Journal of Clinical Epidemiology 68 (2015) 740–751

Journal of Clinical Epidemiology

Odds ratios of treatment response were well approximated from continuous rating scale scores for meta-analysis

Ramona Meister*, Alessa von Wolff, Levente Kriston

Department of Medical Psychology, University Medical Center Hamburg-Eppendorf, Martinistrasse 52, Hamburg 20246, Germany

Accepted 9 February 2015; Published online 18 February 2015

Abstract

Objectives: To empirically evaluate the performance of methods for estimating odds ratios and their corresponding standard errors from continuous end point data for meta-analysis.

Study Design and Setting: A database of randomized controlled trials of chronic depression treatments was used. Trials that reported both continuous and dichotomous end points for symptom improvement were considered. Odds ratios and standard errors were calculated from the dichotomous data and estimated from the continuous data using currently available methods: Hasselblad and Hedges (HH), Cox and Snell (CS), Furukawa (F), Suissa (S), and Kraemer and Kupfer (KK). Single and meta-analytically pooled observed and estimated values were compared.

Results: A total of 26 trials were included. At the trial level, four of five (HH, CS, F, and S) and three of four (HH, F, and S) methods for estimating odds ratios and standard errors performed well, respectively. We found considerable differences in the performance of all methods across trials with more accurate estimates for smaller treatment effects. At the level of meta-analysis, three of four methods (CS, F, and S) performed acceptably.

Conclusion: Odds ratios and standard errors can be approximated from continuous end points, but we recommend sensitivity and subgroup analyses to test robustness of the findings.

Keywords: Meta-analysis; Odds ratio; Standardized mean difference; Response rate; Estimation method; Continuous end point; Dichotomous end point; Depression

1. Introduction

Systematic reviews and meta-analyses combine the results of primary trials on specific research questions to inform treatment-related decisions. Their importance in evidence-based health care is continuously increasing [1]. One of the major prerequisites of standard meta-analyses is that the same effect size (e.g., odds ratio) should be calculated from each included trial [2].

However, although trials in one meta-analysis always assess the same target construct, some may report continuous and others dichotomous end points. For example, in research on antidepressant treatments, a frequently used continuous end point is the severity of depression, generally indicated by a score for each patient on a continuous symptom severity rating scale [e.g., the Hamilton Rating Scale for Depression (HRSD) [3]]. The means and standard deviations of both the experimental and the control group can be calculated and compared as a result of the primary trial. In the case of dichotomous end points, the number of patients who responded to the received treatment in each group is typically described and compared.

For a combination of primary trial results in meta-analysis, effect sizes are calculated from the end points of the primary trials that reflect the differences between the experimental and control groups. Therefore, continuous and dichotomous end points result in different effect sizes. The means and standard deviations reported in primary trials can be combined into (standardized) mean differences, and the numbers of responders in each group can be transformed into binary effect measures such as odds ratios.

In most cases, no individual patient data but only summary statistics are reported in published study reports and are available for meta-analyses. Unfortunately, some trial publications report either continuous or dichotomous end points (but not both), so that the differing effect sizes

Funding: None.

Conflict of interest: None.

* Corresponding author. Tel.: 49-040-7410-52231; fax: 49-040-7410-54965.

E-mail address: r.meister@uke.de (R. Meister).

http://dx.doi.org/10.1016/j.jclinepi.2015.02.006

What is new?

Key findings

- Four out of five estimation methods approximately estimated odds ratios of treatment response from continuous symptom severity rating scale scores (HH, CS, F, S) and three out of four estimation methods (HH, S, F) performed well in estimating standard errors.
- The performance of all methods varied across trials, which allows the conclusion that trial characteristics might influence the accuracy of the estimates.
- The synthesis of the estimated odds ratios and the corresponding estimated standard errors in meta-analyses resulted in acceptable pooled odds ratios but increased standard errors that reflect the additional uncertainty due to estimation.

What this adds to what was known?

- Summarizing all available evidence in systematic reviews and meta-analyses is essential for well-informed clinical decisions.
- Nature and reporting of outcomes can substantially vary between trials, leading to difficulties in pooling all data in a single analysis and a possible loss of information.
- Estimating odds ratios of treatment response with the corresponding standard errors from continuous symptom severity rating scale scores is one possible solution for addressing missing data in meta-analyses.
- Accuracy of the currently available estimation methods has been insufficiently evaluated.
- We empirically evaluated and compared the performance of all currently available methods to estimate odds ratios from continuous symptom severity rating scale scores. Thereby, we considered standard errors directly in addition to effect sizes.

What is the implication and what should change now?

- We encourage researchers to estimate odds ratios and standard errors from continuous endpoints. Though, we recommend sensitivity- and subgroup-analyses to test robustness of the findings.

cannot be synthesized in one meta-analysis. Excluding trials because of missing data might lead to a loss of power, a loss of information, and/or selection bias and should be avoided [4–6]. This is an especially relevant issue for evidence synthesis in the field of chronic depression, in which primary trial results are often inconsistent and till now no clear treatment recommendations can be derived [7–9]. Thus, single trials or the accuracy of single end points can strongly influence the total body of evidence. Synthesizing all available evidence in one meta-analysis is therefore essential [5,6]. One possible solution is to estimate the effect sizes of dichotomous end points from continuous end points [10].

Da Costa et al. [11] identified five methods for estimating odds ratios from continuous end points, four of which considered also the standard error of the odds ratios: Hasselblad and Hedges (HH), Cox and Snell (CS), Furukawa (F), Suissa (S) and Kraemer and Kupfer (KK). The mathematical properties and the underlying assumptions of some of these methods have already been theoretically reviewed or evaluated using simulation studies [12,13]. Beyond statistical theory and simulation studies, however, empirical evaluation in real-world data sets is needed [11]. On that account, Furukawa et al. [10] evaluated the formula that they and S used to derive response rates from means and standard deviations using 51 trials from four meta-analyses of anxiety and depression treatment and found that the observed and estimated numbers of responders in the primary trials as well as the observed and estimated pooled relative risks of the four meta-analyses were nearly identical. Cuijpers et al. [14] determined the performance of HH' method in 49 psychotherapy trials for adult depression and concluded that estimating odds ratios from standardized mean differences or vice versa led to considerable differences in some primary trials but to comparable pooled effect sizes when combined in a meta-analysis. The performance of F's method was investigated by Samara et al. [15] using 16 trials on schizophrenia treatment. They concluded that the estimated values reflected the observed values to a reasonable extent. Furukawa and Leucht [16] conducted a meta-analysis of individual patient data from 10 antipsychotic trials and found that F's method, but not KK's method, was accurate in estimating the number needed to treat from the standardized mean differences. To date, only da Costa et al. [11] have examined the performance of all five methods, using 29 hip and/or knee osteoarthritis trials; they found the methods of HH, CS, F, and S to be suitable and the method of KK to be unsuitable for estimating odds ratios from continuous end points.

However, none of the existing studies evaluated the accuracy of estimated standard errors directly. Da Costa et al. [11] reported 95% confidence intervals (CIs) of observed and estimated odds ratios, which allow an indirect comparison of precision across methods. Here, we followed a different approach by comparing the estimated standard errors directly. Estimating standard errors accurately is essential for the interpretation of the results because wrong conclusions might be drawn otherwise. On the one hand, the standard errors of treatment effects could be underestimated and uncertain treatment effects could be seen as definite. On the other hand, standard