

**DISCUSSION**

# Discussion on ‘Testing small study effects in multivariate meta-analysis’ by Chuan Hong, Georgia Salanti, Sally Morton, Richard Riley, Haitao Chu, Stephen E Kimmel and Yong Chen

James R. Carpenter<sup>1,2</sup> | Gerta Rücker<sup>3</sup> | Guido Schwarzer<sup>3</sup> <sup>1</sup> Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, UK<sup>2</sup> MRC Clinical Trials Unit at UCL, London, UK<sup>3</sup> Medical Faculty and Medical Center, University of Freiburg, Freiburg, Germany**Correspondence**

James R. Carpenter, Department of Medical Statistics, London School of Hygiene, and Tropical Medicine, Keppel Street, London WC1E 7HT, UK.

Email: [james.carpenter@lshtm.ac.uk](mailto:james.carpenter@lshtm.ac.uk)**Funding information**

Deutsche Forschungsgemeinschaft, Grant/Award Number: RU1747/1-2; UK Medical Research Council, Grant/Award Numbers: MC UU 12023/21, MC UU 12023/29

We congratulate the authors on their extension of the univariate Egger test for publication bias to the multivariate setting. The proposed Multivariate Small Study Effect Test (MSSET) raises some interesting questions.

First, it is difficult to interpret the hypothesis that is being tested, which is essentially asymmetry in one or more of the funnel plots. This conflates two separate issues: publication bias and outcome reporting bias. Strictly, publication bias means that—because of their results—some papers are not published at all. Probably it is the primary outcome that triggers this. We doubt this will lead to asymmetry in the funnel plots for *each* outcome. If this is the case, a multivariate test may provide misleading reassurance compared to well-performing univariate test on the primary outcome of the meta-analysis.

On the other hand, outcome reporting bias means that, perhaps in response to the trial results, the authors switch the primary and secondary outcomes around. The consequence is the increased chance of suppression of some outcomes in the publication. But again, it is not clear that this will work in a way that causes a multivariate test to be necessarily more informative than a well-performing univariate test on the primary outcome of the meta-analysis. The authors mention this obliquely, but it would be interesting if they could respond this key point in more detail.

Second the authors’ approach is essentially a score test variant of performing multiple Egger tests in one step using Generalized Estimating Equations, assuming independence across outcomes for estimation, and using robust standard errors (clustered on studies). Looking at it in this way explains how it avoids specification of the correlation, but how much power does it really gain over univariate analysis? Instinct suggests that without knowledge of the correlation between the outcomes—an issue that limits the practical utility of multivariate meta-analysis ((Riley *et al.*, 2017, table 1))—the answer would be little. Although the simulation results suggest quite the reverse, it is unclear whether the results presented can be taken at face value.

Ideally, we would like to see results from both univariate and bivariate tests when the correlation is zero (these should have the same power and type-1 error) and then, as the correlation increases, it would be interesting to see how the results from the two approaches diverge. Instead, the authors’ Table 1 appears to average over the correlation (the key parameter!) in some way that is not described. Then, even in the case of independent outcomes, we appear to be comparing naive application of the Egger test (with known issues of type-1 error) with a corrected sampling distribution for the score test. So, definitive conclusions about the greater power of the new test cannot be drawn. This is

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Biometrics* published by Wiley Periodicals, Inc. on behalf of International Biometric Society.

especially the case when we remember that if the type-1 error of a test is too small, the power is reduced (and vice versa). For this reason, we and others (Rücker *et al.* (2011)) have argued this should be adjusted for before presenting the results. In addition, presenting results of a Bonferroni correction is a straw man: it is well known that Bonferroni-Holm (Holm, 1979) is uniformly better.

The R code provided by the authors to conduct the MSSET is welcome. Unfortunately, though it is not possible to reproduce the MSSET results of the heart failure meta-analysis because key information is omitted. We assume that the (incomplete) reference in the publication refers to a Heart editorial by Inglis *et al.* (2017), which was published online in 2016. This editorial summarizing results of a Cochrane review by Inglis *et al.* (2015) reports two separate meta-analyses for the primary outcome “all-cause mortality”: structured telephone support (STS) versus usual care (UC) in 22 trials and telemonitoring (TM) versus UC with 17 trials, respectively. As two trials compared both STS and TM with UC, the number of independent studies providing data on the primary outcome is 37, which differs from the 34 trials reported here. Furthermore, the Cochrane review only provides qualitative information on the secondary outcome “health-related quality of life” (Inglis *et al.*, 2015, table 2)). Also, did the authors follow Inglis *et al.* in using the risk ratio as a summary? These uncertainties underscore the importance of providing both the analysis code and data sets in order to ensure reproducible research.

Further, it remains unclear whether the comparison of the univariate and bivariate  $P$ -values in the heart failure meta-analysis is sensible, because it is not clear whether the number of trials used in the univariate and bivariate comparisons is the same. Specifically, the authors do not state whether the univariate  $P$ -values are based on the 34 (all-cause mortality) and 12 (mental QoL) trials, or for the 11 trials providing data for both outcomes. In general, and almost certainly in this situation too, the results will be highly dependent on this. Also, we are unsure about the appropriate analysis in this setting. We could envision that a univariate test for funnel plot asymmetry is (a) significant for an outcome in the smaller set of trials providing data for both outcomes, however, (b) non-significant in the larger set of trials providing data for the outcome of interest. This goes back to our initial concern about the actual hypotheses being tested.

Turning to the empirical evaluation in §4.2, the authors state the following requirements for inclusion of a Cochrane review in the evaluation: (a) at least 10 studies in meta-analysis; (ii) common set of two outcomes, and (iii) at least one study should report both outcomes. If correct, this means that  $m = 1$  is possible. But how do you conduct a multivariate meta-analysis in this situation?

Although we welcome the authors’ provision of R code, readers should perhaps be aware this is for bivariate analyses only, not general multivariate analyses.

In summary, while we appreciated the opportunity to read this paper:

- we are not clear about the interpretation of the test;
- the simulation study is limited to the bivariate situation, and its reporting means that the behavior of the text viz-a-viz the correlation parameter is obscured, and
- details of the data analysis leave some key questions unanswered.


We encourage the authors to address these concerns, which we argue need to be resolved for the community to incorporate this methodology into practice.

## ACKNOWLEDGMENTS

James Carpenter is supported by the UK Medical Research Council, grants MC\_UU\_12023/21 and MC\_UU\_12023/29. Gerta Rücker is funded by DFG (German Research Foundation), grant number RU1747/1-2.

## ORCID

James R. Carpenter  <https://orcid.org/0000-0003-3890-6206>

Guido Schwarzer  <https://orcid.org/0000-0001-6214-9087>

## REFERENCES

- Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70.
- Inglis, S.C., Clark, R.A., Dierckx, R., Prieto-Merino, D. and Cleland, J.G. (2015) Structured telephone support or noninvasive telemonitoring for patients with heart failure. *Cochrane Database of Systematic Reviews*, 10, CD007228.
- Inglis, S.C., Clark, R.A., Dierckx, R., Prieto-Merino, D. and Cleland, J.G.F. (2017) Structured telephone support or non-invasive telemonitoring for patients with heart failure. *Heart*, 103(4), 255–257.
- Riley, R.D., Jackson, D., Salanti, G., Burke, D.L., Price, M., Kirkham, J. and White, I.R. (2017) Multivariate and network meta-analysis of multiple outcomes and multiple treatments: rationale, concepts, and examples. *BMJ*, 358, j3932.
- Rücker, G., Carpenter, J.R. and Schwarzer, G. (2011) Detecting and adjusting for small-study effects in meta-analysis. *Biometrical Journal*, 53(2), 351–368.

**How to cite this article:** Carpenter JR, Rücker G, Schwarzer G. Discussion on ‘Testing small study effects in multivariate meta-analysis’ by Chuan Hong, Georgia Salanti, Sally Morton, Richard Riley, Haitao Chu, Stephen E Kimmel and Yong Chen. *Biometrics*. 2020;1–2.

<https://doi.org/10.1111/biom.13344>