Journal of Clinical Epidemiology 80 (2016) 57−67

Journal of Clinical Epidemiology

ORIGINAL ARTICLES

# Meta-analysis on continuous outcomes in minimal important difference units: an application with appropriate variance calculations

Ian Shrier[a,*], Robin Christensen[b], Carsten Juhl[c,d], Joseph Beyene[e,f]

[a]*Centre for Clinical Epidemiology, Lady Davis Institute, Jewish General Hospital, McGill University, 3755 Cote Ste-Catherine Road, Montreal, Quebec H3T 1E2, Canada*

[b]*The Parker Institute, Bispebjerg and Frederiksberg Hospital, Nordre Fasanvej 57, DK-2000 Copenhagen F, Denmark*

[c]*Research Unit for Musculoskeletal Function and Physiotherapy, Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark*

[d]*Department of Physiotherapy and Occupational Therapy, Copenhagen University Hospital, Herlev and Gentofte, Kildegaardsvej 28, 2900 Hellerup, Copenhagen, Denmark*

[e]*Department of Clinical Epidemiology & Biostatistics, Faculty of Health Sciences, McMaster University, 1280 Main Street West, MDCL 3211, Hamilton, ON L8S 4K1, Canada*

[f]*Department of Mathematics & Statistics, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada*

Accepted 2 July 2016; Published online 30 July 2016

## Abstract

**Objective:** To compare results from meta-analyses for mean differences in minimal important difference (MID) units ($MD_{MID}$), when MID is treated as a random variable vs. a constant.

**Study Design and Setting:** Meta-analyses of published data. We calculated the variance of $MD_{MID}$ as a random variable using the delta method and as a constant. We assessed performance under different assumptions. We compare meta-analysis results from data originally used to present the $MD_{MID}$ and data from osteoarthritis studies using different domain instruments.

**Results:** Depending on the data set and depending on the values of rho and coefficient of variation of the MID ($CoV_{MID}$), estimates of treatment effect and *P*-values between an approach considering the MID as a constant vs. as a random variable may differ appreciably. Using our data sets, we provide examples of the potential magnitude. When rho = 0.5 and $CoV_{MID}$ = 0.8, considering MID as a constant overestimated the treatment effect by 33−110% and decreased the *P*-value for heterogeneity from above 0.95 to below 0.08. When rho = 0.8 and $CoV_{MID}$ = 0.5, the magnitude of the effects was similar.

**Conclusions:** Considering MID as a random variable avoids unrealistic assumptions and provides more appropriate treatment effect estimates. 

*Keywords:* Continuous outcomes; Meta-analysis; Minimal important difference; Standardized mean difference; Ratio of means; Variance; Methods

## 1. Introduction

Health care professionals are strongly encouraged to practice evidence-based medicine (EBM) [1−4], where clinical decisions are based on the best evidence addressing a focused clinical question. Practicing EBM requires access to health care evidence, and preferably evidence that is succinctly and systematically summarized. When there is sufficient homogeneity of studies, a meta-analysis fulfills this objective. A meta-analysis may be broadly defined as the quantitative review and synthesis of the results from related but independent studies [5]; a trustworthy meta-analysis is always based on a thorough systematic review of the literature wherein the authors provide an overall quantitative summary statistic for the effect estimate of a group (or subgroup) of studies [6].

When the outcome is binary, investigators commonly combine studies in a meta-analysis by choosing to summarize across the risk difference, risk ratio, or odds ratio scales [7,8]. The magnitude and direction of the overall estimate may be different with different summary statistics because

Funding: This study received no specific funding. I.S. was supported by the Lady Davis Institute, Jewish General Hospital in Montreal, Canada. R.C. (The Parker Institute) was supported by unrestricted grants from the Oak Foundation; C.J. was supported by the University of Southern Denmark and the University Hospital of Copenhagen, Herlev and Gentofte. J.B. is funded by the Canadian Institutes of Health Research. None of the funders had any involvement in the conduct of this study.

* Corresponding author. Tel.: 514-340-8222x4244; fax: 514-340-7564.
*E-mail address*: ian.shrier@mcgill.ca (I. Shrier).

http://dx.doi.org/10.1016/j.jclinepi.2016.07.012
0895-4356/

**What is new?**

**Key findings**

- Current methods to standardize continuous outcomes in minimal important difference (MID) units require unrealistic assumptions.

**What this adds to what was known?**

- We describe a method to standardize continuous outcomes in MID units that allows for greater transparency of assumptions and sensitivity analyses.

**What is the implication and what should change now?**

- When standardizing continuous outcomes using MID units, investigators should incorporate realistic assumptions.
- Investigators should use sensitivity analyses to test the robustness of the results to violations of their assumptions.
- Some examples for Grading of Recommendations Assessment, Development and Evaluation Summary of Findings tables are provided.

the formula for the variances (responsible for the weighting of individual studies) is different; the appropriate summary statistic for a particular meta-analysis may depend on the underlying reasons for variation in control group event rates; in some situations, uncertainty about the choice of summary statistic will remain [9]. Therefore, to avoid introducing reporting bias, investigators should be explicit about why they chose the particular summary statistic for binary data [10].

When the outcome is continuous, systematic reviewers must calculate the treatment effect as either a raw mean difference (MD), or standardize the MD in some way [5]. Standardizing the MD is typically preferable when the construct being measured is the same across studies, but the actual measurement instrument differs. For example, frequently used pain measures for osteoarthritis [11] include the Western Ontario and McMaster University Osteoarthritis Index [12], Knee Injury and Osteoarthritis Outcome Score/Hip Disability and Osteoarthritis Outcome Score [13,14], Visual Analogue Scales, Health Assessment Questionnaire (pain subscale) [15], Lequesne algofunctional index (pain subscale) [16], Arthritis Impact Measurement Scales (pain subscale) [17,18], and McGill Pain Questionnaire (pain intensity) [19]. When constructs are measured using different scales, combining the raw numbers into a weighted average is not meaningful because a result of 10 on one scale might be equivalent to a result of 50 on another scale. Therefore, some form of standardization is necessary before the results can be combined.

Commonly proposed methods for standardization include the standardized mean difference (SMD) [20], ratio of means (RoMs) [21,22], and a more recent method based on standardizing the MD using minimal important differences (MIDs) between groups [23,24]. Although the MID approach has recently been proposed as a simple effect measure to use, it considers MID as a constant. However, different patients will often have different values for MID, just as different people have different heights or weights. For example, if pain is rated on a scale of 0–10, one person might consider 2 as the MID, another 3 as the MID, and another 1 as the MID. If we acknowledge that there is variation in the population, then in statistical terms, the MID is considered a random variable; taking the mean of the values as the one true value would be to treat MID as a constant.

The distinction between treating MID as a constant vs. a random variable is important. As a random variable, there would be an expected correlation between MD and MID, and there would be a coefficient of variation of MID. The value of these two variables will affect confidence intervals and statistical significance testing. The purpose of this article is to highlight the overall benefits of treating MID as a random variable and to illustrate how it can be implemented easily. We illustrate our proposed solution using two different data sets: (1) data originally pooled by Johnston et al. in their original study [23] and (2) data from studies investigating the effects of exercise on knee osteoarthritis [25] that are well known for using different scales to measure the same construct [11].

Finally, once the treatment effect is estimated (whether SMD, RoM, or MID units), authors have different options on how to present the results. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group has suggested that Summary of Findings tables could include a comparative treatment effect such as MD, mean values for each group (by assuming a mean value for the control group and then estimating the mean for the treatment group based on the calculated treatment effect), or converting the continuous scale into categories and reporting proportion of patients who would receive substantive benefit [26]. Although the objective of our article is to estimate a valid treatment effect, the editors have asked us to illustrate how the results could be adapted into different formats for presentation to decision makers.

## 2. Proposed standardization methods

The most commonly recommended and used method of standardization is the SMD [20,27]. In brief, the MD from each study is divided by a pooled standard deviation (SD). Therefore, each study estimate is now expressed as an ‘‘effect size,’’ and these can be combined using standard