ELSEVIER

Journal of Clinical Epidemiology ■ (2015) ■

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Small studies are more heterogeneous than large ones: a meta-meta-analysis

Joanna IntHout[a,*], John P.A. Ioannidis[b,c,d,e], George F. Borm[a], Jelle J. Goeman[a]

[a]*Radboud Institute for Health Sciences (RIHS), Radboud University Medical Center, Mailbox 133, P.O. Box 9101, Nijmegen 6500 HB, The Netherlands*
[b]*Stanford Prevention Research Center, Department of Medicine, Stanford University School of Humanities and Sciences, 1265 Welch Road, Stanford, CA 94305, USA*
[c]*Department of Health Research and Policy, Stanford University School of Medicine, 150 Governor's Lane, Stanford, CA 94305, USA*
[d]*Department of Statistics, Stanford University School of Humanities and Sciences, 390 Serra Mall, Stanford, CA 94305, USA*
[e]*Meta-Research Innovation Center at Stanford (METRICS), Stanford University, 1265 Welch Road, Stanford, CA 94305, USA*

Accepted 27 March 2015; Published online xxxx

Abstract

Objectives: Between-study heterogeneity plays an important role in random-effects models for meta-analysis. Most clinical trials are small, and small trials are often associated with larger effect sizes. We empirically evaluated whether there is also a relationship between trial size and heterogeneity (τ).

Study Design and Setting: We selected the first meta-analysis per intervention review of the Cochrane Database of Systematic Reviews Issues 2009−2013 with a dichotomous (n = 2,009) or continuous (n = 1,254) outcome. The association between estimated τ and trial size was evaluated across meta-analyses using regression and within meta-analyses using a Bayesian approach. Small trials were predefined as those having standard errors (SEs) over 0.2 standardized effects.

Results: Most meta-analyses were based on few (median 4) trials. Within the same meta-analysis, the small study τ_S^2 was larger than the large-study τ_L^2 [average ratio 2.11; 95% credible interval (1.05, 3.87) for dichotomous and 3.11 (2.00, 4.78) for continuous meta-analyses]. The imprecision of τ_S was larger than of τ_L: median SE 0.39 vs. 0.20 for dichotomous and 0.22 vs. 0.13 for continuous small-study and large-study meta-analyses.

Conclusion: Heterogeneity between small studies is larger than between larger studies. The large imprecision with which τ is estimated in a typical small-studies' meta-analysis is another reason for concern, and sensitivity analyses are recommended.

Keywords: Randomized controlled trial; Meta-analysis; Between-study heterogeneity; Random-effects model; Trial size; Cochrane Database of systematic reviews (CDSR)

1. Introduction

In clinical research, many small and possibly underpowered studies are conducted. Among interventional trials registered between 2007 and 2010 in ClinicalTrials.gov, 62% (17,726 of 28,458) enrolled at most 100 participants [1]. In 2008, 70% (10,492 of 14,886) of the meta-analyses with a binary outcome in the Cochrane Database of Systematic Reviews (CDSR), Issue 1, consisted only of studies with less than 50% power to detect a 30% relative risk reduction [2].

There is an ongoing debate on the disadvantages of small trials [3]. Small trials are associated with larger treatment effect estimates [2,4,5], and it is possible that between-study heterogeneity also increases when studies are smaller. Turner et al. [2] observed that removing the underpowered (<50% power) studies from 1,107 meta-analyses resulted in a median 21% decrease in the estimated τ^2. Borm and Donders [6] observed higher heterogeneity between small rheumatoid arthritis studies compared with larger studies. Individual study results are influenced by many, possibly related aspects, such as quality of study, publication bias, and study size [7−10]. Califf et al. [1] observed that small trials contain significant heterogeneity in methodological approaches, including reported use of randomization, blinding, and data

Funding: This study was not supported by any company or grants. The costs were borne by the authors' institutions.

Conflict of interest: None.

* Corresponding author. Tel.: (31) 0-24-361-91-3; fax : (31) 0-24-361-35-05.

E-mail address: joanna.inthout@radboudumc.nl (J. IntHout).

http://dx.doi.org/10.1016/j.jclinepi.2015.03.017
0895-4356/

What is new?

Key findings

- In a sample of 2,009 meta-analyses with a dichotomous outcome and 1,254 meta-analyses with a continuous outcome of the Cochrane Database of Systematic Reviews Issues 2009−2013, the between-study heterogeneity τ was often estimated to be either zero or high, and the imprecision of the estimated τ was large, especially for meta-analyses based on few and/or small studies.
- Small studies had higher mean heterogeneity estimates than medium/large studies of the same meta-analysis.

What this adds to what was known?

- Evidence from small studies tends to show not only larger effect sizes but also larger and less precise estimates of between-study heterogeneity.

What is the implication and what should change now?

- In a random-effects meta-analysis, the estimated between-study heterogeneity directly affects the summary treatment effect and prediction interval. It should be realized that the estimated τ is often imprecise and on average larger for small studies. Sensitivity analyses to check robustness of the pooled effect estimate may be warranted.

monitoring committees. Button et al. [11] argued that underpowered studies are prone to several analytical and reporting biases. Small studies may be of lower quality in other aspects of their design as well. This may affect the between-study heterogeneity.

The current paradigm, in which multiple small studies are conducted and subsequently combined in a meta-analysis, is questioned [3,12]. Especially in random-effects models and with substantial heterogeneity, the influence of small studies will be major and may affect the reliability of meta-analyses. On the other hand, simulations have shown that a meta-analysis containing many, possibly small, studies is better than a single large trial able to estimate the treatment effect [6,13,14], even when there is some publication bias [15]. Roloff et al. [16] showed that in case of cumulative meta-analysis, it is more powerful to add several small studies than one or a few large studies because the between-study heterogeneity can be estimated more precisely when more studies, either small or large, are available. However, a questionable assumption underlying their calculations is that heterogeneity is similar between small and large studies. The same questionable assumption occurs in standard applications of random-effects meta-analysis: one single τ^2 is used in the random-effects weights for all studies.

If there is heterogeneity, treatment effects in individual studies may deviate more from the summary effect than expected by chance. Simulations have shown that when there is heterogeneity but no true treatment effect, the frequency of false statistically significant findings in single trials increases more than 10-fold [15]. When small studies have higher than average heterogeneity, the increase in error rates for small single trials will be even larger. Also, prediction intervals [17] constructed with an average τ will result in too narrow predictions for the expected effect for future small trials.

In summary, if there is a difference in heterogeneity between small and large trials, this can influence both the reliability of the results of single trials and of meta-analyses. Results of the current method for random-effects meta-analysis may be overly drawn toward the small-study results, prediction intervals may be too narrow, and false-positive findings of single trials may occur more frequently than expected.

In this article, we investigate empirically whether the heterogeneity of small and large trials is different. We used meta-analyses from 3,851 reviews on interventions of the 2009−2013 Issues from the CDSR. First, we investigated in a cross-sectional approach the relation between study size and heterogeneity across 3,263 meta-analyses. As Turner et al. [18] showed that the extent of heterogeneity could be related to outcome and intervention type, our primary analysis is a paired-data approach, comparing the between-study heterogeneity of large trials with the small-study heterogeneity of the same meta-analysis.

2. Methods

2.1. Selected data

The UK Cochrane Editorial Unit provided us with the statistical data of the systematic reviews of interventions, included in the CDSR Issues of 2009−2013. We used the mean values and standard deviations per treatment group for meta-analyses with continuous outcomes and counts (with/without event) for those with dichotomous outcomes. Most Cochrane reviews included multiple meta-analyses, and meta-analyses from the same review are often correlated. The first reported analysis in a review is usually one of the primary analyses. Hence, to avoid subjectivity in selecting specific meta-analyses, we used only the first meta-analysis appearing in the data and analyses section that was based on at least two studies. To maximize the number of meta-analyses for our evaluation, we used both the first continuous and the first binary outcome meta-analysis, if available. A selected meta-analysis could