter regarding “Minimal clnically important difference of Berg Balance Scale in People with multiple sclerosis.” Arch Phys Med Rehabil 2019;100:1191-2.
3. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, Group TCSCM. Methods to explain the clinical significance of health status measures. Symposium on Quality of Life in Cancer Patients. Mayo Clinic Proc 2002;77:371-83.
4. Copay AG, Subach BR, Glassman SD, Polly DW, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. Spine J 2007;7:541-6.
5. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008;61:102-9.
6. 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.
7. Wright A, Hannon J, Hegedus EJ, E. KA. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). J Man Manipul Ther 2012;20:160-6.

8. Jayadevappaa R, Cooka R, ChhatreeS. Minimal important difference to infer changes in health-related qualityof life: a systematic review. J Clin Epidemiol 2017;89:188-98.
9. Engel L, Beaton DE, Touma Z. Minimal clinically important difference: a review of outcome measure score interpretation. Rheum Dis Clin N Am 2018;44:177–188.
10. Malec JF, Hammond FM. Minimal clinically important difference for the Rasch Neuropsychiatric Inventory Irritability and Aggression Scale for Traumatic Brain Injury. Arch Phys Med Rehab 2018;99:603-606.
11. Malec JF, Kean J, Monahan PO. (2017). The minimal clinically important difference for the Mayo-Portland Adaptability Inventory. J Head Trauma Rehab 2017;32: E47-E54.
12. Youden WJ. Index for rating diagnostic tests. Cancer 1950:3:32–5.
13. Alma HJ, de Jong C, Jelusic D, Wittmann M, Schuler M, Kollen BJ, Sanderman R, Schultz K, Kocks JWH, Van der Molen T. Assessing health status over time: impact of recall period and anchor question on the minimal clinically important difference of COPD health status tools. Health Qual Life Outcom 2018;16:130-43.
14. Hammond FM, Malec JF Corrigan JD, Ketchum JM, Whiteneck GG, Hart T, Dahdah M, Dams-O'Connor K, Novack TA, Bogner J, Eagye CB, Sevigny M. Functional change from 5 to 10 years following moderate to severe traumatic brain injury. In preparation.
15. Temkin N. Standard error in the Jacobson and Truax Reliable Change Index: The "classical approach" leads to poor estimates. J Int Neuropsych Soc 2004;10:899-901.

Table 1. Agreement between distribution-based indicators and classification values of GIC = Better vs. No Change, Worse or Much Worse.

	Accuracy	Sensitivity	Specificity	Youden's Index
SEM=3	.74	1.00	.62	.62
.5SD=5	.80	.88	.76	.64
1.96SEM=6	.80	.77	.81	.58
RCI=9	.79	.42	.95	.37
1SD=10	.77	.31	.98	.29

Table 2. Agreement between distribution-based indicators and classification values of GIC = Much Better vs. No Change, Worse or Much Worse.

	Accuracy	Sensitivity	Specificity	Youden's Index
SEM=3	.70	1.00	.62	.62
.5SD=5	.81	1.00	.76	.76
1.96SEM=6	.85	1.00	.81	.81
RCI=9	.96	1.00	.95	.95
1SD=10	.97	.94	.98	.92

Table 3. Agreement between distribution-based indicators and classification values of GIC = Better or Much Better vs. No Change, Worse or Much Worse.

	Accuracy	Sensitivity	Specificity	Youden's Index
SEM=3	.78	1.00	.62	.62
.5SD=5	.83	.93	.76	.69
1.96SEM=6	.83	.86	.81	.67
RCI=9	.82	.64	.95	.59
1SD=10	.80	.55	.98	.53

Supplementary Material 1: Detailed Steps for MCID and RCID Determination

1. Obtain a representative sample, i.e., large as possible to represent the relevant patient group.

   Note: The Minimal Clinically Important Difference (MCID) may vary by severity, chronicity, demographic and other factors.

2. Obtain pre-post measurements on the measure of interest (MOI) and compute change scores.
3. At the time of the post-treatment ratings on the MOI, also obtain ratings of overall improvement relative to pre-treatment on a 5-point scale Global Impression of Change scale (GIC) from participants and providers, i.e., (-2) Much Worse, (-1) Worse, (0) No Change, (+1) Better, (+2) Much Better.
4. Compute Spearman correlation coefficient between GIC and MOI change score; value > .5 may be acceptable; >.7, preferred.
5. Compute distribution-based indicators for scale of interest:
   a. SEM = $SD_{baseline}(\sqrt{1} - r)$
   b. ½ baseline (pre-treatment) SD
   c. 1.96 × SEM
   d. Reliable Change Index (RCI) = 1.96 × ($SD_{baseline}(\sqrt{2(1 - r)}$) = 2.77 × SEM
   e. 1 SD (baseline)

   Note: In the above formulas, *r* = a measure of reliability, e.g., test-retest, Cronbach's alpha, or for Rasch or IRT measures, person reliability.
6. Divide the sample between those indicating "Better" on GIC and those indicating No Change, Worse, or Much Worse; do not include those indicating Much Better.
7. With this dichotomized GIC as the classification value, compute sensitivity, specificity, accuracy, and Youden's Index for the MOI change score at each level of the distribution-based indicators, comparing those at or above the distribution-based indicator to those with change scores below the indicator.

8. Select the distribution-based indicator with the highest accuracy and optimal sensitivity and specificity as indicated by Youden's Index as the proposed MCID.
9. Repeat steps 6-7 dividing sample between those indicating Much Better and those indicating No Change, Worse, or Much Worse; do not include those indicating Better.
10. Select the distribution-based indicator with the highest accuracy and optimal sensitivity and specificity as indicated by Youden's Index as the proposed Robust Clinically Important Difference (RCID).
11. Repeat steps 6-7 dividing sample between those indicating Better or Much Better and those indicating No Change, Worse, or Much Worse.
12. Verify or reconsider MCID and RCID values based on results obtained in #11.

For the sake of brevity, we will only describe the calculation of the first row in Table 2 in the main paper. To make these computations, the sample was divided into those whose MOI change was 3 or more, i.e., a SEM, and those with change less than 3. These were compared to those whose GIC was 1 (Better) and whose GIC was 0 or less (No Change, Worse, Much Worse). As described in Step 6, those with a GIC of 2 (Much Better) were not included. The Table below displays the numbers in each of these categories. Applying the formulas in Table 1, Accuracy = (36+26)/84 = 62/84 = .74; Sensitivity = 26/(26+0) = 1.00; Specificity = 36/(36+22) = 36/58 = .62; and Youden's Index = 1.00+.62-1 = .62. All the other rows in in Table 2-4 can be derived in the same fashion.

Case distribution by GIC and Change of 1 SEM on MOI.			
	GIC ≤ 0	GIC = 1	
Change < 3	True Negatives 36	False Negatives 0	36
Change ≥ 3	False Positives 22	True Positives 26	48
	58	26	84

Text Box 1. Definitions and formulas.
Sensitivity [percent of those improved on the GIC correctly identified by selected cutpoint on MOI change score] = # True Positives / [# True Positives + # False Negatives]
Specificity [percent of those <u>not</u> improved on the GIC correctly identified by selected cutpoint on MOI change score] = # True Negatives / [# True Negatives + # False Positives]
Accuracy [overall correct classification rate] = [# True Positives + # True Negatives] / # Total
Youden's Index = Sensitivity + Specificity – 1.00