Ecology, 96(8), 2015, pp. 2056–2063

Reports

Bias and correction for the log response ratio in ecological meta-analysis

Marc J. Lajeunesse[1]

Department of Integrative Biology, University of South Florida, 4202 East Fowler Avenue, Tampa, Florida 33620 USA

Abstract. Ecologists widely use the log response ratio for summarizing the outcomes of studies for meta-analysis. However, little is known about the sampling distribution of this effect size estimator. Here I show with a Monte Carlo simulation that the log response ratio is biased when quantifying the outcome of studies with small sample sizes, and can yield erroneous variance estimates when the scale of study parameters are near zero. Given these challenges, I derive and compare two new estimators that help correct this small-sample bias, and update guidelines and diagnostics for assessing when the response ratio is appropriate for ecological meta-analysis. These new bias-corrected estimators retain much of the original utility of the response ratio and are aimed to improve the quality and reliability of inferences with effect sizes based on the log ratio of two means.

Key words: *delta method; effect size estimator; large sample approximation; linearity of expectation; log ratio; meta-analysis; research synthesis; small sample bias.*

Introduction

Effect sizes offer a practical way to summarize the magnitude and direction of research outcomes, and when they are combined and compared with meta-analysis, they also provide the building blocks for research synthesis (Hunter and Schmidt 1990). One mainstay for summarizing the outcomes of ecological experiments is the log response ratio (RR, sometimes also lnR or lnRR [Curtis and Wang 1998, Hedges et al. 1999]). This effect size metric quantifies the results of an experiment as the log-proportional change between the means ($\bar{X}$) of a treatment (T) and control (C) group

$$\mathrm{RR} = \ln(\bar{X}_\mathrm{T}/\bar{X}_\mathrm{C}). \quad (1)$$

Sampling error plays an important role in introducing variability in experimental outcomes (Hunter and Schmidt 1990), and therefore a central feature of RR is its variance

$$\mathrm{var}(\mathrm{RR}) = \frac{(\mathrm{SD}_\mathrm{T})^2}{N_\mathrm{T}\bar{X}_\mathrm{T}^2} + \frac{(\mathrm{SD}_\mathrm{C})^2}{N_\mathrm{C}\bar{X}_\mathrm{C}^2}. \quad (2)$$

Here, within-study parameters like the sample sizes (N) and standard deviations (SD) are used to help quantify the sampling variability in RR, and as a hallmark of meta-analysis, this variability gets converted into weights to help minimize the influence of studies with low statistical power when analyzing and pooling multiple study outcomes (Hedges and Olkin 1985). Experimental design can also introduce variability, and recently new variances and covariances for RR have been developed to help reduce bias when dependent information is aggregated within and among studies (Lajeunesse 2011).

Other sources of variability and bias are known for the response ratio but their extent and impact are far less understood. For example, inaccuracies in RR will occur when effect sizes are estimated from experiments with small sample sizes (Hedges et al. 1999). Surprisingly, this small-sample issue remains unaddressed; despite the fact that a large portion of ecological experiments will have small N (Jennions and Møller 2003), and that other effect size metrics are routinely corrected for this type of bias. Examples of corrected metrics include Hedges' d (e.g., Hedges 1981, 1982) and the correlation coefficient (e.g., Olkin and Pratt 1958). Problems also exist with the sampling distribution of RR and how the variance estimator of RR (Eq. 2) approximates this distribution. Under certain conditions, such as when one of the control or treatment means is near zero, Hedges et al. (1999) suggested that the distribution of RR is no longer normal. Given that the variance of RR is an asymptotic normal approximation (Lajeunesse 2011), any deviation from normality will destabilize its utility as weights for meta-analysis. To minimize this problem, Hedges et al. (1999) offered a simple heuristic adapted from Geary (1930) to help assess when RR will be normal. However to date, very few meta-analyses in ecology or elsewhere have used this diagnostic to assess the accuracy of effect sizes approximated with RR.

Given that nearly half of all published meta-analyses in ecology use RR to quantify experimental outcomes (Nakagawa and Santos 2012, Koricheva and Gurevitch 2014), a critical revision of this metric is needed to help address issues of bias and variability, and to help improve the reliability of RR for estimating effect sizes.

Manuscript received 16 December 2014; revised 6 April 2015; accepted 8 April 2015. Corresponding Editor: B. D. Inouye.

[1] E-mail: lajeunesse@usf.edu

Here, I first establish with a Monte Carlo experiment the extent of the small-sample bias, and then identify conditions under which RR is no longer reliable for effect size estimation. I then develop two new estimators that help correct RR for small-sample bias, as well as new variance estimators to improve the estimation of its sampling distribution. Finally, I evaluate the small-sample performance and distribution properties of these new estimators relative to RR, and provide new guidelines for when they should be applied as effect sizes for quantifying experimental outcomes.

Exploring Bias in the log Response Ratio and its Variance

Monte Carlo simulation methods

The goal of this simulation is to determine the accuracy of RR when estimating a true population effect size (λ), and to assess when it is appropriate to use var(RR) to approximate the variance of RR. The true population effect, or $\lambda = \ln(\mu_T/\mu_C)$, is the expected population value of the log-proportional change between two independent and normally distributed ($\mathcal{N}$) population means μ_T and μ_C with variances σ_T^2 and σ_C^2. Here, accuracy is evaluated as bias(RR) = $\mathbb{S}$(RR) − λ; which is the difference between $\mathbb{S}$(RR), the mean from k = 100 000 randomly simulated ($\mathbb{S}$) response ratios for a given N, and the true underlying effect λ. Random response ratios can be generated by taking the average ($\bar{X}$) and standard deviation (SD) of N random samples from each μ_T and μ_C, separately. Each random sample (X) can be summarized as $X_i \sim \mathcal{N}(\mu, \sigma^2)$ with $i = 1,\ldots,N$. However, when μ_T and μ_C are near zero, random samples from μ_T and μ_C can yield negative $\bar{X}_T$ and $\bar{X}_C$. Negative values cannot be log transformed; making the calculation of RR impossible. Given that the goal is to enumerate a full range of μ_T and μ_C values, including those near zero when RR is predicted to deviate from normality (Hedges et al. 1999), an alternative way to reliably generate non-negative $\bar{X}$ is to sample from a lognormal distribution ($\ln \mathcal{N}$) as follows: $X_i \sim \ln \mathcal{N}(\ln[\mu] - 0.5\ln[1 + \sigma^2/\mu^2], \ln[1 + \sigma^2/\mu^2])$. Here, taking the average and variance of samples sizes N from this distribution will yield $\bar{X}_T$ and $\bar{X}_C$ with the desired μ and σ^2 values, as well as the non-negative properties necessary for calculating RR. Based on this distribution, a full range of RR and var(RR) were randomly generated using the $\bar{X}_T$ and $\bar{X}_C$ (and their SDs) calculated from N samples of μ spanning from 0.001 to 8 at 0.25 increments. This range of μ was explored for all possible paired values of each control and treatment groups. I also assumed unit variances for μ and equal N for both treatment and control groups, and surveyed a range of small sample sizes typical to ecological studies: $N_T + N_C = 4$, 8, 16, and 32.

The variance and skewness of the simulated RR were also calculated to determine how well var(RR) approximates the sampling variance of RR, and when this distribution is expected to deviate from normality. Anything above two standard errors of skewness was deemed non-normal, and following Tabachnick and Fidell (1996), this non-normal threshold for my simulations was approximated as $2\text{SE} = 2\sqrt{6/k} = 0.01549$. Bias of the variance estimator was estimated as the difference between the mean of the simulated variance estimator $\mathbb{S}$[var(RR)] and the observed variance $\mathbb{S}_{\sigma^2}$ of the simulated RR: bias[var(RR)] = $\mathbb{S}$[var(RR)]. The observed variance, or $\mathbb{S}_{\sigma^2}$(RR), of the simulated RR represents the best possible prediction of the sampling variability in RR under controlled conditions. Further details on the difficulty in characterizing the sampling distribution of RR are found in the Appendix, and the simulation R script is presented as a Supplement.

Monte Carlo results on the small-sample performance of RR *and* var(RR)

When the treatment mean ($\bar{X}_T$) is larger than the control mean ($\bar{X}_C$), RR will estimate a positive effect; when $\bar{X}_C$ is larger than $\bar{X}_T$ the effect will be negative. Further, when $\bar{X}_T$ and $\bar{X}_C$ are close to one another the response ratio will be near zero and estimate a null effect. However, when the treatment and control means have small sample sizes (see ranges of N in Fig. 1), RR will overestimate the expected effect (i.e., have a positive bias) when the treatment mean is larger than the control, or will underestimate the effect and have a negative bias when $\bar{X}_C$ is larger than $\bar{X}_T$. This bias persists even at moderately large sample sizes ($N = 32$; Fig. 1), but appears negligible when RR is predicted to estimate a null effect (e.g., when $\bar{X}_T$ and $\bar{X}_C$ are close to one another). The log response ratio (Eq. 1) is therefore a consistent estimator given that its bias gets minimized at large sample sizes. More crucially, the overall magnitude of bias is dependent on whether at least one of the two means is near zero while the other is non-zero (Fig. 1). Therefore log-ratio effect sizes estimated with RR are at the greatest risk of bias when: (1) the means have small sample sizes, (2) the two means are not close to one another (e.g., when the effect is not null), and (3) at least one of the control and treatment means is near zero.

The variance estimator of RR (Eq. 2) also does not approximate the predicted variability in RR very well. In general, var(RR) will underestimate the variance of RR (Appendix: Fig. A1), and the magnitude of this underestimation will increase as the treatment and control means approach zero. This poor performance is not unusual given that the distribution of ratios and log ratios are challenging to approximate (see Appendix), and that var(RR) is an approximation of that variability (Hedges et al. 1999). This means that var(RR) will conditionally only perform well when the distribution of RR is normal. However, the distribution of RR is consistently skewed (Appendix: Fig. A2), either to the right when the control mean is larger than the treatment mean, or to the left (negative skewness)

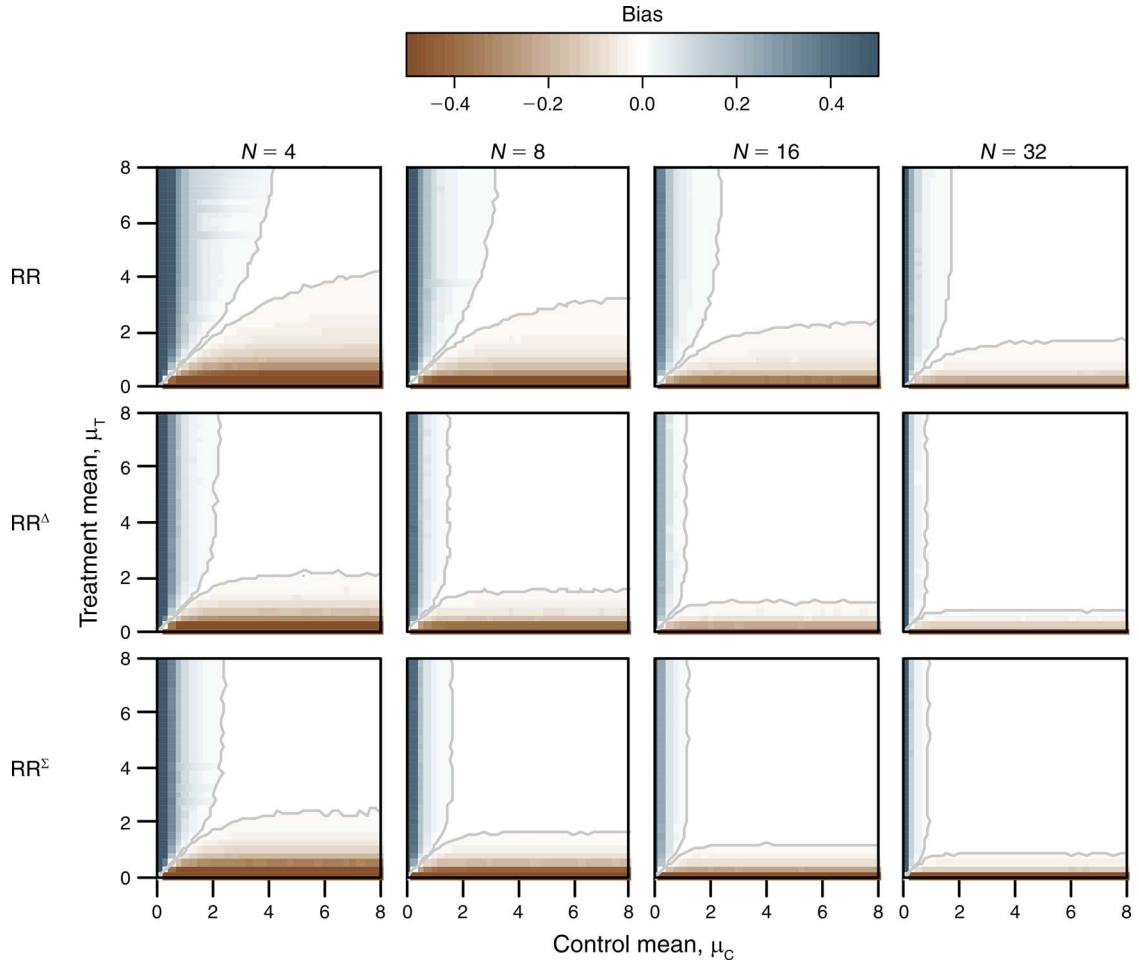

FIG. 1. Results from a Monte Carlo simulation exploring bias in the log response ratio (RR) relative to the two new bias-corrected estimators, RR^{Δ} and RR^{Σ}. Positive bias is emphasized in blue (over-estimation of the effect) and negative bias (under-estimation) in red. Contour lines in light gray indicate effect sizes with bias = |0.01|; values below this range are in white. Sample-means for treatment (T) and control (C) groups were randomly generated from population means (μ) ranging from zero to eight, and were estimated using a range of small to medium sample sizes ($N = N_T + N_C$). Monte Carlo results on the skewness of these estimators, as well as the bias in their variance estimators, are found in the Appendix.

when the opposite is true (i.e., $\bar{X}_T > \bar{X}_C$). Therefore the greatest risk of applying approximations to the variance of RR will occur when at least one of two means is near zero.

Despite this underestimation problem with var(RR), its overall behavior still achieves a weighting scheme useful for meta-analysis. Effect sizes with greater predicted sampling error will get proportionally down-weighted more heavily than those will less predicted sampling error. However, when effect sizes are near the risk area for RR (i.e., when means are close to zero) their weights will be disproportionally small. This is due to the rapid increase in var(RR) within the problematic ranges of μ (Appendix: Fig. A1). Further, this is not an ideal property for weights, as it will introduce variability among var(RR) that is greater than predicted by chance.

Theory and Development of log Response Ratio Estimators

As identified with Monte Carlo simulations, RR is a biased estimator of the log ratio of the treatment and control population means at small sample sizes. The expected direction and magnitude of this bias was determined by taking the difference between the mean of randomly simulated response ratios for a given N and the true underlying effect λ. However, the expected value ($\mathbb{E}$) of the simulated mean with this small-sample bias can also be estimated directly. A direct approach is more practical for calculating effect sizes and developing corrections. Below I derive two ways to directly calculate this expected mean and variance, and apply these to correct the response ratio. These new corrected estimators will be referred to as RR^{Δ} and RR^{Σ}. Here, the superscript Δ indicates an adjustment to RR based on the Delta method (Ver Hoef 2012), and Σ an adjustment

that applies the Linearity of Expectation rule to the sum (Σ) of two normally distributed variables that have been log transformed. Presented below are abbreviations of their derivations. Complete derivations are found in the Appendix.

New estimator based on the Delta method: RR^{Δ}

Typically for meta-analysis, effect size metrics like RR, Hedges' *d*, and the odds ratio use only first-order expansions to approximate asymptotic sampling distributions (Hedges 1981, Lajeunesse 2011). However, calculating higher-order expansions can also be useful given that they can be used to adjust or correct bias in the "naïve" effect size estimator (such as RR). Here, I first show how the mean (Eq. 1) and variance (Eq. 2) of the original RR can be approximated using the multivariate Delta method, and then extend this approach to obtain higher-order terms needed for deriving corrections.

Following Stuart and Ord (1994), the expectation of the simplest estimator of λ based on the first-order Taylor expansion around the population means μ_T and μ_C of $\lambda = \ln(\mu_T/\mu_C)$ is approximately

$$\mathbb{E}(\mathrm{RR}) \approx \lambda + \mathbf{J}^{\top}(\mathbf{x} - \boldsymbol{\mu}) + \varepsilon_{\mathrm{RR}} \tag{3}$$

where the superscript T indicates the transposition of a matrix, ε_{RR} is the remainder (i.e., the ignored higher-order Taylor expansions), $\boldsymbol{\mu}$ is a column vector of the population means μ_T and μ_C (e.g., $\boldsymbol{\mu}^{\top} = [\mu_T, \mu_C]$, and $\mathbf{x}$ is a vector of the sample means $\mathbf{x}^{\top} = [\bar{X}_T, \bar{X}_C]$. Also included is a Jacobian vector ($\mathbf{J}$) containing all the first-order partial derivatives of each variable in λ (see Appendix). Solving Eq. 3, and noting that the expectation of $\bar{X} - \mu$ is zero at large sample sizes (e.g., when sampling error becomes negligible as assumed by the Law of Large Numbers; Stuart and Ord 1994), we get the original response ratio (see Appendix). In a parallel way, we can also apply the multivariate Delta method to approximate the variance of RR using the Law of Propagation of Variances equation:

$$\mathrm{var}(\mathrm{RR}) \approx \mathbf{J}^{\top}\mathbf{V}\mathbf{J} + \varepsilon_{\mathrm{var(RR)}} \tag{4}$$

where $\mathbf{V}$ is the variance–covariance matrix of μ_T and μ_C. Solving Eq. 4, as described in the Appendix, we get the original variance estimator.

However, for both the expected mean and variance of the log ratio (Eqs. 3 and 4, respectively), the remainder portions ε_{RR} and $\varepsilon_{var(RR)}$ of the Taylor expansions were ignored. Here a second-order portion of ε can be added to improve these estimators. The expectation of λ with a second-order Taylor expansion is

$$\mathbb{E}(\mathrm{RR}) \approx \lambda + \mathbf{J}^{\top}(\mathbf{x} - \boldsymbol{\mu}) + \underbrace{\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^{\top}\mathbf{H}(\mathbf{x} - \boldsymbol{\mu})}_{\text{second-order term}} + \varepsilon_{\mathrm{RR}} \tag{5}$$

and the expectation of its variance becomes

$$\mathrm{var}(\mathrm{RR}) \approx \mathbf{J}^{\top}\mathbf{V}\mathbf{J} + \underbrace{\frac{1}{2}\mathrm{tr}[\mathbf{H}(\mathbf{V}\mathbf{V})\mathbf{H}]}_{\text{second-order term}} + \varepsilon_{\mathrm{var(RR)}} \tag{6}$$

with tr indicating the trace of a matrix, and where $\mathbf{H}$ is a Hessian matrix of all the second partial derivatives of λ. These expectations can be used to adjust the RR and its variance as follows:

$$\mathrm{RR}^{\mathrm{adj}} = \mathrm{RR} - \mathrm{bias}(\mathrm{RR}) = \mathrm{RR} - [\mathbb{E}(\mathrm{RR}) - \lambda] \tag{7}$$

However, given that we do not know what λ will be, or the population parameters μ and σ^2, we can substitute the study sample statistics $\bar{X}$ and $(SD)^2$ to approximate these parameters. Given these terms, substituting the original RR as an estimate of λ, and using the expected mean of Eq. 5, the small-sample bias corrected estimator for λ based on the Delta method (Δ) becomes

$$\mathrm{RR}^{\Delta} = \mathrm{RR} + \frac{1}{2}\left[\frac{(\mathrm{SD_T})^2}{N_T\bar{X}_T^2} - \frac{(\mathrm{SD_C})^2}{N_C\bar{X}_C^2}\right]. \tag{8}$$

Likewise, applying Eq. 6 with the general form of Eq. 7, we get its adjusted variance

$$\mathrm{var}(\mathrm{RR}^{\Delta}) = \mathrm{var}(\mathrm{RR}) + \frac{1}{2}\left[\frac{(\mathrm{SD_T})^4}{N_T^2\bar{X}_T^4} + \frac{(\mathrm{SD_C})^4}{N_C^2\bar{X}_C^4}\right]. \tag{9}$$

New estimator based on the Linearity of Expectation rule: RR^{Σ}

The expected value of $\mathbb{E}(RR)$ can also be calculated using the Linearity of Expectation rule, which states that the expected value of a sum of random variables will equal the sum of their individual expectations. According to Stuart and Ord (1994), the individual expected values of μ_T and μ_C in terms of $\ln[\mu_T]$ and $\ln[\mu_C]$ will each have a mean of $\mathbb{E}(\ln[\mu]) = \ln[\mu] - 0.5\ln[1 + \sigma^2/(N\mu^2)]$ and variance of $\mathrm{var}(\ln[\mu]) = \ln[1 + \sigma^2/(N\mu^2)]$. Using the convenient expression of RR based the quotient rule of logarithms, the expected mean of RR becomes $\mathbb{E}(\mathrm{RR}) = \mathbb{E}(\ln[\mu_T]) - \mathbb{E}(\ln[\mu_C])$, which has an expected variance of $\mathrm{var}(\mathrm{RR}) = \mathrm{var}(\ln[\mu_T]) + \mathrm{var}(\ln[\mu_C])$. Finally, applying these expectations to Eq. 7 (as described in the Appendix), we get a new small-sample bias corrected estimator based on the Linearity of Expectation rule

$$\mathrm{RR}^{\Sigma} = \frac{1}{2}\ln\left[\frac{\bar{X}_T^2 + N_T^{-1}(\mathrm{SD_T})^2}{\bar{X}_C^2 + N_C^{-1}(\mathrm{SD_C})^2}\right] \tag{10}$$

which has an approximate variance of

$$\mathrm{var}(\mathrm{RR}^{\Sigma}) = 2 \times \mathrm{var}(\mathrm{RR}) - \ln\left[1 + \mathrm{var}(\mathrm{RR}) + \frac{(\mathrm{SD_T})^2(\mathrm{SD_C})^2}{N_T N_C \bar{X}_T^2 \bar{X}_C^2}\right]. \tag{11}$$

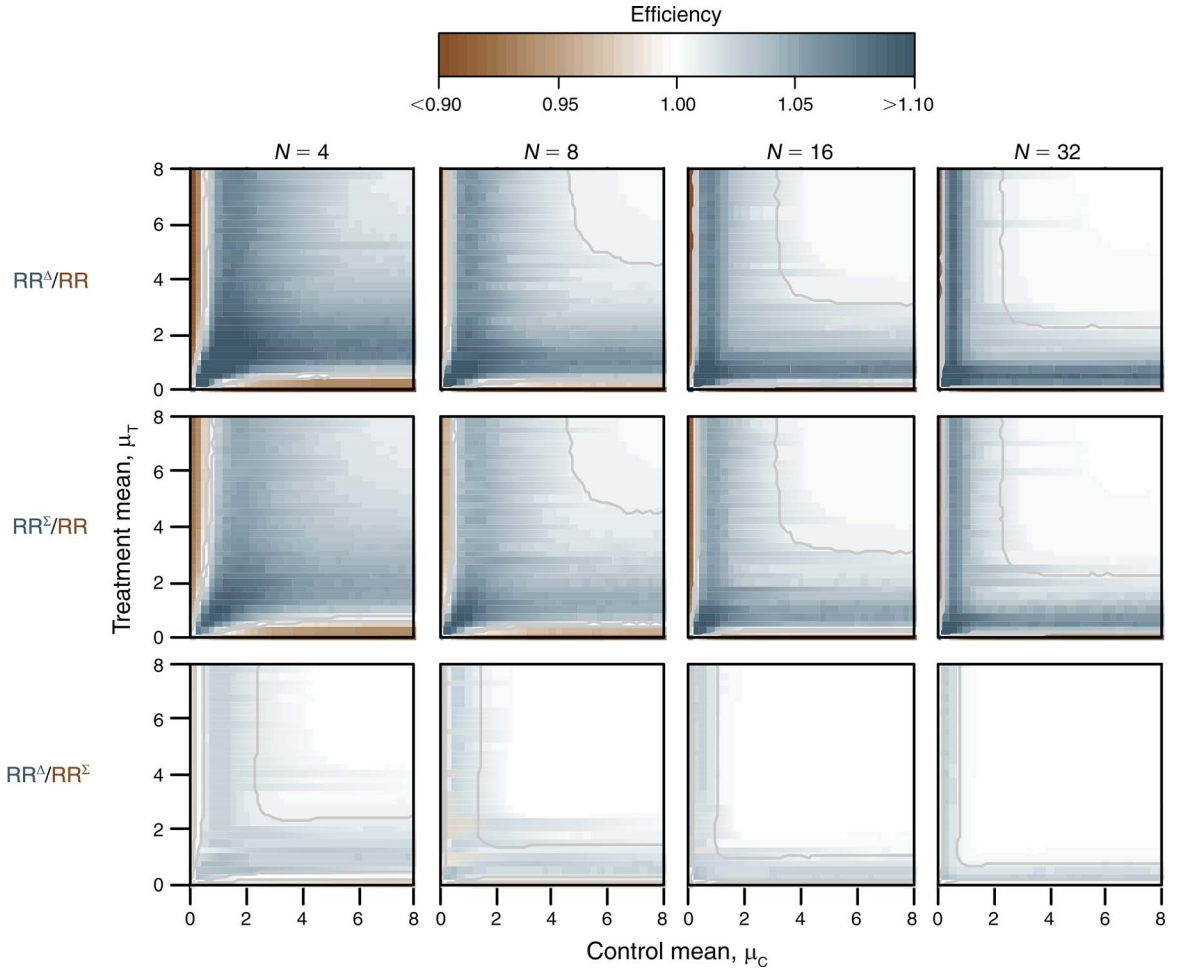

FIG. 2. Monte Carlo results comparing the relative efficiency of estimating effect sizes using the response ratio estimators RR, RR^{Δ}, and RR^{Σ}. Relative efficiencies greater than one (marked in blue) indicate that the numerator estimator is better at estimating log ratio effect sizes; efficiencies less than one (marked in red) indicate that the denominator estimator is a better effect sizes estimator. Contour lines in light gray emphasize when the relative efficiently between two estimators is less than 1%. Overall, both RR^{Δ} and RR^{Σ} perform better than RR; with RR^{Δ} having a slight advantage in estimating effect sizes over RR^{Σ}.

Comparing and ranking the performance of log response ratio estimators

With an aim to balance variability and bias, I used the ratio of the mean squared error (MSE) of each response ratio estimator ($\hat{\theta}$) to compare and rank their pairwise relative efficiency: $\text{eff}(\hat{\theta}_A, \hat{\theta}_B) = \text{MSE}(\hat{\theta}_A)/\text{MSE}(\hat{\theta}_B)$. If the relative efficiency (eff) is >1, then the numerator estimator ($\hat{\theta}_A$) has better MSE properties. Having better mean square error properties is advantageous given that it indicates a smaller combined variance and squared bias: $\text{MSE}(\hat{\theta}) = \sigma^2_{\mathbb{S}(\hat{\theta})} + [\text{bias}(\hat{\theta})]^2$.

Relative to the original response ratio, RR^{Δ} is 5–20% more efficient at estimating log ratio effects with small sample sizes (Fig. 2); the efficiency of RR^{Σ} marks a similar improvement of 6–18% over RR. When comparing the two bias-corrected estimators, RR^{Δ} is slightly more efficient with a 2% to 5% advantage over RR^{Σ} (Fig. 2). This improved efficiency of the two new estimators exists despite having slight skews similar to RR (Appendix: Fig. A2), and having variances that are comparably deficient as var(RR) when estimating the predicted variability of log ratios (Appendix: Fig. A1). Finally, all estimators are unreliable when at least one of the control and treatment mean is near zero.

Revisiting accuracy diagnostics for response ratio estimators

Given that RR and var(RR) are not accurate approximations to the distribution of RR when either the control or treatment mean is near zero (e.g., Fig. 1), Hedges et al. (1999) proposed a simple diagnostic to assess when they can provide correct effects for meta-analysis. Here, effect sizes are deemed valid and accurate approximations when the standardized mean of either the control or treatment group is >3.

TABLE 1. Illustrative examples of using log response ratio (RR) estimators to quantify experimental outcomes based on the means ($\bar{X}$), standard deviations (SD, in parentheses), and sample sizes (N) of control (C) and treatment (T) groups.

Study	Study outcomes						RR effect sizes		
	$\bar{X}_C$	N_C	Geary's test	$\bar{X}_T$	N_T	Geary's test	RR	RR^{Δ}	RR^{Σ}
Allouche and Gaudin (2001)	20.344 (2.257)	10	27.81	1.548 (1.305)	10	3.66	−2.576 [0.0723]	−2.541 [0.0748]	−2.542 [0.0748]
Appleton and Palmer (1988)	0.0438 (0.180)	20	**1.07**	0.159 (0.091)	20	7.72	1.289 [0.8608]	0.875 [1.2175]	0.991 [1.0932]
Black and Dodson (1990)	49.1 (10.752)	10	14.09	36.1 (18.974)	10	6.05	−0.308 [0.0308]	−0.297 [0.0311]	−0.297 [0.0311]
Turner (1997)	0.451 (0.111)	4	7.65	0.347 (0.174)	4	3.75	−0.262 [0.0780]	−0.238 [0.0801]	−0.239 [0.0800]
Turner and Montgomery (2003)	1.254 (1.063)	40	7.42	0.003 (0.202)	40	**0.09**	−6.035 [113.362]	50.628 [6536.844]	−3.675 [221.968]

Notes: All studies compare the outcome of non-lethal predation effects on prey; with controls that were not exposed to any predatory effects (see Preisser et al. 2007). Also presented is Geary's test (eq. 13) to validate the accuracy of each effect size. Here effect sizes derived from Appleton and Palmer (1988) and Turner and Montgomery (2003) should be treated with caution as both had standardized means <3 (emphasized in boldface type). Variance of effect size is given in square brackets.

$$\frac{\bar{X}}{\mathrm{SD}}\sqrt{N} \geq 3. \tag{12}$$

Using 3 as the accuracy boundary was first proposed by Geary (1930) who determined that unlogged ratios tended to be normal if the coefficient of variation (CV) of the denominator was $\leq 1/3$. Here the inverse of Eq. 12 is the CV, and 1/3 is the probability that the denominator can take negative values (Hinkley 1969). More recent studies on the distribution of ratios suggest more conservative boundary values: ≥ 11.11 (Hayya et al. 1975), ≥ 10 (Kuethe et al. 2000), and ≥ 4 (Marsaglia 2006).

Although this diagnostic was not designed for evaluating problematic cases in log ratios, it still works reasonably well for identifying RR that are at risk of providing inaccurate effect sizes. For example, there is a high likelihood that the diagnostic will flag most of the problematic cases when the simulated means are near 0, and when the sample size (N) of the standardized mean is large (see Appendix: Fig. A3). However, at small sample sizes, the sampling variability is too large and the ability of Eq. 12 to detect problematic effects drops considerably (Appendix: Fig. A3). To improve the performance of Eq. 12, I suggest a small modification that includes a small-sample correction to the standardized mean

$$\frac{\bar{X}}{\mathrm{SD}}\left(\frac{4N^{3/2}}{1+4N}\right) \geq 3. \tag{13}$$

With this modified diagnostic, there is a slight 2–3% gain in confirming the accuracy of effect sizes that have small sample sizes but lie outside the problematic ranges near zero (Appendix: Fig. A3). It is also important to emphasize that RR^{Δ} and RR^{Σ} will be more sensitive to violations of Geary's rule (Eqs. 12 and 13). This is because their corrections themselves require that the distribution of ratios to be normal (see Appendix).

Finally, it is good practice to use Eqs. 12 or 13 for both the control and treatment groups: should at least one of two standardized means fail Geary's test, then effect sizes calculated with RR, RR^{Δ}, or RR^{Σ} may be at risk of estimating an incorrect effect and variance. A meta-analysis can then be used to compare the pooled effects among the response ratios that passed and failed Geary's test; should they differ, then this may provide some empirical justification for excluding at-risk effect sizes to help improve inferences with meta-analysis. Alternatively, I advocate using a sensitivity analysis where the meta-analytic results from the complete dataset is compared to those where problematic cases were excluded. This would provide evidence for the robustness of the overall model despite the variability introduced by including potentially inaccurate response ratios (Lajeunesse 2010).

Illustrative examples

Preisser et al. (2007) combined and compared the outcomes of several studies on changes in prey behavior and fitness in response to predation risk. Here I focus on five of these studies to briefly illustrate applications of response ratios (see Table 1); these examples were purposely selected as helpful references for interpreting effect sizes. Overall, among the three studies validated by Geary's test (eq. 13), there was an average bias reduction of 4.6% and 4.5% using RR^{Δ} and RR^{Σ}, respectively. Relative to var(RR^{Δ}) and var(RR^{Σ}), the variance of RR also underestimated the sampling variability of these studies. Also presented are effect sizes that failed Geary's test (Eq. 13). These should be interpreted with caution as their study outcomes lie within the problematic ranges for response ratios (Figs. 1 and A2). In particular, the effect sizes from Turner and Montgomery (2003) are clearly unusual given that they are in an entirely different magnitude relative to other studies (also note the higher sensitivity of RR^{Δ} and RR^{Σ} to these problematic data). These odd effects do necessarily indicate that the study itself is unusual, but given the outcome of Geary's test, they are likely

inaccurate and indicate that log response ratios are not adequate to estimate effect sizes with these data.

Discussion

I found that the log response ratio is a biased but consistent effect size estimator, and that the bias-corrected estimators RR^{Δ} (Eqs. 8 and 9) and RR^{Σ} (Eqs. 10 and 11) are improvements over this traditional metric, in terms of bias reduction and mean square error. Of these two new effect size estimators, RR^{Δ} is the most promising as it behaves well under several often-encountered conditions in experimental ecology. I also identified a range of conditions for when any effect size based on log response ratios will be at risk of estimating inaccurately the effect and predicted variance of that effect. In particular when the means of the control or treatment groups are near zero and the coefficients of variation are large. Therefore, along with RR^{Δ}, I urge ecologists to begin validating the accuracy of effect sizes with Geary's test (Eq. 13) prior to pooling outcomes with meta-analysis. Additional guidelines on how to uphold the accuracy of response ratio estimators are discussed in the Appendix.

Although RR^{Δ} and RR^{Σ} are improvements over RR, both clearly still have issues estimating the log ratio of two means and its predicted variance. For example, at small to moderate sample sizes, both fall short of removing all the small-sample bias (Fig. 1). This may be due to the inability of second-order corrections to fully estimate the rapid (perhaps non-linear) increase in bias and variance when either the treatment or control means are near zero. Including additional higher-order corrections using the Taylor expansion method can perhaps remedy this problem. However, estimators emerging from these expansions would not be practical for meta-analysis since they would require additional study parameters like skewness and kurtosis to reach higher-order corrections. These study parameters are rarely if ever reported in the literature, and given the already difficult challenge of extracting the means and variances from published studies to compute effect sizes (Lajeunesse and Forbes 2003, Lajeunesse 2013), there is no need to further exacerbate these problems by imposing stricter eligibility criteria for study inclusion (Lajeunesse 2010). Different strategies to developing second-order corrections could also be applied. For example, Edgeworth expansions could be used to develop new approximations (van der Vaart 1998). Here, unlike the Delta method, this approach would make use of the probability distribution of RR to achieve corrections (see Appendix). However, much of the intuition and simplicity of effect sizes is lost when probability distributions are used to construct estimators. This is why all effect size metrics used in meta-analysis, despite their probability distributions being known, continue to rely on large sample approximations. For example, Hedges' d and log odds ratio are $O(n^{-1})$ asymptotic normal estimators (Hedges 1981, Phillips and Holland 1987), and so are Beale's (1962) and Tin's (1965) ratio estimators.

On a peripheral but relevant note, research synthesis in medicine also use an independently derived version of the log response ratio abbreviated as RoM, or the ratio of arithmetic means (Friedrich et al. 2005, 2008, 2011). Under this literature, a few interesting caveats have emerged. For example, in the broader context of weighting and pooling multiple effect sizes with meta-analysis, Friedrich et al. (2008) found a small bias towards null-effects when few RR are aggregated (with either fixed- or random-effects models), and a small bias towards non-zero effects when there is large simulated between-study variance. However, these second-order sampling properties (synthesis-level error) are comparable to other established effect sizes like Hedges' d (see Hedges and Olkin 1985).

A second caveat is that RR can show greater heterogeneity than effect sizes estimated with Hedges' d (Friedrich et al. 2011). Hillebrand and Gurevitch (2014) also found similar variability when comparing these two estimators using a large dataset of grazing effects on microalgae. I suspect that one contribution to RR's heterogeneity is its greater sensitivity to the individual sampling errors of each control and treatment group. Unlike RR, Hedges' d aims to minimize this type of sampling error by homogenizing and pooling the SD's from these two groups (Hedges 1981). Another potential contribution to this heterogeneity is the inclusion of inaccurate RR that fail Geary's test. These sources of heterogeneity combined may explain why some ecological meta-analyses using both RR and Hedges' d with the same study parameters (i.e., $\bar{X}$, SD, and N) can yield different synthesis-level outcomes. Here the increased variability in RR would influence two key components that impact inferences with meta-analysis: (1) the relative weighting of each effect sizes and (2) the magnitude of the between-study variance component estimated for random-effects analyses based on these weights. Future research should aim to address these issues, with a direction that emphasizes sources of heterogeneity when estimating log response ratios, and the outcomes of that heterogeneity on ecological meta-analysis.

Acknowledgments

I thank E. Preisser for sharing his "database of fear" as part of the OpenMEE software project. Funding for this study was supported by NSF grants DBI-1262545 and DEB-1451031.

Literature Cited

Allouche, S., and P. Gaudin. 2001. Effects of avian predation threat, water flow and cover on growth and habitat use by chub, *Leuciscus cephalus*, in an experimental stream. Oikos 94:481–492.

Appleton, R. D., and A. R. Palmer. 1988. Water born stimuli released by predatory crabs and damaged prey induce more predatory-resistant shells in a marine gastropod. Proceedings of the National Academy of Sciences USA 85:4387–4391.

Beale, E. M. L. 1962. Some use of computers in operational research. Industrial Organisation 31:27–28.

Black, A. R., and S. I. Dodson. 1990. Demographic costs of *Chaoborus*-induced phenotypic plasticity in *Daphnia pulex*. Oecologia 83:117–122.

Curtis, P. S., and X. Wang. 1998. A meta-analysis of elevated CO_2 effects on woody plant mass, form, and physiology. Oecologia 113:299–313.

Friedrich, J. O., N. K. J. Adhikari, and J. Beyene. 2008. The ratio of means method as an alternative to mean differences for analyzing continuous outcome variables in meta-analysis: a simulation study. BMC Medical Research Methodology 8: 32.

Friedrich, J. O., N. K. J. Adhikari, and J. Beyene. 2011. Ratio of means for analyzing continuous outcomes in meta-analysis performed as well as mean difference methods. Journal of Clinical Epidemiology 64:556–564.

Friedrich, J. O., N. Adhikari, M. S. Herridge, and J. Beyene. 2005. Meta-analysis: low-dose dopamine increases urine output but does not prevent renal dysfunction or death. Annals of Internal Medicine 142:510–524.

Geary, R. C. 1930. The frequency distribution of the quotient of two normal variates. Journal of the Royal Statistical Society 93:442–446.

Hayya, J., D. Armstrong, and N. Gressis. 1975. A note on the ratio of two normally distributed variables. Management Science 21:1338–1341.

Hedges, L. V. 1981. Distribution theory for Glass's estimator of effect size and related estimators. Journal of Educational Statistics 6:107–128.

Hedges, L. V. 1982. Estimation of effect size from a series of independent experiments. Psychological Bulletin 92:490–499.

Hedges, L. V., J. Gurevitch, and P. S. Curtis. 1999. The meta-analysis of response ratios in experimental ecology. Ecology 80:1150–1156.

Hedges, L. V., and I. Olkin. 1985. Statistical methods for meta-analysis. Academic Press, New York, New York, USA.

Hillebrand, H., and J. Gurevitch. 2014. Meta-analysis results are unlikely to be biased by differences in variance and replication between ecological lab and field studies. Oikos 123:794–799.

Hinkley, D. V. 1969. On the ratio of two correlated normal random variables. Biometrika 56:634–639.

Hunter, J. E., and F. L. Schmidt. 1990. Methods of meta-analysis: correcting errors and bias in research findings. Sage, Newbury Park, California, USA.

Jennions, M. D., and A. P. Møller. 2003. A survey of the statistical power of research in behavioral ecology and animal behavior. Behavioral Ecology 14:438–445.

Koricheva, J., and J. Gurevitch. 2014. Uses and misuses of meta-analysis in plant ecology. Journal of Ecology 102:828–844.

Kuethe, D. O., A. Caprihan, H. M. Gach, I. J. Lowe, and E. Fukushima. 2000. Imaging obstructed ventilation with NMR using inert fluorinated gases. Journal of Applied Physiology 88:2279–2286.

Lajeunesse, M. J. 2010. Achieving synthesis with meta-analysis by combining and comparing all available studies. Ecology 91:2561–2564.

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

Lajeunesse, M. J. 2013. Recovering missing or partial data from studies: a survey of conversions and imputations for meta-analysis. Pages 195–206 *in* J. Koricheva, J. Gurevitch, and K. Mengersen, editors. Handbook of meta-analysis in ecology and evolution. Princeton University Press, Princeton, New Jersey, USA.

Lajeunesse, M. J., and M. R. Forbes. 2003. Variable reporting and quantitative reviews: a comparison of three meta-analytical techniques. Ecology Letters 6:448–454.

Marsaglia, G. 2006. Ratios of normal variables. Journal of Statistical Software 16:1–10.

Nakagawa, S., and E. S. A. Santos. 2012. Methodological issues and advances in biological meta-analysis. Evolutionary Ecology 26:1253–1274.

Olkin, I., and J. W. Pratt. 1958. Unbiased estimation of certain correlation coefficients. Annals of Mathematical Statistics 29: 201–211.

Phillips, A., and P. W. Holland. 1987. Estimation of the variance of the Mantel-Haenszel log-odds-ratio estimate. Biometrics 43:425–431.

Preisser, E. L., J. L. Orrock, and O. J. Schmitz. 2007. Predator hunting mode and habitat domain alter nonconsumptive effects in predator–prey interactions. Ecology 88:2744–2751.

Stuart, A., and J. K. Ord. 1994. Kendall's advanced theory of statistics. Volume 1: distribution theory. Griffin, London, UK.

Tabachnick, B. G., and L. S. Fidell. 1996. Using multivariate statistics. Third edition. HarperCollins College Publishers, New York, New York, USA.

Tin, M. 1965. Comparison of some ratio estimators. Journal of the American Statistical Association 60:294–307.

Turner, A. M. 1997. Contrasting short-term and long-term effects of predation risk on consumer habitat use and resources. Behavioral Ecology 8:120–125.

Turner, A. M., and S. L. Montgomery. 2003. Spatial and temporal scales of predator avoidance: experiments with fish and snails. Ecology 84:616–622.

van der Vaart, A. 1998. Asymptotic statistics. Cambridge University Press, New York, New York, USA.

Ver Hoef, J. M. 2012. Who invented the delta method? American Statistician 66:124–127.

Supplemental Material

Ecological Archives

The Appendix and the Supplement are available online: http://dx.doi.org/10.1890/14-2402.1.sm