Can we estimate the minimal important difference from the distribution-based method?

Yasushi Tsujimoto[1, 2], Tomoko Fujii[3,4], Yusuke Tsutsumi[1], Yuki Kataoka[1], Aran Tajika[4], Alonso Carrasco-Labra[5], Tahira Devji[5], Gordon H Guyatt[5], Toshi A Furukawa[4]

1. Department of Healthcare Epidemiology, Kyoto University Graduate School of Medicine, Japan
2. Department of Nephrology and Dialysis, Kyoritsu Hospital, Japan
3. The Australian and New Zealand Intensive Care Research Centre, the School of Public Health and Preventive Medicine, Monash University, Australia
4. Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine, Japan
5. Department of Health Research Methods, Evidence and Impact, McMaster University, Canada

Introduction

A number of studies have reported a minimally important difference (MID) for patient-reported outcome measures (PROMs). The MID is defined as “the smallest difference that patients perceive as beneficial and that would mandate, in the absence of troublesome side effects and excessive cost, a change in the

patient's management" [1]. To estimate an MID, an anchor-based method that uses external criteria for the PROM is considered to be the gold standard as opposed to distribution methods that use a statistical parameter to provide a cutoff for an MID [2]. Several studies have reported MID estimates using anchor-based methods that agree with nearly half of the standard deviation (SD) of a PROM [3, 4], whereas other studies have shown otherwise [5, 6]. However, these studies did not take the methodological quality of the primary studies estimating MIDs or types of PROMs into account, and their examples were more anecdotal than systematic or comprehensive. Another question is which SD is applicable to estimate an MID of a PROM. In longitudinal studies, several SDs can be calculated from a PROM (e.g. baseline SD, post-intervention SD, or SD of change in baseline and post-intervention).

The current study, therefore, aims to determine i) whether we can yield the same estimate of the MID calculated through an anchor-based approach using SD units of MIDs, ii) which SD, if any, can be used across various types of PROMs for this purpose, and iii) which subtypes/subgroups of PROMs, if any, show such relationships. If we can establish certain relationships between the anchor-based MID and MID in SD units for certain subgroups or as regression equations, such knowledge will greatly facilitate the interpretation of PROMs for which we currently lack anchor-based MIDs.

Methods

Eligibility criteria

We will use the dataset of the previous MID inventory research, which included 3389 MID estimates derived from 338 studies and that we will be adding more [7]. The inclusion and exclusion criteria for the MID inventory can be found elsewhere [7]. In summary, the study included MID estimates reported in the studies estimating anchor-based MIDs for PROMs in adolescents (≥13 to 17) or adult (≥18) populations. PROMs of interest included health-related quality of life, functional ability, symptom severity and psychological distress and well-being. We included any MID irrespective of the participants' condition or disease, type of intervention used in the eligible studies, or nature of the anchor. We excluded systematic reviews of studies examining MIDs; conference abstracts; studies in which authors explicitly targeted a moderate or large important difference as opposed to an MID; a combined anchor and distribution-based approach; and estimates obtained using pooled data from multiple cohorts (e.g. different primary investigations).

In addition to the eligibility criteria in the MID inventory study [7], we will exclude studies estimating MIDs by anchor-based methods other than mean change methods using a global rating of change in the present study. The reason for this exclusion is to focus on properly estimated MIDs. We will also exclude studies not reporting SDs of MIDs, or studies not reporting any other variablity measures such as standard error (SE), confidence interval (CI), and interquartile range (IQR), and the number of participants for the MID estimation as it is hard to calculate the SDs.

Search methods and selection in the previous study

We searched MEDLINE, EMBASE, and PsycINFO for studies published between 1989 and April 2015. To complement this search, we retrieved additional relevant citations from the PROQOLID internal library, relevant reviews and eligible studies.

Teams of two reviewers independently screened titles and abstracts for potentially eligible references. Any citations identified as relevant by either screener were selected for full text evaluation, again conducted in duplicate. Reviewers resolved disagreement by discussion or, if needed, by consultation with a third reviewer. This process resulted to include 338 studies consisting 3389 MID estimates [7]. In the current study, we will further assess eligibility according to additional exclusion criteria above.

Data extraction

For eligible studies, we will use the following variables in the previous study (ref): the country of the study; population demographics; types of PROM; interventions administered in the context of the MID estimation; anchor details (i.e. type, construct(s), range of options/categories/values, threshold selected to represent a “small but important change”, specific anchor-based method); MID estimate, its associated measure of variability and direction; details regarding MID determination (e.g. number of patients informing the MID estimate, duration of follow up (if applicable), correlations between the PROM and anchor), and credibility ratings of the MID study. We will classify types of PROM in two main categories with two and four subcategories: 1) generic (health profiles and utility measures), and 2) specific (disease/condition specific, symptom specific,

function specific, and population specific) according to the previous taxonomy [8]. PROMs categorized to health profiles are instruments that attempt to measure all important aspects of HRQL. Utility measures are derived from economic and decision theory that reflect the preferences of patients for treatment process and outcome. Specific measures focus on aspects of health status that are specific to the area of primary interest.

In addition to the extraction of the MID inventory study above, we will further extract the means and the SDs of PROMs from all participants in the primary studies. If authors report other variability measures such as SE, CI, IQR, or range but not SD, we will calculate SD using the following formulae.

SD = SE × $\sqrt{N}$.

SD = $\sqrt{N}$ × (upper limit of CI − lower limit of CI) / 3.92.

SD = IQR / 1.35 [9].

SD = Range / 2 [9].

If authors report the SDs separately in subgroups of the participants (e.g. improvement, not change, or deterioration), we will calculate the overall SD using the following formula in the Cochrane handbook [9].

$$\mathbf{SD} \quad \sqrt{\frac{(N_1-1)\mathrm{SD}_1^2+(N_2-1)\mathrm{SD}_2^2+\frac{N_1N_2}{N_1+N_2}(M_1^2+M_2^2-2M_1M_2)}{N_1+N_2-1}}$$

When there are more than two groups to combine, we will apply the above formula sequentially. However, if the authors reported the variability in subgroups using 95% CI, IQR, or Range, we will not calculate the overall SD

because the formulae above do not produce consistent value when the number of participants is small [10].

We will classify the SDs into the following categories: the SD of the baseline score (baseline SD), the SD of change in pre- and post-intervention scores (change SD), and the SD of post-intervention score (post SD).

Statistical analysis

To describe the distribution of MIDs in SD units, we will first calculate their mean, SD, 95 % confidence interval (CI), and 95 % prediction interval (PI). If the distrtibution is skewed, we will present the result with median and IQR as well. We will use the three SDs; baseline SD, change SD, and endpoint SD in calculating the MIDs in SD units

Our primary outcome is a proportion of MIDs in SD units that falls in the range of the mean (or median if the distribution is skewed) calculated above ± 0.2 among the PROMs. This range has been defined from Cohen [11] and the clinical perspective, suggesting that such MIDs estimated from the SD are approximately correct and can be substituted for the MIDs based on the anchor-based method.

We will then repeat the same analysis above in the following subgroups using the three SD units: published year, intervention (pharmacological or not), types of PROM (generic (health profiles and utility measures) and specific (disease/condition, symptom, function, and population specific)), and MID

direction (improvement, worsening, improvement/worsening, and other), and credibility rating (definitely yes or not for each core credibility item).

If possible, we will also try to build a regression equation predicting MIDs in SD units, using the variables found meaningful in the above analyses. We will assess the performance of such regression equations by calculating the proportion of predicted MIDs in SD units that fall in the range of the mean calculated above ± 0.2 among the PROMs.

Continuous variables will be expressed as mean (SD) and categorical variables will be shown as numbers with the percentage. A two-sided p value smaller than 0.05 will be considered as a statistically significant difference. We will use Stata/SE, V.14.0 (StataCorp, College Station, TexasX, USA) for all analyses.

Table 3. (Expected result) Proportion of MIDs that are within the range of mean ± 0.2 mean

		SD type		
		Baseline SD (%)	Post-intervention SD (%)	Transision SD (%)
PROM type	generic	79	67	78
	specific	69	75	70
Quality	High	84	79	56
	Moderate	72	71	74
	Low	69	68	71
Intervention	Pharmacological	79	77	75

Non-pharmacological	70	61	68

Reference:

1. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Controlled clinical trials. 1989;10(4):407-15.
2. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. Journal of Clinical Epidemiology. 2003;56(5):395-407.
3. Sloan JA, Loprinzi CL, Kuross SA, Miser AW, O'Fallon JR, Mahoney MR, et al. Randomized comparison of four tools measuring overall quality of life in patients with advanced cancer. J Clin Oncol. 1998;16(11):3662-73.
4. 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(5):582-92.
5. Testa MA. Interpreting quality-of-life clinical trial data for use in the clinical practice of antihypertensive therapy. J Hypertens Suppl. 1987;5(1):S9-13.
6. Feinstein AR. Indexes of contrast and quantitative significance for comparisons of two groups. Stat Med. 1999;18(19):2557-81.
7. Johnston BC, Ebrahim S, Carrasco-Labra A, Furukawa TA, Patrick DL, Crawford MW, et al. Minimally important difference estimates and methods: a protocol. BMJ Open. 2015;5(10):e007953.
8. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med. 1993;118(8):622-9.
9. Higgins JPT TJ, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0
: Cochrane; 2019 [updated July 2019. Available from: www.training.cochrane.org/handbook.
.
10. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. BMC Med Res Methodol. 2005;5:13.
11. Cohen J. Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: Erlbaum; 1988.