



Minimal Clinically Important Difference: don't just look at the "p-value"

Marko Ostojic¹ · Philipp W. Winkler² · Jon Karlsson³ · Roland Becker^{4,5} · Robert Prill^{4,5} 

Received: 1 July 2023 / Accepted: 10 July 2023 / Published online: 17 July 2023

© The Author(s) under exclusive licence to European Society of Sports Traumatology, Knee Surgery, Arthroscopy (ESSKA) 2023

Clinicians and clinical scientists often simplify statistics as much as possible and rely much on the statistical numbers. One of the authors mentors during residency explained: "Once the statistician hands you the results, just look at the p values!". In some instance, oversimplifying helps a clinician not to get lost in the plethora of different statistical methods and numbers. On the other instance, it has a high potential leading to misleading interpretations.

Understanding the principles of statistics and going for statistically significant findings is crucial, but it is a bit more complicated than that [8]. There is a substantial qualitative difference between a statistically significant difference and a clinically relevant difference. Statistically significant difference is a mathematical term that means unlikeliness of a difference occurring by chance and is amplified in studies with large sample sizes. It does not necessarily mean that it will have a clinical impact on an individual. Clinical difference means an actual change for the patient that he/she can perceive as relevant. For example, a surgical procedure that has been performed on many subjects and brought a small benefit in patients score, can show a statistically significant difference due to the high number of study participants (i.e. the study is overpowered). The statistical power of the study made significant difference despite the actual low or not even measurable benefit of the procedure for the single

participant. The procedure should therefore not be implemented in clinical practice [10]. As clinicians, both when performing a study or critically reviewing one, beside other aspects mentioned in several checklists [9], clinical difference should have an advantage in contrary to pure statistical data. The combination of statistical analysis and clinical judgement should be used to elucidate Minimal clinically important difference (MCID), with a role to "bridge the gap between clinical and statistical significance" [1].

By definition, MCID would be the threshold value for the smallest change in a patient-reported outcome measure (PROM) that is considered worthwhile to patients [3]. In other words, it helps to understand the slightest difference in a result of a treatment or intervention that would actually make a meaningful difference (increase or decrease) to a patient's status. The term MCID was first mentioned in 1989 by Jaeschke who defined it 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" [5]. Similar measures to MCID like minimal clinically important improvement (MCII) or patient acceptable symptom state (PASS), focus only on a positive change and acceptable level of functioning, respectively. The MCID evaluates if a patient truly feels better or worse after the treatment [6].

Clinicians have been trying to adequately define how MCID will be derived most precisely. There are two methods of obtaining this derivation: the anchor-based and distribution-based methods. They are different by their origin, with anchor-based being more oriented to clinical status and distribution-based being more mathematical, with foundation in statistics [1].

Anchor-based derivations rely on subjective or objective measures that are known to be clinically meaningful, such as Global Transition Question or clinical endpoint. For example, when deriving an MCID for a PROM used to assess knee joint pathology and outcome, other PROMs and clinical outcomes (e.g. need for revision or return to sports) may be considered as anchors [6]. Another example is the assessment of lower-back pain, where PROMs can be anchored to

✉ Robert Prill
Robert.Prill@mhb-fontane.de

¹ Department of Orthopaedics, University Clinical Hospital Mostar, Mostar, Bosnia and Herzegovina

² Department of Orthopaedics and Traumatology, Kepler University Hospital Linz, Linz, Austria

³ Department of Orthopaedics, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

⁴ Center of Orthopaedics and Traumatology, University Hospital Brandenburg/Havel, Brandenburg Medical School Theodor Fontane, Brandenburg a.d.H., Germany

⁵ Faculty of Health Sciences Brandenburg, Brandenburg Medical School Theodor Fontane, Brandenburg a.d.H., Germany

socioeconomic outcomes or taken the impact of pain medication [3, 4]. In many cases, the Global Transition Question, where patients assess their overall level of functioning, is used as an anchor [1]. The MCID is computed as the difference between the baseline value and the value at which the patient reports a meaningful change on the relevant anchor question [3]. For example, one of the anchor-based methods for deriving MCID is the 75th percentile. In this method, a defined value is chosen when 75% of the patients demonstrated clinically important increase in their PROMs, according to the anchor question. These patients are considered as “responders”, with a positive visible effect [6]. All anchor based approaches use external criterion, and different variations of the approach may be identified: “within-patients” score change, “between-patients” score change, sensitivity- and specificity-based approach and social comparison approach [3].

Distribution-based derivation rely purely on statistical data. Different measures of variability can be used: the standard deviation, standard error of measurement, the effect size, or the minimum detectable change [3]. Most commonly the standard deviation is used [1]. A systematic review has shown that a standard deviation of 0.5 of the observed change in PROMs approximated the published MCID in many cases. The authors in this review concluded that a standard deviation of 0.5 represents the limit of the human mental discriminative capacity [7]. Critics have argued that this approach neglects to consider the clinical significance change in clinical outcome.

The first drawback of MCID definition is that each derivation method can produce different values which shows inconsistency in definitions. The anchor-based methods are mostly reliant on clinical status and suffer from a lack of standardization [1]. They depend largely on the anchor scales and how accurate are the differences between the levels on the scale, which determines how large can MCID be. Some authors state that it is a statistical fallacy comparing one subjective self-report to another, like PROMs to other PROMs [4]. Also, one anchor question can hardly cover all changes in PROMs [6]. Distribution-based derivations are statistically sound, but do not address the clinical importance entirely, which is neglecting the very purpose of MCID [3]. Actually, they mostly just show the minimum value below which a change score on a self-report may likely be because of a measurement error [1]. Another drawback is that clinical changes are associated with the baseline levels; *i.e.* MCID is population specific. For example, patients with lesser disability cannot show a substantial improvement (*i.e.* floor effect) [3].

MCID can be of help in analyzing the PROMs, and can also be used to calculate power analysis, and estimate adequate sample sizes in clinical trials [2, 6]. By knowing the expected MCID for a particular outcome measure,

researchers can determine the sample size needed to detect a statistically significant difference between treatment groups.

After some years, another mentor cited a quote by a philosopher William James that changed first authors perspective: “A difference which makes no difference, is no difference.”. A statistical difference can just be a numerical value with little or no difference to the patient’s quality of life or the efficacy of some procedure.

The overall aim of scientists is to improve patient care by improving orthopaedics research quality. Trying to see through the statistical mist and elucidate if the treatment reaches a clinical threshold of relevance will have a major contribution to this process. And once again, do not just look at the “Ps” and remember well the P.S. of this editorial.

P.S. Find a great mentor, for this will surely make an important difference.

Data availability No data available.

References

1. Bloom DA, Kaplan DJ, Mojica E, Strauss EJ, Gonzalez-Lomas G, Campbell KA et al (2023) The minimal clinically important difference: a review of clinical significance. *Am J Sports Med* 51:520–524
2. Clement ND, Weir D, Deehan D (2022) Meaningful values in the Short Form Health Survey-36 after total knee arthroplasty—an alternative to the EuroQol five-dimension index as a measure for health-related quality of life : minimal clinically important difference, minimal important change, patient-acceptable symptom state thresholds, and responsiveness. *Bone Joint Res* 11:477–483
3. 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* 7:541–546
4. Gatchel RJ, Mayer TG (2010) Testing minimal clinically important difference: consensus or conundrum? *Spine J* 10:321–327
5. Jaeschke R, Singer J, Guyatt GH (1989) Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 10:407–415
6. Mabrouk A, Nwachukwu B, Pareek A, Karlsson J, Mouton C, Ollivier M, et al. (2023) MCID and PASS in Knee Surgeries. *Theoretical Aspects and Clinical Relevance References. Knee Surg Sports Traumatol Arthrosc*; <https://doi.org/10.1007/s00167-023-07359-2>
7. Norman GR, Sloan JA, Wywich KW (2003) Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 41:582–592
8. Prill R, Królikowska A, Becker R, Karlsson J (2023) Why there is a need to improve evaluation standards for clinical studies in orthopaedic and sports medicine. *Knee Surg Sports Traumatol Arthrosc* 31:4–5
9. Prill R, Królikowska A, de Girolamo L, Becker R, Karlsson J (2023) Checklists, risk of bias tools, and reporting guidelines

for research in orthopedics, sports medicine, and rehabilitation. *Knee Surg Sports Traumatol Arthrosc.* <https://doi.org/10.1007/s00167-023-07442-8>

10. Prill R, Mouton C, Klugorová J, Królikowska A, Karlsson J, Becker R (2023) Implementation of evidence-based medicine in everyday clinical practice. *Knee Surg Sports Traumatol Arthrosc.* <https://doi.org/10.1007/s00167-023-07468-y>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.