HHS Public Access

Author manuscript

Epidemiol Methods. Author manuscript; available in PMC 2018 April 01.

Published in final edited form as:

Epidemiol Methods. 2017 April ; 6(1): . doi:10.1515/em-2016-0018.

A Bias in the Evaluation of Bias Comparing Randomized Trials with Nonexperimental Studies

Jessica M. Franklin[1,*] **[Assistant Professor]**, **Sara Dejene**[1] **[Research Assistant]**, **Krista F. Huybrechts**[1] **[Assistant Professor]**, **Shirley V. Wang**[1] **[Assistant Professor]**, **Martin Kulldorff**[1] **[Professor]**, and **Kenneth J. Rothman**[2,3] **[Professor and Distinguished Fellow]**

[1]Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

[2]Research Triangle Institute, Research Triangle Park, NC, USA

[3]Department of Epidemiology, Boston University School of Public Health, Boston, MA USA

Abstract

In a recent *BMJ* article, the authors conducted a meta-analysis to compare estimated treatment effects from randomized trials with those derived from observational studies based on routinely collected data (RCD). They calculated a pooled relative odds ratio (ROR) of 1.31 (95% confidence interval [CI]: 1.03–1.65) and concluded that RCD studies systematically over-estimated protective effects. However, their meta-analysis inverted results for some clinical questions to force all estimates from RCD to be below 1. We evaluated the statistical properties of this pooled ROR, and found that the selective inversion rule employed in the original meta-analysis can positively bias the estimate of the ROR. We then repeated the random effects meta-analysis using a different inversion rule and found an estimated ROR of 0.98 (0.78–1.23), indicating the ROR is highly dependent on the direction of comparisons. As an alternative to the ROR, we calculated the observed proportion of clinical questions where the RCD and trial CIs overlap, as well as the expected proportion assuming no systematic difference between the studies. Out of 16 clinical questions, 50% CIs overlapped for 8 (50%; 25 to 75%) compared with an expected overlap of 60% assuming no systematic difference between RCD studies and trials. Thus, there was little evidence of a systematic difference in effect estimates between RCD and RCTs. Estimates of pooled RORs across distinct clinical questions are generally not interpretable and may be misleading.

Keywords

Bias; Meta-analysis; Observational studies; Randomized trials; Routinely collected data

[*]Address correspondence to Dr. Jessica Franklin, 1620 Tremont St., Suite 3030, Boston, MA 02120. JMFranklin@partners.org. Ph: 617-278-0675.

Contributors

JMF and KJR conceived of the study. JMF and SD extracted data and performed statistical analysis. JMF drafted the manuscript and web appendix, and all authors made revisions. All authors read and approved the final version of the manuscript. JMF is the guarantor.

Competing interests

No competing interests.

Ethical approval

This study does not require IRB approval.

Introduction

Routinely collected data (RCD), such as health insurance claims and electronic health records (EHRs), have become increasingly popular as data sources for studies of the comparative effectiveness and safety of treatments as used in routine care.(1) For many clinical questions, randomized controlled trials (RCTs) are unlikely to be conducted, leaving RCD as a primary source of evidence capable of informing clinical decision-making.(2) Because the treatments in RCD have not been randomly allocated to patients, these studies are subject to confounding bias, a problem that is limited in RCTs, where random assignment of treatments ensures estimates are unbiased on average.(3) Although methods exist to control for confounding in study design and data analysis, the existence of unmeasured confounding variables as well as data errors in variables that are measured could theoretically result in residual bias in studies from RCD.

Several studies have attempted to evaluate the agreement of estimated treatment effects between observational data and RCTs through estimation of the relative odds ratio (ROR), defined as the estimated odds ratio (OR) from RCT data divided by the estimated OR from observational data.(4,5). In a recent meta-analysis by Hemkens et al. (4), the authors compared published RCD studies and subsequent RCTs using the ROR, but inverted the clinical question and corresponding treatment effect estimates for all study questions where the RCD estimate was >1, thereby ensuring that all RCD estimates indicated protective effects. They found a meta-analytic ROR of 1.31 (95% confidence interval [CI]: 1.03–1.65) across 16 distinct clinical questions, from which they inferred that RCD studies "systematically and substantially overestimated the mortality benefits of medical treatments compared to subsequent trials". However, this estimate of bias is itself biased.

In this paper, we explain how the pooled ROR estimate was biased as a result of selectively inverting the direction of the clinical questions under study. We then discuss problems in interpretation of the pooled ROR, even in the absence of inversion of some study results. Finally, we illustrate an alternative approach to evaluating the agreement between RCTs and observational studies based on the overlap in CIs between study types, and we use it to reanalyze the data of Hemkens et al. We also apply both approaches to the RCT data alone in order to compare the agreement among RCTs on the same clinical question.

Bias resulting from selective odds ratio inversion

Hemkens et al. extracted estimates of treatment effect from 16 RCD studies that utilized propensity scores to adjust for confounding and reported the comparative effect of interventions on mortality. They then compared these estimated effects with estimated effects from one or more RCTs that investigated the same clinical question and were published after the corresponding RCD study. In total, they included 36 RCTs.(4) When there was more than one available RCT for a given clinical question, the several RCT estimates were pooled to obtain a single RCT estimate for each study question. The pooled RCT estimate divided by the RCD estimate provided the ROR for each clinical question. These RORs were then combined in a final meta-analysis to obtain a single ROR across all

clinical questions. According to Hemkens et al., an overall ROR near 1 would indicate good agreement between the RCD and RCT estimates, while values that departed from 1 would indicate poor agreement.

To understand the difficulty with this interpretation, assume that there is no systematic bias in the RCD studies and that the RCD and RCTs are estimating the same treatment effect parameter. In this case, if we have one RCD estimate and one RCT estimate for a given clinical question, sometimes the RCT estimate will be greater than the RCD estimate (ROR>1), sometimes the reverse (ROR<1), but on average across many RCD-RCT pairs, there would be no systematic difference between the RCD and the RCT, yielding a meta-analytic ROR of approximately 1. However, this property does not hold when the ORs for some RCD-RCT pairs are inverted, as they were in the analysis of Hemkens et al.

For all study questions where the RCD OR was greater than 1, Hemkens et al. inverted the clinical question and the OR value so that if the original OR estimate compared treatment A with treatment B, the new inverted OR (calculated as 1/OR) compared treatment B with treatment A, where either treatment could refer to a “control” treatment or no treatment. Corresponding RCT estimates were also inverted to ensure the direction of comparison matched that of the RCD estimates. This inversion causes the ROR for the associated clinical question also to be inverted, as seen in the equation:

$$\mathrm{ROR}_{AvsB}=\frac{OR_{AvsB,RCT}}{OR_{AvsB,RCD}}=\frac{\frac{odds_{A,RCT}}{odds_{B,RCT}}}{\frac{odds_{A,RCD}}{odds_{B,RCD}}}=\frac{\frac{odds_{B,RCD}}{odds_{A,RCD}}}{\frac{odds_{B,RCT}}{odds_{A,RCT}}}=\frac{OR_{BvsA,RCD}}{OR_{BvsA,RCT}}=\frac{1}{\mathrm{ROR}_{BvsA}}$$

where ROR_{AvsB} refers to the ROR for the comparison of treatment A to treatment B, $OR_{AvsB,RCT}$ ($OR_{AvsB,RCD}$) refers to the OR comparing treatment A to treatment B estimated from RCTs (RCD), and $odds_{A,RCT}$ ($odds_{A,RCD}$) is the odds of outcome on treatment A, as estimated from RCTs (RCD). For example, an ROR of 0.8 becomes an ROR of 1.25 after inversion. In the case of no systematic bias, this corresponds to a negative random error in the ROR being converted into a positive random error or vice versa. However, this inversion is not random with respect to the ROR because a ROR<1 is more common when the RCD OR estimate is >1.

For example, as shown in Figure 1, if the expected value of the OR is 1 for both the RCD and RCT studies, then $\Pr(ROR > 1|OR_{RCD}) > 0.5$ whenever $OR_{RCD} < 1$. In this case, the ROR is not inverted, so the ROR will be greater than 1 more than 50% of the time. When $OR_{RCD} > 1$, then $\Pr(ROR < 1|OR_{RCD}) > 0.5$. However, in this scenario we invert the ROR, and since $\Pr(ROR<1|OR_{RCD})=\Pr(\frac{1}{ROR}>1|OR_{RCD})$, the inverted ROR will be greater than 1 more than 50% of the time. Thus, when applying selective inversion, we will more often have RORs in the meta-analysis that are greater than 1, even in the absence of systematic bias in the RCD studies. This bias may at least partially explain the findings of Hemkens et al. Details and proof of the bias stemming from the inversion method is given in Supplement A.

Estimating the ROR without inversion

Given the problems created by selectively inverting some clinical questions, one could instead pursue estimation of the pooled ROR without inversion, simply using the OR estimates as they originally appeared in the RCD studies. However, the pooled ROR is still a flawed metric even without inversion, since it is entirely dependent on the direction of the comparisons under study, and in studies of two active treatments, the direction is arbitrary, depending on which treatment is chosen as the referent. For example, one clinical question included in the review of Hemkens et al. compared coronary artery bypass grafting (CABG) to placement of a drug eluting stent, and the associated ROR was 2.08, indicating over-estimation of the relative effect.(6) However, if the authors had instead reported drug eluting stent versus CABG, the ROR would instead be 0.48, indicating under-estimation of the effect. Since each of the clinical questions included in the analysis of Hemkens et al. could have been reported in either direction, there is a wide range of potential outcomes. Thus, there would be no reasonable basis for combining results across different, unrelated clinical questions.

To demonstrate the dependence of the estimated ROR on the direction of the comparisons under study, we extracted the number of patients and deaths in each treatment arm from each study reported in the Hemkens et al. paper and reproduced their meta-analysis, which inverted study questions whenever the RCD OR>1 (Figure 2). We then inverted clinical questions whenever the RCT OR>1, and applied the same meta-analysis model to these newly inverted data. This method results in a pooled ROR of 0.98 [0.78 to 1.23] (Figure 3). We also evaluated the most extreme results possible within these data, inverting to achieve either ROR > 1 or ROR < 1 for all clinical questions, yielding pooled ROR values of 1.47 (1.16–1.85) and 0.68 (0.54–0.86), respectively. Thus, many conclusions regarding bias are possible, depending simply on the direction of reported results.

Furthermore, even in cases where there is a clear directionality of comparison for all study questions, such as active treatment versus control, interpretation of the ROR is difficult. A single ROR value greater than 1 may be the result of 3 distinct possibilities: 1) the RCT evidence indicates that the intervention is harmful (OR>1), and the observational study indicates that it is protective (OR<1); 2) both types of study indicate that the intervention is harmful (OR>1), and the RCT estimate is larger in magnitude; or 3) both types of study indicate that the intervention is protective (OR<1), and the RCT estimate is smaller in magnitude. When RORs addressing different clinical questions are combined, these different possibilities become blended in the summary result, thwarting a sensible interpretation.

Interpretation of the ROR also depends on the assumption that bias operates in the same direction for all studies, which may be unlikely in most scenarios. For example, patients nearing the end of life may be more likely to get some treatments, such as rescue procedures, but less likely to receive preventive therapy, such as statins.(7,8) Since end of life status is often not captured in RCD, this important predictor of mortality could bias observational study results in either direction, depending on the study. Similarly, RCTs can be biased in either direction, resulting in increases or decreases in the distance between the observational and randomized estimates.

Finally, in cohort studies it is preferable to estimate a risk ratio rather than an OR.(9) Although many scientists analyzing cohort studies report OR, perhaps because of the popularity of logistic regression, it is not the best metric on which to base an evaluation of study biases.

Confidence interval overlap

In addition to the meta-analytic ROR, Hemkens et al. also reported the proportion of clinical questions where the 95% CIs overlapped between the RCD study and the pooled RCTs. This metric has some appeal because it does not depend on the direction of comparison in the clinical question under study. However, this proportion must be compared with the expected proportion overlapping if there is no systematic difference between the RCD and RCT studies. It is also more informative to use a shorter confidence interval in an exercise like this, since two 95% confidence intervals will almost always overlap even when there are slight differences in the true risks. Thus, we calculated the observed proportion of clinical questions with overlapping CIs at the 25%, 50%, 75%, and 95% levels, along with a 95% CI for this proportion. We also calculated the expected proportion of overlap for each confidence level under the assumption that there was no difference in the expected value between the RCD and RCT studies. Because the expected proportion depends on the standard error of each study, we calculated it separately for each RCD/RCT pair, as described in Schenker and Gentleman (10), and then averaged across the clinical questions to calculate the overall expected proportion. Code for reproducing this analysis is given in Supplement B.

The 95% CIs from the RCD studies overlapped the 95% CIs for the pooled RCT estimates for all 16 clinical questions (100%; 95% CI: 79–100%) (Table 1). This agrees with the expected 98% proportion of overlap when there is no bias in the RCD studies. Agreement between observed and expected overlap was good for other confidence levels as well. For example, the 50% CIs overlapped for 8 questions (50%; 25–75%), which is close to the expected 60% overlap.

Agreement among randomized trials

To further illustrate the bias in the inversion method, we used that method and the data of Hemkens et al. to compare agreement between the first randomized trial for a clinical question with the remainder of trials reporting on the same question. The first RCT to be published for a clinical question was thus substituted for the RCD study in the primary analysis. Subsequent RCTs were pooled and used to validate the findings of the index RCT. If a clinical question had only one RCT published, meaning there were no later trials to use for comparison, it was dropped from this analysis, leaving a total of 12 clinical questions and 32 randomized trials. We then applied both the method of Hemkens et al. and the CI overlap approach described above. When applying the Hemkens et al. approach, we inverted the direction of comparison whenever the first RCT OR was > 1. We then calculated the ROR for each clinical question, comparing the pooled RCT OR to the first RCT OR, and combined the ROR across clinical questions using meta-analysis.

When inverting estimates based on the OR from the first RCT, the pooled ROR is 1.46 (0.97–2.18) (Figure 4), again indicating strong positive bias, based on the interpretation provided in the Hemkens et al. paper. This bias likely stems from the method used, however, since the comparison is between one randomized trial and other randomized trials of the same question. In contrast, the 50% CIs overlapped for 8 clinical questions (67%; 35–90%), compared with the expected overlap of 63%. Again, results using other confidence levels gave similar results.

Discussion

In this paper, we found that the previously published meta-analysis comparing estimates of treatment effect from RCTs versus RCD was flawed, creating positive bias in the pooled estimate of the ROR. Moreover, we demonstrated that the estimated ROR is highly dependent on the direction of comparisons under study, making it an unreliable measure of the agreement between treatment effect estimates when combining across multiple clinical questions, each of which could be inverted. When using a more appropriate analysis that does not depend on the direction of comparisons, we found that agreement between RCT and RCD estimates was similar to what would be expected if there were no bias in the RCD studies. In general, there was no evidence of systematic bias in the estimation of mortality effects from RCD studies.

Hemkens et al. have created an excellent and valuable data set for evaluating the scientific reliability of routinely collected health data with propensity score adjustment. It is a particular strength that only RCD studies conducted before the corresponding randomized trials were included, so that RCD results could not have been influenced by previous trial results. Unfortunately, a simple but fatal flaw in the statistical method invalidated their conclusions. Our re-analysis has shown that there is no more divergence between the results from the propensity score adjusted RCD studies versus the subsequent RCTs than what one would expect by chance in the absence of systematic bias. While there may be bias in some or all of the studies derived from RCD, this bias appears to be minor compared with the uncertainty due to the randomness that is inherent in the effect estimates from the RCTs. In addition, some or all of the RCTs may have bias as well, increasing the probability of discrepancies between the studies.

Despite the thorough literature review, this dataset does have some limitations. First, the studies chosen cover only a very small fraction of all published studies of treatment effects estimated from RCD. Online responses to the paper of Hemkens et al. have pointed out additional criticisms of the underlying data. For example, Suissa noted that several of the RCD studies included in the review were subject to immortal time bias, an important design flaw that can cause severe bias.(11) Hankins et al. further note that RCD studies estimate effects in populations that are often very different from the highly restrictive populations included in RCTs.(12) Finally, an important obstacle to drawing an inference about the differences between estimates of treatment effect from RCTs versus RCD from this study is the low precision attached to the estimates of nearly all the clinical questions that were included. The most precisely estimated question-specific ROR still had a CI that extended

from 0.88 to 2.43, covering a wide range of possible conclusions about the relative estimation from RCTs versus RCD for that question.

In an important 2005 paper, Ioannidis argued that limited study size is one of six reasons why “most published research findings are false”.(13) An advantage of RCD is the ability to greatly increase study size without breaking the bank. This scenario leads to a trade-off between what is typically a higher risk of bias in RCD versus wider CIs in RCTs. To understand better the nature and magnitude of this trade-off, Hemkens’ data set should be augmented either with more RCD/RCT pairs or with RCTs that are much larger.

With the recent focus on comparative effectiveness research, many studies compare one active treatment with another.(10–12) Thus, finding that observational studies over-estimate the comparative effect of treatment A to treatment B is the same as finding that they underestimate the comparative effect of treatment B to treatment A, and the direction of the ROR will vary accordingly. Therefore, a meta-analytic ROR that combines information across many clinical questions, each of which could be inverted, cannot inform the potential direction of bias in any given RCD study. Although we focused on the overlap in CIs as an alternative analysis, some of the other metrics reported by Hemkens et al. could also be modified to compare more explicitly the observed agreement with what would be expected under the hypothesis of no bias. For example, the proportion of clinical questions with a RCD effect CI that includes the pooled effect estimate from subsequent trials (44%) could be compared with the relevant expected proportion (52%), which accounts for the substantial imprecision in the RCT estimates under study (see Supplement Table 2 for additional calculations).

The binary nature of the CI overlap approach may inadvertently lead investigators to use it as a hypothesis test of differences, which we do not recommend. A better approach to assessment of differences would be to estimate the average magnitude of the difference between observational studies and RCTs (regardless of the direction of the distances), accounting for the substantial variation in studies. However, this approach requires additional development.

Prior meta-analyses comparing randomized and observational evidence on treatment effects have also used the ROR to quantify the agreement between study types. A Cochrane review identified 14 such studies, of which 11 were reported to have found no difference (based on statistical significance) between the RCT and observational estimates.(5) Of the 3 studies that were described as finding a difference, 2 indicated lower estimates from observational studies(17,18) and 1 indicated higher estimates from observational studies.(19) None of the studies included in the review appeared to employ the selective inversion rule used by Hemkens et al., but their use of the ROR to quantify bias is just as flawed and dependent on the direction of comparison chosen for each study by investigators. In addition, several of the earliest papers comparing observational studies and RCTs used simple graphical displays without any quantitative assessment of differences.(20–22)

Conclusion

Routine collection of health data for the purposes of observational medical research is increasing, and databases are growing in terms of the number of patients and the amount and type of available information.(1,23–25) Based on the results of this study, the claims that RCD studies are especially unreliable and should be viewed skeptically are unwarranted. This is an important issue, as RCD can provide evidence on a wide variety of clinical questions and patient populations for which there are no randomized trials.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding

This study was internally funded by the Division of Pharmacoepidemiology and Pharmacoeconomics at Brigham and Women's Hospital.

Krista Huybrechts is supported by a career development grant K01MH099141 from the National Institute of Mental Health. Shirley Wang is supported by grant number R00HS022193 from the Agency for Healthcare Research and Quality.

References

1. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol. 2005; 58(4):323–337. [PubMed: 15862718]
2. Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ. 1996; 312(7040):1215. [PubMed: 8634569]
3. Rothman, KJ., Greenland, S., Lash, TL. Modern epidemiology. Lippincott Williams & Wilkins; 2008.
4. Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JPA. Agreement of treatment effects for mortality from routinely collected data and subsequent randomized trials: meta-epidemiological survey. BMJ. 2016; 352:i493. [PubMed: 26858277]
5. Anglemyer, A., Horvath, HT., Bero, L. The Cochrane Collaboration. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd; 2014. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. (http://doi.wiley.com/10.1002/14651858.MR000034.pub2) [Accessed May 12, 2016]
6. Wu C, Hannan EL, Walford G, et al. Utilization and outcomes of unprotected left main coronary artery stenting and coronary artery bypass graft surgery. Ann Thorac Surg. 2008; 86(4):1153–1159. [PubMed: 18805151]
7. Glynn RJ, Knight EL, Levin R, et al. Paradoxical relations of drug treatment with mortality in older persons. Epidemiology. 2001; 12(6):682–689. [PubMed: 11679797]
8. Glynn R, Schneeweiss S, Wang P, et al. Selective prescribing led to overestimation of the benefits of lipid-lowering drugs. J Clin Epidemiol. 2006; 59(8):819–828. [PubMed: 16828675]
9. Knol MJ, Le Cessie S, Algra A, et al. Overestimation of risk ratios by odds ratios in trials and cohort studies: alternatives to logistic regression. Can Med Assoc J. 2012; 184(8):895–899. [PubMed: 22158397]
10. Schenker N, Gentleman JF. On judging the significance of differences by examining the overlap between confidence intervals. Am Stat. 2001; 55(3):182–186.

11. Suissa, S. [Accessed May 17, 2016] Re: Agreement of treatment effects for mortality from routinely collected data and subsequent randomized trials: meta-epidemiological survey. The BMJ. 2016. (http://www.bmj.com/content/352/bmj.i493/rr-3)
12. Hankins, MC., Buysse, B., Chatzitheofilou, I., et al. [Accessed May 17, 2016] Re: Agreement of treatment effects for mortality from routinely collected data and subsequent randomized trials: meta-epidemiological survey. The BMJ. 2016. (http://www.bmj.com/content/352/bmj.i493/rr-1)
13. Ioannidis JP. Why most published research findings are false. PLos Med. 2005; 2(8):e124. [PubMed: 16060722]
14. Dreyer NA, Tunis SR, Berger M, et al. Why observational studies should be among the tools used in comparative effectiveness research. Health Aff (Millwood). 2010; 29(10):1818–1825. [PubMed: 20921481]
15. Sullivan P, Goldmann D. The promise of comparative effectiveness research. JAMA. 2011; 305(4): 400–401. [PubMed: 21266687]
16. Goldberg NH, Schneeweiss S, Kowal MK, et al. Availability of comparative efficacy data at the time of drug approval in the United States. JAMA. 2011; 305(17):1786–1789. [PubMed: 21540422]
17. Bhandari M, Tornetta P III, Ellis T, et al. Hierarchy of evidence: differences in results between non-randomized studies and randomized trials in patients with femoral neck fractures. Arch Orthop Trauma Surg. 2004; 124(1):10–16. [PubMed: 14576955]
18. Beynon, R., Harris, R., Sterne, J. The quantification of bias in randomised and non-randomised studies: the BRANDO NRS Study [poster]. Freiburg im Breisgau; Germany: 2008.
19. Furlan AD, Tomlinson G, Jadad AAR, et al. Examining heterogeneity in meta-analysis: comparing results of randomized trials and nonrandomized studies of interventions for low back pain. Spine. 2008; 33(3):339–348. [PubMed: 18303468]
20. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med. 2000; 342(25):1887–1892. [PubMed: 10861325]
21. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. N Engl J Med. 2000; 342(25):1878–1886. [PubMed: 10861324]
22. Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. Jama. 1995; 273(5):408–412. [PubMed: 7823387]
23. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med. 2015; 12(10):e1001885. [PubMed: 26440803]
24. de Lusignan S, van Weel C. The use of routinely collected computer data for research in primary care: opportunities and challenges. Fam Pract. 2006; 23(2):253–263. [PubMed: 16368704]
25. Powell AE, Davies HTO, Thomson RG. Using routine comparative data to assess the quality of health care: understanding and avoiding common pitfalls. Qual Saf Health Care. 2003; 12(2):122–128. [PubMed: 12679509]

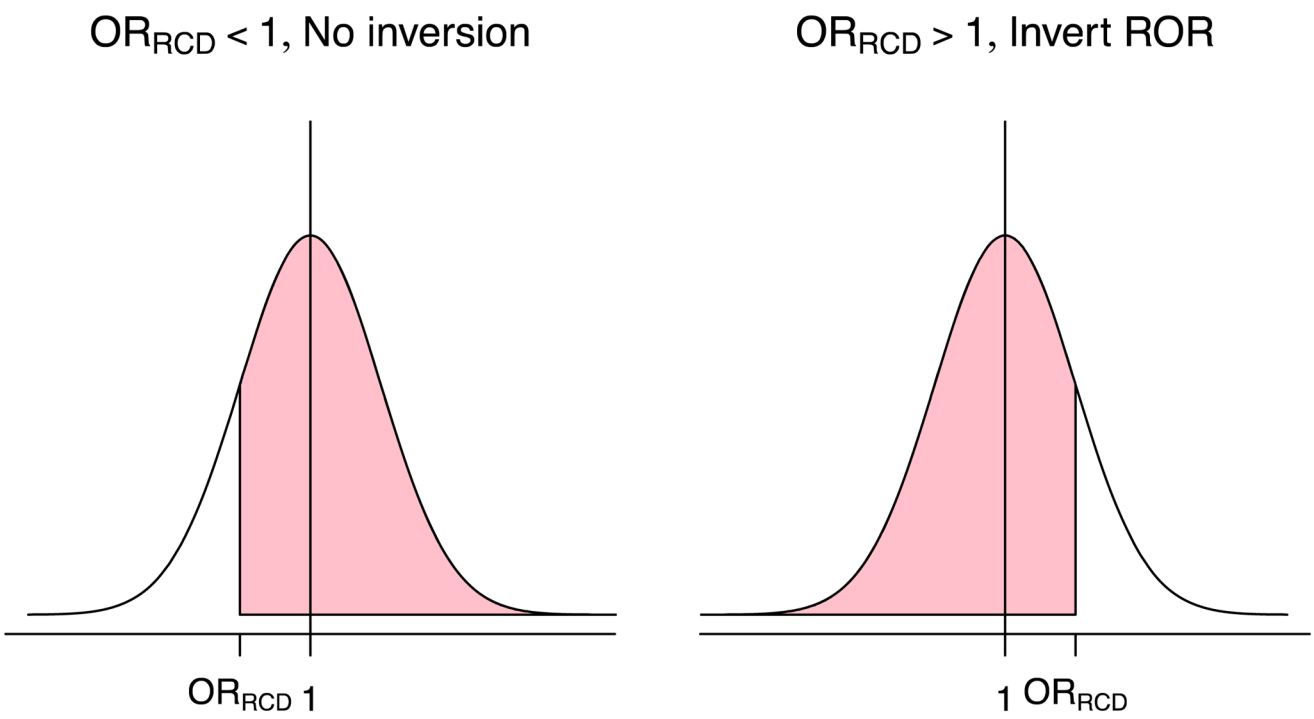

Figure 1.

A hypothetical distribution for the OR_{RCT} centered at a true value of 1. The shaded area represents the probability, after implementing the inversion rule, that $OR_{RCT} > OR_{RCD}$, which implies ROR > 1. Note that when the ROR is not inverted ($OR_{RCD} < 1$), it is more likely that ROR > 1. When the ROR is inverted ($OR_{RCD} > 1$), it is more likely that ROR < 1, which means after inversion, ROR is likely > 1.

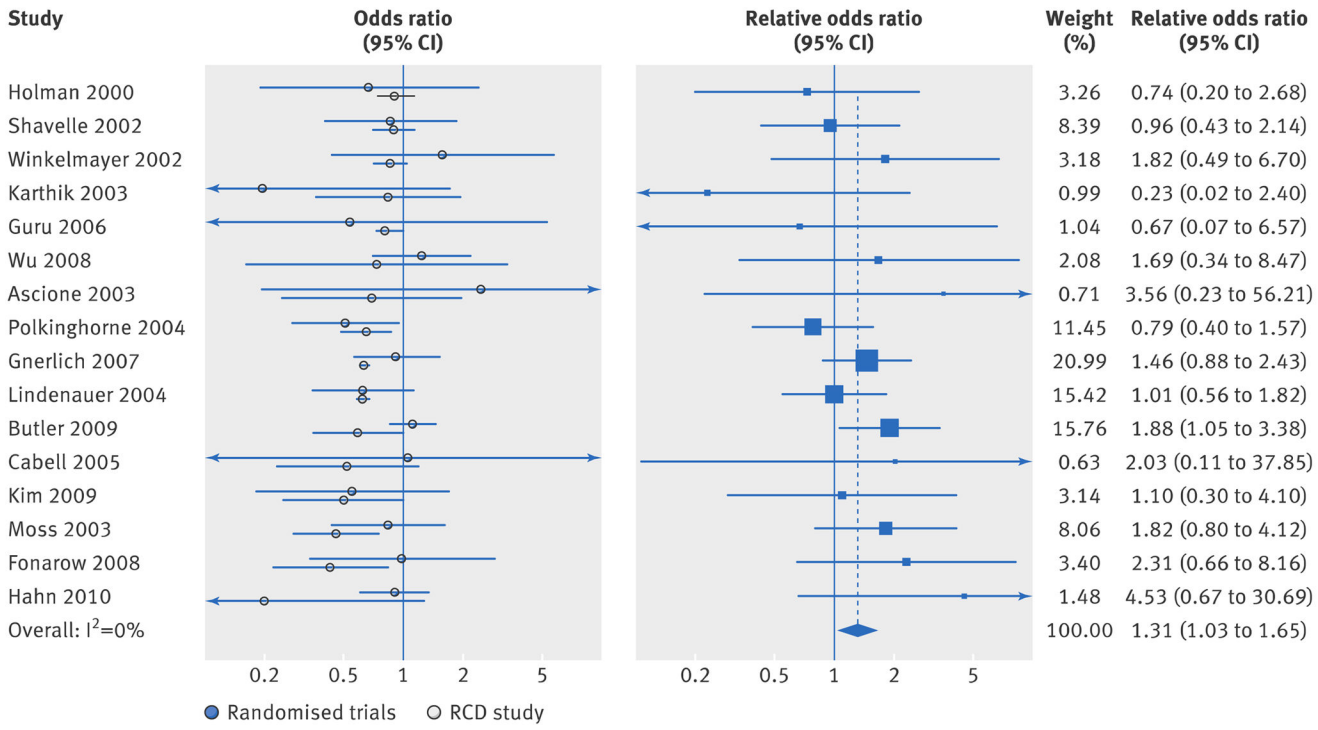

Figure 2.

Original analysis (reproduced from Hemkens et al.) of treatment effects on mortality in RCD studies and RCTs. The left panel shows the comparative effect of medical interventions on mortality reported in RCD studies and results of subsequently published trials on the same treatment comparisons. The right panel shows for each clinical question the relative odds ratio reported in trials versus the corresponding RCD study. Effect estimates are presented when inverting ORs whenever the **RCD** OR>1.

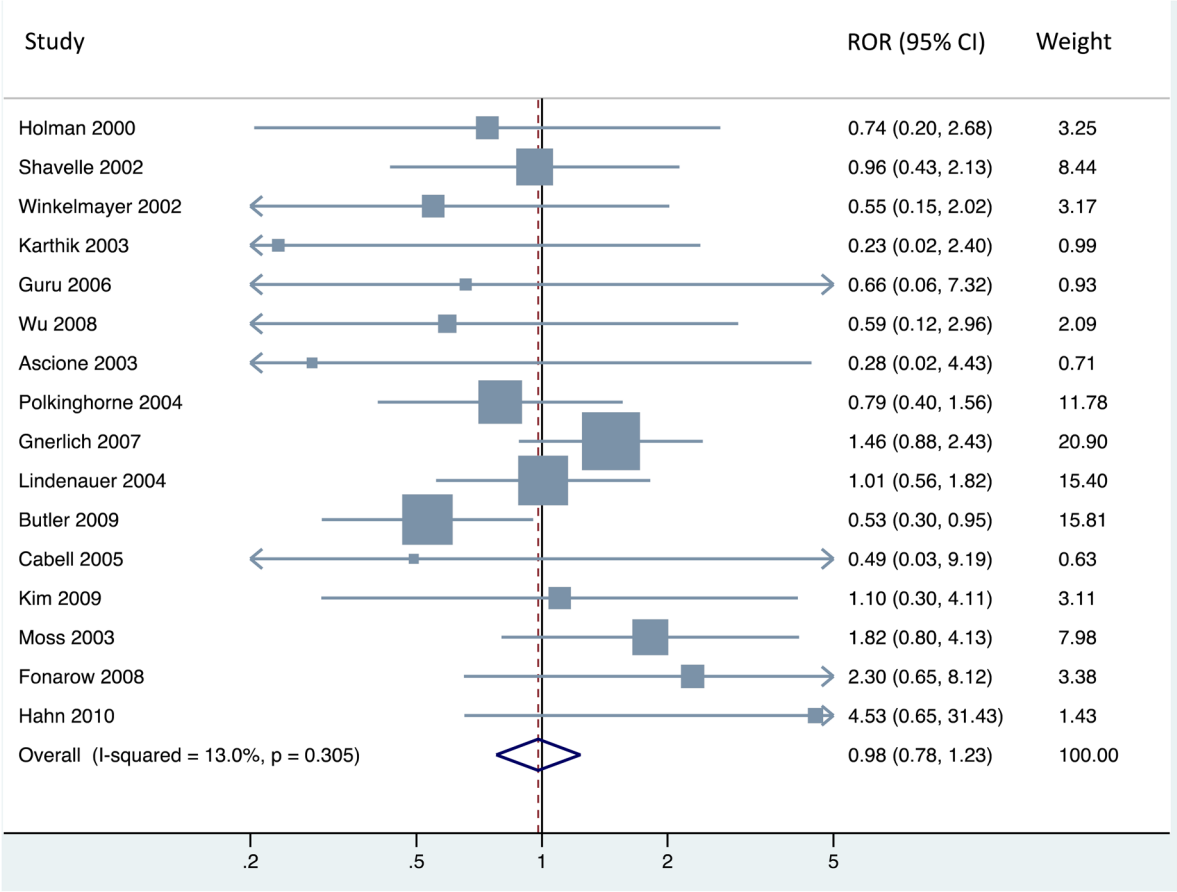

Figure 3.

Re-analysis of treatment effects on mortality in RCD studies and RCTs. For each clinical question, we present the relative odds ratio reported in trial evidence versus the corresponding RCD study. Effect estimates are presented when inverting treatment groups and ORs whenever the **RCT** OR>1.

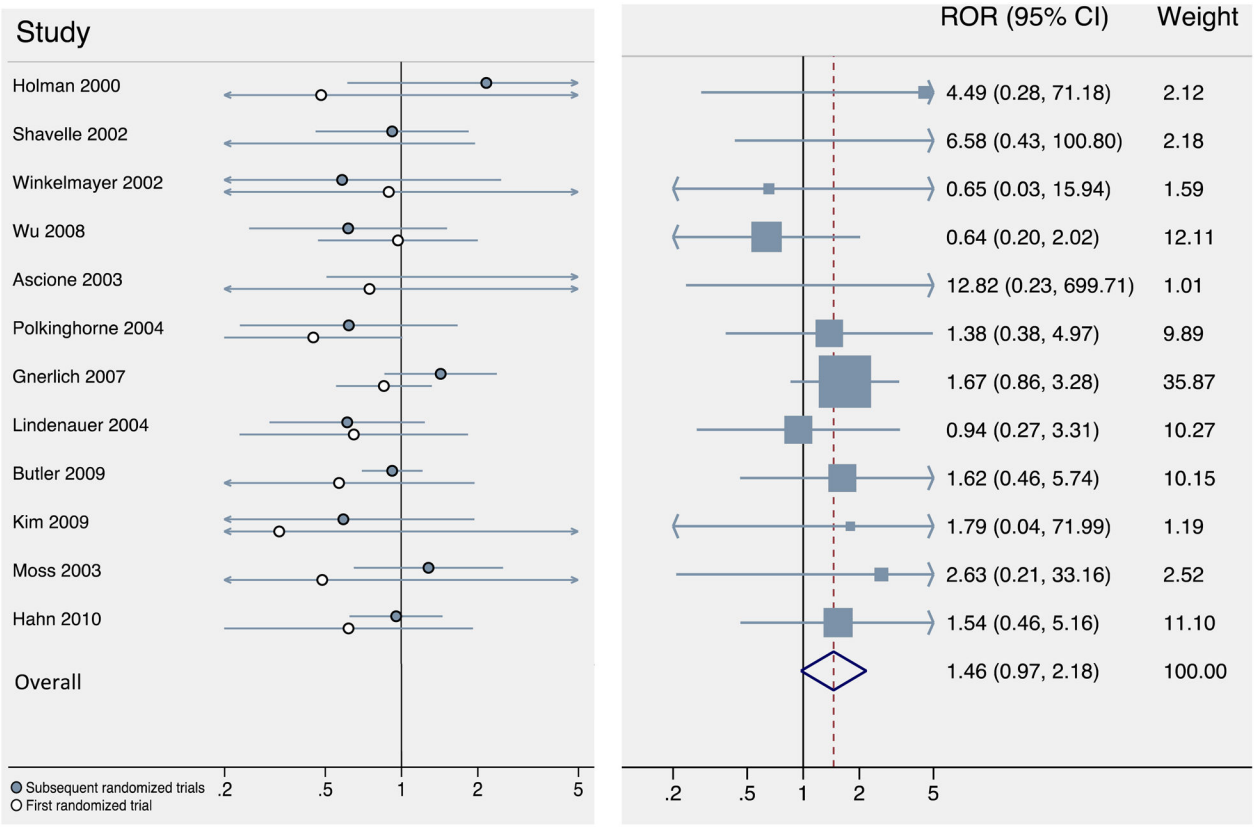

Figure 4.

Agreement among randomized trials on the same clinical question. The left panel shows the comparative effect of medical interventions on mortality reported in the first RCT for a given clinical question and results of subsequently published trials on the same treatment comparisons. Labels refer to the original RCD study for each clinical question (to be comparable with earlier figures). The right panel shows the relative odds ratio reported in subsequent clinical trials versus the first trial for each clinical question. Effect estimates are presented when inverting treatment groups and ORs whenever the **first RCT** OR>1.

Table 1

Observed and expected percent overlap of confidence intervals comparing the RCD/first RCT study and subsequent RCTs.

	RCD vs RCTs		First RCT vs others	
Confidence level	**Observed**	**Expected**	**Observed**	**Expected**
25	19 (4, 46)	31	25 (5, 57)	33
50	50 (25,75)	60	67 (35, 90)	63
75	81 (54, 96)	84	100 (74, 100)	87
95	100 (79, 100)	98	100 (74, 100)	99