Reviewer Feedback and Response Letter

Alessandra Gaia Saracini[1] and Leonhard Held[2]

We would like to express our sincere thanks to the reviewer and editor of Statistics in Medicine for their thoughtful and constructive feedback on our manuscript. We greatly appreciate the time and effort dedicated to the review process. In the following, we respond to each comment individually, providing detailed explanations and justifications. Where applicable, we have revised the manuscript accordingly, and we indicate these changes by explaining the rationale and referencing the relevant sections of the revised version.

To facilitate the review of changes, we have included marked-up versions of the main article and the supplementary material using `latexdiff`. These documents highlight the differences between the original submission and the revised versions.

Thank you again for your feedback and for the opportunity to improve our manuscript. We look forward to hearing your thoughts on the revised version and are happy to make any further changes as needed.

Reviewer 1

Summary

The paper addresses the problem of outcome reporting bias (ORB) in meta-analysis. Often outcomes are left unreported due to their direction, significance, size of study, etc. This is oftentimes informative. Post hoc corrections exist but the current paper proposes use of a selection model for missing/unreported outcomes. Selection models can provide unbiased estimates of pooled treatment effect in meta analysis. The models applied to address ORB are standard and have existed in the missing data literature for decades. Simulation are reasonable although I had some pointed comments to improve them, including using a comparative approach. A comparative example is shown (Copas' approach of likelihood adjustment for non reported outcomes) but this comparison is limited. Further innovation is limited – the manuscript's approach is not innovative but is still useful in an applied sense. Please see major comments below that should be addressed.

Background

1)

Some existing literature is missing from the review of this topic. I suggest the team re-review the literature. Below are two approaches for dealing with ORB in NMA from a group engaged in this work, one published in Stat med. One approach uses copulas to model correlated outcomes, another borrows information from multivariate outcomes under an MAR type model, but also appear to work for MNAR data. Other multivariate MA approaches like these are in your citations but are not explained. Both should be cited.

---

[1]Corresponding author: Alessandra Gaia Saracini, ETH Zurich, Department of Mathematics
alessandragaia.saracini@gmail.com

[2]Leonhard Held, Professor, University of Zurich, Epidemiology, Biostatistics and Prevention Institute
leonhard.held@uzh.ch

Liu, DeSantis, Chen, 2017. Bayesian mixed treatment comparisons meta-analysis for correlated outcomes subject to reporting bias, Journal of the Royal Statistical Society Series C.

Thank you for the suggestion. Lu, DeSantis, and Chen (2017) extend prior work (e.g., var Aert & Wicherts, 2023; Kirkham et al., 2012; Bay et al., 2021) by using a Bayesian NMA approach with a Clayton copula to model correlated outcomes and address reporting bias. This allows borrowing strength across outcomes, which can be highly beneficial. However, while correlations can be powerful, they are not always applicable – e.g., in the data example used by Copas et al. (2019), where outcomes may not be strongly correlated. Our approach, in contrast, allows for separate adjustments for each outcome and places more emphasis on explicitly modeling the missing data mechanism rather than relying on information from another outcome. Please refer to Section 1: Introduction, paragraphs 4-5 for the revised sources mentioned in the manuscript.

Hwang H, DeSantis SM. Multivariate network meta-analysis to mitigate the effects of outcome reporting bias. Stat Med. 2018 Sep 30;37(22):3254-3266. doi: 10.1002/sim.7815. Epub 2018 Jun 7. PMID: 29882392; PMCID: PMC7259375.

Thank you for the reference. Hwang et al. (2018) propose a multivariate network meta-analysis using MCMC with predictive distributions to estimate missing values. This approach is useful when many outcomes are available, as in the Copas et al. (2019) example dataset with multiple outcomes affected by ORB in adverse events, but it also increases model complexity by adding more parameters. A common solution is to use a single global correlation coefficient, as done in Riley et al. (2007)'s work and later in Kirkham et al. (2012), Bay et al. (2021), and var Aert & Wicherts (2023). While this simplifies estimation, it assumes a uniform correlation structure, which may not always hold. These challenges have been noted in the multivariate meta-analysis approaches, cited in our work. Furthermore, as per answer in 1), although this method effectively borrows strength across outcomes, it does not explicitly model the missing data mechanism. Our approach instead prioritizes direct modeling of missingness through selection models, making it more adaptable when correlations are weak or uncertain. A comparison of multivariate approaches vs. selection model approached could offer valuable insights. Please refer to Section 1: Introduction, paragraphs 4-5 of the revised the manuscript for these relevant updates.

Lunny and other use risk of bias tools, which you discuss:

Lunny C, Veroniki AA, Higgins JPT, Dias S, Hutton B, Wright JM, White IR, Whiting P, Tricco AC. Methodological review of NMA bias concepts provides groundwork for the development of a list of concepts for potential inclusion in a new risk of bias tool for network meta-analysis (RoB NMA Tool). Syst Rev. 2024 Jan 12;13(1):25. doi: 10.1186/s13643-023-02388-x. PMID: 38217041; PMCID: PMC10785511.

In the context of ORB, bias is outcome-specific rather than study-specific, since we have information on which outcomes have missing or incomplete data within a study, and the focus should be more on the possible reasons why that specific outcome was not included (Copas et al, 2014, 2019). This necessitates a more targeted approach. To this end, the ORBIT methodology used by Copas is designed specifically for ORB, assessing bias at the outcome level. The Copas method then builds on this by making statistical adjustments based on these assessments, assuming that certain categorizations (e.g., high risk of bias) correspond to specific missing data mechanisms. While RoB tools are valuable for identifying biases, our approach aims to provide a generalized statistical framework for ORB rather than relying on predefined categorizations. Other RoB tools could help inform the choice of selection functions in our method (e.g., indicating higher or lower bias risk), but we seek a more flexible approach independent of such classifications. In Copas et al. (2014), a sensitivity analysis was conducted on the probability of misclassification, and we extended this to other outcomes in our previous work (Saracini, 2023, MSc thesis). We observed that misclassification can influence results, highlighting the need for a more adaptable framework.

Cochrane Collaboration also has additional references that may be useful. Please carefully review work that has been done.

We also include the below reference, however, as previously noted, we highlight that these methods are used for PB to address bias in an overall study and are more difficult to apply directly in ORB.

Higgins, J., Altman, D., Gøtzsche, P., Jüni, P., Moher, D., Oxman, A., Savović, J., Schulz, K., Weeks, L., & Sterne, J. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ, 343, d5928. https://doi.org/10.1136/bmj.d5928

2)

I think the background could benefit from the type of ORB scenarios your method, vs existing methods is useful for. I see no mention of MNAR, MAR, MCAR in the usual forms.

Thank you for emphasizing the importance of clarifying the missing data mechanisms, i.e., MCAR, MAR, and MNAR, in the context of outcome reporting bias in meta-analysis. MCAR refers to missingness that is unrelated to any data, while MAR allows missingness to depend on other observed outcomes or study characteristics. MNAR arises when missingness depends directly on the unreported outcome itself, such as its significance. Because our method focuses on adjusting bias for a single outcome independently, whose significance directly affects their (un)reporting probability, MNAR is the primary scenario our approach addresses. This differs from settings with multiple outcomes, where MAR-related missingness can be handled by borrowing information across outcomes (see Responses 1 and 2). We have added an explanation of these concepts and their relevance to our method in the Introduction (third paragraph) and now refer to this terminology across the manuscript (e.g., Section 2, last paragraph; Section 3, first paragraph, Section 3.1 last paragraph; Section 3.2.5). We appreciate your helpful suggestion.

3)

Use of a selection model in the context of a GLMM is not novel; neither is it a novel application in meta-analysis.

Thank you for your comment. While the use of random effects models in meta-analysis and in publication bias adjustment is indeed well-established, their application within the context of ORB adjustments using selection models has been less explored. Even Copas et al. (2019) develop their approach exclusively under a fixed-effects framework, and their implementation is limited to that setting. In contrast, our method extends the selection model approach to incorporate random effects, allowing for a more realistic modeling of between-study heterogeneity. Furthermore, our work uniquely investigates the interplay between heterogeneity and ORB, including how each impacts the other, particularly in the estimation of treatment effects. While our primary focus remains on treatment effect estimation, this secondary focus on heterogeneity's role in ORB represents a novel contribution to the literature, to the best of our knowledge.

4)

No mention of multivariate meta-analysis except in important citations. Please review this literature as it is relevant to what is being proposed. I have a sense it would produce similar estimates in many or most applied scenarios.

Thank you for your comment. As per answer to point 1), we have added more background on the matter (see Introduction, paragraphs 4-5 and Discussion, paragraph 4) and we thank you for your helpful references. We acknowledge the strengths of multivariate meta-analysis and noted that it may be more appropriate in a setting of MAR, while we focus on MNAR (see also answer to point 2) and on direct modelling of the missing data mechanism, which is more aligned with the ORB method of Copas et al., (2019). In our previous work (Saracini, 2023) we included a detailed discussion of alternative methods in Chapter 4 (we aimed for conciseness in the current manuscript), particularly multivariate methods, i.e., Krikham et al. (2012), Bay et al. (2021) and var Aert & Wicherts (2023), where we noted the major limitations being the estimation of the correlation coefficients, and no direct modelling of the missing data mechanism. Bay et al. (2021) use a latent variable model for publication propensity and jointly model treatment effects with

a selection function typical in PB methods, assuming a constant within-study correlation across all studies and a between-study correlation. Interestingly, when we implemented this model with our example data from Copas at al. (2021), we observed different and contrasting results, to, e.g., the ORB-adjusted estimates obtained with Copas et al. (2019)'s method. As per answer to point 1), the revised manuscript includes further details and references to Liu et al. (2017) and Hwang et al. (2018).

Methods, Simulation, Application

5)

Re analysis of the same data set in Copas is useful for comparisons however your simulation paradigm is more useful to compare different methods and in my opinion should appear first. We do not know which is correct – Copas' original results or your approach.

Thank you for your comment. We agree that the focus should be on our methodology, which is why we have restructured the manuscript to present the simulation first (Sections 3.1-3.2), followed by the application to the example data (Section 3.3). While it would indeed be insightful to compare our vs. Copas et al. (2019)'s method, a direct comparison is challenging due to the different underlying approach in ORB adjustment. We explain this rationale in detail in a new pararaph added to the manuscript: Section 3.1, paragraph 4.

Of note, we elaborate on the similarities between our methodology, and specifically, our choice of selection functions, and the assumptions made in Copas et al. (2019). Specifically, in Section 2.1.1, we highlight how, using our generalized ORB-adjusted log-likelihood, with a piecewise constant function (Equation 8), only on a subset of unreported studies, i.e., those at High Risk (HR) of bias per ORBIT classification, and for a two-sided p-value instead of one-sided, we obtain the ORB-adjusted log-likelihood of Copas et al. (2019). In our simulation studies, we use the piecewise constant selection function (Equation 8), on all unreported outcomes and with a one-sided p-value, in alignment with other estimators/selection functions. Although this is not a direct comparison with Copas et al. (2019), it could be seen as a benchmark comparison to their missing data assumption which justifies the selection function of Equation (8).

Ultimately, we believe the methodology and simulation should be our focus, with the real data application as an illustration. The comparison to Copas et al. (2019) provides useful context but is secondary to the focus on our method and would rise difficulty in their direct comparison. This shift in the structuring of our manuscript reflects our decision to move away from the baseline Copas et al. (2019) approach in favor of our method, which can be seen as a generalized and more broadly applicable approach.

6)

Please consider simulation scenarios first, to show your approach outperforms existing approaches, in this case perhaps that is Copas' approach. You have not compared your approach to any baseline approach despite the fact they exist in the literature

Thank you for your comment. Similarly to the above answer, we wish to highlight that, while we take inspiration from the Copas et al. (2019) method, we chose not to compare it to ours as a baseline in the simulation framework because Copas et al. (2019)'s method relies on different/more information available. Specifically, Copas et al. (2019) classifies missing outcomes into High Risk (HR) and Low Risk (LR), and their method is applied only to the HR outcomes, assuming that the classification is accurate. This classification is based on the assumption that it corresponds to non-significant outcomes. In contrast, our approach eliminates the need for such classification, offering a more general framework that can be applied without knowing the degree of bias. We do not assume that we know the cause of missingness for each missing outcome; instead, we make an assumption about the overall selection function that describes the behavior of all studies with unreported treatment effects. To compare our method with Copas et al. (2019)

would require applying Copas only to the HR outcomes, which would not fully reflect the flexibility of our approach. Comparing the two would require different levels of knowledge available, for more details see response to comment (6) and (8).

7)

Simulation set up is simplistic and always under MNAR, which may not be the mechanism. Please consider adding an MAR scenario to show that if the mechanism were MNAR, you still recover parameters.

Thank you for your comment and insightful suggestion. We have conducted a new simulation, with the same set-up and estimation, but in which the unreported study outcomes are removed based on a MCAR setting. We obtained this by removing the same number of studies as per the MNAR simualtions, but at random and independently of the p-value. This was insightful so as to understand the implications of incorrect model specification (i.e., a MCAR data generating mechanism but a MNAR approach to estimation). We have added a new paragrah at the end of Section 3.1, elaborating on this in further detail. Furthermore, please refer to the new section Section 3.2.5 synthesizing the results of the MCAR simulations. We would also like to point you towards the revised supplementary material, where we added plots for the results of the secondary simulation, and elaborated on such results in the new section "Measures of Secondary Simulation (MCAR)".

We would like to thank you for this input as it shed light on the robustness of our results, on the importance of sensitivity analyses and on the connections to MCAR, MAR, MNAR.

8)

If you use the comparative example only from Copas (and not a second data example) I think simulation are important to show the bias, 95% coverage probability, etc parameters are appropriate for the selection model in Eqn 11, vs under COPAS (or other working) model applied to your simulations.

Thank you for your comment. As discussed in answers (5-6) above, it is challenging to directly compare the Copas et al. (2019) method to our simulation results, as Copas et al. (2019) use different available information (HR/LR classification by ORBIT), which in our method, is assumed to be unknown. Please refer to the Section 3.1, paragraph 4 for further details on this rationale. We would like to highlight again that in our simulations we compare various selection functions, and among these is the piecewise constant, which is closest to the Copas et al. (2019), with regards to the missing data assumption made. Please refer to answer (5) and Section 2.1.1 fo further details.

9)

The fact that imputation of missing variance is involved seems to move this from an informative missingness approach, or an approach for MNAR data, to an approach for MAR data. Please comment on that.

Thank you for your insightful comment. In line with the methodology proposed by Copas et al. (2019), we impute the missing variances using a design factor approach. The key rationale is that, unlike in the case of publication bias, in outcome reporting bias (ORB), studies are published and thus their sample sizes (n) are typically available. Rather than simply approximating the variances of missing outcomes using inverse sample sizes, the design factor approach leverages the relationship between sample size and observed variances across reported outcomes. This allows us to estimate the likely variances of unreported outcomes in a way that reflects study-specific characteristics and borrows strength from the reported data. We added more details on this in Section 2.2 "Imputation of missing variance" in the revised version of the manuscript.

10)

5

The wording in 2.1.1 regarding the different parameterizations of the selection model makes it unclear what you actually did. In many cases you say "one can do" and in others "we can do" – please firm this language up because the reader will be unsure what from 2.1.1 you actually adapted into practice thus this needs clarification. Eg, was only the selection function formulation under Eqn 11 used?

Thank you for the important observation. In the revised manuscript we avoid using impersonal verbs (i.e., "we used/implemented" instead of "one can do"). Furthermore, we changed the naming of the sections, so that after "Section 1: Introduction", we now directly have "Section 2: Selection Models for ORB", (the focus on ORB was previously a subsection of Section 2). This way, "Section 2.1: Selection Functions" is a new section title highlighting what was effectively considered, implemented and used in the simulations. We further subset Section 2.1 into each of the selection functions implemented (which we also renamed for better clarity), i.e.: "Section 2.1.1: Piecewise constant", "Section 2.1.2: Constant-decreasing", "Section 2.1.3: Decreasing-constant", and "Section 2.1.4: Piecewise decreasing". We hope these reformulations provide more clarity to the reader.

11)

Fig 1 presents selection functions for the epilepsy data before describing any data. Please reconsider presentation format with simulations appearing before data analysis.

Thank you for your comment. Figure 1: Possible Selection Functions for ORB-adjustment, shows the selection functions considered in our methodology, i.e., the ones from Sections 2.1.1-2.1.4 listed above. Figure 2: Application of ORB-adjustment to example epilepsy data, shows how these selection functions are applied to the example data. As noted in answers to comment (5-6), we rearranged the manuscript so as to present simulations first, and only subsequently the application of the same functions used in the simulations, to the example data. Please refer to answers (5-6) for further details.

12)

Usually selection models would be applied under a sensitivity analytic framework, and raw results compared to those results. As someone who may implement your approach, I do not know from this paper how to present findings. Please elaborate or clarify the steps to fitting the model to crude data (observed) and then applying the selection model, and guidelines for reporting the MA.

Thank you for your insightful comment. We agree that a sensitivity analysis framework is the best practice in this ORB-adjustment scenario. After restructuring our manuscript (see answers 5,6,12), we moved the application to the example data (with the same parameters as those of the simulation) to Section 3.3: Application to Example Data. Furthermore, in this section we added a sensitivity analysis (see Figure 2 and paragraph 5 onward in Section 3.3). Here we elaborated on the importance of sensitivity analyses, recommendations on how to conduct, report and interpret such analyses. We further emphasize the recommendation of applying a sensitivity analysis framework in the revised paragraph 5 of the Section 4: Discussion.

Associate Editor

The reviewer of the manuscript finds merit in the manuscript and suggests a few major revisions that would improve the quality of the manuscript.

Thank you for your review. We have addressed all comments from the reviewer as detailed above.