Simulation protocol:
Comparison of confidence intervals summarizing the uncertainty of the combined estimate of a meta analysis

Florian Gerber, Leonhard Held, Lisa Hofer, Philip Heesen

For the present protocol is inspired by [Burton et al., 2006] and [Morris et al., 2019].

Contents

1 Aims and objectives

Comparison of confidence intervals summarizing the uncertainty of the combined estimate of a meta analysis.

1. Considered methods are
   (a) DerSimonian and Laird (DL) analysis [DerSimonian and Laird, 1986]
   (b) Hartung Knapp Sidik Jonkman (HK) analysis [IntHout et al., 2014]
   (c) Harmonic mean analysis with alternative `none` [Held, 2020]
   (d) Harmonic mean analysis with alternative `two.sided` [Held, 2020]
   (e) Harmonic mean analysis with alternative `none`, variance adjusted with $\hat{I}^2$. An extension of the idea in [Held, 2020].
   (f) Henmi and Copas (HC) [Henmi and Copas, 2010].
2. Assess the CIs by the following metrics
   (a) CI coverage
   (b) CI width
   (c) CI score [Gneiting and Raftery, 2007]
   (d) number of CIs (only for Harmonic mean method with alternative `none`).

2 Simulation procedures

2.1 Level of dependence between simulated datasets

2.2 Allowance for failures

We expect no failures, i. e. , for all simulated datasets all type of CI methods should lead to a valid CI and all valid CIs should lead to valid CI metrics. If a failure occurs, we stop the simulation study and investigate the reason for the failure.

2.3 Software to perform simulations

The simulation study is performed using the statistical software R [R Core Team, 2021]. We will save the output of `sessionInfo()` giving information on the used version of R, packages, and platform with the simulation results.

2.4 Random number generator to use

We use the package 'doRNG' with its default random number generator to ensure that random numbers generated inside parallel for loops are independent and reproducible.

3 Methods for generating the datasets

4 Scenarios to be investigated

The 60 simulated scenarios consist of all combinations of the following parameters:

- Higgin's I^2 heterogeneity measure $\in \{0, 0.3, 0.6, 0.9\}$.
- Number of studies summarized by the meta analysis $k \in \{2, 3, 5, 10, 20\}$.
- Publication bias is $\in \{$'none', 'moderate', strong'$\}$ following the terminology of [Henmi and Copas, 2010]. The average study effect also influences the publication bias, and we set it to $\theta = 0.2$ to obtain a similar scenario as used in [Henmi and Copas, 2010].
- The distribution the draw the true study values δ_i is either 'Gaussian' or 't'. The latter leads to more 'outliers'.

Sample size of the individual studies (number of patients per study) is fixed to $n = 50$.
[IntHout et al., 2014] used a similar setup.

4.1 Simulation of the data

For each scenario in Section 4 we

1. simulate 10000 meta-analysis datasets
2. compute the CI's listed in Section 1 for each meta-analysis
3. summarize the performance of the CI's by the metrics listed in Section 1

For the **Gaussian model without publication bias**, the simulation of one meta-analysis dataset is performed as follows:

1. Compute the within-study variance $\epsilon^2 = \frac{1}{k}\sum_{i=1}^{k}\frac{2}{n_i} = \frac{2}{n}$.
2. Compute the between-study variance $\tau^2 = \epsilon^2 \frac{I^2}{1-I^2}$.
3. For a trial i of the meta-analysis with k trials, $i = 1, \dots, k$:
   (a) Simulate the true effect size using the Gaussian model: $\delta_i \sim \mathcal{N}(\theta, \tau^2)$ or using a students-t distribution such that the samples have mean θ and variance τ^2.
   (b) Simulate the effect estimates of each trial $y_i \sim \mathcal{N}(\delta_i, \frac{2}{n_i})$.
   (c) $\mathrm{se}_i \sim \sqrt{\frac{\chi^2(2n_i-2)}{(n_i-1)n_i}}$ are the standard errors of trial outcomes

To simulate studies under **publication bias**, we follow the suggestion of [Henmi and Copas, 2010] and accept each simulated study with probability

$$\exp(-4\,\Phi(-\theta_i/\mathrm{se}_i)^{\gamma}),$$

where $\gamma = 3$ and $\gamma = 1.5$ correspond to *moderate* and *strong* publication bias, respectively. This is, accepted studies are kept and for a rejected study we replace θ_i and se_i by newly simulated values, which are then again accepted with the given probability above. This procedure is repeated until the required number of studies is simulated.

5 Statistical methods to be evaluated

1. DerSimonian and Laird (DL) analysis [DerSimonian and Laird, 1986]
2. Hartung Knapp Sidik Jonkman (HK) analysis [IntHout et al., 2014]
3. Harmonic mean analysis with alternative `none` [Held, 2020]
4. Harmonic mean analysis with alternative `two.sided` [Held, 2020]
5. Harmonic mean analysis with alternative `none`, variance adjusted with $\hat{I}^2$. An extension of the idea in [Held, 2020].
6. Henmi and Copas (HC) [Henmi and Copas, 2010].

6 Estimates to be stored for each simulation and summary measures to be calculated over all simulations

7 Number of simulations to be performed

We perform the simulation of each scenario 10000 times. We repeat the final simulation several times to assess the variability in the simulation results.

8 Criteria to evaluate the performance of statistical methods for different scenarios

9 Presentation of the simulation results

[Rücker et al., 2008]

References

[Burton et al., 2006] Burton, A., Altman, D. G., Royston, P., and Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, 25(24):4279–4292. 1

[DerSimonian and Laird, 1986] DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3):177–188. 2, 3

[Gneiting and Raftery, 2007] Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378. 2

[Held, 2020] Held, L. (2020). The harmonic mean χ^2-test to substantiate scientific findings. *Journal of the Royal Statistical Society Series C*, 69(3):697–708. 2, 3

[Henmi and Copas, 2010] Henmi, M. and Copas, J. B. (2010). Confidence intervals for random effects meta-analysis and robustness to publication bias. *Statistics in Medicine*, 29(29):2969–2983. 2, 3

[IntHout et al., 2014] IntHout, J., Ioannidis, J. P., and Borm, G. F. (2014). 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). 2, 3

[Morris et al., 2019] Morris, T. P., White, I. R., and Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38(11):2074–2102. 1

[R Core Team, 2021] R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. 2

[Rücker et al., 2008] Rücker, G., Schwarzer, G., Carpenter, J. R., and Schumacher, M. (2008). Undue reliance on i2 in assessing heterogeneity may mislead. *BMC Medical Research Methodology*, 8(1):79. 4