

# ORIGINAL RESEARCH

## Decision thresholds and minimal important difference estimates for evidence-based practice and policy

*Part 2: Threshold concepts in biostatistics and epidemiology*

Beth McDougall, MSc, PhD Candidate<sup>1</sup>, Mike Reid, MMM, MSc, PhD Candidate<sup>1</sup>  
Souvik Mitra, MD, MSc, PhD Candidate<sup>1,2,3</sup>, Bradley Johnston, PhD<sup>1</sup>

1. Department of Community Health and Epidemiology, Dalhousie University, Halifax, Nova Scotia, Canada

2. Department of Pediatrics, Dalhousie University, Halifax, Nova Scotia, Canada

3. IWK Health Centre - Halifax, Nova Scotia, Canada

---

### Abstract

Understanding core concepts in epidemiology and biostatistics is crucial for evidence-based clinical practice and policy. In this second installment of our two-part series on threshold concepts, we transition from understanding the ubiquitous p-value to tools and measures for decision making among clinicians-in-training, highlighting the growing importance of utilizing explicit and evidence-based approaches to make appropriate and efficient decisions. We review two related decision-making concepts: (1) Minimal Important Difference (MID) estimates and (2) Decision Thresholds, focusing specifically on patient-reported outcome measures (PROMs). These terms and many other related expressions are used regularly, and often interchangeably, but what are they? Why are they valuable? And how can they be used to support evidence-based decision-making in clinical contexts and develop strong clinical practice guidelines? We conclude our brief review on the utility of these measures with a spotlight on a local example of how the theory underlying MID estimates and decision thresholds is currently being embedded in electronic platforms in primary care contexts targeting depression in Nova Scotia.

---

### Introduction

Understanding and applying threshold concepts such as minimal important difference (MID) estimates and decision thresholds is crucial for evidence-based clinical practice and policy. In the first installment of this two-part series on core biostatistical and epidemiological concepts<sup>1</sup>, we learned about the ubiquitous but often misunderstood p value<sup>2</sup> and how it simply serves as a starting point when appraising research to better inform clinical decision making. In this second installment, we move beyond statistical significance, and address two related decision-making concepts: (1) MID estimates and (2) decision thresholds, focusing specifically on patient-reported outcome measures (PROMs). These terms and many other related expressions are used regularly, and often interchangeably<sup>3</sup>, but what are they? Why are they valuable? And how can they be used to support clinical decisions? Clinicians and researchers acknowledge the critical role of disease symptoms, as well as the function (e.g. mental, physical) and perceptions of general well-being for evidence-based decision-making. Typically measured by direct patient inquiry, these outcomes, previously referred to as 'health-related quality of life' (HRQL), are now most commonly

referred to as PROMs, and measure patient perception of health and well-being, such as pain or depression severity. PROMs provide patients' perspectives on treatment benefits and harms and are often the outcomes of most importance to patients<sup>4</sup>. While there is significant research on the reliability and validity of established PROMs<sup>5</sup>, less is known regarding how to interpret changes or differences in scores, and what indicates a potentially clinically meaningful difference, which can also serve as a decision threshold.

### MID Estimates

The MID can generally be described as the *smallest difference* in an outcome of interest perceived to be important and that would lead a patient or clinician to consider a change in treatment or management<sup>6</sup>. The MID is exclusively used in the context of outcomes that are being examined on a *continuous scale*. For example, the patient-rated Beck Depression Inventory (BDI)<sup>7</sup> measures depression severity on a continuous scale from 0 to 63. What minimal change in score would a patient (ideally) or clinician deem important to suggest a potential change in treatment or management?

As outlined by King et al.<sup>3</sup>, there is no universal MID that can be used for all PROMs across all populations.

However, an effect size (a quantitative estimate of the magnitude of change) based on distribution-based methods between 0.2 and 0.5 standard deviation (SD) units has been offered by Cohen as a rule of thumb<sup>8</sup>, while Norman et al.<sup>9</sup> has suggested that the universal MID is half the SD of the baseline score. In addition to these suggestions, since criticized<sup>10</sup>, multiple methods to estimate MIDs have been developed<sup>3,11</sup>. Of the many methods that exist, recent research generally advocates for the anchor-based approach, methods that typically involve working with actual patients to arrive at the best estimate of change that matters to patients<sup>4</sup>. This approach incorporates patient values by comparing the relationship between the target PROM and an anchor relevant to patients, such as global ratings of change in symptom severity<sup>4</sup>. Applied to the context of evidence-based decision-making for practice or policy, the choice of MID should be supported by a primary study that establishes an estimate of the MID using anchor-based methods, or a systematic review of such studies<sup>12</sup>. For the BDI, in particular, new research using anchor based methods has suggested that the MID can range anywhere from 17.5 to a 32% reduction in scores from baseline, with this percentage change being highly dependent on the baseline severity of depression<sup>13</sup>.

Unfortunately, using MIDs for clinical decision-making, and using established MIDs to develop practice guideline recommendations is often challenging as there is currently no definitive resource that describes all established PROM-specific MIDs<sup>4</sup>. There is, however, research detailing a framework for appraising the quality of anchor-based MID estimates underway, and the application of the framework to over 300 PROMs is under-review<sup>14</sup>. Until this definitive resource is published, those who wish to search for MIDs are left to conduct their own search of the literature and to interpret the estimates based on their own judgement. Moreover, we are unaware of similar work for distribution-based MID methods. Despite these challenges, the best MID estimates should be sought for PROMs as they provide a clear indicator of the minimal change that patients consider important for a change in treatment or management<sup>13</sup>. Further, MIDs are important estimates for sample size calculations in clinical trials and for making clinical guideline recommendations to evaluate interventions evaluated for PROMs.

## Decision Thresholds

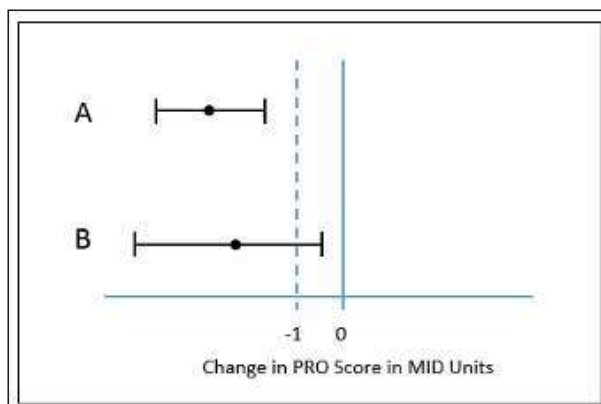
When we think of a decision threshold in the context of clinical practice decisions, we are considering the boundary to sway a decision and related action in one way versus another<sup>3,6</sup>. For an outcome that has been dichotomized, a threshold, or cut-off value is used to

assign a person to one of the two categories. Clinically, a decision made using a binary decision threshold can strongly impact the approach to treatment. For example, a patient who no longer meets the threshold for clinical depression may be viewed as responding favorably to a recent treatment, such as an antidepressant, and may be monitored less frequently by the treating clinician. Within clinical research, not meeting the threshold for clinical depression may exclude a potential participant from a randomized clinical trial of alternative management strategies for depression. What decision threshold estimate should be used to determine if a patient is likely to be clinically depressed or not?

For health-related status such as depression symptoms measured on a continuous instrument, there are various methods for arriving at a decision threshold. Ideally, if the instrument in question is a PROM, the threshold used should be the best estimate of the MID derived using anchor-based methods<sup>4</sup>. If the instrument is not patient-reported (e.g. Hamilton Rating Scale for Depression, Montgomery Asberg Depression Rating Scale), the threshold is likely to be determined by clinical consensus or through the use of distribution-based MID methods such as half the baseline standard deviation (SD)<sup>9</sup>, a method with universality but that may be oversimplistic<sup>15</sup>. Regardless, clinicians should be aware of the minimum and maximum scores of the instrument, whether a higher score is considered desirable or undesirable, and, when available, the best estimate of the decision threshold or MID. For instance, the Montgomery Asberg Depression Rating Scale<sup>16</sup> contains 10 items, each scored 0-6, with a total possible score between 0 and 60. Mild depression is associated with a score of 12 to 23. Therefore, a threshold score of 12 is often used for diagnosis or a change in clinical management. However, the distribution-based MID estimate suggests a minimum change of 3 in the overall score for determining treatment decisions<sup>16-18</sup>.

The decision threshold estimate will differ depending on the measure being used, including the validity and reliability of the measure and the severity and heterogeneity of the condition itself<sup>19</sup>. Decision thresholds on established rating scales for depression with sound measurement properties not only vary by measure, but can also vary by patient characteristics such as age, sex, and clinical profile<sup>20,21</sup>.

In the context of clinical decision-making, decision thresholds are attractive as they provide a simple tool to allow the translation of a continuous score on a measure to a 'yes' or 'no' decision. However, with multiple methods of generating decision thresholds being published, the threshold could vary greatly<sup>22,23</sup>. Typically, the optimal decision threshold for a continuous measure such as depression severity is



**Figure 1.** Two hypothetical interventions in which a lower PRO score is deemed better. Intervention A demonstrates a treatment effect that is to the left of the decision threshold (the associated MID), and represents a significantly large and precise treatment effect. Intervention B includes the MID decision threshold, and the quality of evidence for this comparison would be rated down for imprecision (wide confidence interval).

derived from the best estimate of the MID<sup>24</sup>.

### *The value of MID in primary mental health care*

There has been a notable increase in cases of depression over the past ten years<sup>25,26</sup>, with the highest rates of mood and anxiety disorders being found in Canadians between the age of 15 to 24<sup>27</sup>. Most patients who seek treatment for depression are treated in primary care contexts through a trial and error process of identifying the appropriate antidepressant medication or behavioral or psychological support(s). Primary care clinicians are faced with the challenge of making evidence-based clinical decisions collaboratively with their patients typically in very short consultations. Approaches to simplify decision making, such as using established MIDs can support this challenging task.

For example, in the primary health care context in Nova Scotia, the Maritimes Depression Registry is being piloted among primary care providers and patients with depression<sup>28</sup>. The freely available web-based application houses a battery of standardized measurement tools and allows clinicians and patients to regularly complete PROMs, such as depressive severity using the BDI. Instruments are automatically scored, and established MIDs and decision thresholds, where available, are highlighted for clinicians when considering clinical treatment decisions. Standard clinical practice guidelines are also embedded in the tool and individualized recommendations can be made tailored to patients' self-identified personal goals for recovery.

Capitalizing on the notion of measurement-based person-centered care, the tool is a local demonstration of the potential value that can be added by incorporating

the best estimates of MIDs for routine clinical care for patients with depression. Challenging clinical decisions, such as starting or discontinuing antidepressants, are facilitated by visual cues of MIDs and decision thresholds on established measures, where available.

Ultimately, if it exists for a given PROM, decision makers should find the best estimate of the MID and consider where the MID lies in relation to the best estimate of effect and the corresponding 95% confidence interval (CI). If the effect estimate surpasses the MID, a treatment or management decision is well-supported (Figure 1). Figure 1 offers a visual representation of this using two hypothetical interventions. Intervention A demonstrates a treatment effect that is to the left of the decision threshold (the associated MID) and represents a significantly large and precise treatment effect. Intervention B includes the MID decision threshold, and the quality of evidence for this comparison would be rated down for imprecision.

### **Conclusion**

We have reviewed important concepts, namely decision thresholds and MIDs, and have highlighted a local example of how the theory underlying these concepts is currently being embedded in electronic platforms in primary care contexts targeting depression in Nova Scotia. Decision thresholds and MIDs – particularly anchor-based MIDs for PROMs – are important concepts and can provide valuable information for decision makers.

### **References**

1. Reid M, McDougall B, Mitra S, Johnston B. The problem with "p". An examination of the use of p-values as a foundation for evidence-based practice and policy. Submitted for publication to Dalhousie Medical Journal 2019.
2. Cristea IA, Ioannidis JPA. P values in display items are ubiquitous and almost invariably significant: A survey of top science journals. *PLoS ONE* 2018;13(5):e0197440.
3. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res* 2011;11(2):171-84.
4. Johnston BC, Ebrahim S, Carrasco-Labra A, Furukawa TA, Patrick DL, et al. Minimally important difference estimates and methods: a protocol. *BMJ Open* 2015;5:e007953.
5. University of Oxford. Patient Reported Outcomes Measurement Group. <http://phi.uhce.ox.ac.uk/about.php>. Accessed 14 May 2019.
6. Robays J, Vlayen J. Clinical decision threshold and minimally important difference. In: *KCE Process Book, 2012* <<http://processbook.kce.fgov.be/node/118>> (14 May 2019).
7. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Archives of General Psychiatry* 1961;4:561-571.
8. Cohen J. *Statistical Power Analysis for the Behavioural Sciences* (2nd Edition). Lawrence Erlbaum Associates, NJ, USA (1988).
9. Norman GR, Sloan JA, Wyrwich KW. The truly remarkable

- universality of half a standard deviation: confirmation through another look, *Expert Review of Pharmacoeconomics & Outcomes Res* 2004;4(5):581-585.
10. Johnston BC, Thorlund K, da Costa BR, Furukawa TA, Guyatt GH. New methods can extend the use of minimal important difference units in meta-analyses of continuous outcome measures. *J Clin Epidemiol*. 2012;65(8):817-26.
  11. Redelmeir DA, Guyatt GH, Goldstein RS. Assessing the minimal important difference in symptoms: A comparison of two techniques. *Journal of Clinical Epidemiology* 1996;49(11):1215-1219.
  12. Jayadevappa R, Cook R, Chhatre S. Minimal important difference to infer changes in health-related quality of life—a systematic review. *J Clin Epidemiol*. 2017;89:188-198.
  13. Button KS, Kounali D, Thomas L, et al. Minimal clinically important difference on the Beck Depression Inventory--II according to the patient's perspective. *Psychol Med*. 2015;45(15):3269-3279.
  14. Carrasco-Labra A, Devji T, Johnston BC, et al. Minimally important difference estimates and assessment of their credibility for patient-reported outcomes in adults. A systematic survey. *Global Evidence Summit*. Cape Town, South Africa (September 2017).
  15. Farivar SS, Liu H, Hays RD. Half standard deviation estimate of the minimally important difference in HRQOL scores?. *Expert Rev Pharmacoecon Outcomes Res*. 2004;4(5):515-523.
  16. Montgomery SA, Asberg M. New depression scale designed to be sensitive to change. *British Journal of Psychiatry* 1979;134:382-389.
  17. Duru G, Fantino B. The clinical relevance of changes in the Montgomery-Asberg Depression Rating Scale using the minimum clinically important difference approach. *Current Medical Research and Opinion Journal* 2008;24(5):1329-1335.
  18. Bandelow B, Baldwin DS, Dolberg OT, Andersen HF, Stein DJ. What is the threshold for symptomatic response and remission for major depressive disorder, panic disorder, social anxiety disorder, and generalized anxiety disorder? *Journal of Clinical Psychiatry* 2006;67(1):428-434.
  19. Monroe SM., Anderson SF. Depression: The shroud of heterogeneity. *Current Directions in Psychological Science* 2015;24(3):227-231.
  20. Schulte-van Maaren YWM, Carlier IVE, Zitman FG, van Hemert AM, de Waal MW, et al. Reference values for major depression questionnaires: The Leiden Routine Outcome Monitoring Study. *Journal of Affective Disorders* 2013;149(1-3):342-349.
  21. Van Steenberghe-Weijnenburg KM, De Vroeghe L, Ploeger R, Brals JW, Vloedveld MG, et al. Validation of the PHQ-9 as a screening instrument for depression in diabetes patients in specialized out-patient clinics. *BMC Health Services Research* 2010;10:235.
  22. Ebell MH, Locatelli I, Senn N. A novel approach to the determination of clinical decision thresholds. *Evid Based Med* 2015;20(2):41-7.
  23. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry* 1993;39(4):561-577.
  24. Schünemann HJ, Guyatt GH. Commentary--goodbye M(C) ID! Hello MID, where do you come from? *Health Serv Res*. 2005;40(2):593-597.
  25. The World Health Organization. "Mental Health: Depression, Let's Talk." <[http://www.who.int/mental\\_health/management/depression/en/](http://www.who.int/mental_health/management/depression/en/)> (24 May 2019).
  26. Statistics Canada. "Mood and Anxiety Disorders in Canada: Fast Facts from the 2014 Survey on Living with Chronic Diseases in Canada." <[https://www.canada.ca/en/public-health/services/publications/diseases-conditions/mood-anxiety-disorders-canada.html?\\_ga=2.39418275.1893200471.1537882937-796303240.1536244507](https://www.canada.ca/en/public-health/services/publications/diseases-conditions/mood-anxiety-disorders-canada.html?_ga=2.39418275.1893200471.1537882937-796303240.1536244507)> (24 May 2019).
  27. Statistics Canada. Table 105-1101 Mental health profile, Canadian Community Health Survey - Mental Health (CCHS), by age group and sex, Canada and provinces. CANSIM (database). Last updated September 17, 2013.
  28. Uher, R. Person-centered measurement-based care for depression. *World Psychiatry* 2016, 15(3):238-239.