Journal of Clinical Epidemiology 82 (2017) 128–136

Journal of Clinical Epidemiology

The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies

Felix Angst[a,*], André Aeschlimann[a], Jules Angst[b]

[a]*Rehabilitation Clinic ("RehaClinic"), Department of Research, Quellenstrasse 34, 5330 Bad Zurzach, Switzerland*

[b]*Department of Psychiatry, Psychotherapy and Psychosomatics, Psychiatric Hospital Burghölzli, University of Zurich, Lenggstrasse 31, 8008 Zurich, Switzerland*

Accepted 30 November 2016; Published online 14 December 2016

Abstract

Objective: To illustrate and discuss current and proposed new concepts of effect size (ES) quantification and significance, with a focus on statistical and clinical/subjective interpretation and supported by empirical examples.

Study Design and Settings: Different methods for determining minimal clinically important differences (MCIDs) are reviewed, applied to practical examples (pain score differences in knee osteoarthritis), and further developed. Their characteristics, advantages, and disadvantages are illustrated and discussed.

Results: Empirical score differences between verum and placebo become statistically significant if sample sizes are sufficiently large. MCIDs, by contrast, are defined by patients' perceptions. MCIDs obtained by the most common "mean change method" can be expressed as absolute or relative scores, as different ES parameters, and as the optimal cutoff point on the receiver operating characteristic curve. They can further be modeled by linear and logistic regression, adjusting for potential confounders.

Conclusion: Absolute and relative MCIDs are easy to interpret and apply to data of investigative studies. MCIDs expressed as effect sizes reduce bias, which mainly results from dependency on the baseline score. Multivariate linear and logistic regression modeling further reduces bias. Anchor-based methods use clinical/subjective perception to define MCIDs and should be clearly differentiated from distribution-based methods that provide statistical significance only.

Keywords: Osteoarthritis; WOMAC; Outcome measurement; Effect size; Receiver operating characteristics curve; Statistics; Significance; Standardized mean difference; Minimal clinically important difference; Regression; Confounding

1. Introduction

Whether consciously or not, we are constantly measuring changes in health dimensions, such as pain, physical function, social function, or depression. Especially if affected by symptoms, we are alert to daily alterations in our state. In medicine, particularly in rheumatology, the last 30 years have seen considerable progress in the standardization of the measurements and methods of outcome effect quantification [1–4]. It should be possible for results that are qualitatively and quantitatively identical or very similar to mean the same across cultures and languages and to be similarly interpreted.

Human perceptions of most health dimensions are subjective and individual, posing an inherent problem for the standardization of assessment, interpretation, and comparison. Time, place, health state, and internal and external circumstances all affect perception and can lead to wide variations. Any changes measured may disappear in the noise of variability, hampering the use of conventional statistical, analytical methods.

Many studies, especially older pharmacological trials, confine themselves to quantifying the size and the significance of differences in health dimensions by conventional statistical methods, for example, by the *t*-test. However, statistically significant differences are mainly dependent on

Conflict of interest: None.

Funding: This study was supported by the Zurzach Rehabilitation Foundation SPA, Bad Zurzach, Switzerland.

* Corresponding author. Tel.: ++41-56-269-53-97; fax: ++41-56-269-51-72.

E-mail address: fangst@vtxmail.ch (F. Angst).

http://dx.doi.org/10.1016/j.jclinepi.2016.11.016

What is new?

Key findings

- This study provides an overview of the most important methods for determining minimal clinically important differences (MCIDs) and presents further developments.

What this adds to what is known?

- It illustrates the strengths and weaknesses of different MCID parameters and the relationship of MCIDs to statistically significant differences.

What is the implication and what should change now?

- Multivariate regression modeling of MCIDs may open up new prospects for less-biased estimates of MCIDs.

the number of persons examined (n), as will be demonstrated later [1,5–7].

Given these shortcomings, an alternative "significance" has been identified to characterize the sizes of effects [1]. Because medicine measures outcome in humans, who feel and communicate, the patient alone can define an identifiable difference. Responding to the relevant assessments, the patient can only assess an effect that is felt. This led to the development of the concept of the smallest subjectively perceptible effect that is "clinically" important, named the minimal clinically important difference (MCID) [1,8–10].

An instructive overview of the history, concepts, and characteristics of methods which estimate "clinical significance" is provided by Kamath et al. in their fundamental textbook "Methods and applications in clinical trials" [1]. "Anchor-based" methods use external criteria (the anchor) to quantify differences measured by an outcome instrument (example: the mean change method). "Distribution-based" methods define different statistical parameters to assess clinical significance (example: t-test). Kamath et al. include our previous investigation (2001) on statistically detectable differences and MCIDs as one of seven exemplary studies [5].

In this article, we will compare current concepts of effect size quantification and significance, with a focus on statistical and clinical/subjective meaning. Based on those models, new concepts for quantifying MCIDs will be developed, discussed, and illustrated by specific examples from empirical studies. The concepts of our earlier study are expanded and combined with new approaches which are particularly relevant for randomized controlled trials (RCTs).

2. Smallest detectable difference (SDD)—statistical definition for quantifying the significance of differences

In empirical outcome research, the t-test formula is used in many analyses of continuous parameters as a suitable approximation, although some of the necessary assumptions (normal distribution, homoscedasticity, etc.) are often not met by the data [5–7]. Choosing the simplest example of pairwise differences in one sample, for example, within-person differences between baseline and follow-up, the statistic to examine significance is [6] $t = \Delta/(s/\sqrt{n})$, where Δ = difference of the means = mean of the differences (baseline to follow-up), s = standard deviation of the differences, and $s/\sqrt{n}$ is the standard error of Δ, n = sample size.

To reach statistical significance, t has to be large. It can then be assumed with a low probability of error (type I error P) that the difference really exists. Δ and s are finite parameters, especially in closed scales, such as the visual analogue scale (VAS) for pain between 0 and 100 (mm). The sample size n may increase to high, almost infinite numbers. Therefore, t mainly grows by n, together with the probability of significance. A minimal Δ, whose corresponding t reaches a predefined significance level P, is defined as the smallest (statistically) detectable difference [5,6]. A common example of the principal application of the t-test is the measurement of effects in RCTs.

3. Standardized mean difference (SMD)—statistical parameter for quantifying differences in RCTs

The SMD is expressed as the difference Δ in the two mean score differences (baseline to follow-up) between the verum and placebo groups divided by the so-called "pooled" or "within" standard deviation s_Δ of the two groups [7].

$$\mathrm{SMD} = \frac{\Delta}{s_\Delta} \; \textit{and} \; s_\Delta = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

Thus s_Δ is the square root of the mean of the variances (of the score differences), weighted by the number of subjects, of the verum group (index 1) and the placebo group (index 2). Where sample sizes are equal ($n_1 = n_2$), the pooled variance s_Δ^2 is simply the mean of the two variances.

In other words, the SMD is the difference in mean pain relief between verum and placebo in number of pooled standard deviations and is dimensionless. Positive SMDs reflect the superiority of the verum, negative SMDs the superiority of the placebo. The larger the SMD, the greater the probability of attaining statistical significance to support the conclusion that verum is more effective than placebo. Nowadays, the SMD is the standard effect size parameter for RCTs [7]. In meta-analyses, the SMDs of different RCTs are themselves pooled to give a global effect size.

The 95% confidence interval (95% CI) of the SMD based on the standard error (se) is given by