

# Statistical Development on Determining the Minimum Clinically Important Difference

## Abstract

In medical research, the concept of minimum clinically important difference (MCID) has gained its popularity among clinical practitioners and health policy makers. Over the past 20 years, intensive research has been conducted to explore how clinical significance could be interpreted from the patient reported outcomes (PROs) by using MCID. This article aims to provide a review on the statistical development on the determination of MCID.

## Mini Review

Volume 2 Issue 5 - 2015

**Tu Xu\***

*Gilead Sciences Inc., USA*

**\*Corresponding author:** Tu Xu, Gilead Sciences Inc., 333 Lakeside Dr., Foster City, CA, 94404, USA, Tel: 3122061073; Email: Tu.Xu@gilead.com

**Received:** May 10, 2015 | **Published:** May 18, 2015

## Introduction

In clinical research, statistical significance is widely used to evaluate the effectiveness of new drugs or medical devices. However, there has been growing recognition that statistical significance could be misleading when evaluating treatment effect [1]. It is known that statistical significance only infers the existence of treatment effect, regardless of the effect size, and therefore the true treatment effect may have little to do with its clinical significance. The statistical significance could result from small sample variability or a huge sample size. For instance, with  $t$ -test statistics  $t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$ , statistical significance could always be declared when the sample size  $n$  is large enough, such as  $n = o(1/d^2)$ , where  $d = E(X) - \mu$ .

The clinically important difference (MCID) was first proposed by Guyatt et al. [2] to provide an appropriate assessment of the clinically meaningful benefit from a post-treatment evaluation. The most influential definition of MCID was established by Jaeschke et al. [3] as "the smallest difference 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". The concept of MCID has quickly gained popularity among clinicians and health policy makers with its interpretation of clinical significance based on health-related patient-reported outcomes (PROs) [4, 5]. King [6] described its popularity by the fact that the number of literature on MCID has got trebled every 5 years over the past 20 year. From the regulatory perspective, FDA's final guidance for industry on PRO measures rolled out in 2009 [7]. In November 2012, FDA hosted a special conference on the MCID for orthopedic devices; c.f.

<http://www.fda.gov/MedicalDevices/NewsEvents/Workshops/Conferences/ucm327292.htm>

for more details.

## Existing methods on determining MCID

Although the importance of MCID has been widely recognized, only a few ad-hoc approaches have been proposed for its estimation. The existing methods are classified into two major types: anchor-based approaches and distribution-based approaches [7].

## Distribution-based approaches

Distribution-based approaches are methods that compare the change in PRO scores to certain measure of its variability such as the standard error of measurement (SEM), minimal detectable change (MDC), the standard deviation (SD), and the effect size (ES). All these methods are trying to capture the individual level of change that is considered above the ranges of measurement error. Wyrwich et al. [8] defined MCID equal to SEM, where  $r$  is the sample correlation coefficient. Beaton et al. [9] further defined MCID equal SEM with 1.96 representing the Z-score of 95% confidence interval. The ES is a statistics defined as  $\frac{d}{s}$ , where  $d$  denotes the individual change of PRO scores from baseline to post-treatment. In literature, the MCID was considered to be  $0.2 \cdot ES$  [10]. Distribution-based approaches thoroughly examine the distribution of PROs, and define the MCID through the magnitude of change in PROs above the measurement error. However, as pointed out in [11, 12], the limitation of distribution-based approaches lies in its unclear relatedness with clinical meaningfulness. The subjective bias in the PROs or unreliability of a poorly designed questionnaire could cause the measured MCID problematic. Copay et al. [13] even suggested that distribution-based approaches do not really address the clinical significance and ignore the purpose of MCID. The FDA guidance on PROs [7] also mentioned that the primary evidence should be provided by the analysis based on PROs combined with clinical anchors, and the distribution-based approaches are considered to be supportive.

## Anchor-based approaches

The anchor-based approaches establish the MCID by exploring the association between the target PROs and some other external criterion (anchors) [6, 7, 12, 13]. The commonly used anchors are clinical criteria that clinicians adopted to measure treatment efficacy, such as the widely used Eastern Cooperative Oncology Group Performance Status. The properly selected anchors naturally strengthen the connection between the interpretation of PRO results and clinical significance, and the anchor-based approaches quickly get favored by clinicians and health policy makers [7, 11]. Among the existing anchor-based approaches, the within-patients score change method and between-patients score change method select certain group of subjects as anchors, and

defined the MCID as the average PRO changes with in the selected groups and difference on PRO changes between selected group, respectively. However, the arbitrariness on the selection of subject groups makes these two methods less attractive [13]. For anchor-based approaches, the major statistical development lies in a class of methods called sensitivity and specificity-based approaches. In statistics literature, the concept of sensitivity and specificity has been widely used in the research of diagnostic tests. In the context of MCID re-research, sensitivity is the proportion of subjects who have an improvement on the anchor criterion and PRO scores above the MCID; specificity is the proportion of subjects who have no improvement on the anchor criterion and PRO scores below the MCID. The MCID is typically defined as the threshold with equal sensitivity and specificity among researchers [13]. Similarly, Bennett [14], Leisenring and Alonzo [15] discussed the connection between the MCID and positive predictive value (PPV) and negative predictive value (NPV). Shiu and Gatsonis [16] defined MCID as the maximizer of the sum of PPV and NPV based on the argument that the sum reflects a distance from ideal situation, where the ideal situation refers to a perfect match of responses based on PROs and anchor criterion. Recently, Hedayat et al. [17] proposed to define the MCID as the threshold that minimizes the mismatch between responses based on PRO scores and anchor criterion. Also in their paper, the definition and estimation of personalized MCID were established, which allows MCID to vary according to subjects' clinical profiles.

## Conclusion

Although many statistical approaches have become available, different methods lead to different estimations of MCID and no agreement has been reached regarding the suitability of estimating MCID. As discussed in [7], the measurement of MCID will be reviewed by FDA based on the context of each individual clinical study. Therefore, to deliver a legitimate interpretation of PROs with MCID, close collaboration between statisticians and clinicians is critical and necessary.

## Acknowledgement

None.

## Conflict of Interest

No conflict of interest

## References

- Jacobson NS, Truax P (1991) Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 59(1): 12-19.
- Guyatt G, Walter S, Norman G (1987) Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis* 40(2): 171-178.
- Jaeschke R, Singer J, Guyatt GH (1989) Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 10(4): 407-415.
- Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR (2002) Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 77(4): 371-383.
- Wyrwich KW, Bullinger M, Aaronson N, Hays RD, Patrick DL, Clinical Significance Consensus Meeting Group (2005) Estimating clinically significant differences in quality of life outcomes. *Qual Life Res* 14(2): 285-295.
- King MT (2011) A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res* 11(2): 171-184.
- U.S. Department of health and human services, food and drug administration, center for drug evaluation and research, center for biologics evaluation and research, center for devices and radiological. Health (2009) Guidance for industry Patient-report outcome measures: use in medical product development to support labeling claims. U.S. Department of Health and Human Services, Rockville MD.
- Wyrwich KW, Tierney WM, Wolinsky FD (1999) Further evidence supporting an SEM-based criterion for identifying meaningful criterion for identifying meaningful intra-individual changes in health-related quality of life *J Clin Epidemiol* 52(9): 861-873.
- Beaton DE, Bombardier C, Katz JN, Wright JG, Wells G, et al. (2001) Looking for important change/differences in studies of responsiveness. OMERACT MCID Working Group. Outcome Measures in Rheumatology. Minimal Clinically Important Difference. *J Rheumatol* 28(2): 400-405.
- Taylor SJ, Taylor AE, Foy MA, Fogg AJ (1999) Responsiveness of common outcome measures for patients with low back pain. *Spine (Phila Pa 1976)* 24(17): 1805-1812.
- Hays RD, Farivar SS, Liu H (2005) Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. *COPD* 2(1): 63-67.
- Cappelleri JC, Bushmakin AG (2014) Interpretation of patient-reported outcomes. *Stat Methods Med Res* 23(5): 460-483.
- Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC (2007) Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J* 7(5): 541-546.
- Bennett BM (1985) On tests for equality of predictive values for diagnostic procedures. *Stat Med* 4(4): 535-539.
- Leisenring W, Alonzo T, Pepe MS (2000) Comparisons of predictive values of binary medical diagnostic tests for paired designs. *Biometrics* 56(2): 345-351.
- Shiu SY, Gatsonis C (2008) The predictive receiver operating characteristic curve for the joint assessment of the positive and negative predictive values. *Philos Trans A Math Phys Eng Sci* 366(1874): 2313-2333.
- Hedayat AS, Wang J, Xu T (2015) Minimum clinically important difference in medical studies. *Biometrics* 71(1): 33-41.