Published in final edited form as:

Med Care. 2017 April ; 55(4): 328–335. doi:10.1097/MLR.0000000000000696.

Meta-analysis of Odds Ratios: Current Good Practices

Bei-Hung Chang, ScD¹ and David C. Hoaglin, PhD¹

¹Department of Quantitative Health Sciences, University of Massachusetts Medical School

Abstract

Background—Many systematic reviews of randomized clinical trials lead to meta-analyses of odds ratios. The customary methods of estimating an overall odds ratio involve weighted averages of the individual trials' estimates of the logarithm of the odds ratio. That approach, however, has several shortcomings, arising from assumptions and approximations, that render the results unreliable. Although the problems have been documented in the literature for many years, the conventional methods persist in software and applications. A well-developed alternative approach avoids the approximations by working directly with the numbers of subjects and events in the arms of the individual trials.

Objective—We aim to raise awareness of methods that avoid the conventional approximations, can be applied with widely available software, and produce more-reliable results.

Methods—We summarize the fixed-effect and random-effects approaches to meta-analysis; describe conventional, approximate methods and alternative methods; apply the methods in a meta-analysis of 19 randomized trials of endoscopic sclerotherapy in patients with cirrhosis and esophagogastric varices; and compare the results. We demonstrate the use of SAS, Stata, and R software for the analysis.

Results—In the example, point estimates and confidence intervals for the overall log-odds-ratio differ between the conventional and alternative methods, in ways that can affect inferences. Programming is straightforward in the three software packages; an appendix gives the details.

Conclusions—The modest additional programming required should not be an obstacle to adoption of the alternative methods. Because their results are unreliable, use of the conventional methods for meta-analysis of odds ratios should be discontinued.

Keywords

Meta-analysis; Odds ratio; SAS; Stata; R

Introduction

Many disciplines synthesize evidence on research questions that can be stated in terms of a numerical measure of effect for an intervention (e.g., the rate of myocardial infarction [MI]

Correspondence to: Bei-Hung Chang, ScD, 368 Plantation Street, Worcester, MA 01605, 508-856-8959 (O), bei.chang@umassmed.edu.

Potential conflicts of interest: None

in a comparison of drugs). The process of synthesizing evidence involves specifying the question, searching for all studies that reported relevant evidence, retaining those that have adequate quality, extracting their reported data on the focal effect, and, finally, analyzing those data to assess the available evidence on the research question.¹ The term “meta-analysis” is sometimes applied to the entire process of research synthesis. More often, however, it refers to the statistical analysis of the assembled data. One goal of meta-analysis is to estimate the overall effect of the intervention by combining data from individual studies. This article focuses solely on randomized controlled trials (RCTs).

Often the data on the effect of the intervention are the frequencies of an event (e.g., MI or death) in the intervention group and the control group of each study (Table 1). Among the measures of effect calculated from these frequency data, perhaps the most common is the odds ratio ($OR = ad/bc$). Meta-analyses have used various statistical techniques to estimate an overall effect that summarizes the effect sizes from the individual studies. The majority of these methods involve approximations that rely on the assumption that each study has a large sample size, but actual sample sizes are often not large. Alternative methods have been developed that do not use those approximations.

Using a dataset from the literature, this article demonstrates the use of one alternative method and conventional methods (for comparison) when the available data are the number of events and the number of subjects in each arm of each RCT. We show how to carry out the analysis in SAS, Stata, and R. In the sections that follow we

- introduce an example involving 19 randomized trials that studied endoscopic sclerotherapy for the prevention of first bleeding and reduction of mortality in patients with cirrhosis and esophagogastric varices;
- briefly review the two main categories of modeling approaches (fixed-effect and random-effects models) in meta-analysis;
- discuss conventional statistical methods that use study-level odds ratios (and hence assume that the individual studies have large samples) to estimate the overall effect;
- discuss an established alternative statistical method that directly uses the studies' numbers of events and numbers of subjects to estimate the overall effect;
- apply both the conventional methods and the alternative method to the data from the studies of sclerotherapy;
- describe programming of the alternative statistical method in SAS, Stata, and R, with attention to choices of options other than the defaults for some commands; and
- provide recommendations and some concluding discussion.

Methods

Example

To make the details of the approaches concrete, we use data from Pagliaro et al.,² who used meta-analysis to assess the effectiveness of endoscopic sclerotherapy in the prevention of first bleeding and reduction of mortality in patients with cirrhosis and esophagogastric varices. The data came from 19 randomized trials. For the treatment and control groups in each trial Table 2 lists the total number of subjects, the number of deaths, and the number of subjects with bleeding. The number of trials (19) is moderate. The numbers of subjects in each of the groups in those trials are also not large. Except for 143 in one trial, the number in the treated group ranges from 13 to 73; the corresponding numbers in the control group are 138 and 16 to 72. Two studies have zero events in one of the groups. A number of other authors have used these data in examples, including Higgins and Whitehead,³ Thompson et al.,⁴ Thompson and Sharp,⁵ Whitehead,⁶ Lu and Ades,⁷ and Simmonds and Higgins.⁸

Statistical Methods

1. Approaches: fixed effect and random effects—Approaches to meta-analysis generally separate into two categories, according to their assumption about the study-level effects. Fixed-effect methods assume that a single overall effect underlies all of the studies and that the observed study-level effects differ from that true effect only because of sampling variation within each study. This assumption may be appropriate when the studies have essentially the same design, intervention, patient population, and outcome measures. Usually, however, designs, patients' characteristics, and details of the intervention vary among studies; and it is more realistic to account for variation in the true effects by regarding them as coming from a distribution of study-level effects. Thus, random-effects methods focus on estimating the mean of that distribution and also the distribution's variance (as a summary of the heterogeneity of the study-level effects).

2. Conventional methods using study-level effect sizes—When the outcome for each subject is binary, the common methods of meta-analysis (both fixed-effect and random-effects models) use the logarithm of OR as the effect size in each study, along with an estimate of its variance (usually $\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$). Those estimated variances are usually assumed to be equal to the true variances. In practice this assumption is made routinely, regardless of the sample sizes, with the consequence that the results of the meta-analysis may be unreliable. Technically, the formula $\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$ cannot be correct: whenever any of a , b , c , and d has a positive probability of being 0, the true variance of $\log(\text{OR})$ is not finite. The formula corresponds to the variance of the normal distribution that the distribution of $\log(\text{OR})$ approaches as n_I and n_C become large. Adding a positive constant to the count in each cell avoids non-finite variance, but hardly any empirical evidence is available on how closely the distribution of the modified $\log(\text{OR})$ resembles a normal distribution.

To describe the conventional methods for fixed-effect and random-effects models more specifically, we denote the estimated effect from Study i (such as $\log(\text{OR})$) by y_i and its estimated variance by s_i^2 ($i=1, \dots, k$). The usual fixed-effect model relates y_i to the overall effect, μ :

$$y_i = \mu + e_i,$$

where e_i has (at least approximately) a normal distribution with mean 0 and variance σ_i^2 . That is, except for sampling variation within studies, each y_i is an estimate of μ . Such situations seldom arise in practice, so the random-effects model allows μ to vary among studies:

$$y_i = \mu_i + e_i$$

where the μ_i come from a distribution with mean μ and variance τ^2 . (Some methods further assume a normal distribution for the μ_i .)

The conventional fixed-effect estimate of the overall effect, μ , is a weighted mean with $w_i = 1/s_i^2$:

$$\bar{y}_w = \frac{\sum w_i y_i}{\sum w_i}.$$

It is customary to estimate the variance of $\bar{y}_w$ by $1/\sum w_i$, but this approach relies heavily on the assumption that s_i^2 is the true variance of y_i (which we denote by σ_i^2) or the assumption that the number of subjects in each group is large.

In the conventional random-effects model each observed study-level effect has two sources of variation: within study (σ_i^2) and between studies (τ^2). That is, $\text{var}(y_i) = \sigma_i^2 + \tau^2$. The conventional random-effects estimate of μ is a weighted mean, with weights that reflect both sources of variation, $w_i^* = 1/(s_i^2 + \hat{\tau}^2)$:

$$\bar{y}_w^* = \frac{\sum w_i^* y_i}{\sum w_i^*}.$$

The estimate of τ^2 , denoted by $\hat{\tau}^2$, is derived from the y_i and the s_i^2 . The customary estimate of the variance of $\bar{y}_w^*$ is $1/\sum w_i^*$. As in the fixed-effect method, it substitutes s_i^2 for the unknown σ_i^2 . Another potential source of complications is its use of $\hat{\tau}^2$ as if it were the true value, τ^2 (i.e., without making allowance for the variability of $\hat{\tau}^2$).

The random-effects approach summarized above is the basis for the procedure described by DerSimonian and Laird.⁹ They obtain an estimate of τ^2 by using the method of moments: setting $Q = \sum w_i (y_i - \bar{y}_w)^2$ equal to an expression for its expected value and solving to produce

$$\hat{\tau}^2 = \max\{0, [Q - (k-1)] / [\sum w_i - (\sum w_i^2 / \sum w_i)]\}$$

(the max operation avoids negative estimates of between-study variance, but $\tau^2 = 0$ is still possible; then $w_i^* = w_i$ and $\bar{y}_w^* = \bar{y}_w$). Unfortunately, the derivation of the expected value of Q assumes that $s_i^2 = \sigma_i^2$. A number of studies, using extensive simulations, have found that the resulting $1/\sum w_i^*$ tends to underestimate the variance of $\bar{y}_w^*$ and leads to confidence intervals (based on a normal distribution) that have lower than nominal coverage (e.g., at the 95% level) of μ .^{10–12} Also, some studies have found substantial bias in $\bar{y}_w^*$ as an estimate of μ .^{11,13} Despite these shortcomings, the DerSimonian-Laird procedure (DL) is the default method in many meta-analysis software programs. We discuss software further below.

Methods have been developed to address the shortcomings of the DL method. To estimate the variance of $\bar{y}_w^*$, Hartung and Knapp and, separately, Sidik and Jonkman developed an alternative method (the HKSJ method). The HKSJ confidence interval has been shown to perform better than DL even when the number of studies is small and sample sizes are unequal.¹² However, the HKSJ method uses the same estimate of μ as DL.

Hardy and Thompson¹⁴ proposed the profile-likelihood method to estimate μ and τ^2 and also a 95% CI for τ^2 . By assuming a normal distribution for each y_i and a normal distribution (with mean μ and variance τ^2) for the random effects, they obtain a joint likelihood and thus maximum-likelihood estimates for μ and τ^2 . They use profiles of the likelihood surface to determine a 95% confidence interval (CI) for μ and a 95% CI for τ^2 , each of which reflects the fact that the other parameter is being estimated. Also, those CIs are not forced to be symmetric about the point estimates. Against these favorable properties one must consider that the s_i^2 (which usually are only approximate) are still used as if they were the σ_i^2 and that a normal distribution for each y_i is a strong assumption, especially when y is log(OR).

For log(OR) and some other measures of effect, all of these conventional methods do not take into account correlation between the y_i and the s_i^2 , which violates one of their assumptions (that, in the random-effects model, the μ_i and the e_i are independent). From the expressions given earlier, log(OR) and its estimated variance are both functions of a , b , c , and d . Tang shows, mathematically and graphically, the relation between the underlying risk in the treated group and the study's fixed-effect weight (i.e., the reciprocal of s_i^2).¹⁵

3. Alternative methods using number of events and number of subjects in each group—When the numbers of events (and the sample sizes) in the intervention and control groups can be extracted for each of the studies, the meta-analysis can use those data directly and avoid the difficulties associated with using the studies' sample log(OR) and its estimated variance.⁸ Letting n_{Ii} denote the sample size in the intervention group of Study i and p_{Ii} denote the probability of an event for an individual in that group, the analysis models the number of events, x_{Ii} , as an observation from the binomial distribution $\text{bin}(n_{Ii}, p_{Ii})$. Similarly, the number of events in the control group, x_{Ci} , is modeled as an observation from $\text{bin}(n_{Ci}, p_{Ci})$.

For a fixed-effect meta-analysis a straightforward logistic regression model uses the logit transformation, $\text{logit}(p) = \log_e[p/(1-p)]$, to connect binomial probabilities with effects:

$$\begin{aligned}\text{logit}(p_{Ci}) &= g_i \\ \text{logit}(p_{Ii}) &= g_i + \mu.\end{aligned}$$

Here g_i is the log-odds for the control group in Study i , and μ is the common (i.e., overall) log-odds-ratio. The data enter through the binomial likelihoods for x_{Ci} and x_{Ii} . This approach incorporates the possibility that one or more of the cell counts in Table 1 equal 0.

Turner et al.¹⁶ approach the corresponding random-effects meta-analysis as a multilevel model (a type of generalized linear mixed model); the two levels are within-study variability and between-study variability. When the measure of effect is log(OR), the result is also known as a mixed-effects logistic regression model:

$$\begin{aligned}\text{logit}(p_{Ci}) &= g_i \\ \text{logit}(p_{Ii}) &= g_i + \mu + u_i \\ u_i &\sim N(0, \tau^2)\end{aligned}$$

(as in the fixed-effect model, the g_i are nuisance parameters). Stijnen et al. survey approaches to random-effects meta-analysis of event outcomes and discuss the disadvantages of the conventional approach.¹⁷

The approach of using a generalized linear mixed model does not readily extend to the log of the rate ratio, $\log(\text{RR}) = \log(p_I) - \log(p_C)$, or to the risk difference (RD), $p_I - p_C$. Complications arise because the calculations must ensure that estimated values of p_I and p_C remain between 0 and 1. Research on alternative approaches is ongoing.

One can also consider a Bayesian approach to meta-analysis, as described, for example, by Dias et al.¹⁸ For log(OR) the Bayesian analysis uses the same basic model as the mixed-effects logistic regression. It also requires prior distributions for μ and τ^2 (or τ). Markov chain Monte Carlo calculations then yield entire posterior distributions for μ , τ^2 , and other quantities. At present a Bayesian analysis requires specialized (but freely available) software and more programming than a mixed-effects logistic regression. The assistance of a statistician who has experience with Bayesian meta-analysis is likely to be essential.

Software

Fixed-effect and random-effects meta-analyses based on the study-level sample log(OR) and s_i^2 are available in a number of software environments, including the Cochrane Collaboration's Review Manager (RevMan, tech.cochrane.org) and the Stata user-contributed commands `metan` and `metaan`.¹⁹ For random-effects meta-analysis the DerSimonian-Laird procedure is often the default (as in RevMan) or the only option. In view of the documented shortcomings of DL, this situation is unfortunate. The profile-likelihood method is an option for the command `metaan` in Stata. To analyze the example dataset by the conventional approach, we used `metan` and `metaan` in Stata to fit the fixed-effect model by

the weighted-mean method and to fit the random-effects model by the DL and profile-likelihood methods. These two commands add 0.5 to all four cells for a study when at least one cell count is 0.

For working directly with the numbers of events and the sample sizes in the studies' two groups, a fixed-effect meta-analysis can use ordinary software for logistic regression (most software accepts either binomial data or individual binary data). Random-effects meta-analysis based on a mixed-effects logistic regression model, however, requires more-complex software, because estimation involves integrating out the random effect (u_j). Current software handles the numerical integration using several approximation methods, including Laplace approximation and adaptive Gauss-Hermite quadrature, and offers sophisticated users a variety of options to control the process. Because the software is designed to fit many types of mixed-effects logistic regression models, the default options may not be satisfactory for meta-analysis. Thus, we offer advice on expressing random-effects meta-analyses as mixed-effects logistic regression models in several software environments and on choosing the appropriate options.

We use the software packages SAS, Stata, and R to demonstrate fitting the mixed-effects logistic regression model for meta-analysis of the sclerotherapy data, using the numbers of events and numbers of subjects in each trial. We used the procedure glimmix in SAS and the commands melogit in Stata and glmer in R. Each has the necessary options, but a different default method. Maximum likelihood with adaptive Gauss-Hermite quadrature (AGHQ) is the default for melogit in Version 14 of Stata and also for glmer in R. The default number of quadrature points is 7 in melogit; in glmer, however, it is 1, which reduces to the Laplace approximation. Higher numbers of quadrature points provide better approximation of the likelihood function, but the improvement levels off.²⁰ The SAS procedure glimmix has many more options, and its default is restricted maximum likelihood (REML) using Taylor-series expansions.²¹ To obtain estimates of the variance parameters of the random effects, REML maximizes the likelihood that has the fixed effects subtracted off (called residual or restricted likelihood). The estimates obtained via ML are biased when the samples are small, whereas REML estimates are more nearly unbiased. Therefore REML is a good alternative to ML when the sole focus is on estimation of the variance components.²² For consistency we use the same estimation method, maximum likelihood and AGHQ with 7 quadrature points, in the three software procedures/commands to demonstrate the use of mixed-effects logistic regression for meta-analysis. The AGHQ method usually provides a good approximation to the likelihood and is considered the most reliable approximation method for single-level random-effects models. The Appendix lists the coding for this analysis in the three software packages. Other procedures and commands, such as PROC nlmixed in SAS and glm and meglm in Stata, can also be used to fit fixed-effect and mixed-effects logistic regression models for meta-analysis.

Results

The odds ratios of intervention vs. control for each of the 19 studies are displayed in Figure 1 (death outcome) and Figure 2 (bleeding outcome). In Table 3 (death outcome) and Table 4 (first bleeding outcome) the top panel shows the results for the random-effects model, and

the bottom panel shows the results for the fixed-effect model; each panel includes conventional meta-analysis approaches using study-level log(OR) and its estimated variance and meta-analyses based directly on the numbers of subjects and the numbers of events. We first compare the results from conventional vs. alternative methods in the random-effects model. We then compare those results with the results of the fixed-effect model for both conventional and alternative methods.

Death outcome (Table 3)

Random-effects model—For the direct random-effects meta-analysis that uses mixed-effects logistic regression, Stata, R, and SAS produced the same estimates (to three decimal places) of the overall treatment effect, μ [expressed as log(OR)], and its standard error (SE). However, the 95% CI produced by Stata is based on the normal distribution, whereas SAS uses a t-distribution (on $k-1=18$ d.f., to reflect the uncertainty in estimating τ^2). The estimate of μ obtained from the mixed-effects logistic regression model differs from those obtained from the DL and profile-likelihood methods (−0.374 vs. −0.349 and −0.342). Further, the 95% CIs from the mixed-effects logistic model (−0.683 to −0.065 and −0.706 to −0.042) do not contain zero, whereas the CIs from the conventional approaches do (DL: −0.704 to 0.007, Profile likelihood: −0.704 to 0.000).

Mixed-effects logistic regression in Stata, R, and SAS also produced the same estimate of the between-study variance, τ^2 . This estimate is substantially smaller than those obtained from the conventional random-effects meta-analysis based on study-level summaries (0.191 vs. 0.324 and 0.258). Proc glimmix in SAS and melogit in Stata gave the same estimated SE of τ^2 , but glmer in R does not provide an estimated SE. Stata also provided a 95% CI for τ^2 (0.053 to 0.680). The conventional method using profile likelihood also provided a 95% CI of τ^2 (0.000 to 0.781).

Fixed-effect model—The fixed-effect logistic regression model produced an estimate of the overall treatment effect slightly different from that produced by the conventional meta-analysis approach (−0.287 vs. −0.260). Both of these estimates differ substantially from the corresponding random-effects estimates. Since the 95% CIs for the between-study variance do not include zero, these differences are to be expected. Usually, a random-effects model is more appropriate than a fixed-effect model.

First bleeding outcome (Table 4)

Random-effects model—Similar to the death outcome, fitting a mixed-effects logistic regression model in Stata, R, and SAS produced the same estimates (to two decimal places) of the overall treatment effect, μ , and its standard error (SE) for the bleeding outcome. These estimates differ slightly from those produced by the DL method (μ : −0.649 vs. −0.610, $se(\mu)$: 0.276 vs. 0.270).

Mixed-effects logistic regression in SAS, R, and Stata also produced the same estimate (to two decimal places) of the between-study variance, τ^2 , and its SE. Although this estimate of τ^2 is quite similar to that produced by the profile-likelihood method (1.06 vs. 1.04), the two 95% CIs are quite different (0.421 to 2.684 vs. 0.000 to 2.651).

Fixed-effect model—For the bleeding outcome the fixed-effect logistic regression model produced a different estimate of the overall treatment effect, μ , than that from the conventional approach (−0.564 vs. −0.487). Again, the results of the fixed-effect model are very different from those of the random-effects model. The forest plot in Figure 2 shows substantial variation in the study-level treatment effects (expressed as odds ratio). The random-effects model is more appropriate than the fixed-effect model for the bleeding outcome as well.

Discussion

We have demonstrated an alternative approach for meta-analysis of paired binomial outcomes that does not rely on the assumption of known within-study variances. The approach uses the numbers of events and subjects in the two groups in the studies to fit logistic regression models (fixed-effect or mixed-effects). We applied the method to data from 19 randomized trials of endoscopic sclerotherapy, using SAS, Stata, and R. Most of the results from the alternative approach are quite different from those produced by the conventional meta-analysis approaches, which use study-level summaries that reduce the data to log-odds-ratio and its estimated SE. Those approaches also assume that within-study variances are known.

In some, perhaps many, instances the conventional approaches and the alternative approach may give similar results. Such anecdotal evidence, however, is outweighed by extensive simulation results showing that the conventional approaches are unreliable. Unfortunately, no criteria are yet available that characterize situations in which those methods are safe to use.

The alternative approach is straightforward and can be easily implemented in two commonly used commercial software packages, SAS and Stata, and one free software package, R. The Appendix includes the commands for the example. Some other alternative approaches, such as Bayesian methods, are available.

Additional care is needed for meta-analyses of rare events, when some studies report zero events in one or, especially, both groups. Kuss reviewed and compared the available methods for handling “double-zero” studies.²³

On the surface our comparison of the results from several models resembles an approach advocated by Stoto²⁴ for meta-analyses of the risk of rare adverse events of drugs. He recommends that analysts use multiple statistical models for a meta-analysis (preferably by random effects) in order to assess the sensitivity of the results to the choice of model (and associated assumptions). Such an approach may be instructive, but criteria for making an objective choice among models require careful attention. We would start by excluding models that are known to be unreliable or are generally considered inappropriate for the particular meta-analysis. Some of the examples cited by Stoto include such models. Interestingly, none of those examples include the alternative approach that we recommend, although it would avoid or mitigate some of the difficulties that arise with rare events.

Although the randomization in an RCT generally produces groups that are balanced on patient-level covariates (both measured and unmeasured), studies sometimes report odds ratios that have been adjusted for selected covariates. Before using adjusted odds ratios (and their SEs) in a meta-analysis, it is important to consider the covariates involved in the adjustment in each study. If studies adjusted for different sets of covariates, their adjusted odds ratios are not comparable, and it may not be appropriate to combine them.

Adjustment for differences in study-level characteristics, however, may offer a way of accounting for some of the heterogeneity that would otherwise be summarized in the between-studies variance (τ^2). By incorporating study-level covariates, mixed-effects logistic regression models can readily handle such a meta-regression.

Recommendations

1. When the number of events and the sample size of each study group in the studies are available, we recommend alternative approaches based on fitting logistic regression models (fixed or mixed) directly to those data. This choice avoids the approximations and assumptions in the conventional meta-analysis approach adopted by commercial software such as RevMan and the `metan` and `metaan` commands in Stata; for odds ratios, the conventional approach should be discontinued. Stijnen et al. make this point strongly: “With the increased availability of GLMM [generalized linear mixed model] programs in the main statistical packages there is no compelling reason anymore for not using the exact within-study likelihood.”¹⁷ Commands for mixed-effects logistic regression often have a rich menu of options that control various aspects of the fitting process. When the default options are not the best choice for meta-analysis, our Appendix explains what to do instead.
2. Because meta-analyses can involve unexpected features and subtle complications, it is advisable to have expert assistance. Choices built into software should not replace thoughtful judgment.
3. When the outcome data are available only as study-level summaries such as odds ratio and rate ratio, and conventional meta-analysis approaches must be used (for example, the rate ratio is not compatible with logistic regression), methods that account for the sampling variation in the estimate of the between-study variance (e.g., profile likelihood) are preferable. The method used by RevMan and by another commercial system, Comprehensive Meta-Analysis (Biostat, Inc., www.Meta-Analysis.com), is inappropriate.
4. Because studies generally have some degree of heterogeneity, meta-analyses should use a random-effects model, in the absence of clear reasons for using a fixed-effect model (e.g., all studies had the same design and the same population). Formal tests for homogeneity based on the statistics Q and I^2 are problematic.²⁵ If studies are actually homogenous, the results from a random-effects model will be similar to those from a fixed-effect model.

5. It is essential to document the methods of analysis so that others can reproduce the results. For many years the International Committee of Medical Journal Editors has offered the following guidance on good practice: “Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results.”

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Jeroan Allison, MD, for his encouragement and comments.

Research reported in this publication was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under award number UL1-TR001453. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

1. Cooper, H., Hedges, LV., Valentine, JC. The Handbook of Research Synthesis and Meta-Analysis. New York: Russell Sage Foundation; 2009.
2. Pagliaro L, D’Amico G, Sorensen TI, et al. Prevention of first bleeding in cirrhosis. A meta-analysis of randomized trials of nonsurgical treatment. *Ann Intern Med.* 1992; 117:59–70. [PubMed: 1350716]
3. Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med.* 1996; 15:2733–49. [PubMed: 8981683]
4. Thompson SG, Smith TC, Sharp SJ. Investigating underlying risk as a source of heterogeneity in meta-analysis. *Stat Med.* 1997; 16:2741–58. [PubMed: 9421873]
5. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med.* 1999; 18:2693–708. [PubMed: 10521860]
6. Whitehead, A. Meta-Analysis of Controlled Clinical Trials. Chichester, England: John Wiley & Sons, Ltd; 2002.
7. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med.* 2004; 23:3105–24. [PubMed: 15449338]
8. Simmonds MC, Higgins JP. A general framework for the use of logistic regression models in meta-analysis. *Stat Methods Med Res.* 2016; 25:2858–77. [PubMed: 24823642]
9. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials.* 1986; 7:177–88. [PubMed: 3802833]
10. Raghunathan T, Li Y. Analysis of binary data from a multicentre clinical trial. *Biometrika.* 1993; 80:127–39.
11. Sidik K, Jonkman JN. Robust variance estimation for random effects meta-analysis. *Comput Stat Data Anal.* 2006; 50:3681–701.
12. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Med Res Methodol.* 2014; 14:25. [PubMed: 24548571]
13. Hamza TH, van Houwelingen HC, Stijnen T. The binomial distribution of meta-analysis was preferred to model within-study variability. *J Clin Epidemiol.* 2008; 61:41–51. [PubMed: 18083461]
14. Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random effects. *Stat Med.* 1996; 15:619–29. [PubMed: 8731004]
15. Tang JL. Weighting bias in meta-analysis of binary outcomes. *J Clin Epidemiol.* 2000; 53:1130–6. [PubMed: 11106886]

16. Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Stat Med.* 2000; 19:3417–32. [PubMed: 11122505]
17. Stijnen T, Hamza TH, Ozdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Stat Med.* 2010; 29:3046–67. [PubMed: 20827667]
18. Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med Decis Making.* 2013; 33:607–17. [PubMed: 23104435]
19. Palmer, TM., Sterne, JAC. *Meta-Analysis in Stata: An updated collection from the Stata Journal.* 2. College Station: TX: Stata Press; 2016.
20. Lesaffre E, Spiessens B. On the effect of the number of quadrature points in a logistic random effects model: an example. *J Roy Statist Soc Ser C.* 2001; 50:325–35.
21. SAS/STAT® 13.1 User's Guide. Cary, NC: SAS Institute Inc; 2013.
22. Fitzmaurice, GM., Laird, NM., Ware, JH. *Applied Longitudinal Analysis.* 2. Hoboken, NJ: John Wiley & Sons; 2011.
23. Kuss O. Statistical methods for meta-analyses including information from studies without any events-add nothing to nothing and succeed nevertheless. *Stat Med.* 2015; 34:1097–116. [PubMed: 25446971]
24. Stoto MA. Drug safety meta-analysis: promises and pitfalls. *Drug Saf.* 2015; 38:233–43. [PubMed: 25687791]
25. Hoaglin DC. Misunderstandings about Q and 'Cochran's Q test' in meta-analysis. *Stat Med.* 2016; 35:485–95. [PubMed: 26303773]

Figure 1.
 Forest Plot of Death Outcome
 The area of each square is proportional to the sample size of the study.

Figure 2.

Forest Plot of Bleeding Outcome

The area of each square is proportional to the sample size of the study.

Table 1

Frequencies of event (and no event) in the two groups of a study

	Intervention Group	Control Group
Event	a	b
No Event	c	d
	n_I	n_C

a,b,c,d: Frequency in the cell

Data on deaths and patients with bleeding in 19 trials of sclerotherapy in patients with cirrhosis and esophagogastric varices

Table 2

Trial ID	Total Subjects		Number of Death		Number of Bleeding	
	Control	Intervention	Control	Intervention	Control	Intervention
1	36	35	14	2	22	3
2	53	56	29	12	30	5
3	18	16	6	6	6	5
4	22	23	6	4	9	3
5	46	49	34	30	31	11
6	60	53	14	13	9	19
7	60	53	27	15	26	17
8	69	71	26	16	29	10
9	41	41	19	10	14	12
10	20	21	2	0	3	0
11	41	42	18	18	13	9
12	35	33	21	20	14	13
13	138	143	23	46	23	31
14	51	55	24	19	19	20
15	72	73	14	18	13	13
16	16	13	4	2	12	3
17	28	21	8	6	5	3
18	19	18	6	7	0	4
19	24	22	5	5	2	6

Table 3

Analysis Results for Death Outcome

	Log(Odds Ratio)			Between-Study Variance		
	Estimate	Standard Error	95% Confidence Interval	Estimate	Standard Error	95% Confidence Interval
Random-Effects Model						
Conventional Approaches using study-level data (Stata: metaan)						
DerSimonian and Laird method	-0.349	0.181	(-0.704, 0.007) ^a	0.324	NA	NA
ML: Profile-Likelihood method	-0.342	NA	(-0.704, 0.000) ^a	0.258	NA	(0.000, 0.781)
Alternative Approach using mixed-effects logistic regression (ML, AGHQ, 7 quadrature points)						
Stata: melogit	-0.374	0.158	(-0.683, -0.065) ^a	0.191	0.124	(0.053, 0.680)
R: glmer	-0.374	0.158	(-0.683, -0.065) ^a	0.191	NA	NA
SAS: glimmix	-0.374	0.158	(-0.706, -0.042) ^b	0.191	0.124	NA
Fixed-Effect Model						
Conventional Approach using study-level data: weighted-mean method (Stata: metaan)	-0.260	0.112	(-0.479, -0.041) ^a	0.000		
Alternative Approach using logistic regression (Stata: melogit)	-0.287	0.109	(-0.500, -0.073) ^a	0.000		

^a: Based on normal distribution; b: Based on t distribution (DF=18)

ML: Maximum Likelihood

AGHQ: Adaptive Gauss-Hermite Quadrature

NA: Not available (from either the estimation method or the software)

Table 4

Analysis Results for Bleeding Outcome

	Log(Odds Ratio)			Between-Study Variance		
	Estimate	Standard Error	95% Confidence Interval	Estimate	Standard Error	95% Confidence Interval
Random-Effects Model						
Conventional Approaches using study-level data (Stata: metaan)						
DerSimonian and Laird method	-0.610	0.270	(-1.140, -0.080) ^a	0.980	NA	NA
ML: Profile-Likelihood method	-0.611	NA	(-1.187, -0.036) ^a	1.039	NA	(0.000, 2.651)
Alternative Approach using mixed-effects logistic regression (ML, AGHQ, 7 quadrature points)						
Stata: melogit	-0.649	0.276	(-1.190, -0.108) ^a	1.063	0.502	(0.421, 2.684)
R: glmer	-0.650	0.276	(-1.190, -0.108) ^a	1.060	NA	NA
SAS: glimmix	-0.649	0.276	(-1.229, -0.069) ^b	1.063	0.502	NA
Fixed-Effect Model						
Conventional Approach using study-level data: weighted-mean method (Stata: metaan)	-0.487	0.119	(-0.721, -0.253) ^a	0.000		
Alternative Approach using logistic regression (Stata: melogit)	-0.564	0.113	(-0.786, -0.342) ^a	0.000		

^a: Based on normal distribution; b: Based on t distribution (DF=18)

ML: Maximum Likelihood

AGHQ: Adaptive Gauss-Hermite Quadrature

NA: Not available (from either the estimation method or the software)