

LETTER

Open Access



Empirical comparisons of heterogeneity magnitudes of the risk difference, relative risk, and odds ratio

Yuxi Zhao¹, Elizabeth H. Slate¹, Chang Xu^{2,3}, Haitao Chu⁴ and Lifeng Lin^{1*}

Introduction

In epidemiology and medical research, the choices of effect measures for binary outcomes have been long debated. Common choices include the risk difference (RD), relative risk (RR), and odds ratio (OR). The RD is often considered more heterogeneous than the ratio measures, RR and OR [1, 2]. Nevertheless, the arguments supporting this claim have been challenged [3]. For example, more rejections of homogeneity in hypothesis testing of RDs are expected than those of ORs. This article empirically compares the heterogeneity magnitudes between the RD, RR, and OR.

Methods

We applied heterogeneity measures to a large Cochrane database of meta-analyses [4]. The Cochrane Library publishes systematic reviews on a wide range of health-care-related topics. We searched for all Cochrane reviews available online from issue 1 in 2003 to issue 1 in 2020. The search strategy for an older version of the Cochrane database was used in our earlier work [5–7]. In the Cochrane Library, each issue was published monthly, and it included systematic reviews on new topics with formal meta-analyses as well as protocols without formal analyses. An issue may also publish notices to withdraw outdated or flawed reviews and protocols. In this study, we iteratively included all published reviews that reported

statistical data in each issue and excluded all withdrawn reviews. In total, we identified 64,929 meta-analyses.

In addition, a Cochrane review could investigate multiple disease outcomes and/or multiple intervention comparisons. Therefore, the meta-analyses within the review may not be independent due to the correlations between outcomes or intervention comparisons. For removing the impact of such potential correlations on heterogeneity, we also conducted sensitivity analyses, which were restricted to the meta-analyses with the largest number of studies from each Cochrane review. A total of 3125 meta-analyses were included in the sensitivity analyses.

We focused on the heterogeneity measure I^2 and also considered the CV_B statistic as a supplemental measure. We reanalyzed each Cochrane meta-analysis and obtained the heterogeneity measures using each effect measure. The RR and OR were analyzed on the logarithmic scale. The I^2 is widely used and is interpreted as a percentage of total variation due to heterogeneity rather than sampling error [4]. The CV_B is the between-study coefficient of variation used for providing further insight into heterogeneity magnitudes; it is calculated as the ratio of the between-study standard deviation τ over the absolute value of the overall effect size [8]. In this article, we estimated the between-study variance τ^2 using both the DerSimonian–Laird (DL) and restricted maximum likelihood (REML) methods; the former is the most popular while the latter is recommended with better statistical performance.

*Correspondence: linl@stat.fsu.edu

¹ Department of Statistics, Florida State University, 411 OSB, 117 N

Woodward Avenue, Tallahassee, FL, USA

Full list of author information is available at the end of the article



Results

Figure 1 and Fig. S1 present the histograms of $\hat{\tau}$ on a logarithmic scale for the RD, RR, and OR based on the REML and DL estimation methods. Because τ that truly equals 0 may not be exactly estimated as 0, depending on the tolerance of the REML algorithm’s convergence, the histograms in Fig. 1 shows small peaks at very small $\hat{\tau}$ values. As the RD, RR, and OR are on different scales, the magnitudes of their corresponding $\hat{\tau}$ may not be directly comparable. In general, the RR and OR led to $\hat{\tau} < 0.01$ in more meta-analyses than the RD (Table S1).

Among the 64,929 Cochrane meta-analyses, 48.09% of RDs had $I^2 = 0\%$ based on the DL method, while about 56% of RRs and ORs had $I^2 = 0\%$. The REML algorithm failed to converge in a few meta-analyses ($\leq 0.22\%$) and I^2 was not calculable; for the remainder, 43.56% of RDs had $I^2 = 0\%$, while about 50% of RRs and ORs had $I^2 = 0\%$. About 6% of RDs, RRs, and ORs had $0\% < I^2 \leq 1\%$; their REML estimates of τ were very close, but not exactly equal, to 0. Fewer DL estimates ($\leq 0.40\%$) led to $0\% < I^2 \leq 1\%$, while the DL and REML methods produced similar numbers of meta-analyses with $0\% \leq I^2 \leq 1\%$ (Table S2). In about 40% of meta-analyses, the RDs’ I^2 were larger than the RRs’ or ORs’ by over 1%, while in about 10 to 15% of meta-analyses, the RDs’ I^2 were smaller than the RRs’ and ORs’ by over 1% (Table S3). Based on the Q test, there were more meta-analyses (about 10%) with significant heterogeneity for RDs and non-significant heterogeneity for RRs or ORs than meta-analyses (about 1%) with non-significant for RDs and significant heterogeneity for RRs or ORs (Table S4). The RDs’ histogram was right-skewed, with a peak around $I^2 = 70\%$; the RRs’ and ORs’ histograms were less skewed, with peaks around $I^2 = 50\%$ (Figs. 2A and S2). Table S5 presents the mean and quantiles of I^2 ; they were based on 23,966 meta-analyses with $I^2 > 0\%$ for all three measures and both the DL and REML methods to avoid the impact of many $I^2 = 0\%$. The RDs’

descriptive statistics of I^2 were noticeably larger than the RRs’ and ORs’.

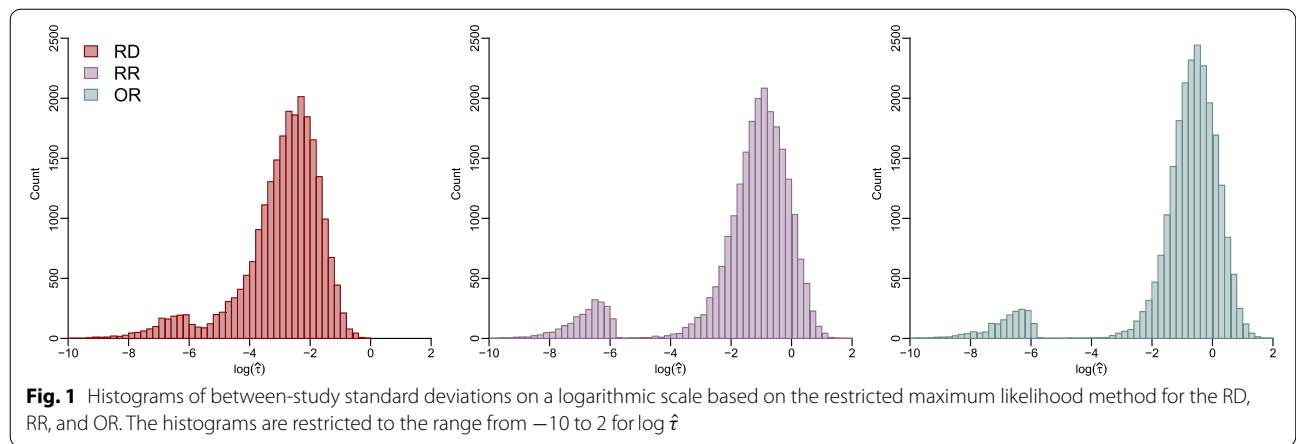
Categorized by the number of studies, the average study size, and the total number of events in a meta-analysis, RDs continued to have larger I^2 than RRs and ORs in each category (Fig. S3). The I^2 slightly decreased as the number of studies increased, consistent with previous findings [9]. It remained nearly unchanged as the average study size increased and noticeably increased as the total number of events increased.

Similar to the trends of I^2 , the histograms in Fig. 2B and S4 indicate that RDs generally had greater CV_B values than RRs and ORs. The conclusions regarding CV_B by categories of number of studies, average study size, and the total number of events in a meta-analysis were also consistent with those regarding I^2 (Figure S5).

In sensitivity analyses using the 3125 meta-analyses with the largest number of studies from each review, the histograms’ overall trends were similar to those based on the complete datasets (Figs. S6 and S7).

Discussion

Our findings consistently supported that the RD seems more heterogeneous than the RR and OR. Yet, large uncertainties in I^2 may confound these findings. The accuracy of I^2 may also be questionable in meta-analyses with few studies and/or rare events [10]. In addition, I^2 has several limitations; for example, it increases as sample sizes increase for the same τ^2 . The CV_B overcomes this drawback, while it is also subject to some disadvantages, as it increases rapidly for the overall effect size approaching 0. Nevertheless, they are arguably the appropriate tools with intuitive interpretations available in the current research synthesis literature to compare heterogeneity of measures across different scales. We intend our findings as supporting evidence rather than an assertion about heterogeneity magnitudes.



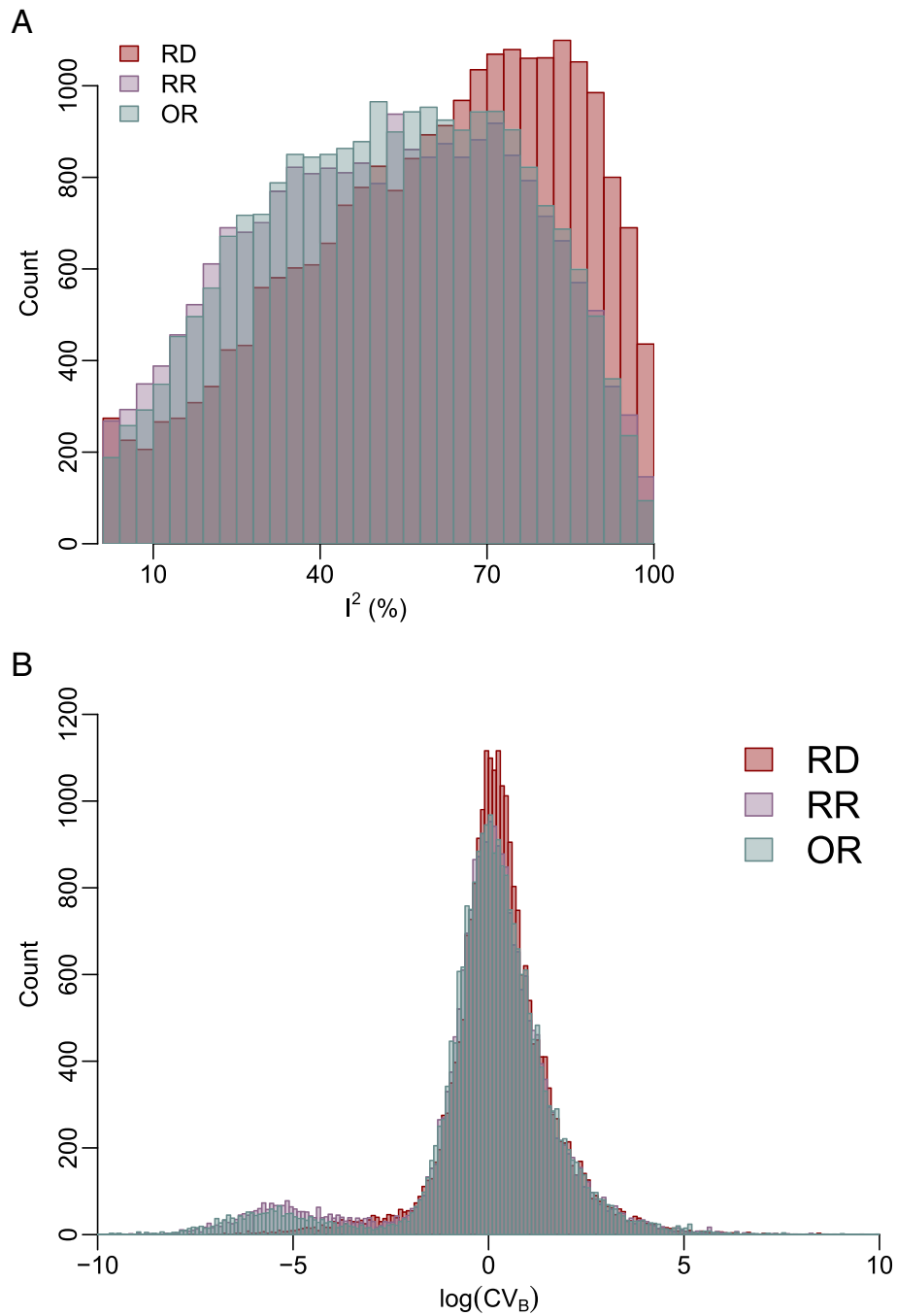


Fig. 2 Histograms of I^2 (**A**) and CV_B on a logarithmic scale (**B**) based on the restricted maximum likelihood method for the RD, RR, and OR. **A** Restricted to $I^2 > 1\%$. **B** Restricted to the range from -10 to 10 for better visualizations

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13643-022-01895-7>.

Additional file 1: Table S1. Summary of situations where $\hat{\tau}$ is not calculable or takes very small values (<0.01) among the 64,929 meta-analyses. **Table S2.** Summary of situations where I^2 is not calculable, equals 0%, or takes very small values ($\leq 1\%$) among the 64,929 meta-analyses. **Table S3.** Comparisons between I^2 of the RD, RR, and OR within the 64,929 meta-analyses. **Table S4.** Q test results (with the significance level at 0.05) among the pairs of RD, RR, and OR within the 64,929 meta-analyses. **Table S5.** Summary of descriptive statistics of I^2 (%) among the 23,966 meta-analyses with $I^2 > 0\%$ for all three effect measures based on both the DL and REML methods. **Figure S1.** Histograms of between-study standard deviations on a logarithmic scale based on the DerSimonian–Laird method for the RD, RR, and OR. The histograms are restricted to the range from -8 to 2 for $\log \hat{\tau}$. **Figure S2.** Histogram of I^2 based on the DerSimonian–Laird method for the RD, RR, and OR, restricted to $I^2 > 1\%$ for better visualizations. **Figure S3.** Boxplots of I^2 for the RD, RR, and OR categorized by the number of studies (panels a and b), average study size (panels c and d), and total number of events (panels e and f), restricted to $I^2 > 1\%$. The left panels a, c, and e are based on the DerSimonian–Laird method, and the right panels b, d, and f are based on the restricted maximum likelihood (REML) method. **Figure S4.** Histogram of CV_B on a logarithmic scale based on the DerSimonian–Laird method for the RD, RR, and OR. **Figure S5.** Boxplots of CV_B on a logarithmic scale for the RD, RR, and OR categorized by the number of studies (panels a and b), average study size (panels c and d), and total number of events (panels e and f). The left panels a, c, and e are based on the DerSimonian–Laird method, and the right panels b, d, and f are based on the restricted maximum likelihood (REML) method. **Figure S6.** Histograms of I^2 for the RD, RR, and OR, restricted to $I^2 > 1\%$ for better visualizations, among the meta-analyses with the largest number of studies from each Cochrane review. Panel a is based on the DerSimonian–Laird method, and panel b is based on the restricted maximum likelihood (REML) method. **Figure S7.** Histograms of CV_B on a logarithmic scale for the RD, RR, and OR among the meta-analyses with the largest number of studies from each Cochrane review. Panel a is based on the DerSimonian–Laird method, and panel b is based on the restricted maximum likelihood (REML) method.

Authors' contributions

YZ: methodology, software, formal analysis, writing - original draft, writing - review & editing, visualization. EHS: writing - review & editing. CX: writing - review & editing. HC: conceptualization, writing - review & editing, funding acquisition. LL: conceptualization, methodology, data curation, writing - original draft, writing - review & editing, project administration, funding acquisition. All authors read and approved the final manuscript.

Funding

This research was supported in part by the US National Institutes of Health/ National Library of Medicine grant R01 LM012982 (LL and HC) and the National Institutes of Health/ National Institute of Mental Health grant R03 MH128727 (LL). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Availability of data and materials

The datasets and code for this study are available upon reasonable request from the corresponding author.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Statistics, Florida State University, 411 OSB, 117 N Woodward Avenue, Tallahassee, FL, USA. ²Key Laboratory of Population Health Across Life Cycle, Ministry of Education of the People's Republic of China, Hefei, Anhui, China. ³Anhui Provincial Key Laboratory of Population Health and Aristogenetics, Anhui Medical University, Hefei, Anhui, China. ⁴Division of Biostatistics, University of Minnesota, Minneapolis, MN, USA.

Received: 17 May 2021 Accepted: 25 January 2022

Published online: 12 February 2022

References

- Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. *Stat Med*. 2000;19(13):1707–28.
- Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med*. 2002;21(11):1575–600.
- Poole C, Shrier I, VanderWeele TJ. Is the risk difference really a more heterogeneous measure? *Epidemiology*. 2015;26(5):714–8.
- Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557–60.
- Lin L, Chu H, Murad MH, Hong C, Qu Z, Cole SR, et al. Empirical comparison of publication bias tests in meta-analysis. *J Gen Int Med*. 2018;33(8):1260–7.
- Ma X, Lin L, Qu Z, Zhu M, Chu H. Performance of between-study heterogeneity measures in the Cochrane Library. *Epidemiology*. 2018;29(6):821–4.
- Lin L, Shi L, Chu H, Murad MH. The magnitude of small-study effects in the *Cochrane Database of Systematic Reviews*: an empirical study of nearly 30 000 meta-analyses. *BMJ Evid Based Med*. 2020;25(1):27–32.
- Takkouche B, Khudyakov P, Costa-Bouzas J, Spiegelman D. Confidence intervals for heterogeneity measures in meta-analysis. *Am J Epidemiol*. 2013;178(6):993–1004.
- Int'Hout J, Ioannidis JPA, Borm GF, Goeman JJ. Small studies are more heterogeneous than large ones: a meta-meta-analysis. *J Clin Epidemiol*. 2015;68(8):860–9.
- Xu C, Furuya-Kanamori L, Zorzela L, Lin L, Vohra S. A proposed framework to guide evidence synthesis practice for meta-analysis with zero-events studies. *J Clin Epidemiol*. 2021;135:70–8.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

