

RESEARCH

Hartung-Knapp-Sidik-Jonkman approach and its modification for random-effects meta-analysis with few studies

Christian Röver^{1*}, Guido Knapp² and Tim Friede¹

Abstract

Background: Random-effects meta-analysis is commonly performed by first deriving an estimate of the between-study variation, the *heterogeneity*, and subsequently using this as the basis for combining results, i.e., for estimating the *effect*, the figure of primary interest. The heterogeneity variance estimate however is commonly associated with substantial uncertainty, especially in contexts where there are only few studies available, such as in small populations and rare diseases.

Methods: Confidence intervals and tests for the effect may be constructed via a simple normal approximation, or via a Student-*t* distribution, using the Hartung-Knapp-Sidik-Jonkman (HKSJ) approach, which additionally uses a refined estimator of variance of the effect estimator. The modified Knapp-Hartung method (mKH) applies an *ad hoc* correction and has been proposed to prevent counterintuitive effects and to yield more conservative inference. We performed a simulation study to investigate the behaviour of the standard HKSJ and modified mKH procedures in a range of circumstances, with a focus on the common case of meta-analysis based on only a few studies.

Results: The standard HKSJ procedure works well when the treatment effect estimates to be combined are of comparable precision, but nominal error levels are exceeded when standard errors vary considerably between studies (e.g. due to variations in study size). Application of the modification on the other hand yields more conservative results with error rates closer to the nominal level. Differences are most pronounced in the common case of few studies of varying size or precision.

Conclusions: Use of the modified mKH procedure is recommended, especially when only a few studies contribute to the meta-analysis and the involved studies' precisions (standard errors) vary.

Keywords: random-effects meta-analysis; Knapp-Hartung adjustment; small populations; rare diseases

Background

Random-effects meta-analysis is most commonly performed based on an underlying hierarchical model including two unknowns as parameters: the *effect* μ , which is the figure of primary interest, and the *between-study variance (heterogeneity)* τ^2 , which is a nuisance parameter. Inference then is usually done sequentially, by first deriving an estimate of the heterogeneity variance, $\hat{\tau}^2$, and then determining the effect estimate $\hat{\mu}$ by conditioning on the estimate $\hat{\tau}^2$ [1, 2]. A large number of different estimators for the heterogeneity variance is available (see e.g. [3–6]), and effect

estimation may be done based on a simple normal approximation, or by utilizing a Student-*t* distribution [7] with an additionally refined estimator of the variance of $\hat{\mu}$ [8–12]. While the normal model may be motivated by asymptotic arguments, in actual applications the number of estimates to be combined is commonly small [13, 14] and hence the estimation uncertainty in the between-study variance τ^2 is substantial, so that an adjustment is appropriate and in fact improves operating characteristics [7–12, 15].

The problem of deriving estimates from only a small number of data sources is a common problem especially in fields of application where empirical information is sparse due to the rarity of the condition in question. The rarity of a disease is often accompanied with a low (commercial) interest or incentive, which is

* Correspondence: christian.roever@med.uni-goettingen.de

¹Department of Medical Statistics, University Medical Center Göttingen, Humboldtallee 32, 37073 Göttingen, Germany

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

why such diseases are also known as *orphan diseases*. According to the European Commission, a disease is designated orphan status when the prevalence is ≤ 5 in 10 000 [16]. While by definition any individual rare disease has a low prevalence, there is a large number of these, eventually affecting a substantial fraction of an estimated 6–8% of the total population [17], and with that posing a challenge to health care systems worldwide.

The European Medicines Agency acknowledges the particular obstacles in rare diseases research but points out that there is no fundamental difference between rare and more common diseases and hence no “paradigm change” when it comes to regulatory issues. Because of the common small-sample settings, the importance of sophisticated methods is emphasized, and meta-analyses of good quality randomised controlled clinical trials are still considered the highest level of evidence [18]. The problems encountered in rare diseases research often call for special statistical methods, especially with respect to study designs [17, 19, 20]. Meta-analyses are particularly important in this field due to the lack of large trials, while these will commonly still be faced with the problem of small numbers of available studies. Between-study heterogeneity then is anticipated, since the gathered pieces of evidence are likely to differ with respect to study designs, types of control groups or treatment allocation [17, 19–21]. Small studies have in fact empirically been found to exhibit more heterogeneity than large trials [22]. Consequently, the use of methods suitable for few studies and marginally significant findings is of crucial importance here.

With an estimated incidence of 2–20 cases per 100 000 population, juvenile idiopathic arthritis (JIA) is an example of a rare disease [23]. In the following, we will use a meta-analysis in JIA [24] as a case study to illustrate the different methods discussed below.

In the following sections, we will first describe the methods used, then show the results of a simulation study, and demonstrate the different types of analyses in an example data set, before closing with conclusions and recommendations.

Methods

Random-effects meta-analysis

Meta-analysis is very commonly performed via a *random-effects* approach, utilizing the *normal-normal hierarchical model*. Here the data are given in terms of a number k of estimates $y_i \in \mathbb{R}$ that are associated with some uncertainty given through standard errors $s_i > 0$ that are taken to be known without uncertainty. The estimates are assumed to measure trial-specific parameters $\theta_i \in \mathbb{R}$:

$$y_i \sim N(\theta_i, s_i^2) \quad \text{for } i = 1, \dots, k. \quad (1)$$

The parameters θ_i vary from trial to trial around a global mean $\mu \in \mathbb{R}$ due to some *heterogeneity variance* between trials that constitutes an additive variance component to the model,

$$\theta_i \sim N(\mu, \tau^2) \quad \text{for } i = 1, \dots, k, \quad (2)$$

where $\tau^2 \geq 0$. The model may then be simplified by integrating out the parameters θ_i , leading to the marginal expression

$$y_i \sim N(\mu, s_i^2 + \tau^2) \quad \text{for } i = 1, \dots, k. \quad (3)$$

Among the two unknowns in the model, the overall mean μ , the *effect*, usually is the figure of primary interest, while the heterogeneity variance τ^2 constitutes a nuisance parameter. When $\tau^2 = 0$, the model simplifies to the so-called *fixed-effect* model [1, 2].

The “relative amount of heterogeneity” in a meta-analysis may be expressed in terms of the measure I^2 , which is defined as

$$I^2 = \frac{\tau^2}{\tau^2 + \tilde{s}^2} \quad (4)$$

where \tilde{s} is some kind of “average” standard error among the study-specific s_i [25]. In the following simulation studies, we will determine \tilde{s}^2 as the arithmetic mean of squared standard errors.

Parameter estimation

If the value of the heterogeneity variance parameter τ^2 were known, the (conditional) maximum-likelihood effect estimate would result as the weighted average

$$\hat{\mu} = \frac{\sum_i w_i y_i}{\sum_i w_i} \quad (5)$$

with “inverse variance weights” defined as

$$w_i = \frac{1}{s_i^2 + \tau^2} \quad \text{for } i = 1, \dots, k. \quad (6)$$

A common approach to inference within the random-effects model is to first estimate the heterogeneity variance τ^2 , and subsequently estimate the effect μ *conditional on the heterogeneity estimate*. Note that the w_i are effectively treated as “known” while in fact both the s_i^2 as well as τ^2 are only measured with some uncertainty (that depends on the size/precision of the i th individual study and the number of studies k). There is a wide range of different heterogeneity estimators available (see e.g. [3–6] for more details). In the following we will concentrate on some of the most common ones, the DerSimonian-Laird (DL) estimator, a

moment estimator [26] with acknowledged shortcomings [27], the restricted maximum likelihood (REML) estimator [3, 28], and the Paule-Mandel (PM) estimator, an essentially heuristic approach [29, 30]. Software to compute the different estimates is provided e.g. in the `metafor` and `meta` R packages [31, 32].

Confidence intervals and tests

Normal approximation

Confidence intervals and, equivalently, tests for the effect μ are commonly constructed using a normal approximation for the estimate $\hat{\mu}$. The standard error of $\hat{\mu}$, conditional on a fixed heterogeneity variance value τ^2 , is given by

$$\hat{\sigma}_\mu = \sqrt{\frac{1}{\sum_i w_i}}. \quad (7)$$

A confidence interval for the effect μ then results via a normal approximation as

$$\hat{\mu} \pm \hat{\sigma}_\mu z_{(1-\alpha/2)} \quad (8)$$

where $z_{(1-\alpha/2)}$ is the $(1-\alpha/2)$ -quantile of the standard normal distribution, and $(1-\alpha)$ is the nominal coverage probability [1, 2]. The normal approximation based on a heterogeneity variance estimate $\hat{\tau}^2$ usually works well for many studies (large k) and small standard errors (small s_i), or negligible heterogeneity (small variance τ^2), but tends to be anticonservative otherwise [8, 9, 33].

The Hartung-Knapp-Sidik-Jonkman (HK SJ) method
Hartung and Knapp [8, 9] and Sidik and Jonkman [10] independently introduced an adjusted confidence interval. In order to determine the adjusted interval, first the quadratic form

$$q = \frac{1}{k-1} \sum_i w_i (y_i - \hat{\mu})^2 \quad (9)$$

is computed [8–10]. The adjusted confidence interval then results as

$$\hat{\mu} \pm \sqrt{q} \hat{\sigma}_\mu t_{(k-1);(1-\alpha/2)} \quad (10)$$

where $t_{(k-1);(1-\alpha/2)}$ is the $(1-\alpha/2)$ -quantile of the Student- t distribution with $(k-1)$ degrees of freedom. Note that $q\hat{\sigma}_\mu^2$ is derived from a non-negative and unbiased estimator of $\frac{1}{\sum_i w_i}$ [34]. Confidence intervals based on the normal approximation may easily be converted to HK SJ-adjusted ones [12]. The method has also been generalized to the cases of multivariate meta-analysis and meta-regression [35].

The modified Knapp-Hartung (mKH) method

The HK SJ confidence interval (10) tends to be wider than the one based on the normal approximation (8), since the Student- t quantile is larger than the corresponding normal quantile, while q will tend to be somewhere around unity. However, q may in fact also turn out arbitrarily small, and if $\sqrt{q} < \frac{z_{(1-\alpha/2)}}{t_{(k-1);(1-\alpha/2)}}$, then the modified interval will be shorter than the normal one, which may be considered counter-intuitive. A simple *ad hoc* modification to the procedure results from defining

$$q^* = \max\{1, q\} \quad (11)$$

[11] and using q^* instead of q to construct confidence intervals and tests. This will ensure a more conservative procedure. The modification was originally proposed in the meta-regression context, but the simple meta-analysis here constitutes the special case of an “intercept-only” regression.

Note that the PM heterogeneity variance estimator is effectively defined by choosing $\hat{\tau}^2$ such that q (equation (9)) is = 1 (or less, if no solution exists) [29, 30], so that for the PM estimator the corresponding q^* value always equals $q^* = 1$.

Simulations

Since a $(1-\alpha)$ confidence interval is supposed to cover the true parameter value with probability $(1-\alpha)$, the calibration of such intervals may be checked by repeatedly generating random data based on known parameter values and then determining the empirical frequency with which true values are actually covered [36]. We performed such a *Monte Carlo* simulation comparing the HK SJ and mKH approaches using the setup that was introduced by IntHout et al. [12]. Data were simulated on a continuous scale, according to the random-effects model described above, with study-specific standard errors s_i set to reflect certain scenarios with respect to the relative size of studies and their variation due to estimation uncertainty. Many meta-analyses are based on (discrete) count data, where the random-effects model assumptions only hold to an approximation that works well unless event probabilities are very low. Alternative methods have been proposed to deal with low event probabilities [37, 38], but low-event-rate effects were not considered in the present investigation.

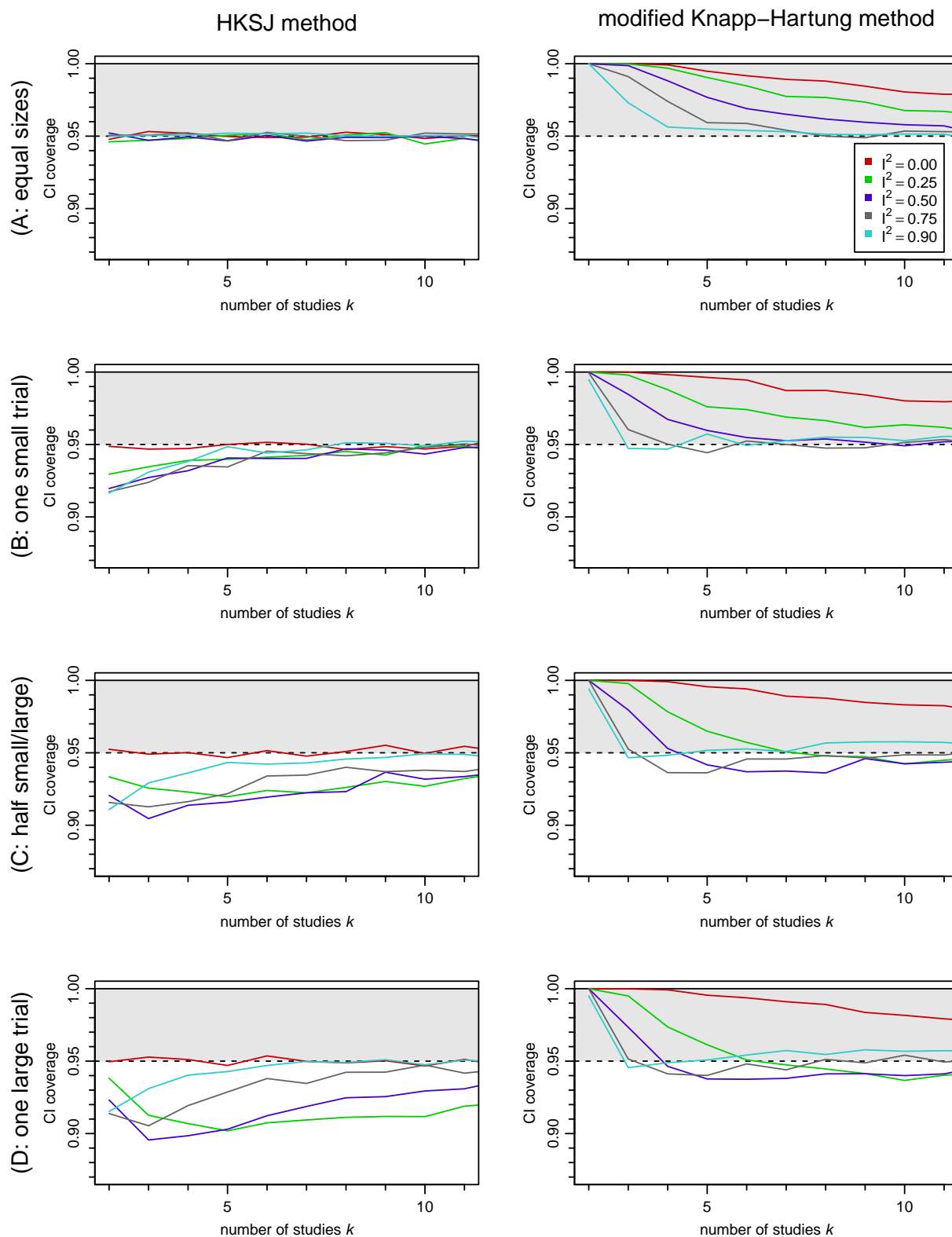


Figure 1 Coverage probabilities of HKSJ and mKH 95% confidence intervals. Probabilities are shown in dependence of the number k of studies and the amount of heterogeneity I^2 . The four different scenarios A–D correspond to different amounts of imbalance between the study-specific standard errors s_i . The DerSimonian-Laird (DL) method was used for estimation of the heterogeneity variance τ^2 .

These simulations considered four different scenarios, namely meta-analyses (A) with trials of equal size, (B) with equally sized trials but including one small trial, (C) with 50% large and small trials, and (D) equally sized trials and one large trial. The sizes of “small” and “large” trials (and hence, squared standard errors s_i^2) here differ by a factor of ten, so that the associated standard errors differ by roughly a factor of 3; for more details see [12, Appendix 2]. Numbers of studies k considered here are in the range of 2–11, and the true levels of heterogeneity were $I^2 \in \{0.00, 0.25, 0.50, 0.75, 0.90\}$. At each combination of parameters 10 000 meta-analyses were simulated. All simulations were performed using R [39].

Results

Simulations

Figure 1 shows the estimated coverage probabilities of the different confidence intervals based on the DL heterogeneity variance estimate. The corresponding figure for REML and PM look essentially the same, which is in line with the findings in [33, 40]. The HKSJ method works very well when the analyzed studies are of equal size (i.e., have equal standard errors), as can also be shown analytically [41], but coverage decreases in more imbalanced settings, especially for small numbers of studies. For the case of no heterogeneity ($I^2 = 0$) the HKSJ method also works fine, but if τ^2 was known a priori, in this case the fixed-effect model should work as well. The mKH procedure on the other hand is rather conservative for small k , but does not tend to inflate the type-I error substantially regardless of the underlying study sizes or true heterogeneity. For both methods, the dependence on the amount of heterogeneity is mostly a matter of whether τ^2 is $= 0$ or > 0 .

Application of the modification obviously tends to widen the resulting confidence intervals. The ratio of interval lengths (which is equal to $\frac{\sqrt{q^*}}{\sqrt{q}}$) is shown in Figure 2. Most notably, the effect is largest for small heterogeneity and for few studies.

The modification eventually only makes a difference in those cases where q turns out smaller than one. The fraction of intervals affected by the modification ranges between 31% and 82% for the DL heterogeneity variance estimator and for the scenarios investigated here, with an overall average of 61%. For the other two estimators the fractions are 29–82% with a mean of 62% (REML), and 31–91% with a mean of 74% (PM). Again, the differences between the different estimators are rather small. With respect to the underlying simulation scenario, the probability decreases with increasing heterogeneity, since larger heterogeneity also leads to larger values of q .

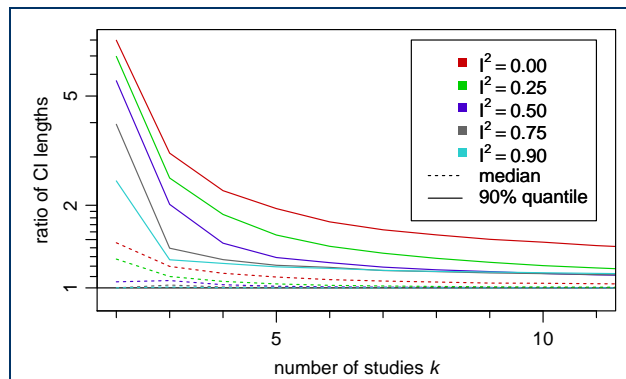


Figure 2 Ratios of lengths of HKSJ and mKH confidence intervals. (Same as $\frac{\sqrt{q^*}}{\sqrt{q}}$). Lengths are shown in dependence of the number k of studies and the amount of heterogeneity I^2 . Numbers are averaged over all 4 scenarios. The DL method was used for τ^2 estimation.

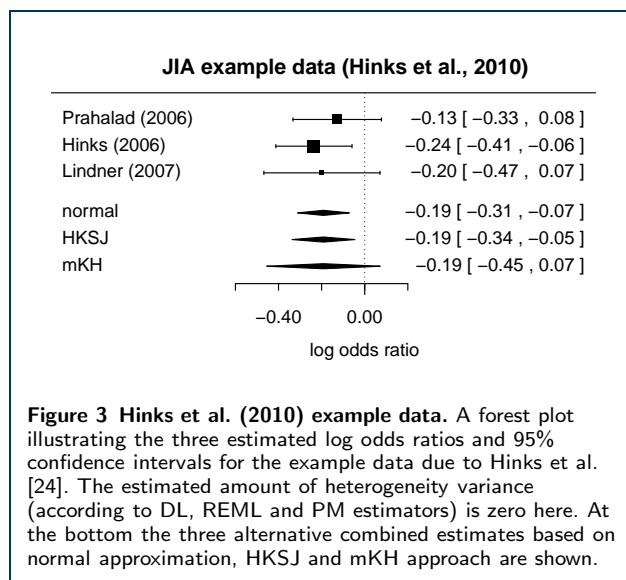


Figure 3 Hinks et al. (2010) example data. A forest plot illustrating the three estimated log odds ratios and 95% confidence intervals for the example data due to Hinks et al. [24]. The estimated amount of heterogeneity variance (according to DL, REML and PM estimators) is zero here. At the bottom the three alternative combined estimates based on normal approximation, HKSJ and mKH approach are shown.

Application to JIA example data

Hinks et al. [24] studied the occurrence of a particular genetic variant, CCR5, in juvenile idiopathic arthritis (JIA) patients in comparison with the general population. Their investigation included a meta-analysis of a small number ($k = 3$) of available controlled studies looking into the association of JIA with this particular biomarker. The analysis was based on logarithmic odds ratios; the three estimates along with their standard errors are shown in a forest plot in Figure 3. Here the largest standard error is 50% larger than the smallest one. For these data, all three (DL, REML and PM) heterogeneity variance estimates turn out as $\hat{\tau}^2 = 0$. A zero heterogeneity variance estimate is not uncommon, even when the actual heterogeneity is in fact substantial [14]. The associated q value is also

small with $q = 0.31$ ($\sqrt{q} = 0.55$). The resulting three confidence intervals based on normal approximation, HKSJ adjustment and mKH method all differ in their lengths. The mKH interval is longest, and includes the zero log odds ratio (indicating no association between genetic marker and disease), while the other two intervals do not include zero. So the choice of procedure directly affects conclusions in this example.

In this example, the three standard errors are rather similar (the largest is 50% larger than the smallest), but while the heterogeneity variance estimates $\hat{\tau}^2$ are all at zero, the 95% Q-profile confidence interval [42] ranges up to $\tau^2 = 0.33^2$, corresponding to $I^2 = 0.90$. Within the context of our simulations (Fig. 1), we cannot tell from the data which of the heterogeneity (I^2) scenarios we are in fact in, and as the standard errors are not exactly the same, it remains a matter of balancing the potential consequences whether one would rather risk losing on the side of (type I) error probability or power.

Discussion

The HKSJ procedure ensures the coverage probability only when the included studies' standard errors s_i are similar; for unbalanced settings, the actual error probability tends to exceed the targeted one. With the standard definition of the correction factor q the results may sometimes be counterintuitive, since the corresponding CIs may turn out shorter than using the simple normal approximation; in fact they may get arbitrarily short. In case of no heterogeneity ($\tau = 0$) the HKSJ method also works well, however practically this is of limited relevance, as one can rarely tell (or convincingly argue) whether this condition holds.

The *ad hoc* modification of the mKH method aims at fixing these shortcomings and results in type-I error probabilities that are not grossly in excess of the pre-specified ones. Especially when the standard errors s_i are of dissimilar magnitude, the mKH method can therefore be recommended. For few studies (small k), the modified procedure however tends to be very conservative, with very small error probabilities especially in the extreme case of meta-analysis of only $k = 2$ studies. In this extreme case the choice of methods may therefore be considered a matter of a power vs. type-I error probability tradeoff.

While meta-analyses of few studies are a particular problem in indications where there is only little evidence available (such as rare diseases), such circumstances are not as uncommon as one might expect. Turner et al. [13] and Kontopantelis et al. [14] investigated the analyses archived in the Cochrane Database and actually found a *majority* of them to be based on as few as $k = 2$ or $k = 3$ studies; so these constitute

highly relevant cases for which the proper control of error rates is crucial.

The properties of either unmodified or modified method for the extreme case of $k = 2$ may be considered unsatisfactory, as it seems one has the choice of either falling short of or exceeding the targeted error probability; the problem has in fact been regarded as effectively unsolved [43]. The poor behaviour may be explained by the fact that performing a random-effects meta-analysis effectively means the estimation of first- and second-order statistics, and it is not overly surprising to find that this is a hard task when the data consist of as few as two samples that are only measured with uncertainty. Bearing this in mind, the use of Bayesian methods [44] and the consideration of external evidence on the likely magnitude of the heterogeneity [45] may be the way forward.

List of abbreviations

DL	DerSimonian-Laird
HKSJ	Hartung-Knapp-Sidik-Jonkman
JIA	juvenile idiopathic arthritis
mKH	modified Knapp-Hartung
PM	Paule-Mandel
REML	restricted maximum likelihood

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

TF conceived the concept of this study. CR carried out the simulations and drafted the manuscript. GK critically reviewed and made substantial contributions to the manuscript. All authors commented on and approved the final manuscript.

Acknowledgements

This work has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number FP HEALTH 2013-602144 through the InSPiRe project [46].

Author details

¹Department of Medical Statistics, University Medical Center Göttingen, Humboldtallee 32, 37073 Göttingen, Germany. ²Department of Statistics, TU Dortmund University, 44221 Dortmund, Germany.

References

- Hedges, L.V., Olkin, I.: *Statistical Methods for Meta-analysis*. Academic Press, San Diego, CA, USA (1985)
- Hartung, J., Knapp, G., Sinha, B.K.: *Statistical Meta-analysis with Applications*. Wiley, Hoboken, NJ, USA (2008)
- Viechtbauer, W.: Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics* **30**(3), 261–293 (2005). doi:10.3102/10769986030003261
- Sidik, K., Jonkman, J.N.: A comparison of heterogeneity variance estimators in combining results of studies. *Statistics in Medicine* **26**(9), 1964–1981 (2007). doi:10.1002/sim.2688
- Panitikul, T., Bumrungrsup, C., Knapp, G.: On estimating heterogeneity in random-effects meta-regression: A comparative study. *Journal of Statistical Theory and Applications* **12**(3), 253–265 (2013). doi:10.2991/jsta.2013.12.3.4
- Veroniki, A.A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., KuB, O., Higgins, J.P.T., Langan, D., Salanti, G.: *Methods to estimate the between-study variance and its uncertainty in meta-analysis*. *Research Synthesis Methods* (2015). doi:10.1002/jrsm.1164. (in press)

7. Follmann, D.A., Proschan, M.A.: Valid inference in random effects meta-analysis. *Biometrics* **55**(3), 732–737 (1999). doi:10.1111/j.0006-341X.1999.00732.x
8. Hartung, J., Knapp, G.: On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Statistics in Medicine* **20**(12), 1771–1782 (2001). doi:10.1002/sim.791
9. Hartung, J., Knapp, G.: A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Statistics in Medicine* **20**(24), 3875–3889 (2001). doi:10.1002/sim.1009
10. Sidik, K., Jonkman, J.N.: A simple confidence interval for meta-analysis. *Statistics in Medicine* **21**(21), 3153–3159 (2002). doi:10.1002/sim.1262
11. Knapp, G., Hartung, J.: Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine* **22**(17), 2693–2710 (2003). doi:10.1002/sim.1482
12. Int'Hout, J., Ioannidis, J.P.A., Borm, G.F.: The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Medical Research Methodology* **14**, 25 (2014). doi:10.1186/1471-2288-14-25
13. Turner, R.M., Davey, J., Clarke, M.J., Thompson, S.G., Higgins, J.P.T.: Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *International Journal of Epidemiology* **41**(3), 818–827 (2012). doi:10.1093/ije/dys041
14. Kontopantelis, E., Springate, D.A., Reeves, D.: A re-analysis of the Cochrane Library data: The dangers of unobserved heterogeneity in meta-analyses. *PLoS ONE* **8**(7), 69930 (2013). doi:10.1371/journal.pone.0069930
15. Higgins, J.P.T., Thompson, S.G.: Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine* **23**(11), 1663–1682 (2004). doi:10.1002/sim.1752
16. European Commission: Communication from the Commission on Regulation (EC) No 141/2000 of the European Parliament and of the Council of 16 December 1999 on orphan medicinal products. *Official Journal of the European Union* **46**(C178), 2–8 (2003)
17. Gagne, J.J., Thompson, L., O'Keefe, K., Kesselheim, A.S.: Innovative research methods for studying treatments for rare diseases: methodological review. *BMJ* **349**, 6802 (2014). doi:10.1136/bmj.g6802
18. European Medicines Agency (EMA): Guideline on clinical trials in small populations. CHMP/EWP/83561/2005 (2006). http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003615.pdf
19. Korn, E.L., McShane, L.M., Freidlin, B.: Statistical challenges in the evaluation of treatments for small patient populations. *Science Translational Medicine* **5**(178) (2013). doi:10.1126/scitranslmed.3004018
20. Kesselheim, A.S., Myers, J.A., Avorn, J.: Characteristics of clinical trials to support approval of orphan vs nonorphan drugs for cancer. *Journal of the American Medical Association* **305**(22), 2320–2326 (2011). doi:10.1001/jama.2011.769
21. Unkel, S., Röver, C., Stallard, N., Benda, N., Posch, M., Zohar, S., Friede, T.: Systematic reviews in paediatric multiple sclerosis and Creutzfeldt-Jakob disease exemplify shortcomings in methods used to evaluate therapies in rare conditions. (submitted for publication) (2015)
22. Int'Hout, J., Ioannidis, J.P.A., Borm, G.F., Goeman, J.J.: Small studies are more heterogeneous than large ones: a meta-meta-analysis. *Journal of Clinical Epidemiology* **68**(8), 860–869 (2015). doi:10.1016/j.jclinepi.2015.03.017
23. Prakken, B., Albani, S., Martini, A.: Juvenile idiopathic arthritis. *The Lancet* **377**(9783), 2138–2149 (2011). doi:10.1016/S0140-6736(11)60244-4
24. Hinks, A., Martin, P., Flynn, E., Eyre, S., Packham, J., *et al.*: Association of the CCR5 gene with juvenile idiopathic arthritis. *Genes & Immunity* **11**(7), 584–589 (2010). doi:10.1038/gene.2010.25
25. Higgins, J.P.T., Thompson, S.G.: Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* **21**(11), 1539–1558 (2002). doi:10.1002/sim.1186
26. DerSimonian, R., Laird, N.: Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**(3), 177–188 (1986). doi:10.1016/0197-2456(86)90046-2
27. Böhning, D., Malzahn, U., Dietz, E., Schlattmann, P., Viwatwongkasem, C., Biggeri, A.: Some general points in estimating heterogeneity variance with the DerSimonian-Laird estimator. *Biostatistics* **3**(4), 445–457 (2002). doi:10.1093/biostatistics/3.4.445
28. Raudenbush, S.W.: Analyzing effect sizes: random-effects models. In: Cooper, H., Hedges, L.V., Valentine, J.C. (eds.) *The handbook of research synthesis and meta-analysis*, pp. 295–315. Russell Sage Foundation, New York (2009)
29. Paule, R.C., Mandel, J.: Consensus values and weighting factors. *Journal of Research of the National Bureau of Standards* **87**(5), 377–385 (1982)
30. Rukhin, A.L., Biggerstaff, B.J., Vangel, M.G.: Restricted maximum-likelihood estimation of a common mean and the Mandel-Paule algorithm. *Journal of Statistical Planning and Inference* **83**(2), 319–330 (2000). doi:10.1016/S0378-3758(99)00098-1
31. Viechtbauer, W.: Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* **36**(3) (2010). doi:10.18637/jss.v036.i03
32. Schwarzer, G.: *Meta: Meta-analysis with R*. (2014). R package version 3.7-1. <http://CRAN.R-project.org/package=meta>
33. Friede, T., Röver, C., Wandel, S., Neuenschwander, B.: Meta-analysis of few small studies in orphan diseases (2015). (submitted)
34. Hartung, J.: An alternative method for meta-analysis. *Biometrical Journal* **41**(8), 901–916 (1999). doi:10.1002/(SICI)1521-4036(199912)41:8;901::AID-BIMJ901i3.0.CO;2-W
35. Jackson, D., Riley, R.D.: A refined method for multivariate meta-analysis and meta-regression. *Statistics in Medicine* **33**(4), 541–554 (2014). doi:10.1002/sim.5957
36. Dawid, A.P.: The well-calibrated Bayesian. *Journal of the American Statistical Association* **77**(379), 605–610 (1982). doi:10.1080/01621459.1982.10477856
37. Bradburn, M.J., Deeks, J.J., Berlin, J.A., Localio, A.R.: Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Statistics in Medicine* **26**(1), 53–77 (2007). doi:10.1002/sim.2528
38. Kuß, O.: Statistical methods for meta-analyses including information from studies without any events—add nothing to nothing and succeed nevertheless. *Statistics in Medicine* **34**(7), 1097–1116 (2015). doi:10.1002/sim.6383
39. R Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2014). R Foundation for Statistical Computing. <http://www.r-project.org/>
40. Sánchez-Meca, J., Marín-Martínez, F.: Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods* **13**(1), 31–48 (2008). doi:10.1037/1082-989X.13.1.31
41. Sidik, K., Jonkman, J.N.: Authors' reply. *Statistics in Medicine* **23**(1), 159–162 (2004). doi:10.1002/sim.1729
42. Viechtbauer, W.: Confidence intervals for the amount of heterogeneity in meta-analysis. *Statistics in Medicine* **26**(1), 37–52 (2007). doi:10.1002/sim.2514
43. Gonnermann, A., Framke, T., Großhennig, A., Koch, A.: No solution yet for combining two independent studies in the presence of heterogeneity. *Statistics in Medicine* **34**(16), 2476–2480 (2015). doi:10.1002/sim.6473
44. Sutton, A.J., Abrams, K.R.: Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research* **10**(4), 277–303 (2001). doi:10.1177/096228020101000404
45. Turner, R.M., Jackson, D., Wei, Y., Thompson, S.G., Higgins, J.P.T.: Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. *Statistics in Medicine* **34**(6), 984–998 (2015). doi:10.1002/sim.6381
46. Innovative methodology for small populations research (InSPiRe). URL: <http://www.warwick.ac.uk/inspire>