ELSEVIER

Journal of Clinical Epidemiology ■ (2015) ■

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Choice of effect measure for meta-analyses of dichotomous outcomes influenced the identified heterogeneity and direction of small-study effects

Spyridon N. Papageorgiou[a,b,c,*], Elli Tsiranidou[d], Georgios N. Antonoglou[e], James Deschner[c,f], Andreas Jäger[a,c]

[a]Department of Orthodontics, School of Dentistry, University of Bonn, Welschnonnenstr. 17, Bonn 53111, Germany
[b]Department of Oral Technology, School of Dentistry, University of Bonn, Welschnonnenstr. 17, Bonn 53111, Germany
[c]Clinical Research Unit 208, University of Bonn, Welschnonnenstr. 17, Bonn 53111, Germany
[d]Medical Research Methodology, Medical School, Aristotle University of Thessaloniki, University Campus, PO 17034, Thessaloniki 54210, Greece
[e]Department of Periodontology and Geriatric Dentistry, Institute of Dentistry, University of Oulu, Medical Campus, Aapistie 3A, PO 5281, Oulu 90014, Finland
[f]Experimental Dento-Maxillo-Facial Medicine, University of Bonn, Welschnonnenstr. 17, Bonn 53111, Germany

Accepted 6 January 2015; Published online xxxx

Abstract

Objectives: The aim of this study was to compare the use of odds ratio (OR), risk ratio (RR), and risk difference (RD) in meta-analyses of dichotomous outcomes and assess their influence on their results.

Study Design and Setting: Initially, we included meta-analyses from a meta-epidemiologic database and reanalyzed them with OR, RR, and RD as summary metric. The primary outcomes were the effects of metric choice on the (1) statistical significance, (2) heterogeneity, and (3) Egger's test for publication bias. Additionally, meta-analyses that originally used OR were reanalyzed using RR to assess the differences in their results.

Results: In the 235 meta-analyses (147 reviews) that were included, the conclusions in terms of significance rarely changed. On the other hand, use of OR displayed the lowest I^2 values (median 42%), followed by RR (+5.1%) and RD (+15.0%). The Egger's test was most often significant with RR (32%), followed by RD (29%) and OR (24%). Substitution of RR for OR led to a change of the observed effects in 3%, change of between-study heterogeneity in 6% to 24%, and change in Egger's test results in 7% of the cases, respectively.

Conclusion: The choice of metric for meta-analyses of dichotomous outcomes might influence the identified between-study heterogeneity and the results of Egger's test.

Keywords: Dentistry; Meta-analysis; Systematic reviews; Effect measure; Odds ratio; Risk ratio; Risk difference

1. Introduction

1.1. Background

Systematic reviews and meta-analyses of controlled clinical evidence are fundamental in evidence-based decision making. One of the first steps of quantitative synthesis of data is selecting an effect measure with which to express the observed effects of each trial and pool them together. For binary outcomes, the three most widely used metrics are the odds ratio (OR), the risk ratio (RR), and the risk difference (RD), with or without the number needed to treat (NNT).

Conflict of interest: None.

* Corresponding author. Tel.: +49-(0)228-287-22449; fax: +49-(0) 228-287-22588.

E-mail address: snpapage@gmail.com (S.N. Papageorgiou).

The selection of the most appropriate effect measure has been long debated, but no general guidelines have yet been developed. Consistency [1], interpretation ease [2,3], and favorable mathematical properties [4,5] play an important role in this choice. OR has the best mathematically properties and a desirable symmetry. However, with event rates greater than 10−15%, the OR does not approximate well the RR [6]. Furthermore, it is not as intuitive compared with the other two metrics and is often misinterpreted as an RR both by readers and researchers [2,6,7]. RRs provide useful and easily understood effect estimates but lack symmetry and differ according to the calculation of harms/benefits [8]. Contrary to the other two relative metrics, RD expresses the absolute difference and can be easily translated clinically with the NNT [9]. However, the individualized NNT is based on a constant RR, which is an

http://dx.doi.org/10.1016/j.jclinepi.2015.01.004
0895-4356/

What is new?

Key findings

- Choice of effect measure for meta-analyses of dichotomous outcomes does not seem to influence the statistical significance of the results.
- Between-study heterogeneity and the results of Egger's test might be influenced by the choice of effect measure.
- Changing the effect measure from odds ratio to risk ratio in meta-analyses of dichotomous outcomes might alter their results.

What this adds to what was known?

- The influence of effect measure in the results of meta-analyses of dichotomous outcomes might not be negligible.

What is the implication and what should change now?

- Caution is needed in the choice of effect measure for meta-analyses of dichotomous outcomes, as this might have an impact on their results.

assumption [10] and might lead to the underestimation of benefits among low-risk patients [11].

A number of concerns have been pointed out against the use of ORs [3,12], cautioning in particular that ORs should only be used in case–control studies and regression analyses [3]. However, selection of an ideal meta-analysis metric should be supported by empirical evidence [13,14] and not be solely based on theoretical grounds.

1.2. Aim

The aim of this study was to compare OR, RR, and RD as summary effect estimates in meta-analyses of dichotomous outcomes with regard to the meta-analyses results. Secondarily, OR is widely used in meta-analyses, but caution has been advocated regarding its use. Therefore, we planned to substitute RR for OR in meta-analyses that originally used the latter metric to assess how this metric change affects the results on the basis of significance and identified heterogeneity.

2. Methods

2.1. Protocol

This study's protocol was constructed *a priori*, based on the guidelines of the PRISMA statement [15], the Cochrane Handbook for Systematic Reviews of Interventions [16], and previous studies [17], and it was circulated and accepted by all authors.

2.2. Inclusion criteria

Eligible for this study were meta-analyses in any field of dentistry or oral medicine with binary outcomes. A minimum of five included studies per meta-analysis was arbitrarily adopted, as meta-analysis methods, including estimations of pooled effects, identified heterogeneity, and reporting bias, have been shown to perform poorly with few studies [16,18–20]. Additionally, the original meta-analysis should report raw data for the included studies. Excluded were all other study types and meta-analyses that did not meet the eligibility criteria. From systematic review articles with more than one meta-analysis, the meta-analysis with the most studies was included, or if they had the same number, the one with the higher heterogeneity estimate.

2.3. Search and selection procedures

The search and selection procedures have been described previously [21]. In short, we searched seven general, open-access, regional, or gray literature databases up to December 2012 without language, publication year, or publication status restrictions. After study selection, data were extracted into predefined and piloted worksheets.

2.4. Analysis

As a first step, each meta-analysis was reanalyzed three times with OR, RR, and RD as metrics in RevMan version 5.2 (Copenhagen: The Nordic Cochrane Center, The Cochrane Collaboration).

There are many methods to calculate and pool effect measures in RevMan. For the fixed-effect model, the inverse variance method is a straightforward method that can be used generally in most situations by weighing studies according to their precision. The Mantel–Haenszel method [22,23] is a good method for reviews with few events or small studies (default method in RevMan). For ORs, there is also the Peto method [24], which is a good method for studies with few events, small effects (OR close to 1), and similar numbers in the experimental and control group. For the random-effects model, there are two options (Mantel–Haenszel and inverse variance). As this study aimed to compare only the choice of metric, pooling was carried out with the Mantel–Haenszel method for both fixed-effect and random-effects models for consistency reasons.

The DerSimonian and Laird heterogeneity estimator was chosen, although it might be inferior to other estimators [19], as it is used most often and is included in RevMan. The extent of between-study heterogeneity was assessed with the Q heterogeneity statistic (hereon plainly Q) and the associated chi-square test [25,26]. RevMan uses a hybrid method for Mantel–Haenszel random-effects