Received: 13 October 2020 | Revised: 2 February 2021 | Accepted: 4 February 2021

DOI: 10.1002/jrsm.1482

BRIEF METHOD NOTE

Research Synthesis Methods WILEY

A confidence interval robust to publication bias for random-effects meta-analysis of few studies

Masayuki Henmi[1] | **Satoshi Hattori**[2] | **Tim Friede**[3]

[1]Institute of Statistical Mathematics, Tachikawa, Tokyo, Japan

[2]Department of Biomedical Statistics, Graduate School of Medicine, and Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Osaka, Japan

[3]Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany

Correspondence
Tim Friede, Department of Medical Statistics, University Medical Center Göttingen, Germany.
Email: tim.friede@med.uni-goettingen.de

Abstract

In meta-analyses including only few studies, the estimation of the between-study heterogeneity is challenging. Furthermore, the assessment of publication bias is difficult as standard methods such as visual inspection or formal hypothesis tests in funnel plots do not provide adequate guidance. Previously, Henmi and Copas (Statistics in Medicine 2010, 29: 2969–2983) proposed a confidence interval for the overall effect in random-effects meta-analysis that is robust to publication bias to some extent. As is evident from their simulations, the confidence intervals have improved coverage compared with standard methods. To our knowledge, the properties of their method have never been assessed for meta-analyses including fewer than five studies. In this manuscript, we propose a variation of the method by Henmi and Copas employing an improved estimator of the between-study heterogeneity, in particular when dealing with few studies only. In a simulation study, the proposed method is compared to several competitors. Overall, we found that our method outperforms the others in terms of coverage probabilities.

KEYWORDS

between-trial heterogeneity, confidence interval, coverage probability, meta-analysis, publication bias

1 | INTRODUCTION

Recently, the topic of meta-analysis of few studies, say 2–5, got more attention, since this case is very common in practice.[1] With few studies, however, confidence intervals of the overall effect based on normal quantiles tend to be too short as they ignore the uncertainty in estimating the between-trial heterogeneity. As remedies, methods based on *t*-quantiles have been proposed.[2-5] With few studies only, however, they are often conservative and so long that they are uninformative.[6] Between-trial heterogeneity estimates often result in zero,[7] with the notable exception of the method proposed by Chung et al.[8] Chung et al. suggested the so-called Bayes modal (BM) estimator, which uses in a Bayesian framework a weakly informative prior for the between-trial heterogeneity to avoid zero estimates of the heterogeneity. Furthermore, a fully Bayesian approach has some advantages in this situation.[7,9]

A number of methods have been proposed to deal with publication bias in meta-analyses including visual inspection of funnel plots as well as formal tests.[10] For funnel plots, trim-and-fill methods have been proposed to correct the overall effect for potential publication bias.[11] Following an alternative approach, several sensitivity analysis methods have been suggested based on selection functions describing the selective publication process.[12-14] In contrast, Henmi and Copas[15] proposed a method for random-effects meta-analysis that is robust to the selection of studies. The problem with these methods is that they become more powerful with larger

number of studies, but are less sensitive with few studies only.

In this paper, we propose a modification of the Henmi-Copas method by replacing the estimator of the between-study heterogeneity in the computation of the quantiles by the one developed by Chung et al.[8] The properties of the new approach are assessed and compared to alternative methods including the Henmi-Copas approach and a proposal by Doi et al.[16] in Monte Carlo simulation studies. Our method is not conditional on having detected publication bias, for example, in a funnel plot, since this would be very difficult with only few studies included in the meta-analysis. But it is robust to the selection of studies even with few studies.

Highlights

What is already known?

- Estimated overall effects from meta-analyses might be impacted by publication bias
- A confidence interval for the overall effect has been proposed that is to some extent robust to the selection of studies

What is new?

- The performance of the robust confidence interval previously proposed is assessed in meta-analyses with few studies and found not to work well in this setting
- The approach is refined resulting in improved coverage probabilities of the confidence intervals in particular in meta-analyses with few studies

Potential impact for RSM readers outside the authors' field

- The refined approach is recommend for application in meta-analyses with few studies yielding more reliable results

2 | METHODS

In the following, we provide a summary of the main ideas; a more detailed description of the methods can be found in Data S1. Adopting the notation by Henmi and Copas,[15] the true effect of an individual study i out of n independent studies is denoted by θ_i. Here, we consider the normal-normal hierarchical model (*NNHM*), the standard model for random-effects meta-analysis. In the *NNHM*, it is assumed that the θ_i are from a normal distribution with expectation θ and variance τ^2. Furthermore, the effect estimators Y_i follow (at least approximately) a normal distribution with expectation θ_i and variance σ_i^2. A standard method to construct an estimator and a $(1-\alpha)$ confidence interval for θ was proposed by DerSimonian and Laird[17] (DL).

Henmi and Copas[15] tackled the two problems that (a) the distribution of the pivot statistic is quite different from the standard normal distribution when the number of studies n is small, and (b) the estimators of θ are biased due to selective publication of smaller studies with less favorable results (publication bias). With respect to the latter they note that the common (or fixed) effect estimator $\hat{\theta}_F$ is more robust to publication bias than the random-effects estimator $\hat{\theta}_R$ simply because smaller studies, which are less likely to be published when their outcome is not favorable, have a smaller weight in the construction of $\hat{\theta}_F$ than in $\hat{\theta}_R$. To address the problem of the normal approximation they derive the distribution of the pivot statistic based on the fixed effect estimator under the random-effects model. The point in the derivation of the distribution of the pivot statistic by Henmi and Copas[15] is to take into account the random variation of $\hat{\tau}^2_{DL}$ in addition to $\hat{\theta}_F$. The $(1-\alpha)$ confidence interval for θ is given by

$$\left(\hat{\theta}_F - u^{(DL)}_{\alpha/2}\sqrt{V\left(\hat{\tau}^2_{DL}\right)},\ \hat{\theta}_F + u^{(DL)}_{\alpha/2}\sqrt{V\left(\hat{\tau}^2_{DL}\right)}\right). \quad (1)$$

For definitions of the γ quantile $u^{(DL)}_{\gamma}$ and the variance $V\left(\hat{\tau}^2_{DL}\right)$, which both depend on the DL estimator $\hat{\tau}^2_{DL}$ of τ^2, see Data S1.

The use of weakly informative priors for the between-study heterogeneity to avoid zero estimates has been advocated for some time, whereas an uninformative, for example, improper uniform, prior is used for the effect θ.[18,19] Here, we follow Chung et al.,[8] who proposed to use a gamma distribution with shape η and rate λ as a prior for τ. This choice means that the logarithm of the posterior of θ and τ is equal to the log likelihood plus a term depending only on τ but not θ. Rather than using the mean or median of the posterior, Chung et al.[8] considered the mode, which can be computed by numerical optimization. This estimator of τ is referred to as the Bayes Modal (BM) estimator $\hat{\tau}_{BM}$. The BM estimator $\hat{\tau}_{BM}$ can be interpreted as a penalized maximum likelihood (ML) estimator.[8]

Here, we propose to use the DL estimator $\hat{\tau}^2_{DL}$ in the construction of the pivot statistic in the same way as Henmi and Copas,[15] but we use the BM estimator $\hat{\tau}^2_{BM}$ in

the approximate calculation for the quantile of the pivot statistic instead of $\hat{\tau}^2_{DL}$. In the following, we refer to this approach as HC-BM. The choice of the BM estimator is motivated by its performance in comparison to other estimators in recent simulation studies (see, eg, Figures 2 and 3 in Reference 7). The resulting γ quantile is denoted by $u^{(BM)}_{\gamma}$. The $(1-\alpha)$ confidence interval for θ is then given by

$$\left(\hat{\theta}_F - u^{(BM)}_{\alpha/2}\sqrt{V\left(\hat{\tau}^2_{DL}\right)},\ \hat{\theta}_F + u^{(BM)}_{\alpha/2}\sqrt{V\left(\hat{\tau}^2_{DL}\right)}\right). \quad (2)$$

TABLE 1 Summary of the scenarios considered in the simulation study

Parameter	Values
Treatment effect θ	0.3,0.5
Between-trial heterogeneity τ^2	0.05,0.15,0.25
Number of trials included in the meta-analysis n	3,6,9,12,15
Selection model	
No publication bias	
Moderate publication bias	$\beta = 4, \gamma = 3$
Severe publication bias	$\beta = 4, \gamma = 1.5$

Note: The results for $\theta = 0.3$ are reported in Data S1.

3 | SIMULATION STUDY

As comparators for the proposed approach (HC-BM) the methods by Henmi and Copas[15] (HC), Chung et al.[8] (BM), Doi et al.,[16] (IVH) and DerSimonian and Laird[17] (DL) were included. The first one is known to be robust to publication bias to some extent, but its performance in meta-analyses with few studies only is unknown. The approach by Chung et al.[8] was developed for the scenario of few studies but might not be robust to publication bias. Doi et al.[16] proposed the inverse variance heterogeneity model. As with the HC approach, the interval is centered around an estimator assuming the common-effect model. Therefore, it might have attractive properties in settings with publication bias. In contrast to the HC approach, however, it is based on normal approximation. This approach was not included in recent method comparison studies.[20] The DL approach was included as it is often considered to be the standard approach. The simulation model by Brockwell and Gordon[21] formed the basis for our simulation study. To account for publication bias, we used the same selection function (probability that a study with an outcome y and associated standard error σ is selected in the meta-analysis)

$$P(\text{selected}\,|y,\sigma) = \exp\left[-\beta\left\{\Phi\left(-\frac{y}{\sigma}\right)\right\}^{\gamma}\right] \quad (3)$$

as in Reference 15 with the same sets of the parameters β and γ for moderate and severe publication bias. Here, Φ is the cumulative distribution function of the standard normal distribution. Table 1 summarizes the simulation scenarios considered ($N = 2{,}000$ simulation replications per scenario).

Figure 1 presents the simulated coverage probabilities for the different confidence intervals in the various scenarios with $\theta = 0.5$. In all scenarios considered, the proposed method performs at least as well as the HC method in terms of the coverage probability. With larger number of studies, say $n \geq 9$, and more pronounced between-trial heterogeneity, say $\tau^2 \geq 0.15$, the performance of both approaches is fairly similar. With smaller numbers of studies or only low levels of heterogeneity, however, there is a clear advantage for the new proposal as it improves the coverage probability considerably. In scenarios with few studies, $n = 3$ or $n = 6$, and only low levels of between-trial heterogeneity, $\tau^2 = 0.05$, the coverage probabilities of the BM approach are slightly higher than those of the proposed method. In the scenarios with publication bias, however, the coverage probabilities of the BM approach rapidly decrease well below the nominal level of 0.95 with increasing numbers of studies included in the meta-analysis and increasing levels of between-trial heterogeneity. Without publication bias, the coverage of the IVH interval is similar to the coverage of the DL interval, that is, poor for small numbers of studies n and closer to the nominal level for larger n. In the settings with publication bias, the coverage probabilities of the IVH intervals are generally larger than those of the DL approach, in particular with more pronounced heterogeneity τ^2 and larger numbers of studies n. However, the coverage probabilities are below those achieved by the HC and HC-BM approaches. Overall, the coverage probabilities of the proposed approach are closest to the nominal level, whereas the coverages for the DL approach are well below the nominal level for several scenarios characterized by publication bias and small numbers of studies included in the meta-analysis. The coverage probabilities for $\theta = 0.3$ are included in Data S1. The trends are overall very similar to those observed for the scenarios with $\theta = 0.5$. In scenarios with selection bias, however, the coverage probabilities tend to be lower with $\theta = 0.3$ than with $\theta = 0.5$ as the selection bias becomes more severe.

In scenarios, where different methods resulted in similar coverage probabilities close to the nominal level, it is of interest to compare the length of the intervals. In

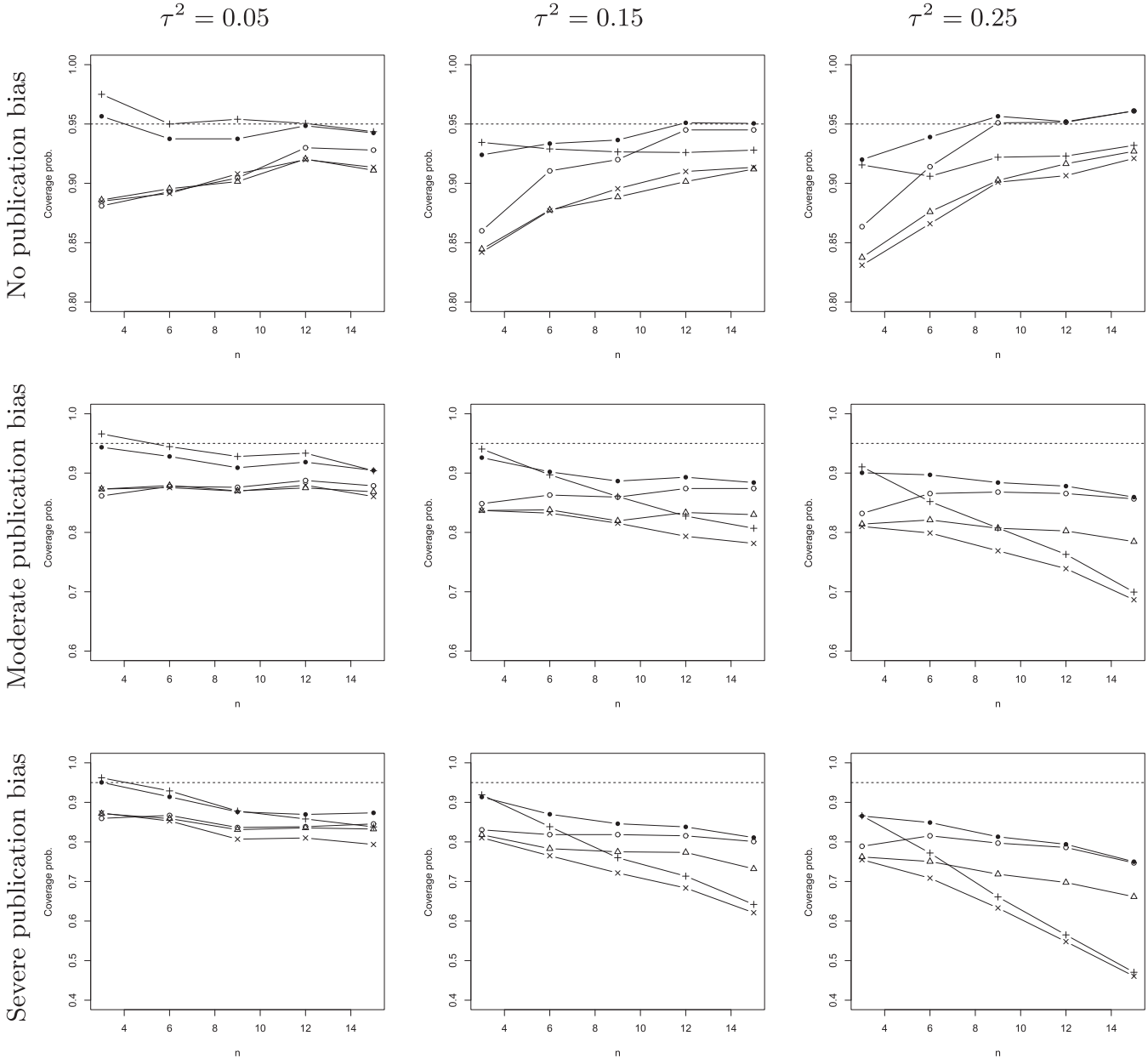

FIGURE 1 Coverage probabilities of the various confidence intervals (circle: HC, cross: DL, dot: HC-BM, plus: BM, triangle: IVH) depending on the number of studies n included in the meta-analysis for no, moderate and severe publication bias and for different degrees of between-trial heterogeneity $\tau^2 = 0.05,0.15,0.25$. The overall effect is $\theta = 0.5$

Data S1, the median interval lengths of the different confidence intervals are given.

4 | DISCUSSION

Meta-analyses of only a few studies are very common, but pose a number of challenges. These include the estimation of between-trial heterogeneity as well as the assessment of publication bias. Here, we proposed a method that faces both challenges successfully. The confidence interval of the overall effect proposed by Henmi and Copas[15] was improved by replacing the DerSimonian-Laird estimator by the Bayes Modal estimator of Chung et al.[8] in the computation of the quantiles to construct the confidence interval. The use of a weakly informative prior biases the Bayes Modal estimator away from zero. This resulted in larger quantiles, in particular in situations with few studies and only small to moderate levels of between-trial heterogeneity, which improved the coverage of the confidence intervals.

There are a number of limitations. We focused on properties related to estimating the overall effect and did not consider other parameters such as the heterogeneity τ^2.[22] Based on a previous evaluation,[7] which did not consider publication bias, we picked the heterogeneity estimator by Chung et al, although a variety of estimators have been proposed.[23] We consider this more an example demonstrating how the approach by Henmi and Copas can be improved. However, other estimators might lead to similar improvements. Furthermore, we refrained from investigating other selection functions.[15] Also, we did not include other comparators such as the Knapp-Hartung-Sidik-Jonkman approach[3-5] and did not evaluate the proposed estimator in the setting of a fixed-effect (or common-effect) meta-analysis.

The normal-normal hierarchical model considered here is a standard model for random-effects meta-analyses. This model is very general but not without limitations, since effect estimates are modeled and not the data directly implying a two-step procedure. For instance, considering binary outcomes and treatment effects summarized by odds ratios Jackson et al.[24] discuss more efficient one-step procedures. Modeling the data directly can have particular benefits when dealing with rare events.[25-27]

ACKNOWLEDGMENTS

The authors are grateful to Professor John Copas (Warwick) for discussions during his visit to Tokyo and Osaka in spring 2019. Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The author reported no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Satoshi Hattori https://orcid.org/0000-0001-5446-2305
Tim Friede https://orcid.org/0000-0001-5347-7441

REFERENCES

1. Bender R, Friede T, Koch A, et al. Methods for evidence synthesis in the case of very few studies. *Res Synth Methods*. 2018; 9:382-392.
2. Follmann DA, Proschan MA (1999). Valid inference in random effects meta-analysis. *Biometrics* 1999; 55: 732–737.
3. Hartung J, Knapp G. On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Stat Med*. 2001;20:1771-1782.
4. Hartung J, Knapp G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Stat Med*. 2001; 20:3875-3889.
5. Sidik K, Jonkman JN. A simple confidence interval for meta-analysis. *Stat Med*. 2002;21:3153-3159.
6. Röver C, Knapp G, Friede T. Hartung-Knapp-Sidik-Jonkman approach and its modification for random-effects meta-analysis with few studies. *BMC Med Res Methodol*. 2015;15:99.
7. Friede T, Röver C, Wandel S, Neuenschwander B. Meta-analysis of few small studies in orphan diseases. *Res Synth Methods*. 2017;8:79-91.
8. Chung Y, Rabe-Hesketh S, Choi I-H. Avoiding zero between-study variance estimates in random-effects meta-analysis. *Stat Med*. 2013;32:4071-4089.
9. Friede T, Röver C, Wandel S, Neuenschwander B. Meta-analysis of two studies in the presence of heterogeneity with applications in rare diseases. *Biom J*. 2017;59:658-671.
10. Jin ZC, Zhou XH, He J. Statistical methods for dealing with publication bias in meta-analysis. *Stat Med*. 2015;34:343-360.
11. Duval S, Tweedie R. A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *J Am Stat Assoc*. 2000;95:89-98.
12. Copas J, Shi JQ. Meta-analysis, funnel plots and sensitivity analysis. *Biostatistics*. 2000;1:247-262.
13. Copas J, Jackson D. A bound for publication bias based on the fraction of unpublished studies. *Biometrics*. 2004;60:146-153.
14. Henmi M, Copas JB, Eguchi S. Confidence intervals and p-values for meta-analysis with publication bias. *Biometrics*. 2007;63:475-482.
15. Henmi M, Copas JB. Confidence intervals for random effects meta-analysis and robustness to publication bias. *Stat Med*. 2010;29:2969-2983.
16. Doi SAR, Barendregt JJ, Khan S, Thalib L, Williams GM. Advances in the meta-analysis of heterogeneous clinical trials. I: the inverse variance heterogeneity model. *Contemp Clin Trials*. 2015;45:130-138.
17. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7:177-188.
18. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester, UK: Wiley; 2004.
19. Gelman A. Prior distributions for variance parameters in hierarchical models (Comment on article by Browne and Draper). *Bayesian Anal*. 2006;1:515-534.
20. Veroniki AA, Jackson D, Bender R, et al. Methods to calculate uncertainty in the estimated overall effect size from a random-effects meta-analysis. *Res Synth Methods*. 2019;10:23-43.
21. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. *Stat Med*. 2001;20:825-840.
22. Jackson D. The implications of publication bias for metaanalysis' other parameter. *Stat Med*. 2006;25:2911-2921.
23. Veroniki AA, Jackson D, Viechtbauer W, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res Synth Methods*. 2015;7:55-79.
24. Jackson D, Law M, Stijnen T, Viechtbauer W, White IR. A comparison of 7 random-effects models for meta-analyses that estimate the summary odds ratio. *Stat Med*. 2018;37: 1059-1085.
25. Kuss O. Statistical methods for meta-analyses including information from studies without any events—add nothing to nothing and succeed nevertheless. *Stat Med*. 2015;34:1097-1116.
26. Günhan BK, Röver C, Friede T. Meta-analysis of few studies involving rare events. *Res Synth Methods*. 2020;11:74-90.

27. Gronsbell J, Hong C, Nie L, Lu Y, Tian L. Exact inference for the random-effect model for meta-analyses with rare events. *Stat Med*. 2020;39:252-264.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Henmi M, Hattori S, Friede T. A confidence interval robust to publication bias for random-effects meta-analysis of few studies. *Res Syn Meth*. 2021;1–6. https://doi.org/10.1002/jrsm.1482