*A Method article prepared for Ecology Letters*

**A robust and readily implementable method for the meta-analysis of response ratios with and without missing standard deviations**

Shinichi Nakagawa\*1, Daniel W. A. Noble2, Malgorzata Lagisz1, Rebecca Spake3, Wolfgang Viechtbauer4 & Alistair M. Senior\*5

Address:

1 Evolution & Ecology Research Centre and School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia

2 Division of Ecology and Evolution, Research School of Biology, The Australian National University, Canberra, ACT, Australia

3 XXXX

4 Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Faculty of Health, Medicine, and Life Sciences, Maastricht University, 6200 MD Maastricht, The Netherlands

5 Charles Perkins Centre and School of Life and Environmental Sciences, University of Sydney, Camperdown, NSW 2006, Australia.

Running head: Handling missing SDs in meta-analysis

\* Correspondence: Shinichi Nakagawa ([s.nakagawa@unsw.edu.au](mailto:s.nakagawa@unsw.edu.au)); Alistair M. Senior ([alistair.senior@Sydney.edu.au](mailto:alistair.senior@Sydney.edu.au))

# Abstract

(do it in the end)

**KEYWORDS**

Missing data, multiple imputation, meta-regression,

Dan N to create a function to take effect size, sampling variance and study\_ID

# INTRODUCTION

Meta-analyses are routinely used to test the generality of ecological phenomena and explain inconsistencies in findings between studies ([Nakagawa *et al.* 2015](#_ENREF_20); [Gurevitch *et al.* 2018](#_ENREF_6)). However, incomplete reporting of necessary data in the primary literature threatens the validity of meta-analytic evidence. Many empirical papers fail to report standard deviations (SDs) or their derivatives, such as standard errors (SEs) and confidence intervals (CIs). SDs are required to calculate effect sizes and/or their precision for comparing means between two groups in ecological meta-analyses: standardised mean difference, SMD (well-known estimators include Cohen’s *d* and Hedges’ *g*) and the natural logarithm of response ratio, lnRR ([Hedges *et al.* 1999](#_ENREF_8)).

A recent review of 505 ecological meta-analytic studies showed nearly 70% of the datasets included studies with missing SDs ([Kambach *et al.* 2020](#_ENREF_11)). While sample sizes were sometimes unreported, SDs were around 8 times more likely to be missing. The same review also showed that the majority of meta-analyses went on to exclude studies with missing SDs (i.e., complete-case meta-analysis). Using simulations, Kambach et al. ([2020](#_ENREF_11)) demonstrated that excluding studies with missing data can bias meta-analytic results. However, multiple imputation (MI) of missing SDs (and also sample sizes) was an effective means of providing unbiased meta-analytic results.

MI was introduced to ecologists more than a decade ago ([Nakagawa & Freckleton 2008](#_ENREF_18)). However, the uptake of this method has been very slow, especially so for meta-analyses (cf. [Ellington *et al.* 2015](#_ENREF_4); [Kambach *et al.* 2020](#_ENREF_11)). There are, we believe, two major reasons for this slow uptake. First, for many ecologists, the implementation of MI is tedious, perhaps because it involves three steps: 1) creating multiple datasets with imputed missing data (e.g., *Ndataset* = 100 with all missing SD data imputed), 2) analysing each dataset separately, and 3) aggregating parameter estimates (i.e., regression coeffects) to get unbiased estimates using Rubin’s rules ([Rubin 1987](#_ENREF_25)) (for details, see [Nakagawa 2015](#_ENREF_17); [van Buuren 2018](#_ENREF_30)). We note, however, that these steps can be largely automated using the *R* packages *mice* ([van Buuren & Groothuis-Oudshoorn 2011](#_ENREF_31)) and *metafor* [[Viechtbauer 2010](#_ENREF_32)]. The second reason the lack of uptake of MI maybe uncertainty around its use with the more complex analyses that are often necessary for ecological questions. For example, it is unclear if Rubin’s rules are always appropriate for aggregating variance components and related quantities (e.g., *I*2 and *R*2) and information criteria (e.g., AIC, BIC; cf. [Nakagawa & Freckleton 2011](#_ENREF_19)). Furthermore, MI cannot be easily implemented for multilevel (mixed-effects / hierarchical) models unless they have only two levels (i.e. one random factor) ([van Buuren 2018](#_ENREF_30)). Therefore, the use of MI for meta-analytic studies using multilevel models is still seldom applied, reenforcing age-old approaches of excluding studies with missing SDs.

Here, we propose alternatives to MI for handling studies with missing SDs in meta-analyses that use lnRR, the most widely used effect size statistic in ecological meta-analyses ([Nakagawa & Santos 2012](#_ENREF_21); [Koricheva & Gurevitch 2014](#_ENREF_12); [Kambach *et al.* 2020](#_ENREF_11)). Several studies recommend lnRR over SMD for ecological meta-analysis because of the assumptions of SMD are most likely to be violated (e.g., homoscedasticity) in ecological and evolutionary meta-analyses ([Nakagawa *et al.* 2015](#_ENREF_20); [Spake *et al.* 2021](#_ENREF_28); see also [Osenberg *et al.* 1997](#_ENREF_23)). Here we propose three new methods to deal with missing SDs in a multilevel meta-analytic models as well as traditional random-effects models. First, we introduce a method developed by Doncaster and Spake ([2018](#_ENREF_3)) that uses an adjusted sampling variance formula for lnRR values. We, then, improve and extend this approach for missing SDs; we provide two procedures within this method: only using this adjustment method for effect sizes with missing SD (Method 1A), and using this adjustment method for all effect sizes regardless of missingness (Method 1B). Second, we describe a method based on traditional weighted regression (Method 2). Next, we combine these two methods to provide a hybrid method (Method 3). We conduct a simulation study to compare these methods to the control method (a standard meta-analytic procedure without missing SDs) and show that Method 1B performs best. Surprisingly, for many cases, Method 1B (and Method 2) with missing SDs outperforms the control method without missing SDs (i.e., full data analysis). We make recommendations for future meta-analyses accordingly. We illustrate these new methods using the R package, *metafor* (all relevant data and code are available at XXX and Github repository: XXX).

# NEW STATISTICAL METHODS

## Estimating sampling variances prior to meta-analysis: Method 1A and 1B

The effect size statistic, lnRR was first proposed by Hedges et al. ([1999](#_ENREF_8)) as follows:

where *m*1 and *m*2 is the mean of group 1 and group 2, respectively (e.g., experimental and control groups), *v* represents the sampling variance, *sd* and *n* is the SD and sample sizes, respectively, and CV (*sd*/*m*) is the coefficient of variation.

Using simulation, Doncaster and Spake ([2018](#_ENREF_3)) showed that when *n* is small (*n* = 3–10 observations, meaning *n*1 + *n*2 = 6–20), the sampling variance of lnRR using Equation 2 is underestimated by 15–20%, with consequences for the accuracy and precision of the overall mean estimate (i.e., meta-analytic mean). This is because when sample sizes are small, SDs themselves can be underestimated (in Equation 2). To correct for this bias, they propose using the following as the sampling variance for lnRR, instead of Equation 2:

where and are from the *i*th effect size (study; *i* = 1, 2, …, *Nstudy* = *Neffect-size*, as we assume *Nstudy* is independent) and note that this formula uses the cross-study average of CV2 rather than study-wise CV2, which could be underestimated when sample size is small. Doncaster and Spake ([2018](#_ENREF_3)) demonstrated that the use of Equation 3 substantially reduced the bias in sampling variance, improving the overall (meta-analytic) mean estimate. This seems intuitive because when SD is poorly estimated, the average CV (or CV2) may well be a good substitution assuming CV stays fairly constant (cf. [Nakagawa *et al.* 2015](#_ENREF_20)). Notably, Doncaster and Spake also suggested this formula could be used when SDs are missing from some studies, although this possibility was not statistically investigated.

Here we propose two improvements to Equation 3. Lajeunesse (2015) conducted a simulation study to show that Equation 1 and 2 are both biased when sample size is small; even when *n*1 + *n*2 = 32, such biases were still substantial. Lajeunesse ([2015](#_ENREF_15)) showed the following estimators – based on the second-order Taylor expansion – can reduce these biases (see also [Senior *et al.* 2020](#_ENREF_27)):

Therefore, Doncaster and Spake’s variance estimator (Equation 3) can be further improved by using Equation 5. Also, rather than using the average CV2, we can use the weighted average of CV2 for both point estimate and sampling variance as follows:

It should be clear that one can use Equation 7 (along with Equation 6) where SDs are missing, because the weighted cross-study CV2 can be applied to those studies that do report SDs, allowing the inclusion of these studies in downstream analyses (Method 1A; a mixture of Equations 4-7). Alternatively, one may use Equation 7 throughout regardless of the missingness of SDs (Method 1B with Equation 6 & 7)

## Using weighted-regression-like method: Method 2

In the absence of SDs, it has been suggested that information on sample sizes can be used to approximate a study’s sampling variance in meta-analyses of lnRR or SMD, using the inverse of the following (e.g., [Lajeunesse 2013](#_ENREF_14); [Kambach *et al.* 2020](#_ENREF_11)):

However, treating the version of Equation 8 as sampling variance (or Equation 8 as weight) is clearly erroneous, ignoring other terms such as mean and SDs, for example, in Equation 2 & 5, although some ecological meta-analyses have done so (see the review by [Kambach *et al.* 2020](#_ENREF_11)). A more realistic assumption is to treat 1/ as proportional to the sampling error; indeed, Equation 2 reduces to the inverse of Equation 8 when we set both CVs to 1. Weighted regression makes this assumption of proportionality. Furthermore, we believe many ecologists are likely to be familiar with weighted regression models fitting sample sizes as the weights ([Fletcher & Dixon 2012](#_ENREF_5)).

The simplest random-effects meta-analytic model using lnRR can be written as follows:

where is the overall estimate (or meta-analytic mean); *s*i is the between-study effect for the *i*th effect size, normally distributed with a mean of zero and a variance of (sometimes, referred to as ), *mi*is the sampling error for the *i*th effect size, distributed with the *i*th sampling variance (note that *i* = 1, 2, …,  *Neffect-size*, the number of effect sizes = *Nstudy*, the number of studies). Note we assume here that the sample variance of lnRR is known; that is, we use either Equation 2 or 5 as true sampling variances () in Equation 9. Also, the ratio between and the total variance is often used to quantify heterogeneity (*I*2), written as:

where is known as a ‘typical’ (or ‘average’) sampling variance (originally referred to as ‘typical within-study variance’; *sensu* REF), which has several ways of estimating (REF).

In a weighted regression, the following is assumed:

where , which is estimated by the model, works as a multiplicative parameter fulfilling the assumption of proportionality (i.e., is proportional to the sampling variance ). Doncaster and Spake ([2018](#_ENREF_3)) recommended that we use Equation 3 as sampling variance for lnRR, which is . However, Doncaster and Spake’s simulation suggests that Equation 3 may still underestimate sampling variance when sample sizes are small (e.g., *n*1 + *n*2 = 6 – 20). Therefore, it may be advisable to assume: . In a similar vein, we can apply the same assumption to Equation 7, as , and we use this as our Method 2; taking Equation 7 for the sampling variance, and either Equation 4 or 6 as the point estimate of the effect size.

## Combining Methods 1 and 2: Method 3

In the second method, Equation 7 can be used regardless of whether SDs are missing or not. We can, however, combine Method 1 and Method 2. When SDs are available, we can use Equation 5 to obtain the sampling variance of lnRR (along with Equation 4 for the point estimate). When SDs are missing, we can use Equation 7 (and Equation 6), and combine this with Method 2 in a weighted regression (i.e., ). We can write such a model (Method 3), using a multilevel meta-analysis (modelling multiple effect sizes per study) as follows:

where *sj* is the between-study effect for the *j*th study, normally distributed with a mean of 0 and variance of (often referred to as tau2); *ui* is the between-effect-size effect, or within-study effect, for the *i*th effect size, distributed with a mean of zero and variance of (*j* = 1, 2, …, *Nstudy*, the number of studies, and *i* = 1, 2, …, *Neffect-size*, the number of effect sizes; *Neffect-size* > *Nstudy*; cf. Equation 8), **M** is a diagonal matrix with when SDs are not missing and when SDs are missing (i.e., Method 3). For example, when we have five effect sizes, **M** can be:

where 1st, 2nd and 5th cases (effect sizes) have SDs while 3rd and 4th did not have SDs and as above, is estimated in the model. Because this model accounts for non-independence it is appropriate in ecological meta-analyses that include correlations among-effect sizes such as when there is more than one effect size per study or species ([Nakagawa & Santos 2012](#_ENREF_21)),

# SIMULATION

## Simulation design

To compare Methods 1A, 1B, 2 and 3 with different levels of missing SDs with the baseline (a standard meta-analysis with full data), we conducted a simulation study We simulated hierarchical datasets where each study within the dataset contained ≥1 correlated effect sizes (i.e., an intra-class correlation for study; ICC*s* = /( + ) using the terms in Equation 12), as is commonly seen in eco-evolutionary contexts. For each simulated dataset we analysed the complete cases, before deleting the SDs for 5%, 15%, 25% … or 55% of the studies (i.e., missingness was on a study-wise, rather than effect size-wise basis, which seems the most likely scenario) and re-analysing with the four proposed methods for handling missing SDs (Method 1A, 1B, 2 and 3). Datasets were analysed using a multi-level meta-analytic model, which included a study-level random effect and was specified using the ‘rma.mv’ function in *metafor*. We evaluated bias in the overall estimated effect size (difference between estimated and parametrised value), coverage of 95% confidence intervals (CIs), bias in the estimated total heterogeneity (*τ*2 = + in Equation 12 and *τ*2 = in Equation 8; log ratio of estimated and parametrised value) and bias in the estimated ICC*s* (difference between estimated and parametrised value). CIs were calculated as the estimated effect ± *t*-value × SE, where for t-values the degrees of freedom was the number of effect sizes minus 1, when ICC*s* = 0, and the number of studies minus 1 when ICC*s* > 0.

Each simulated dataset contained *K* studies (*K* = 12, 30 and 100 was tested). Because studies often vary in the number of effect sizes they contain, the number of effect sizes per study, *L*, was treated as a random variable. We simulated *L* using a double Poisson distribution, which is a discrete probability distribution that can be under/over dispersed relative to a Poisson distribution *via* a multiplicative dispersion parameter. In the main text, the simulation drew values from a random double Poisson distribution (‘rDPO’ function in the *gamlss.dist* package) with a mean of 2 and a multiplicative dispersion parameter of 2.88, before adding 1 (to prevent 0 values). This resulted in *L* having a minimum of 1, a mean of 3, and SD of 2.4 (i.e., dispersion of 1.92). We termed this set of parameters Set I. We also simulated a second set where *L* is fixed to 1 (i.e. each study had only one effect size; *L* = 1, dispersion = 0), which we called Set II. Set II is equivalent to a meta-analysis with just one effect size per study (i.e., no dependency), and which would be assessed using a standard random-effects meta-analysis.

To simulate effect sizes that were correlated in a hierarchal manner we assumed an overall lnRR (*θ*) of 0.3 (*e*0.3 = 1.35, or a 35% increase in the mean) with either negligible (*τ*2 = 9×10-6 or *τ* / *θ* = 0.01) or a high level of total heterogeneity (*τ*2 = 0.09 or *τ* / *θ* = 1). This heterogeneity was partitioned between among-, and within-study level effects assuming a given intra-class correlation (ICC*S*; values of 0 and 0.5 were tested) such that the *l*th effect size (*l* = 1 … *L*) in the *k*th (*k* = 1 … *K*) study, *θkl*, was drawn from a hierarchical pair of random normal distributions (‘rnorm’ function in *base* R) as:

To simulate variation in the precision of the studies in the dataset we treated the sample size of the underlying studies as a random variable, *N*. We assumed *N* varied at the level of the study such that each group/effect size within the same study had the same sample size. In our experience it is common for experimental designs to vary among, more than within, studies. We drew the simulated sample size for study *k* by drawing a random value from double Poisson distribution before adding a value of 3. The double Poisson distribution was parametrised with a mean of either 2 or 27 coupled with dispersion parameters of either 3.65 or 1.66. After adding the constant of 3, this resulted in two different distributions of *N* both with a minimum of 3, and (over) dispersion of 1.5, but with a means (*μN*) of either 5 or 30. The smaller mean value is more akin to the kinds of studies observed in terrestrial/ecosystem ecology (or some pre-clinical biomedical studies), while the larger mean value is more like evolutionary/behavioural ecology studies (or even clinical trials).

The underlying data in control and treatment groups in each effect size were drawn from random normal distributions ‘shifted’ to ensure both groups had a positive mean as is required for analysis using lnRR. From these individual simulated values, we calculated the mean and SD in each group for calculation of lnRR and downstream meta-analysis. The observations for the control group in effect size *l* in study *k* weredrawn from the random normal distribution, , and the paired treatment group from the random normal distribution, , where *σk* is the SD in the underlying individual observations in study *k*.

Because we are assessing the performance of methods to deal with missing SD values, we chose to treat the within-group (among-observation) SD as a random variable, *S*. The SD for study *k* was drawn from a random Gamma distribution (‘rgamma’ function in *base* R) with shape and scale , where *μS* is the mean of *S* (i.e., mean SD of studies; here 15), and *σS* is the SD in *S*. This latter parameter thus specifies how heterogeneous the within-study (among-observation) variances are; we tested values of 10-10 (~0), 3.75 and 7.5 (i.e., entirely homogeneous variances, or the CV for the SD among studies is 0.25 or 0.5). A summary of the key parameters and their values is given in table S1. Each combination of parameter values was simulated 10,000 times for both Set I and Set II. We note that for Set I presented in the main text we used the model in Equation 12 (the multilevel meta-analytic model). For Set II we used the model in Equation 8 (the random-effects meta-analytic model), where the results are presented in the supplementary materials. For all the new methods, we needed to calculate the average CV2 as in Equation 6 and 7. In Set I, this calculation was done by averaging CV2 within studies and then taking weighted-average CV2 across studies (using mean *n* per study as the weight), disregarding rows containing missing SDs. For Set II, we calculated weighted CV2 among studies (using *n* per study as the weight) as we only had only one CV2 value per study.

## Simulation results

Elsewhere (REFREF), we advocated the use of multilevel meta-analytic models as they are much more realistic and practical than the random-effects model. Therefore, we focus on presenting the results from Set I, using the multilevel meta-analysis (Equation 12; the results for Set II are very similar and presented in the Supporting Information). Fig. 1A shows the distribution of the mean bias in the estimated effect under each simulated condition with complete data and using the four different methods for handling missing SDs. Even with full data, there could be both upward and downward bias in the estimated effect size, and this was reflected in the analyses using Method 1A and 3 to handle missing SDs. Notably, even at its most extreme, this bias only amounted to a little over 2% of the simulated effect and was usually ~0.5%. Nonetheless, Method 1B and 2, both of which use the weighted CV2 to estimate the sampling variance for all effect sizes regardless of missingness, displayed the least bias on average (Fig. 1A). The degree of bias across conditions in the full data analysis correlated very strongly with bias using Method 1A and 3, while bias in Method 1B and 2 correlated strongly (Fig 1B). This suggests that the methods fall in to two classes that perform similarly across situations: Method 1A with Method 3, and Method 1B with Method 2. Contrasting methods 1A and 1B directly, in almost all cases the absolute level of bias in Method 1A was higher than that for method 1B (Fig. 1C). Further where Method 1B had higher bias than method 1A, this difference was small (Fig. 1C). Although Method 1B and 2 outperformed the other approaches on average, they were prone to producing excessively large bias on rare occasions; Fig. 1D shows the range in bias among the individual replicates under each simulated condition as a function of the different methods. Large ranges in bias occurred when the SDs among different studies were very heterogeneous, and the individual studies themselves had low sample size (Fig. 1E).

All methods for handling missing data, and the full data analyses, could produce 95% CIs that were too narrow, or too wide under different scenarios (Fig. 2A). The full data, and Methods 1A and 3 tended to typically produce Cis that were slightly too narrow, whereas method 1B and 2 were prone to producing wider Cis (Fig 2A). Again, contrasting Method 1A and 1B, it is clear what Method 1B’s tendency to produce too-wider CI occurs when the total heterogeneity among studies is low (Figs 2B and 2C). However, where total heterogeneity is high Method 1B performs as well as, or better than Method 1A (Figs 2B and 2C).

Fig. 3A shows the mean bias in the estimated heterogeneity under each condition and method. Under most conditions all methods estimated heterogeneity with little bias, but the full data, as well as Method 1A and 3 were prone to substantially overestimating the total heterogeneity (Fig. 3A). All methods were prone to underestimating heterogeneity under some conditions (Fig. 3A). Method 1B displayed a distinct bimodal distribution, either estimating heterogeneity perfectly or underestimating it substantially (Fig. 3A), the latter occurring when the total level of heterogeneity was low (Fig. 3B). Method 2 most inaccurately partitioned the variance between the within- and among-study levels (Fig. 3C). Across the different scenarios tested, the other methods typically partitioned the heterogeneity with little bias, although again, bias in Method 1B displayed a distinct bimodal distribution (Fig. 3C). Method 1B was prone to overestimating the ICC for the study-level effect when the simulated study effect was absent and when the total heterogeneity was low (Fig 3D).

In summary, on average, Method 1B performs with the least bias under the broad range of simulated conditions tested. This method may be prone to overestimating the effect size on very rare occasions; this happened on a handful of individual replicates when studies were small and there was a high degree of variation among study SDs (in our simulation, all SDs are on the same scale so for in real dataset, it is more meaningful to look at differences in CVs among studies). Method 1B may underestimate total heterogeneity and overestimate the width of 95% CIs, although only when the heterogeneity is low. Finally, Method 1B was prone to overestimating the ICC for study-level effects, although again this only occurred when heterogeneity was low. However, almost all meta-analyses in ecology and evolution have substantial total heterogeneity ([Senior *et al.* 2016](#_ENREF_26)). Therefore, we recommend the use of Method 1B over other methods and this is the case for even with no missing SDs (see more in section ‘Discussion’ below).

# IMPLIMENTATION

## Extending the proposed methods to more complex situations

In many ecological meta-analyses, meta-analytic models are often made more complex by the need to account for phylogenetic relatedness when a meta-analytic dataset contains multiple species. Moreover, almost all meta-analytic studies test a number of moderators (predictors) to explain heterogeneity among effect sizes (i.e., meta-regression). Furthermore, effect sizes from the same studies can be correlated at the level of sampling error (e.g., the same individuals are used to calculate 2 effect sizes, see Noble et al. ([2017](#_ENREF_22)). We can write a meta-analytic model which can accommodate these three points above, as follows ([Nakagawa & Santos 2012](#_ENREF_21); [Cinar *et al.* 2022](#_ENREF_2)):

where is the *k*th moderator’s value and is the regression coefficient of the *k*th moderator (*h* = 1, 2, …, *Nmod*, the number of moderators), *ah* is the phylogenetic effect for the *h*th species, considered multivariate normally distributed with a covariance of (**A** is a correlation matrix derived from a phylogeny; see [Hadfield & Nakagawa 2010](#_ENREF_7)); *qh* is the non-phylogenetic effect for the *h*th species, distributed with the variance of (*h* = 1, 2, …, *Nspecies*, the number of species, which is different from *Nstudy* > *Neffect-size*); and the other notations are the same as above. **M**\* is a variance-covariance matrix of the sampling variance, which may result, say, when effect sizes share a common-control (e.g., Noble et al. 2017); for example, when we have 5 cases from 3 studies, **M**\* can be written as:

where and are the sampling variances for the 1st and 2nd effect sizes from the same study, and and are the co-variances between the two effect sizes (the 1st study), , and come from the same study (the 2nd study; if we want to make this equation similar to Equation 12, then and ), and is the sampling variance of the 5th effect size from another study. The correlation needs to be provided, but can often be assumed to be 0.5 or 0.8 ([Noble *et al.* 2017](#_ENREF_22); for a formula for the direct estimation of the sampling covariance for lnRR, see [Lajeunesse 2011](#_ENREF_13)).

In reality, constructing **M\*** may be as challenging as just doing MI for many ecologists, because the actual value of is often unknown ([Noble *et al.* 2017](#_ENREF_22)). Fortunately, Hedges *et al*. ([2010](#_ENREF_9)) derived the robust variance estimator (RVE), which bypasses these challenges by estimating from the data. By using RVE we need only construct **M**, rather than **M\*** (see also [Tipton 2013](#_ENREF_29)). We show an implementation of this procedure in Supporting Information, using the *clubSandwitch* package ([Pustejovsky 2017](#_ENREF_24)), which implements the RVE method with a multilevel meta-analysis in *metafor*.

## Worked examples 1

Here we provide a pair of worked examples to demonstrate the application of the methods described. Our first example is from McDonald et al. (2019). McDonald et al. (2019) studied the effects of strategic-rest grazing (SRG) regimes on both ungrazed and constantly grazed (CG) systems. They looked at number of different ecological outcomes, and here we focus on their data on the effects of SRG vs CG on biomass; this dataset contains 173 effect sizes from 67 studies. In their original analysis McDonald et al. (2019) find that the biomass of CG systems is significantly reduced relative to that SRG, but do not report the total heterogeneity. The dataset contains two dimensions of non-independence that are common to eco-evolutionary meta-analyses; 1) multiple effect sizes per study and 2) several effect sizes within the same study are computed as relative to the same control group (sometimes termed ‘stochastic dependency’; REF to Noble et al. and also Gleser & Olkin).

Of the 173 effect sizes in the biomass dataset, 35.8% have missing SD data. Where missing, SDs were missing for both the CG and SRG treatment groups in the effect size. In their original analyses McDonald et al. (2019) handle these missing data by calculating the average CV from all studies without missing data. They then use the reported mean value for studies with missing SDs coupled with the average CV to impute the missing SD value. This method is similar to single imputation of missing SDs, by predicting their value from a mean-SD linear regression. Non-independence was handled by McDonald et al. using MLMA, which included a variance-covariance matrix to account for stochastic dependency.

In table 1 we present the results of re-analysis of the biomass data from McDonald et al. (2019), again using MLMA, but with the four different methods to handles missing SDs. For reference we also include the results of a complete cases analysis where studies with missing SDs have been excluded. The effect sizes for the different methods are all very similar, although the CI for the complete cases analysis is wider than for those that include studies with missing SDs. Method 1B estimated slightly lower heterogeneity than the other methods.

Table 1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| aMethod | lnRR | CI Lower | CI Upper | *τ* |
| Complete Cases | -0.20 | -0.31 | -0.09 | 0.41 |
| Method 1A | -0.21 | -0.30 | -0.13 | 0.40 |
| Method 1B | -0.21 | -0.29 | -0.13 | 0.38 |
| Method 2 | -0.21 | -0.30 | -0.13 | 0.42 |
| Method 3 | -0.21 | -0.29 | -0.12 | 0.41 |

**Worked example 2: Quantifying the effects of competition between herbivorous insect species**

Our second example features data from Bird et al. (2019), who conducted a meta-analysis exploring the impacts of competition on herbivorous insect fitness when occupying the same host plant with another species or in isolation. Building on work by Kaplan & Denno (2007), they collected data on a series of fitness measurements (e.g., abundance, body size, development time, fecundity; see Table 1 in Bird et al. 2019) and explored the overall impacts of competition on the various fitness measures independently and in composite analyses. Bird et al. (2019) also tested the importance of a series of moderators they predicted would impact the magnitude of competition between species including population density, phylogenetic distance, diet breadth and spatial and temporal separation. A phylogeny was constructed using DNA sequence data and this gene tree was used to control for phylogenetic non-independence within analyses.

For demonstration purposes, we focus on a subset of fitness data, abundance, and use a simple multilevel meta-analytic model to estimate the overall impact of competition on focal insect fitness (i.e., intercept or overall meta-analytic mean) while controlling for phylogeny, research group, and research year (as per the analysis by Bird et al. 2019). Our use of log response ratio (lnRR) meant that we could only use a subset of abundance data from Bird et al. (2019) because of lnRR requiring ratio scale data. In addition, the ratio of minimum to largest sampling error variance calculated from the raw data was high suggesting some errors in the original published papers. To avoid model convergence issues we excluded these data and used a sample comprised of 293 effect sizes across 67 unique focal insect species with known phylogenetic relationships. We then introduced missing data at the paper level so that ~20% of papers had effect sizes with missing SD in the control and experimental treatment; a scenario that is typical of many meta-analyses.

An analysis of these data applying the different methods compared to the whole data is provided in Table 2. We can see that the complete case scenario (excluding all data with missing SD) results in slightly larger confidence intervals and a slight reduction in the meta-analytic mean effect size. Method 1B, 2 and 3 all suggest the overall meta-analytic is slightly larger and result in greater precision around this estimated effect size. Using this example, we show how each approach is implemented in the supplement.

**Table 2**: Re-analysis of a subset of data from Bird et al. 2019 using the methods we propose to deal with missing SD data estimating the overall effects of competition on focal insect abundance. All models controlled for phylogeny, research group and research year as per Bird et al. 2019. We show the whole (complete data), complete case (missing SD excluded) and each of the methods (Method 1A, B, 2, 3) for imputing missing SD.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Method** | **Est.** | **SE** | **95% LCI** | **95% UCI** |
| Whole Data | 0.338 | 0.354 | -0.356 | 1.031 |
| Complete Case | 0.272 | 0.364 | -0.440 | 0.985 |
| Method 1A | 0.360 | 0.346 | -0.317 | 1.038 |
| Method 1B | 0.365 | 0.186 | 0.001 | 0.730 |
| Method 2 | 0.391 | 0.133 | 0.130 | 0.653 |
| Method 3 | 0.328 | 0.149 | 0.036 | 0.621 |

# DISCUSSION

In this study, we have developed new methodological procedures that allows the inclusion of studies that do not report SDs in meta-analyses of lnRR. Our simulation has suggested the best performance was achieved by Method 1B that uses the average CV2 to calculate lnRR’s point estimates and sampling variances for all effect sizes regardless of missingness in SD. In terms of implementation, this is the easiest method among all (see Supporting Information). Further, we were surprised to see Method 1B (along with Method 2) outperforms the analysis with full data. This is especially so given Method 1B along with other new proposed methods use ‘single imputation’ rather than ‘multiple imputation’, and theatrically speaking, analysis with single imputation should always fare worse than analysis with full data ([Nakagawa & Freckleton 2008](#_ENREF_18); [Nakagawa 2015](#_ENREF_17); [van Buuren 2018](#_ENREF_30); [Kambach *et al.* 2020](#_ENREF_11)).

Yet, given the simulation work by Doncaster and Spake ([2018](#_ENREF_3)), we might have expected Method 2 would do well (see also [Lin & Aloe 2021](#_ENREF_16)). It is important to note that our simulation work clearly differed at least in two respects. First, Doncaster and Spake never tested how their method fare with missing data. Second, their simulation was restricted to non-multilevel models, which we believe is non-realistic models. The main reason Method 1B and Doncaster and Spake’s procedure perform well (i.e. using Equation 3 rather than Equation 2) is because SDs from studies are often estimated poorly with small sample size (i.e., replicates per effect size), leading to poor estimates of sampling variances in lnRR. However, by using a pooled CV2, estimates of sampling variance improves, which lead to better estimation of almost all parameters in our simulation, which has been demonstrated by Doncaster and Spake ([2018](#_ENREF_3)) as well as our simulation. This improved estimation via CV2 is the very reason that these methods perform better than analysis with full data, despite the use of ‘single imputation’.

As mentioned earlier, there are two occasions where Method 1B performs worse than other methods. Firstly, it is when CVs are very different between studies and the number of studies are relatively small (e.g., *K* < 20). In such cases, we recommend conducting sensitivity analysis using Method 1A (or alternatively Method 3, but more difficult to implement) to check meta-analytic results are consistent between Method 1A and Method 1B. Also, it may be a good practice to check how heterogenous between studies, using the meta-analysis of lnCVR (log CV ratio; [Nakagawa *et al.* 2015](#_ENREF_20); [Senior *et al.* 2020](#_ENREF_27)). Large variation between-study CVs would violate our assumption that CV stays fairly constant for our new methodologies (cf. [Nakagawa *et al.* 2015](#_ENREF_20)) although our simulation shows this assumption is less important when we have relative large number of studies (e.g., *K* > 20). Secondly, it is when there is very low total heterogeneity (*τ*2 = + , which usually translates to low *I*2; see [Higgins *et al.* 2003](#_ENREF_10); [Nakagawa & Santos 2012](#_ENREF_21); also see [Borenstein *et al.* 2017](#_ENREF_1)). As eluded earlier, in meta-analyses in ecology and evolution, heterogeneity is often high. Indeed, Senior et al ([2016](#_ENREF_26)) have shown that on average, ecological and evolutionary meta-analyses have high heterogeneity with *I*2 around 90%. Nonetheless, when there is very low heterogeneity (or heterogeneity test *Q* is non-significant), we also recommend using Method 1B for sensitivity analysis. Taken together, it may be advisable to conduct meta-analysis using Method 1A and 1B in tandem, as it is straightforward to do both (see Supplementary Information).

Finally, emphasize that is an alternative method and is not a replacement for multiple imputation. Indeed, if there are missing values in moderators, the only way to deal with such missing data is to use multiple imputation. However, we believe, our proposed method (i.e., Method 1B) is easier to implement and readily extendable to complex models, as we showed above, especially when we do not have any missing data in moderators or are less concerned about missing data in moderators. We wish meta-analysts in ecology and evolution will start adopting this new method (Method 1B) to improve their meta-analytic estimation, while checking the sensitivity of results using Method 1A. We also remind that we have provided easy-to-follow examples and the function to calculate pooled CV2 to facilitate adoption in Supporting Information.

# ACKNOWLEDGEMENTS

SN, …. were supported by an ARC (Australian Research Council) Discovery grant (DP200100367).

# AUTHORS’ CONTRIBUTIONS

# DATA AVAILABILITY

# REFERENCES

**Figure Legends**

**Figure 1**

A. Violin plot showing the distribution of mean bias in the estimated effect under each simulated condition as a function of the method used to handle missing data (distribution assuming full data shown for reference). B. Pairwise correlations between degree of bias under each simulated condition for each method. C. Distribution of the difference between methods 1.1 and 1.2 in the absolute degree of bias under each condition (positive values indicate greater mean bias under method 1.1). D. Violin plot showing the distribution of range bias (log10 transformed) in the estimated effect under each simulated condition as a function of the method used to handle missing data. E. Violin plot showing the distribution of range bias (log10 transformed) in the estimated effect under each simulated condition as a function of the degree of heterogeneity in SDs among studies and typical size of studies in the meta-analysis.

**Figure 2**

A. Violin plot showing the distribution of coverage of 95% CIs under each simulated condition as a function of the method used to handle missing data (distribution assuming full data shown for reference). B. Violin plot showing the distribution of coverage under each simulated condition as a function of the simulated level of total heterogeneity and the ICC for study using method 1.1 to handle missing SDs. C. Violin plot showing the distribution of coverage under each simulated condition as a function of the simulated level of total heterogeneity and the ICC for study using method 1.2 to handle missing SDs.

**Figure 3**

A. Violin plot showing the distribution of mean bias in the estimated heterogeneity under each simulated condition as a function of the method used to handle missing data (distribution assuming full data shown for reference). Bias in heterogeneity is calculated as the log ratio of the estimated and parametrised value. B. Box plot showing the mean bias in estimated heterogeneity under each simulated condition as a function of the method used to handle missing data (colours as in panel A), and the simulated level of heterogeneity. C. Violin plot showing the distribution of the mean bias in the estimated ICC for study under each simulated condition as a function of the method used to handle missing data. Bias in the ICC was calculated as the difference between the estimated and parameterised value. D. Violin plot showing the distribution of the mean bias in the estimated ICC for study under each simulated condition as a function of the simulated level of total heterogeneity and the ICC for study using method 1.2 to handle missing SDs.

Table S1. Variables/parameters in simulation.

|  |  |  |
| --- | --- | --- |
| Variable (Notation) | Description and details | Value(s) |
| % Studies Missing SD | Percentage of studies that have missing SDs | 5, 15, 25, 35, 45 or 55 |
| Overall Effect Size (*θ*) | The overall mean lnRR effect size | 0.3 |
| Number of Studies (*K*) | Total number of studies within the meta-analytic dataset | 12, 30, 100 |
| Number of Effect Sizes in Study (*L*) | The number of effect sizes within a study. Values for each study were randomly distributed using a double Poisson distribution | Random with mean of 5, and dispersion 1.5 |
| Total Heterogeneity (*τ*2) | The total heterogeneity among effect sizes | 0.09 |
| Intra-Class Correlation for Study (ICC*s*) | The proportion of total heterogeneity that is attributable to study-level effects | 0 or 0.5 |
| Sample Size in Study (*N*) | The sample size of groups within studies; individual sample sizes were randomly distributed using a double Poisson distribution | Random with mean of either 5 or 30, and dispersion 1.5 |
| Standard Deviation in Study (*S*) | The within-study SDs. Individual within-study SDs were randomly distributed following a Gamma distribution | Random with mean 15 and SD of either 10-10, 3.75 or 7.5 |

**Figure 1**



**Figure 2**



**Figure 3**



1.

Borenstein, M., Higgins, J.P.T., Hedges, L.V. & Rothstein, H.R. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Res Synth Methods*, 8, 5-18.

2.

Cinar, O., Nakagawa, S. & Viechtbauer, W. (2022). Phylogenetic multilevel meta-analysis: A simulation study on the importance of modelling the phylogeny. *Methods Ecol Evol*, 13, 383-395.

3.

Doncaster, C.P. & Spake, R. (2018). Correction for bias in meta-analysis of little-replicated studies. *Methods Ecol Evol*, 9, 634-644.

4.

Ellington, E.H., Bastille-Rousseau, G., Austin, C., Landolt, K.N., Pond, B.A., Rees, E.E. *et al.* (2015). Using multiple imputation to estimate missing data in meta-regression. *Methods Ecol Evol*, 6, 153-163.

5.

Fletcher, D. & Dixon, P.M. (2012). Modelling data from different sites, times or studies: weighted vs. unweighted regression. *Methods Ecol Evol*, 3, 168-176.

6.

Gurevitch, J., Koricheva, J., Nakagawa, S. & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature*, 555, 175-182.

7.

Hadfield, J. & Nakagawa, S. (2010). General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology*, 23, 494 - 508.

8.

Hedges, L.V., Gurevitch, J. & Curtis, P.S. (1999). The Meta-Analysis of Response Ratios in Experimental Ecology. *Ecology*, 80, 1150-1156.

9.

Hedges, L.V., Tipton, E. & Johnson, M.C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Res Synth Methods*, 1, 39-65.

10.

Higgins, J.P.T., Thompson, S.G., Deeks, J.J. & Altman, D.G. (2003). Measuring inconsistency in meta-analyses. *Brit Med J*, 327, 557-560.

11.

Kambach, S., Bruelheide, H., Gerstner, K., Gurevitch, J., Beckmann, M. & Seppelt, R. (2020). Consequences of multiple imputation of missing standard deviations and sample sizes in meta-analysis. *Ecol Evol*, 10, 11699-11712.

12.

Koricheva, J. & Gurevitch, J. (2014). Uses and misuses of meta-analysis in plant ecology. *J Ecol*, 102, 828-844.

13.

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

14.

Lajeunesse, M.J. (2013). Recovering missing or partial data from studies: a survey. In: *The handbook of meta-analysis in ecology and evolution* (eds. Koricheva, J, Gurevitch, J & Mengersen, K). Princeton University Press Princeton, pp. 195-206.

15.

Lajeunesse, M.J. (2015). Bias and correction for the log response ratio in ecological meta-analysis. *Ecology*, 96, 2056-2063.

16.

Lin, L.F. & Aloe, A.M. (2021). Evaluation of various estimators for standardized mean difference in meta-analysis. *Stat Med*, 40, 403-426.

17.

Nakagawa, S. (2015). Missing data: mechanisms, methods and messages. In: *Ecological statistics: contemporary theory and application* (eds. Fox, GA, Negrete-Yankelevich, S & Sosa, VJ). Oxford University Press Oxford, pp. 81-105.

18.

Nakagawa, S. & Freckleton, R.P. (2008). Missing inaction: The dangers of ignoring missing data. *Trends in Ecology & Evolution*, 23, 592-596.

19.

Nakagawa, S. & Freckleton, R.P. (2011). Model averaging, missing data and multiple imputation: a case study for behavioural ecology. *Behav Ecol Sociobiol*, 65, 103-116.

20.

Nakagawa, S., Poulin, R., Mengersen, K., Reinhold, K., Engqvist, L., Lagisz, M. *et al.* (2015). Meta-analysis of variation: ecological and evolutionary applications and beyond. *Methods Ecol Evol*, 6, 143-152.

21.

Nakagawa, S. & Santos, E.S.A. (2012). Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology*, 26, 1253-1274.

22.

Noble, D.W.A., Lagisz, M., O'Dea R, E. & Nakagawa, S. (2017). Nonindependence and sensitivity analyses in ecological and evolutionary meta-analyses. *Molecular Ecology*, 26, 2410-2425.

23.

Osenberg, C.W., Sarnelle, O. & Cooper, S.D. (1997). Effect size in ecological experiments: The application of biological models in meta-analysis. *Am Nat*, 150, 798-812.

24.

Pustejovsky, J. (2017). clubSandwich: Cluster-robust (sandwich) variance estimators with small-sample corrections. R package version 0.2. 3. *R Found. Stat. Comput., Vienna*.

25.

Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley, New York ; Chichester.

26.

Senior, A.M., Grueber, C.E., Kamiya, T., Lagisz, M., O'Dwyer, K., Santos, E.S.A. *et al.* (2016). Heterogeneity in ecological and evolutionary meta-analyses: its magnitude and implications. *Ecology*, 97, 3293-3299.

27.

Senior, A.M., Viechtbauer, W. & Nakagawa, S. (2020). Revisiting and expanding the meta-analysis of variation: The log coefficient of variation ratio. *Res Synth Methods*, 11, 553-567.

28.

Spake, R., Mori, A.S., Beckmann, M., Martin, P.A., Christie, A.P., Duguid, M.C. *et al.* (2021). Implications of scale dependence for cross-study syntheses of biodiversity differences. *Ecol Lett*, 24, 374-390.

29.

Tipton, E. (2013). Robust variance estimation in meta-regression with binary dependent effects. *Res Synth Methods*, 4, 169-187.

30.

van Buuren, S. (2018). *Flexible Imputation of Missing Data, Second Edition*. 2nd edn. CRC Press, Boca Raton.

31.

van Buuren, S. & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45, 1-67.

32.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1-48.