




The minimum clinically important difference: which direction to take

T. H. P. Draak^a , B. T. A. de Greef^{a,b} , C. G. Faber^a  and I. S. J. Merkies^{a,c} on behalf of the PeriNomS study group

^aDepartment of Neurology, Maastricht University Medical Centre+, Maastricht; ^bDepartment of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Centre+, Maastricht, The Netherlands; and ^cDepartment of Neurology, St Elisabeth Hospital, Willemstad, Curaçao

Keywords:

minimum clinically important difference (MCID)

Received 6 November 2018
Accepted 19 February 2019

European Journal of Neurology 2019, **26**: 850–855

doi:10.1111/ene.13941

Over the past decades in modern medicine, there has been a shift from statistical significance to clinical relevance when it comes to interpreting results from clinical trials. A concept that is increasingly being used as a surrogate for clinical relevance and effect size calculation is the minimum clinically important difference (MCID). In this paper, an overview is presented of the most important aspects of the MCID concept used in research trials and a discussion of what this means for the neurological patient in clinical trials and daily practice is given. Is the MCID the best outcome measure cut-off to be implemented?

Introduction

The need for evidence-based medicine is greater than it has ever been before. In accordance, there is a great need for outcome measures which fulfil modern clinimetric requirements such as validity, reliability and responsiveness and which solely represent a level of assessing outcome as postulated by the International Classification of Functioning, Disability, and Health concept [1].

Rating scales are frequently used as an outcome measurement in both daily neurological practice as well as many clinical trials. An example is the Modified Rankin Scale, an ordinal scale consisting of seven items running from no symptoms (0) to death (6), a commonly used outcome measurement in stroke trials like the ‘Mr Clean’ study regarding endovascular treatment of ischaemic stroke [2].

Another example is the Unified Parkinson’s Disease Rating Scale (UPDRS) which was used as a primary outcome in the Adagio study for the effect of rasagiline in Parkinson’s disease [3]. The UPDRS is also an

ordinal scale with a range of 0 to 176 points in which a higher score indicates more severity.

The list with similar examples in daily practice and clinical trials is endless. However, when researchers do not correlate a statistically significant study result with its clinical relevance point of view, this could lead to misinterpreting results showing falsely positive or negative findings and exposing patients unnecessarily to (lack of) therapies [4]. Fortunately, over the past decades there has been a shift from statistical significance to clinical relevance when it comes to interpreting results from clinical trials [5–7].

Although the term ‘clinical relevance’ seems to be straightforward, it is not easy to define and to quantify. Who decides what is clinically relevant, the patient and/or the clinician? And how does one deal with different views on clinical relevance between patients? As an example, imagine two patients (A and B), both bedridden due to Guillain–Barré syndrome (GBS). Patient A is an elderly patient who considers ‘being able to walk with aid’ as a clinically relevant improvement. Patient B is a young adult who considers ‘being able to compete in professional sports again’ as a clinically relevant change. Both are affected by the same disease and are functioning on a similar level. Yet, they have a different interpretation of the term ‘clinical relevance’ and will have different

Correspondence: T. H. P. Draak, Department of Neurology, Maastricht University Medical Centre+ Maastricht, PO Box 5800, 6202 AZ Maastricht, The Netherlands (tel.: +31-43-3877059; fax: +31-43-3877055; e-mail:tim.draak@mumc.nl).

goals for their treatment. The same could be applied to differences between clinicians in terms of interpreting the significance of the minimum clinically important difference (MCID).

Even if a consensus could be reached on what would be a clinically relevant change, a proper outcome measure would still be needed to be able to detect that change. When these changes are vast and our study population is large, then an instrument will have little to no problem detecting such a change. However, changes are more likely to be subtle, and sometimes can be shrouded by the natural fluctuating disease course or other confounding factors. In these cases, proper outcome measures are needed to detect these smaller but clinically relevant changes and differentiate these changes from fluctuating 'noise' variations seen in illnesses. Subsequently, consensus needs to be reached on how to determine when a change is relevant enough to call the trial a success and not simply let this be driven by the *P* value hypothesis approach. In essence, is a trial successful by only looking at a statistically significant *P* value, like the Adagio study [3]? Or should a standardized cut-off point be used, like being able to walk independently as was done in the GBS trials [8]? Another option would be to determine the MCID, as was done in a follow-up analysis of the results of the ICE trial in patients with inflammatory neuropathies [6].

In this paper, an overview is provided of the MCID concept striving to help neurologists become more familiar with this entity. The origin of the MCID, its variable faces, and methods on how to determine its cut-off as well as pitfalls when applying this concept are discussed.

Finally, recommendations regarding its use in future clinical studies and trials are provided.

Minimum clinically important difference: origin

The term MCID was first defined by Jaeschke *et al.* 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' [9]. Over time, a mix of various definitions has been adopted for the MCID concept, such as the minimal important difference, the minimally important change, the minimally detectable difference, the minimum detectable change etc. [10,11]. All these adaptations have a common denominator aiming to quantify changes that are considered clinically relevant.

Another related term is the patient acceptable symptom state (PASS), which is defined as 'the value beyond which patients consider themselves well' [12]. Tubach *et al.* state that the MCID deals with the concept of improvement (feeling better) whereas the PASS deals with the concept of wellbeing (feeling good), thus also being complementary to the MCID.

Minimum clinically important difference: methods

The MCID concept is generally categorized into two main streams: the anchor-based method and the distribution-based method [13,14]. Extensive information regarding these (and newer) methods have been published in several excellent reviews [15–17].

Anchor-based method

Anchor-based methods involve comparing the change in the situation of a patient as captured by an outcome measure to an external criterion. This external criterion is often a patient's own categorization of their personal change, e.g. after an intervention. Examples of this method are using a pain score like the visual analogue scale, or a patient global impression of change scale (much worse, somewhat worse, about the same, somewhat better, much better).

Most often researchers will look at the change in a single patient over time, the so-called 'within-patient' change [18]. In a study population, the group who scored 'somewhat better' or 'much better' is of interest since these people have informed the researcher they have clinical improvement (from the patient's point of view). The next step is to look at the (median) change of the score of the instrument used that is representing the level of assessment of interest, which is often considered as the minimum change that correlates with clinical improvement.

Another anchor-based method is looking between patients at a single point in time, the so-called 'between-patient' difference [18]. Patients are grouped based on their rating on the external criterion: e.g. pain (I have no pain, I have moderate pain, I have extreme pain). Next, one would look at the (median) scores of the instrument of interest in these groups and then determine the MCID as the difference between the median score of the groups 'I have moderate pain' and 'I have no pain'.

Less commonly adaptations are a combination of within-person and between-person and a method in which patients rate their health state in comparison to other patients [19].

Distribution-based method

Distribution-based MCID methods are built upon the statistical properties of a study's result [20]. These include both the effect size, where the mean change of the individual is divided by the variability of either the whole group or the subset of stable subjects, and the reliability change index, a statistic rooted in the standard error of measurement (SEM) [21]. The SEM is a measure of the variation of observed scores due to measurement error compared to the 'true' score. Any changes which are below the SEM could be due to a measurement error 'noise' rather than to a truly occurring change. The number of SEMs needed to qualify a change in a patient's score to be meaningful is not yet fully established [16,17]. The most compelling argument has been provided by Wyrwich *et al.*, who state that the 'one-SEM criterion holds promise for identifying clinically meaningful intra-individual change' [21,22].

The standard deviation (SD) is a measure that is used to quantify the amount of variation or dispersion of a set of data values. In multiple studies with different outcome measures there seemed to be a universally applicable rule of thumb that the MCID is equal to 0.5 SD [23].

Minimum clinically important difference: pitfalls

When using and interpreting MCID values, one should always take several considerations into account, of which some are addressed in the following.

Use of ordinal scales

The use of ordinal scales is widespread across modern medicine, even though their shortcomings have been known for a long time [24]. Modern techniques like the Rasch model can help transform ordinal scales to interval-based scales [25,26]. Ordinal scales are often treated as an interval-based scale. However, one of the major problems with ordinal scales is that the amount of change required to go from 2 to 4 on a 10-point scale is not always the same as to go from 6 to 8. Therefore, the lack of a fixed unit using a metric as such hampers the stability definition of MCID across its range and the proper interpretation of the final results, which could subject patients to false positive or false negative results.

Small but clinically relevant changes

Distribution-based methods define the MCID as a change bigger than the expected variation or error in

measurement. It is not beyond imagination, however, to suggest a clinically relevant change that lies within the variation of a measurement, especially when the variation of a measurement has a wide range. All these relevant changes will not be noted because they are considered changes due to variation of the measurement rather than to clinical real change.

Static value for a dynamic concept

Ideally, one would be able to use a fixed MCID cut-off for each instrument for all patients in clinical studies. However, since the MCID is a dynamic concept, this is not a realistic proposition. Patients' (individual) MCID will vary by the severity of their illness, their social status, their own concepts of health and improvement etc. Furthermore, the MCID can be different for different treatments in the same patient group. For a surgical intervention with high risk and long recovery, a patient will expect more improvement for it to be clinically relevant and to warrant this treatment than for example when the patient only needs to make a minor adjustment in his/her lifestyle. So, when examining a new treatment, one cannot just blindly copy an MCID determined in a different study with different treatments.

The MCID does not account for cost-benefit

Although the classic definition by Jaeschke *et al.* did incorporate cost ('in the absence of ... excessive cost'), the MCID values determined nowadays do not take into account the cost of a specific treatment. In the light of the widespread pressure on funds for healthcare, it seems justified to take the cost of a certain amount of change into account as well, before establishing a change to be an MCID.

Minimum clinically important difference: Rasch improvements

The Rasch model is a modern technique that helps transform ordinal data to interval data. The term modern is relative, the first publication on its theory dating from 1960 [25]. It is being applied increasingly in research in modern medicine over the past decades (Fig. 1). The Rasch model states that 'a person having a greater ability than another person should have the greater probability of solving any item of the type in question, and similarly, one item being more difficult than another means that for any person the probability of solving the second item is the greater one'. So, in short, the person with a higher ability (thus being less ill) has a higher chance of getting a higher score.

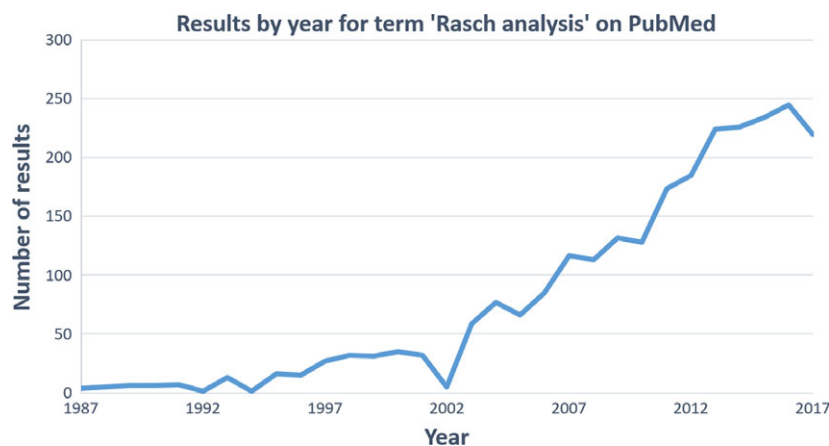


Figure 1 Graphical display of the number of hits in PubMed for the term 'Rasch analysis', showing its increasing use in modern medicine. [Colour figure can be viewed at wileyonlinelibrary.com]

See also Fig. 2 for a graphic illustration of this concept. This means that a confirmation on an item depends on both the difficulty of the item and the ability of the person. For more in-depth information, readers are referred to other publications [27,28].

The Rasch model can help improve the MCID. One way is to transform the ordinal data to interval data, as stated above. This removes the obstacle of score changes being different across the scale (an example is a change from 2 to 4 is not necessarily the same as from 8 to 10 on an ordinal scale), since on an interval scale the increment along the scale is equal for each step. Applying the Rasch model is mostly beneficial if the original (ordinal) scale does not behave in a linear fashion.

Additionally, after applying the Rasch model, one can determine the individual standard error (SE) of the ability estimate of each individual patient. This allows for individually determined MCIDs based on the patient's own SE. This is important because the SE of the ability of a patient varies across the range of said ability. In general, a patient with either a low or high ability score has a higher SE, whilst a patient with an average score will have a lower SE. This

shows that one cannot just take a single MCID cut-off value to be applied to all patients equally. For more detail on this, see an extensive report by Hobart and Cano, in particular Chapter 8 [29].

Recommendations and future perspectives

For clinical research trials to advance, one should always consider if the pre-defined primary outcome truly is clinically relevant to our patients and that the method of determining that outcome is up to current modern clinimetric standards.

Before determining the MCID of a rating scale, one should be cautious of the nature of the scale. If it is an ordinal scale (behaving in a non-linear fashion), one should apply the Rasch model to create an interval-based scale, as mentioned above. This will also allow for individual MCIDs to be determined, furthering the ability of establishing responsiveness in individual patients and capturing their voice more accurately.

Researchers using previously established MCID cut-off values when setting up new trials are also strongly discouraged if the intervention, outcome measure and

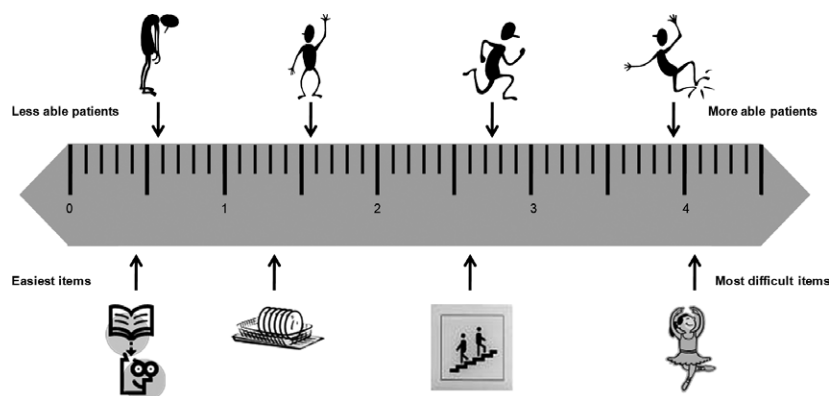


Figure 2 Graphical depiction of the Rasch model. The Rasch model takes both a person's ability (top part) and the item difficulty (bottom part) into account.

patient population are not similar to the settings in which the MCID cut-off was established.

The current paper provides the neurological community a brief overview on the meaning of the MCID. Despite all the limits, pitfalls and controversy on how to determine the MCID, its concept is of great importance in modern clinical trials. As long as no consensus is reached on which method to use, an anchor-based method alongside a distribution-based method is recommended, and they should be seen as complementary to each other rather than separately. Furthermore, by examining both methods, one can also compare the two methods in terms of their dynamics and results. Finally, by applying an anchor-based method with for example a self-evaluation of one's health, one can also capture the opinion of the patient. After all, is not the patient's perspective on their own health, be it improvement, deterioration or maintenance, that which should be the most important outcome measure in clinical trials?

A relatively new trend regarding this aspect is the inclusion of patients (and their caregivers) in clinical trial design as part of the study team. This can lead to a better understanding of the scope over which a disease can impact a patient's life. In doing so, previously neglected outcome domains, such as fatigue or sleep disturbances, can be identified and corrected [30–32].

Perhaps a next step would be to let the patients themselves tell us what they would deem as a sufficient amount of change in their health status for a treatment to be qualified as successful. This would surely be easier than constructing intricate surrogate markers with dubious clinimetric qualities which most clinicians struggle to comprehend. Simply talking and, more importantly, listening to our patients, however, is something every clinician excels at and perhaps researchers should start doing more and more.

Disclosure of conflicts of interest

Thomas H.P. Draak reports no disclosures. Bianca T.A. de Greef reports a grant from Prinses Beatrix Spierfonds (W.OR12-01), outside the submitted work. Catharina G. Faber reports grants from European Union's Horizon 2020 research and innovation programme Marie Skłodowska-Curie grant for PAIN-Net, Molecule-to-man pain network (grant no. 721841), grants from Prinses Beatrix Spierfonds (W.OR15-25), grants from Grifols and Lamepro for a trial on IVIg in small fibre neuropathy, and other from steering committees/advisory board for studies in small fibre neuropathy of Biogen/Convergence, Vertex and Chromocell, outside the submitted work.

Ingemar S.J. Merkies received funding for research from the Talecris Talents programme, CIDP/GBS Foundation International, Prinses Beatrix Spierfonds and the European Union 7th Framework Programme (grant number 602273). Furthermore, a research foundation at the University of Maastricht received honoraria on his behalf for participation in steering committees of the Talecris ICE Study, Laboratoire français du Fractionnement et des Biotechnologies, CSL Behring, Novartis, Grifols and Octapharma. He serves on the editorial board of the *Journal of the Peripheral Nervous System*, is a member of the Inflammatory Neuropathy Consortium and is a member of the Peripheral Nerve Society.

Funding

No funding of any kind was received for this study

References

1. World Health Organization. *International classification of functioning, disability and health : ICF*. World Health Organization: Geneva, Switzerland, 2001. iii, 299 pp.
2. Fransen PS, Berkhemer OA, Lingsma HF, *et al*. Time to reperfusion and treatment effect for acute ischemic stroke: a randomized clinical trial. *JAMA Neurol* 2016; **73**: 190–196.
3. Olanow CW, Rascol O, Hauser R, *et al*. A double-blind, delayed-start trial of rasagiline in Parkinson's disease. *N Engl J Med* 2009; **361**: 1268–1278.
4. Stang A, Poole C, Kuss O. The ongoing tyranny of statistical significance testing in biomedical research. *Eur J Epidemiol* 2010; **25**: 225–230.
5. Kemmler G, Zabernigg A, Gatteringer K, *et al*. A new approach to combining clinical relevance and statistical significance for evaluation of quality of life changes in the individual patient. *J Clin Epidemiol* 2010; **63**: 171–179.
6. Merkies IS, van Nes SI, Hanna K, Hughes RA, Deng C. Confirming the efficacy of intravenous immunoglobulin in CIDP through minimum clinically important differences: shifting from statistical significance to clinical relevance. *J Neurol Neurosurg Psychiatry* 2010; **81**: 1194–1199.
7. Terwee CB, Terluin B, Knol DL, de Vet HC. Combining clinical relevance and statistical significance for evaluating quality of life changes in the individual patient. *J Clin Epidemiol* 2011; **64**: 1465–1467; author reply 1467–1468.
8. van Koningsveld R, Schmitz PI, Meche FG, *et al*. Effect of methylprednisolone when added to standard treatment with intravenous immunoglobulin for Guillain-Barré syndrome: randomised trial. *Lancet* 2004; **363**: 192–196.
9. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertain the minimal clinically important difference. *Control Clin Trials* 1989; **10**: 407–415.
10. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res* 2011; **11**: 171–184.

11. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol* 2002; **14**: 109–114.
12. Tubach F, Ravaud P, Baron G, *et al*. Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state. *Ann Rheum Dis* 2005; **64**: 34–37.
13. Redelmeier DA, Guyatt GH, Goldstein RS. Assessing the minimal important difference in symptoms: a comparison of two techniques. *J Clin Epidemiol* 1996; **49**: 1215–1219.
14. Lydick E, Epstein RS. Interpretation of quality of life changes. *Qual Life Res* 1993; **2**: 221–226.
15. Wells G, Beaton D, Shea B, *et al*. Minimal clinically important differences: review of methods. *J Rheumatol* 2001; **28**: 406–412.
16. Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J* 2007; **7**: 541–546.
17. 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**: 102–109.
18. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis* 1985; **38**: 27–36.
19. Wells GA, Tugwell P, Kraag GR, Baker PR, Groh J, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. *J Rheumatol* 1993; **20**: 557–560.
20. Sloan J, Symonds T, Vargas-Chanes D, Fridley B. Practical guidelines for assessing the clinical significance of health-related quality of life changes within clinical trials. *Drug Inf J* 2003; **37**: 23–31.
21. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 1999; **37**: 469–478.
22. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol* 1999; **52**: 861–873.
23. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003; **41**: 582–592.
24. Stevens SS. On the theory of scales of measurement. *Science* 1946; **103**: 677–680.
25. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. University of Chicago Press: Chicago, IL, 1960.
26. Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil* 1989; **70**: 857–860.
27. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum* 2007; **57**: 1358–1362.
28. Vanhoutte EK, Hermans MC, Faber CG, *et al*. Rasch-ionale for neurologists. *J Peripher Nerv Syst* 2015; **20**: 260–268.
29. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess* 2009; **13**: iii, ix–x, 1–177.
30. de Wit M, Abma T, Koelewijn-van Loon M, Collins S, Kirwan J. Involving patient research partners has a significant impact on outcomes research: a responsive evaluation of the international OMERACT conferences. *BMJ Open* 2013; **3**: e002241.
31. Collins SP, Levy PD, Holl JL, *et al*. Incorporating patient and caregiver experiences into cardiovascular clinical trial design. *JAMA Cardiol* 2017; **2**: 1263–1269.
32. Abma TA, Nierse CJ, Widdershoven GA. Patients as partners in responsive research: methodological notions for collaborations in mixed research teams. *Qual Health Res* 2009; **19**: 401–415.