Title: An assessment of statistical methods for non-independent data in ecological meta-analyses: Reply

Running title: Reply to Nakagawa et al.

Authors: Chao Song[1,*], Scott D. Peacor[1], Craig W. Osenberg[2], James R. Bence[1]

Affiliations:

[1]Department of Fisheries and Wildlife, Michigan State University, East Lansing, Michigan 48824, USA

[2]Odum School of Ecology, University of Georgia, Athens, Georgia 30602, USA.

[*]Correspondence: chaosong@msu.edu

Recently, Nakagawa et al. (2021) provided a timely and insightful comment to our paper on statistical methods for non-independent data in ecological meta-analyses (Song et al. 2020). Their comment highlighted the value of using hierarchical models in meta-analysis to address non-independence, and offered two assertions: 1) that a two-step method that first calculates a weighed mean effect size of each paper and then analyzes the paper mean in a random effect model has limited scope of application; and 2) that several solutions to avoid inflated type I error rates in hierarchical models already exist and can be implemented with existing software packages in R.

Two-step method using weighted paper mean

We fully agree with Nakagawa et al. (2021) that the two-step method using a paper mean cannot be applied in all situations. For example, this method does not allow the analyst to address non-independence due to phylogeny or to analyze the effect of covariates if the value of the covariate varies within a paper. However, that an approach is not always applicable does not mean it is never a useful approach. The frequent occurrence of the two-step method within the ecological literature points to its accessibility and suitability in many contexts. Within the scope of its applicability, the two-step method offers good performance in terms of precision and type I error rates and thus is a viable choice of method for meta-analysts.

Nakagawa et al. (2021) expanded the scope of our analysis by considering cases in which the non-independence within papers arose via correlations among the within-study error (Gleser and Olkin 2009, Lajeunesse 2011). They argue that when the two-step method is used in this situation, the average should not be calculated as a weighted average using inverse variance weights, but rather an unweighted average. They provided a formula for the variance of the unweighted mean that accounts for correlated within-study error. We do not agree with this suggestion because a weighted average yields a more precise estimate of the mean effect size than does an unweighted mean. If the within-study errors are correlated, the weighted average and its variance can be calculated as

$$\widehat{\mu_w} = (\mathbf{J^T V^{-1} J})^{-1} \mathbf{J^T V^{-1} y}, \tag{1}$$

$$\mathrm{var}(\widehat{\mu_w}) = (\mathbf{J^T V^{-1} J})^{-1}. \tag{2}$$

Here, $\widehat{\mu_w}$ is the estimated mean for a paper, $\mathbf{J}$ is a column vector of 1s, $\mathbf{V}$ is the variance-covariance matrix of the within-study error, and $\mathbf{y}$ is a column vector of observed effect sizes from a paper. The term $(\mathbf{J^T V^{-1} J})^{-1} \mathbf{J^T V^{-1}}$ is a row vector of weights. In practice, meta-analysts do not need to manually calculate the weighted average and its variance for each paper using these equations. Instead, analysts can use existing tools to easily make these calculations. For example, in our paper we assumed within-study errors were independent, and we fit a fixed-effect model to observed effect sizes from each paper to obtain the weighted average and its variance using the rma function in R package metafor (Viechtbauer 2010). One can extend this method to cases of non-independent within-study error by incorporating the variance-covariance matrix ($\mathbf{V}$) of the within-study error in the fixed effect model (e.g., using function rma.mv in metafor). Alternatively, one can use function aggregate in metafor to make these calculations.

Hierarchical models in meta-analysis

We fully agree with Nakagawa et al. (2021) that the hierarchical model is a versatile tool that allows analysts to answer a much richer set of ecological questions, including modeling the effects of covariates and partitioning the source of random variation in observed effect sizes. While we embrace a hierarchical approach in principle, our reservation about this method was its consistently high type I error rates when implemented in the metafor package in R.

Any debate about the two-step method would be moot if we could readily fit hierarchical meta-analysis models without inflating type I error rates and thus avoid giving a false sense of confidence in calculated effect sizes. The issue of inflated type I error rate in hierarchical models in Song et al. (2020) occurred because metafor uses the number of observations minus number of model coefficients as its default degrees of freedom for hypothesis testing and confidence interval calculation. We suggested that adjusting the degrees of freedom, which has been applied more generally in linear mixed-effect model, could be a solution. Nakagawa et al. (2021) implemented and evaluated several methods for adjusting the degrees of freedom in hierarchical meta-analysis models. They showed that the Satterthwaite adjustment of degrees of freedom largely resolves the issue of high type I error rate. More simply, using the so-called containment method for degrees of freedom also reduced the type I error rate. This containment method was recently implemented in metafor after the publication of Song et al. (2020), which makes it more accessible to analysts.

However, the methods used to adjust degrees of freedom and thus improve type I error rate vary in their performance. For example, the containment method for degrees of freedom gives the exact degrees of freedom when the design is balanced, i.e., all random effects in the model are nested and sample sizes within each group defined by the random effects are equal. With an unbalanced design, the containment method gives an inflated type I error rate, although this inflation was trivial over the conditions simulated by Song et al. (2020) and Nakagawa et al. (2021). The Satterthwaite method is more generally applicable in these situations. Another commonly used method to adjust the degrees of freedom is the Kenward-Roger method (Kenward and Roger 1997). A simulation study showed that it may perform better than the Satterthwaite method (Schaalje et al. 2002) although both methods appear to give adequate type I error rate in linear mixed-models in general (Luke 2017). Neither method is, however, currently available in metafor although the Satterthwaite method can be implemented with tools suggested by Nakagawa et al. (2021).

Conclusions

We appreciate the helpful clarification and analysis of our paper by Nakagawa et al. (2021). Based on findings in our paper and their comment, we agree that the two-step method is not universally applicable, but could be a viable choice of method when it fits the goal of the application. Hierarchical models provide a more versatile and powerful tool for meta-analysis. However, analysts should be aware of the inflated type I error rate under default methods for degrees of freedom in metafor. Although one might be tempted to dismiss this inflation as minor, error rates were as much as 1.6 times the nominal rate of 0.05, which in certain contexts might be unacceptable. Given that the high type I error rate that can result from the default in metafor, we encourage analysts fitting hierarchical models with metafor to use t- or F-distributions for hypothesis tests with adjustments for the degrees of freedom. While we agree that solutions are already known to statistically savvy analysts, many authors will rely on default options of the software. We encourage developers of readily available meta-analysis software to incorporate these methods for adjusting degrees of freedom, and when appropriate, make them the default method.

Acknowledgements

This work was supported by the U.S. National Science Foundation (DEB–1655394 and DEB–1655426). JRB and SDP received support from AgBioResearch of Michigan State University. This is publication 2021-XX of the Quantitative Fisheries Center at Michigan State University.

Literature cited

Gleser, L. J. and I. Olkin. 2009. Stochastically dependent effect sizes. In H. Coopers, L. V. Hedges, and J. C. Valentine, editors, The handbook of research synthesis and meta-analysis, pages 357–376. Russell Sage Foundation, New York, NY, 2nd edition.

Kenward, M. G. and J. H. Roger. 1997. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics, **53**:983–997.

Lajeunesse, M. J. 2011. On the meta-analysis of response ratios for studies with correlated and multi-group designs. Ecology, **92**:2049–2055.

Luke, S. G. 2017. Evaluating significance in linear mixed-effects models in R. Behavior Research Methods, **49**:1494–1502.

Nakagawa, S., A. Seniro, W. Viechtbauer, and D. Nobel. 2021. An assessment of statistical methods for nonindependent data in ecological meta-analyses: Comment. Ecology.

Schaalje, G. B., J. B. McBride, and G. W. Fellingham. 2002. Adequacy of approximations to distributions of test statistics in complex mixed linear models. Journal of Agricultural, Biological, and Environmental Statistics, **7**:512–524.

Song, C., S. D. Peacor, C. W. Osenberg, and J. R. Bence. 2020. An assessment of statistical methods for nonindependent data in ecological meta-analyses. Ecology, **101**:e03184.

Viechtbauer, W. 2010. Conducting meta-analysis in R with the metafor package. Journal of Statistical Software, **36**:1–48.