Ecology, 92(11), 2011, pp. 2049–2055

On the meta-analysis of response ratios for studies with correlated and multi-group designs

Marc J. Lajeunesse[1]

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

Abstract. A common effect size metric used to quantify the outcome of experiments for ecological meta-analysis is the response ratio (RR): the log proportional change in the means of a treatment and control group. Estimates of the variance of RR are also important for meta-analysis because they serve as weights when effect sizes are averaged and compared. The variance of an effect size is typically a function of sampling error; however, it can also be influenced by study design. Here, I derive new variances and covariances for RR for several often-encountered experimental designs: when the treatment and control means are correlated; when multiple treatments have a common control; when means are based on repeated measures; and when the study has a correlated factorial design, or is multivariate. These developments are useful for improving the quality of data extracted from studies for meta-analysis and help address some of the common challenges meta-analysts face when quantifying a diversity of experimental designs with the response ratio.

Key words: generalized least squares; large-sample theory; log response ratio; multivariate effect size; nonindependence; variance–covariance matrix; weighted regression.

Introduction

The response ratio (RR) has emerged as a common effect size metric for the meta-analysis of ecological research (Hedges et al. 1999), and for quantifying simple two-group experimental designs the calculation of RR is straightforward:

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

Here, RR is the natural-log proportional change in the means ($\bar{X}$) of a treatment (T) and control group (C). Meta-analysis, when pooling RR from multiple studies, also assigns a weight to each RR that is inversely proportional to its sampling variance:

$$\hat{\sigma}^2(\mathrm{RR}) = \frac{(\mathrm{SD_C})^2}{N_\mathrm{C}\bar{X}_\mathrm{C}^2} + \frac{(\mathrm{SD_T})^2}{N_\mathrm{T}\bar{X}_\mathrm{T}^2} \quad (2)$$

where SD and N are the standard deviation and sample size of $\bar{X}_\mathrm{T}$ and $\bar{X}_\mathrm{C}$, respectively (see Hedges and Olkin 1985, Hedges et al. 1999). The process of quantifying study outcomes with RR and the subsequent down-weighting of studies with large variances form the basis for several successful meta-analyses in ecology (see Curtis and Wang 1998, Arnqvist and Nilsson 2000, Schmitz et al. 2000, Ainsworth and Long 2005, Parker et al. 2006).

However, since its formal description by Hedges et al. (1999), RR has had little further development, and one unsolved problem of RR is its inability to properly quantify experiments with varying research designs (Lajeunesse 2010). For example, the above description of $\hat{\sigma}^2(\mathrm{RR})$ assumes that $\bar{X}_\mathrm{T}$ and $\bar{X}_\mathrm{C}$ are from an experiment with an independent-groups design, where one group serves as the control and the other receives an experimental manipulation or treatment. However, when a study has a correlated-groups design, such as pre- and post-test experiments of animal behavior, individuals serve as their own control prior to a treatment. This design is a challenge for estimating $\hat{\sigma}^2(\mathrm{RR})$ because measures before and after treatments on the same individuals will be correlated, and this can influence the accuracy of the variance estimate, perhaps leading to erroneous conclusions when pooling multiple effects. These issues are less a challenge for more established effect sizes like Hedges' d, where 30 years of cross-disciplinary development has resulted in analogues and conversions of d for nearly every conceivable experimental design (e.g., Raudenbush et al. 1988, Gleser and Olkin 1994, Gurevitch et al. 2000, Kim and Becker 2010).

Experimental ecology is much more diverse than the scenarios for which the original formulation of RR was developed: many hypotheses are tested using experiments with partial-random designs and designs with multiple variables. Here, I derive new variance and covariance equations for RR that allow for the pooling of results from these types of experiments. I focus primarily on developing covariance equations for RR because these are essential for describing how multiple effect sizes relate within an experiment. For example, rather than calculating separate RR for each control–treatment contrast in a multi-treatment study, a single

Manuscript received 3 March 2011; revised 27 April 2011; accepted 6 July 2011. Corresponding Editor: B. D. Inouye.

[1] Email: lajeunesse@usf.edu

adjusted RR can be calculated for an entire experiment. This limits the redundancy arising from calculating multiple (nonindependent) RR from a single experiment, while also minimizing the overrepresentation of an experiment in the final meta-analysis (see Hunter and Schmidt 2004).

In particular, I rely heavily on large-sample theory to estimate the variance of RR (Stuart and Ord 1994), and apply generalized least squares (GLS) models to pool within-study effects. The current formulation of $\hat{\sigma}^2(\mathrm{RR})$ is based on the assumption that with large sample sizes the sampling distribution of RR will asymptote to a normal distribution (Hedges et al. 1999). This is a convenient distribution because it has simple and well understood properties, in contrast to the true (exact) distribution of RR, which is difficult to develop statistically (discussed further in the Appendix). Another advantage of this assumption is that the statistical relationships among multiple, correlated RR can be modeled using a GLS framework. This framework itself has a normality assumption required for analysis. Again, I entirely focus on providing adjustments for the variance and covariance of the response ratio; discussion on the advantages\disadvantages of this metric over other effect size metrics is found in Hedges et al. (1999), Osenberg et al. (1999), and Lajeunesse and Forbes (2003).

Aggregating Within-Study Effects

Before outlining the variance and covariance equations for RR, I begin with a brief description on how multiple RR can be aggregated prior to meta-analysis. This aggregate $\overline{\mathrm{RR}}$ and its variance $\hat{\sigma}^2(\overline{\mathrm{RR}})$ will become the data to be pooled in a meta-analysis. For example, multiple pairwise RR can be extracted from a single experiment testing multiple treatments with a control (Gleser and Olkin 1994). These RR should not be treated independently in meta-analysis because they share a common control. An exception is if the meta-analyst groups these multiple effects separately among (mutually exclusive) moderator categories. Here, the potential bias for nonindependence is minimized because these multiple RR will not be pooled together.

To estimate an aggregate within-study response ratio from k number of RR, the following weighted GLS regression (in matrix notation) can be applied:

$$\overline{\mathrm{RR}} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{E} \qquad (3)$$

where matrices with the superscript prime symbol and negative one indicate their transposition and inversion, respectively. Here, $\mathbf{X}$ is a column vector of ones with k number of rows, $\mathbf{V}$ is the variance–covariance matrix with $k \times k$ dimensions, and $\mathbf{E}$ is a column vector of k number effect sizes (here RR as defined in Eq. 1). The variance of Eq. 3 is $\hat{\sigma}^2\left(\overline{\mathrm{RR}}\right) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$. Modifying the elements of the variance–covariance matrix $\mathbf{V}$ is how various forms of nonindependence can be managed among multiple RR from the same experiment (see Lajeunesse 2009). The ith of k elements of $\mathbf{V}$ are typically defined as follows:

$$\mathbf{V}_{ii^*} = \begin{cases} \hat{\sigma}^2(\mathrm{RR}_i) & \text{when} \quad i = i^* \text{(main diagonal)} \\ \mathrm{cov}\left(\mathrm{RR}_i, \mathrm{RR}_{i^*}\right) & \text{when} \quad i \neq i^* \text{(off-diagonals)}. \end{cases} \qquad (4)$$

The inverse of $\mathbf{V}$ becomes the weighting matrix used to aggregate multiple RR; where RR with large variances (e.g., large sampling error) and large covariances (e.g., strong nonindependence) are down-weighted in the final $\overline{\mathrm{RR}}$ estimate. In the following sections, I outline the equations for the variance ($\hat{\sigma}^2$) and covariances (cov) of $\mathbf{V}$ for multiple RR derived from multi-group designs.

Estimation of Variance

Derivation of large-sample variance

A formal derivation of the large-sample variance of RR is missing from the literature; see Hedges et al. (1999) for a partial description on how it was developed. Here, I report this derivation to provide insight as to how more elaborate variances and covariances can be developed for RR. The variance of RR (Eq. 2) is a sample statistic of the "true" population of effect sizes that is approximately normal with a mean of $\lambda = \ln(\mu_T/\mu_C)$ and a variance of $\sigma^2(\lambda) = \sigma_C^2/(N_C\mu_C^2) + \sigma_T^2/(N_T\mu_T^2)$ Here, μ and σ^2/N are the mean and variance for the distributions of the control ($\bar{X}_C$) and treatment groups ($\bar{X}_T$). In the Appendix, I discuss why this distribution of λ is approximate. To derive the population variance$\sigma^2(\lambda)$, begin by applying the Taylor series approximation method to find all the first-order expansions of λ (Stuart and Ord 1994). These are the derivatives (Δ) to each variable in λ:

$$\Delta\lambda/\Delta\mu_T = 1/\mu_T \qquad \Delta\lambda/\Delta\mu_C = -1/\mu_C.$$

And let $\mathbf{A}$ represent a vector of these derivatives:

$$\mathbf{A}' = [\,1/\mu_T \quad -1/\mu_C\,].$$

Then define the large-sample variance–covariance matrix ($\mathbf{\Sigma}$) of μ_T and μ_C as:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_T^2/N_T & 0 \\ 0 & \sigma_C^2/N_C \end{bmatrix}.$$

Here, the main diagonal of $\mathbf{\Sigma}$ contains the large-sample approximations for the population variance (σ^2) of the two means in λ. These approximations are found in Stuart and Ord (1994). The off-diagonals of $\mathbf{\Sigma}$ describe the zero covariance (statistical independence) between μ_T and μ_C. Finally, by using matrix algebra and by substituting μ and σ^2 with the sample statistics $\bar{X}$ and $(\mathrm{SD})^2$, the large-sample variance of RR originally defined by Hedges et al. (1999) is

$$\hat{\sigma}^2(\mathrm{RR}) = \mathbf{A}'\mathbf{\Sigma}\mathbf{A} = \frac{(\mathrm{SD}_C)^2}{N_C\bar{X}_C^2} + \frac{(\mathrm{SD}_T)^2}{N_T\bar{X}_T^2} = \text{Eq. 2.} \qquad (5)$$

RR and variance when the control and treatment means are correlated

In the *Introduction* I described an example of a study with a correlated-groups design, where individuals serve as their own control prior to a treatment. These before and after experiments, or repeated measures, have a partial-random design, and are a challenge for computing the variance of RR because $\bar{X}_T$ and $\bar{X}_C$ will be statistically correlated. However, if this correlation (r) is known then this information can be used to adjust $\hat{\sigma}^2(\mathrm{RR})$ as follows:

$$\hat{\sigma}^2(\mathrm{RR})^r = \frac{(\mathrm{SD_C})^2}{N_C\bar{X}_C^2} + \frac{(\mathrm{SD_T})^2}{N_T\bar{X}_T^2} - \frac{2r\mathrm{SD_C}\mathrm{SD_T}}{\bar{X}_C\bar{X}_T\sqrt{N_CN_T}}. \quad (6)$$

Here, $\hat{\sigma}^2(\mathrm{RR})^r$ was derived by using the matrix algebra in Eq. 5 and by assuming that the predicted covariance (cov) between μ_T and μ_C was

$$\mathrm{cov}(\mu_C, \mu_T) = \rho\sqrt{\sigma_C^2/N_C}\sqrt{\sigma_T^2/N_T}.$$

This covariance is the large-sample covariance equation for two correlated variables (see Stuart and Ord 1994), and becomes the off-diagonals (replacing the zeros) of $\boldsymbol{\Sigma}$. The variance $\hat{\sigma}^2(\mathrm{RR})^r$ requires an estimate of the correlation ρ, such as a Pearson product-moment correlation coefficient r. If it is not available in the study, it may be possible to pool r from other similar studies (see Rosenthal 1984 for how to pool correlations). Alternatively, if a paired (or repeated-measures) t test is reported in the study, this t can be used to estimate $\hat{\sigma}^2(\mathrm{RR})^r$, where $N = N_C = N_T$, and

$$\hat{\sigma}^2(\mathrm{RR})^r = \frac{1}{N}\left[\frac{(\mathrm{SD_C})^2}{\bar{X}_C^2} + \frac{(\mathrm{SD_T})^2}{\bar{X}_T^2} - \frac{(\mathrm{SD_C})^2 + (\mathrm{SD_T})^2 - t^{-2}N(\bar{X}_T - \bar{X}_C)^2}{\bar{X}_C\bar{X}_T}\right]. \quad (7)$$

The derivation of Eq. 6 is found in the Appendix.

Estimation of Covariance

Derivation of the large-sample covariance

Here I derive a general form of the large-sample covariance (cov) needed to describe how two RR behave together statistically. Later, I modify this generalization to determine the covariance of multiple RR from multigroup experiments. As when deriving the variance of RR, let $\mathbf{A}$ and $\mathbf{B}$ represent vectors of the derivatives of all the variable of the response ratios $\lambda^A = \ln(\mu_T^A/\mu_C^A)$ and $\lambda^B = \ln(\mu_T^B/\mu_C^B)$:

$$\mathbf{A}' = \begin{bmatrix} 1/\mu_T^A & -1/\mu_C^A & 0 & 0 \end{bmatrix}$$

$$\mathbf{B}' = \begin{bmatrix} 0 & 0 & 1/\mu_T^B & -1/\mu_C^B \end{bmatrix}.$$

When λ^A and λ^B are independent, their large-sample variance–covariance matrix $\boldsymbol{\Sigma}$ is defined by the diagonal matrix of the Appendix: Fig. A1a, where the main diagonal of $\boldsymbol{\Sigma}$ has the large-sample variances of each mean and all off-diagonals (covariances) are zero. Note that $\mathbf{A}'\boldsymbol{\Sigma}\mathbf{A}$ and $\mathbf{B}'\boldsymbol{\Sigma}\mathbf{B}$ will yield $\sigma^2(\lambda^A)$ and $\sigma^2(\lambda^B)$. These are the population variances of λ^A and λ^B. Further, $\mathbf{A}'\boldsymbol{\Sigma}\mathbf{B}$ and $\mathbf{B}'\boldsymbol{\Sigma}\mathbf{A}$ will yield zero covariance because the off-diagonals of $\boldsymbol{\Sigma}$ do not specify any statistical dependence between the two λ. In the following sections, I describe experimental designs with nonzero covariances.

Covariance when multiple treatments share a common control

In the previous section, I outlined the covariance between two independent RR. Here, I describe an ANOVA-style study where multiple treatments are compared to a single control. One approach to quantify this multi-treatment study is to calculate RR separately for each control–treatment pair, and then treat these RR as independent data in a meta-analysis. Keeping each RR separate may be important if the identity of the treatment itself is a moderator to be tested (e.g., whether studies examining treatment A differ from studies testing B). However, if the meta-analyst is simply interested in quantifying the entire experimental outcome across all treatments, then treating these RR as separate data points violates assumptions of independence for meta-analysis (Gleser and Olkin 1994). To limit this bias, RR^A and RR^B should be aggregated prior to meta-analysis (see *Aggregating within-study effects*). This aggregate can be estimated by pooling RR^A and RR^B using the regression Eq. 3 with this variance–covariance matrix:

$$\mathbf{V}^{\bar{X}_C} = \begin{bmatrix} \dfrac{(\mathrm{SD_C})^2}{N_C\bar{X}_C^2} + \dfrac{(\mathrm{SD_T^A})^2}{N_T^A\left(\bar{X}_T^A\right)^2} & \dfrac{(\mathrm{SD_C})^2}{N_C\bar{X}_C^2} \\ \dfrac{(\mathrm{SD_C})^2}{N_C\bar{X}_C^2} & \dfrac{(\mathrm{SD_C})^2}{N_C\bar{X}_C^2} + \dfrac{(\mathrm{SD_T^B})^2}{N_T^B\left(\bar{X}_T^B\right)^2} \end{bmatrix}. \quad (8)$$

Note that the variance (main diagonal) of each RR remains the same as in Eq. 2, but the covariance (off-diagonals) is now the variance due to the shared control mean $\bar{X}_C$. To evaluate the strength of this nonindependence due to this shared control, it may also be useful to calculate the correlation (r) between RR^A and RR^B:

$$r = \frac{(\mathrm{SD_C})^2/\left(N_C\bar{X}_C^2\right)}{\sqrt{\hat{\sigma}^2\left(\mathrm{RR}^A\right)}\sqrt{\hat{\sigma}^2\left(\mathrm{RR}^B\right)}}. \quad (9)$$

The covariance term in $\mathbf{V}^{\bar{X}_C}$ was derived by including nonzero off-diagonals in $\boldsymbol{\Sigma}$ (the matrix describing how the population means between multiple λ are related) that account for the shared control among multiple RR (see Appendix: Fig. A1a). These new off-diagonals become $\mathrm{cov}\left(\mu_C^A, \mu_C^B\right) = \left(\sigma_C^B\right)^2/N_C^B$, which is a reduced

form of the large-sample covariance

$$\rho\sqrt{\left(\sigma_C^A\right)^2/N_C^A}\sqrt{\left(\sigma_C^B\right)^2/N_C^B}$$

when $\mu_C^A = \mu_C^B$ and $\rho = 1$. Finally, the variance and covariance for RR^A and RR^B described in $\mathbf{V}^{\bar{X}_C}$ were estimated by solving $\mathbf{A}'\mathbf{\Sigma}\mathbf{A}$, $\mathbf{A}'\mathbf{\Sigma}\mathbf{B}$, and $\mathbf{B}'\mathbf{\Sigma}\mathbf{B}$ with the $\mathbf{\Sigma}$ described in the Appendix: Fig. A1b.

I illustrate this method with the Sokol-Hessner and Schmitz (2002) study on the mortality rate of grasshoppers when exposed to spider predators A ($\bar{X} = 0.011$, SD = 0.0099, $N = 14$), B ($\bar{X} = 0.024$, SD = 0.0179, $N = 14$), C ($\bar{X} = 0.041$, SD = 0.0179, $N = 14$), and a control group without predators ($\bar{X} = 0.02$, SD = 0.0162, $N = 14$). The three RR for each species control–treatment pair are $\mathbf{E} = [-0.598 \quad 0.182 \quad 0.718]'$, and their variances/covariances defined by the equations in the main and off-diagonals of $\mathbf{V}^{\bar{X}_C}$ are

$$\mathbf{V} = \begin{bmatrix} 0.105 & 0.047 & 0.047 \\ 0.047 & 0.087 & 0.047 \\ 0.047 & 0.047 & 0.061 \end{bmatrix}.$$

The aggregate response ratio ($\overline{RR}$), adjusted for the nonindependence due to a shared control among the three spider treatments, is $\overline{RR} = 0.4$ with $\hat{\sigma}^2(\overline{RR}) = 0.0558$. The regression Eq. 3, applying $\mathbf{E}$ and $\mathbf{V}$ from the previous paragraph, was used to calculate $\overline{RR}$. Note that pooling RR without accounting for the shared control (off-diagonal covariance) yields an $\overline{RR} = 0.218$ and $\hat{\sigma}^2(\overline{RR}) = 0.0267$. This aggregate effect size underestimates the magnitude of experimental manipulations, and in the context of meta-analysis, will have a greater relative weight in the overall analysis.

Variance and covariance for repeated-measures or matched-pairs designs

The simplest repeated-measures or matched-pairs design compares the experimental response between a control and treatment group, both before (PRE) and after (POST) a treatment effect. This type of a correlated-groups design is meant to control for selection effects of individuals assigned to each group prior to experimentation. For example, with a small sample size, and through random chance, larger individuals could be assigned to the treatment group. This would bias the posttreatment measurements should size itself becomes a response variable. A typical approach for quantifying this type of experiment for meta-analysis is to simply use the last (POST) endpoints as an estimate of the study outcome (see Gurevitch and Hedges 1999). However, this is a lost opportunity to improve the precision of RR because important information that contributed to the outcome of the treatment effect was ignored.

To integrate these data, we can use both the PRE and POST treatment results to estimate an aggregate $\overline{RR}$. These pretreatment $RR^{PRE} = \ln(\bar{X}_T^{PRE}/\bar{X}_C^{PRE})$ and a posttreatment $RR^{POST} = \ln(\bar{X}_T^{POST}/\bar{X}_C^{POST})$ will individually have the same variances as described in Eq. 2, but will have a covariance

$$\operatorname{cov}\left(RR^{PRE}, RR^{POST}\right) = \frac{r_C SD_C^{PRE} SD_C^{POST}}{\bar{X}_C^{PRE}\bar{X}_C^{POST}\sqrt{N_C^{PRE}N_C^{POST}}} + \frac{r_T SD_T^{PRE} SD_T^{POST}}{\bar{X}_T^{PRE}\bar{X}_T^{POST}\sqrt{N_T^{PRE}N_T^{POST}}} \quad (10)$$

where r is the correlation between the PRE and POST treatment means for the control and treatment groups. If there is no reason to expect that r_C and r_T differ, these can be pooled (see Rosenthal 1984). This covariance was derived using the $\mathbf{A}$ and $\mathbf{B}$ defined in the section *Derivation of the large-sample covariance*, but calculating $\mathbf{A}'\mathbf{\Sigma}\mathbf{B}$ with the within-study $\mathbf{\Sigma}$ defined in the Appendix: Fig. A1c.

This example has the simplest repeated-measures design: only PRE and POST treatment periods are compared. However, a series of measurements can also be pooled with the covariance Eq. 10. For example, in the case where multiple measurements are autocorrelated (e.g., are dependent with the only previous measurement), the duration between measurements as a scaled linear distance can be treated as the correlations of Eq. 10. This assumes that measurements with short intervals will show the least amount of change and will be the most similar. In this case, the PRE and POST treatment effects become time interval ONE and TWO, TWO and THREE, and so on. For each of these pairwise intervals, their covariance is then applied to the regression Eq. 3 to pool each RR across all intervals. Typically, however, the outcome of interest from a repeated-measures experiment is not the aggregate effect across multiple time intervals, but the trajectory of curves due to treatment effects. This form of repeated-measures meta-analysis has been described elsewhere (see Peters and Mengersen 2008).

Response ratio and variance for factorial designs

Morris et al. (2007) describe the variance for both the overall effects (RR^O) and interaction effects (RR^I) for studies with a 2×2 factorial design (also see Hawkes and Sullivan 2001). The individual effects among factors use the RR and $\hat{\sigma}^2(RR)$ described in Eqs. 1 and 2. Here, I briefly revisit RR^O and RR^I, but make additional recommendations on how to proceed when studies have a correlated factorial design.

To quantify the overall (O) effects of a 2×2 factorial experiments contrasting the control (C) and treatment means (T) between groups A and P, the following RR^O can be used:

$$RR^O = \ln\left(\frac{\bar{X}_T^A + \bar{X}_C^A}{\bar{X}_T^P + \bar{X}_C^P}\right) \quad (11)$$

which has a variance of

$$\hat{\sigma}^2(\mathrm{RR^O}) = \left(\frac{1}{\bar{X}_\mathrm{T}^\mathrm{A} + \bar{X}_\mathrm{C}^\mathrm{A}}\right)^2 \left(\frac{(\mathrm{SD}_\mathrm{T}^\mathrm{A})^2}{N_\mathrm{T}^\mathrm{A}} + \frac{(\mathrm{SD}_\mathrm{C}^\mathrm{A})^2}{N_\mathrm{C}^\mathrm{A}}\right) + \left(\frac{1}{\bar{X}_\mathrm{T}^\mathrm{P} + \bar{X}_\mathrm{C}^\mathrm{P}}\right)^2 \left(\frac{(\mathrm{SD}_\mathrm{T}^\mathrm{P})^2}{N_\mathrm{T}^\mathrm{P}} + \frac{(\mathrm{SD}_\mathrm{C}^\mathrm{P})^2}{N_\mathrm{C}^\mathrm{P}}\right). \quad (12)$$

To quantify the interaction (I) effect, the response ratio is:

$$\mathrm{RR^I} = \ln\left(\frac{\bar{X}_\mathrm{C}^\mathrm{A}}{\bar{X}_\mathrm{C}^\mathrm{P}}\right) - \ln\left(\frac{\bar{X}_\mathrm{T}^\mathrm{A}}{\bar{X}_\mathrm{T}^\mathrm{P}}\right) \quad (13)$$

with a variance of

$$\hat{\sigma}^2(\mathrm{RR^I}) = \frac{(\mathrm{SD}_\mathrm{C}^\mathrm{A})^2}{(\bar{X}_\mathrm{C}^\mathrm{A})^2 N_\mathrm{C}^\mathrm{A}} + \frac{(\mathrm{SD}_\mathrm{T}^\mathrm{A})^2}{(\bar{X}_\mathrm{T}^\mathrm{A})^2 N_\mathrm{T}^\mathrm{A}} + \frac{(\mathrm{SD}_\mathrm{C}^\mathrm{P})^2}{(\bar{X}_\mathrm{C}^\mathrm{P})^2 N_\mathrm{C}^\mathrm{P}} + \frac{(\mathrm{SD}_\mathrm{T}^\mathrm{P})^2}{(\bar{X}_\mathrm{T}^\mathrm{P})^2 N_\mathrm{T}^\mathrm{P}} \quad (14)$$

which can be simplified to $\hat{\sigma}^2(\mathrm{RR^I}) = \hat{\sigma}^2(\mathrm{RR_A}) + \hat{\sigma}^2(\mathrm{RR_P})$ where $\hat{\sigma}^2(\mathrm{RR_A})$ and $\hat{\sigma}^2(\mathrm{RR_P})$ are the variances defined by Eq. 2, respectively, for factors A and P.

In the Appendix, I outline how the variances of $\hat{\sigma}^2(\mathrm{RR^O})$ and $\hat{\sigma}^2(\mathrm{RR^I})$ were derived, and further modify this derivation to develop analogous variances to Eqs. 12 and 14 for experiments with a correlated factorial design. A correlation can occur between factors A and P if these are repeated measures: again, the correlation coefficient r is integrated in the calculations of the new variances for both the overall $\hat{\sigma}^2(\mathrm{RR}_r^\mathrm{O})$ and interaction $\hat{\sigma}^2(\mathrm{RR}_r^\mathrm{I})$ models. The equation for $\hat{\sigma}^2(\mathrm{RR}_r^\mathrm{O})$ is rather cumbersome and is found in the Appendix, whereas, for the interaction effect, the variance of $\mathrm{RR^I}$ becomes:

$$\hat{\sigma}^2(\mathrm{RR}_r^\mathrm{I}) = \hat{\sigma}^2(\mathrm{RR}_\mathrm{C}^r) + \hat{\sigma}^2(\mathrm{RR}_\mathrm{T}^r) \quad (15)$$

where $\hat{\sigma}^2(\mathrm{RR}_\mathrm{C}^r)$ and $\hat{\sigma}^2(\mathrm{RR}_\mathrm{T}^r)$ are defined in Eq. 6, but here $\hat{\sigma}^2(\mathrm{RR}_\mathrm{C}^r)$ uses the two controls, and $\hat{\sigma}^2(\mathrm{RR}_\mathrm{T}^r)$ the two treatment effects, from factors A and P.

Multivariate RR

It is not uncommon for a single study to report control–treatment contrasts for several dependent variables (e.g., measurement metrics or traits estimates). For example, experiments on the anti-herbivore defenses of plants often explore suites of chemical and structural traits (e.g., Carmona et al. 2011). However, these traits are often related to one another, and may be heavily correlated (Raudenbush et al. 1988). This violates assumptions of meta-analysis because individual effect sizes calculated from each trait will not represent independent pieces of information (Hedges and Olkin 1985). However, if the correlations between traits are available, for example between traits X and Y, then the RR of each trait can be combined using the regression Eq. 3 with the following covariance:

$$\mathrm{cov}(\mathrm{RR^X}, \mathrm{RR^Y}) = \frac{r\mathrm{SD}_\mathrm{T}^\mathrm{X}\mathrm{SD}_\mathrm{T}^\mathrm{Y}}{\bar{X}_\mathrm{T}\bar{Y}_\mathrm{T}\sqrt{N_\mathrm{T}^\mathrm{X}N_\mathrm{T}^\mathrm{Y}}} + \frac{r\mathrm{SD}_\mathrm{C}^\mathrm{X}\mathrm{SD}_\mathrm{C}^\mathrm{Y}}{\bar{X}_\mathrm{C}\bar{Y}_\mathrm{C}\sqrt{N_\mathrm{C}^\mathrm{X}N_\mathrm{C}^\mathrm{Y}}}. \quad (16)$$

Eq. 16 can also be generalized to aggregate RR from multiple, correlated traits. For example, in a study on the phenotypic plasticity of a milkweed, Agrawal et al. (2008) reported treatment and control means (with and without herbivory) for multiple plant traits (trichome density, percentage of water, and C:N ratio), in addition to the correlations between these traits (see data in Appendix: Table A1). These correlations can be used to estimate an aggregate $\overline{\mathrm{RR}}$ across these three plant traits as follows. First, calculate RR for each control–treatment pair: let $\mathbf{E}$ represent a vector of these three effect sizes $\mathbf{E} = [-0.606 \quad 0.155 \quad -0.488]'$. Next, calculate the variances of these RR using Eq. 2 and their covariance with Eq. 16. These variance and covariance are found in $\mathbf{V}$:

$$\mathbf{V} = \begin{bmatrix} 0.148 & -0.004 & 0.008 \\ -0.004 & 0.004 & -0.005 \\ 0.008 & -0.005 & 0.088 \end{bmatrix}.$$

Finally, applying $\mathbf{E}$ and $\mathbf{V}$ to the regression Eq. 3 yields an aggregate $\overline{\mathrm{RR}} = 0.073$ with a variance of $\hat{\sigma}^2(\overline{\mathrm{RR}}) = 0.0026$. If the meta-analyst ignored these correlations, the aggregate $\overline{\mathrm{RR}}$ would have overestimated the effect ($\overline{\mathrm{RR}} = 0.109$) and variance $\hat{\sigma}^2(\overline{\mathrm{RR}}) = 0.0034$.

Discussion

The current formulation of the response ratio quantifies experiments with a simple control–treatment design. However, the composition of ecological studies available for meta-analysis is broad and includes designs with various manipulations on multiple traits (see Jennions and Møller 2003). Here, I provide equations to quantify various experimental designs. These are important because if incorrect formulae are used, the effect size estimate and its variance may exaggerate the outcome of the experimental manipulation (Gleser and Olkin 1994). This will influence the strength of hypothesis tests with ecological meta-analysis (Lajeunesse 2010). Applying these equations to estimate RR will also minimize the effect of incorrect variances influencing the outcome of moderator tests (e.g., homogeneity tests; see Hedges and Olkin 1985). For example, not accounting for differences in experimental design across studies may yield a significant moderator effect when none actually exist.

I particularly focus on modifying the variance of effect sizes as way to minimize bias by adjusting the weight of effect sizes when pooling multiple studies. I also focus on covariance structures for pooling multiple related effects from a single study. By applying these equations to adjust or aggregate within-study effect sizes, meta-analysts are no longer forced to choose a single aspect of a study to quantify. For instance, meta-analysts are often left to select which manipulation effect

to estimate from a study, make judgments on which manipulation is more appropriate for the meta-analysis (Gurevitch et al. 2003), or perhaps even ignore the correlations that exist between multiple effects estimated from the same study. This can result in a loss of information when estimating the magnitude of effect and variance for that study, and ignoring correlations will affect the Type I error rate (likelihood of a false positive result) of nonzero and homogeneity tests (Walsh 1947). These homogeneity tests are important because they evaluate differences among RR beyond the predicted differences due to sampling error (Hedges and Olkin 1985). Extracting the correlations necessary to compute many of the variances described in this paper will also be a challenge. However, because correlations are a very common type of effect size metric (Rosenthal 1984), various ways to extract correlations from experiments using conversions, imputations and approximation methods have been developed (see Lipsey and Wilson 2001, Lajeunesse, *in press b*).

The primary reason to aggregate multiple within-study effect sizes prior to a between-study meta-analysis was to address any possible redundancy of information. Multiple extractions of RR from the same experiment are not independent pieces of information. Addressing this problem is necessary because the amount of information included in a meta-analysis strongly affects the precision of the combined (pooled) effects. This is because the statistical power of meta-analysis is indirectly a function of the number of effect sizes included in the analysis (see Lajeunesse, *in press a*).

Aggregating effect sizes prior to meta-analysis is useful if a meta-analyst is limited to software that cannot easily integrate information on the covariance of effect sizes. However, with some understating of GLS modeling, it is possible to perform a meta-analysis that simultaneously pools all within- and between-study RR. Here the within-study matrices (such as $\mathbf{V}^{\tilde{X}_C}$) are included as sub-matrices in a between-study variance–covariance matrix. For example, the elements of a between-study **V** can describe the relationships among all effect sizes from each study as follows:

$$E = \begin{bmatrix} 0.337 \\ 0.328 \\ 0.319 \\ 0.315 \end{bmatrix}$$

$$V = \begin{bmatrix} 0.0019 & 0.0016 & 0.0000 & 0.0000 \\ 0.0016 & 0.0018 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0021 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0011 \end{bmatrix}.$$

Here, we have an effect size vector (**E**) and variance–covariance matrix (**V**) for a meta-analysis of three independent studies. The first study (shown by the top two rows in **E**) reported two treatment effects with a shared control (see example in Eq. 8). The shared variance due to the common control is defined in the upper left 2×2 submatrix of **V**.

Modeling all the within- and between-study variances simultaneously has the advantage of expanding the domain of testable hypotheses (Becker and Schram 1994). For example, this model can be used to systematically explore each individual response outcome (e.g., dependent variables or traits used to estimate RR), or specific treatment effects, and whether these treatments effects depend on factors such as study design or other biological predictors that may moderate effects. This exploration is possible because all of the calculated RR (not the aggregate $\overline{RR}$) are included in the model: This retains the identity of each treatment or trait used to calculate effect sizes. For example, with a full model, we can pool all the RR across studies that explore the same treatment effect, with the following design matrix:

$$X' = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}.$$

The first study (shown by the first two columns) had two RR from treatments A and B, whereas the second study (third column of **X**) reported only the manipulation with treatment B, and the third only with A. Integrating **X**, **E,** and **V** in the regression model of Eq. 3 will yield two pooled effect sizes: one for all the RR based on treatment A, and the other for all the B effects. This highlights the ability to include studies with different designs within a single meta-analysis, even if they report only a subset of all the treatment effects explored across studies.

The disadvantage of this type of mixed-experiment modeling is the difficulty in estimating the between-study variance (τ^2) required for random-effects meta-analysis. Here, the commonly used method of moments approach will not suffice (see Hedges and Olkin 1985). However, maximum likelihood and restricted maximum likelihood (REML) approaches are emerging as powerful alternatives to estimating τ^2 (Berkey et al. 1998); only recently have advances allowed for fitting more elaborate variance–covariance designs with random-effects models (see Jackson et al. 2010).

Conclusion and Prospectus

Dealing with multiple effect sizes from a single experiment is one of many challenges meta-analysts face when they use the response ratio. Another challenge is when studies do not report the means, SD and *N* needed to compute RR; but the test statistics themselves, such as *t* tests or *F* tests, are available. These test statistics are still useful to estimate an effect size based on a metric like Hedges' *d* because conversions are available (Lipsey and Wilson 2001, Nakagawa and Cuthill 2007). These conversions are possible because Hedges' *d* has a *t* distribution whose properties and relationship with other statistical distributions are well understood (Hedges 1982). However, it is unclear to me whether similar conversions can be formulated for RR given its unwieldy distribution (the Appendix has additional discussion on

this problem). Further exploration of this issue, and other issues that affect the overall utility of the response ratio, will be necessary to improve inferences from published experiments for ecological meta-analysis.

Acknowledgments

I thank Anurag Agrawal and two anonymous reviewers for valuable feedback on this manuscript. This study and my research are funded by the College of Arts and Sciences, University of South Florida.

Literature Cited

Agrawal, A. A., A. C. Erwin, and S. C. Cook. 2008. Natural selection and predicted response for ecophysiological traits of swamp milkweed (*Asclepias incarnata*) in the field. Journal of Ecology 96:536–542.

Ainsworth, E. A., and S. P. Long. 2005. What have we learned from 15 years of free-air CO_2 enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO_2. New Phytologist 165:351–372.

Arnqvist, G., and T. Nilsson. 2000. The evolution of polyandry: multiple mating and female fitness in insects. Animal Behaviour 60:145–164.

Becker, B. J., and C. M. Schram. 1994. Examining explanatory models through research synthesis. Pages 357–381 in H. Cooper and L. V. Hedges, editors. The handbook of research synthesis. Russell Sage Foundation, New York, New York, USA.

Berkey, C. S., D. C. Hoaglin, A. Antczak-Bouckoms, F. Mosteller, and G. A. Colditz. 1998. Meta-analysis of multiple outcomes by regression with random effects. Statistics in Medicine 17:2537–2550.

Carmona, D., M. J. Lajeunesse, and M. T. J. Johnson. 2011. Plant traits that predict resistance to herbivores. Functional Ecology 5:358–367.

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.

Gleser, L. J., and I. Olkin. 1994. Stochastically dependent effect sizes. Pages 339–355 in H. Cooper and L. V. Hedges, editors. The handbook of research synthesis. Russell Sage Foundation, New York, New York, USA.

Gurevitch, J., P. S. Curtis, and M. H. Jones. 2001. Meta-analysis in ecology. Advances in Ecological Research 32: 199–247.

Gurevitch, J., and L. V. Hedges. 1999. Statistical issues in ecological meta-analyses. Ecology 80:1142–1149.

Gurevitch, J., J. A. Morrison, and L. V. Hedges. 2000. The interaction between competition and predation: a meta-analysis of field experiments. American Naturalist 155:435–453.

Hawkes, C. V., and J. J. Sullivan. 2001. The impact of herbivory on plants in different resource conditions: a meta-analysis. Ecology 82:2045–2058.

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., 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.

Hunter, J. E., and F. L. Schmidt. 2004. Methods of meta-analysis: Correcting error and bias in research findings. Second edition. Sage Publications, Thousand Oaks, California, USA.

Jackson, D., I. R. White, and S. G. Thompson. 2010. Extending Dersinomian and Laird's methodology to perform multivariate random effects meta-analysis. Statistics in Medicine 29:1282–1297.

Jennions, M. D., and A. P. Møller. 2003. A survey of the statistical power of research in behavioral ecology and animal behaviour. Behavioural Ecology 14:438–455.

Kim, R.-S., and B. J. Becker. 2010. The degree of dependence between multiple-treatment effect sizes. Multivariate Behavioral Research 45:213–238.

Lajeunesse, M. J. 2009. Meta-analysis and the comparative phylogenetic method. American Naturalist 174:369–381.

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

Lajeunesse, M. J. *In press a*. Power statistics for meta-analysis: test for mean effects and homogeneity. *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. *In press b*. Recovering missing or partial data from studies: a survey of conversions and imputations for meta-analysis. *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.

Lipsey, M. W., and D. B. Wilson. 2001. Practical meta-analysis. Sage Publications, Thousand Oaks, California, USA.

Morris, W. F., R. A. Hufbauer, A. A. Agrawal, J. D. Bever, V. A. Borowicz, G. S. Gilbert, J. L. Maron, C. E. Mitchell, I. M. Parker, A. G. Power, M. E. Torchin, and D. P. Vázquez. 2007. Direct and interactive effects of enemies and mutualists on plant performance: a meta-analysis. Ecology 88:1021–1029.

Nakagawa, S., and I. C. Cuthill. 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biological Reviews 82:591–605.

Osenberg, C. W., O. Sarnelle, S. D. Cooper, and R. D. Holt. 1999. Resolving ecological questions through meta-analysis: goals, metrics, and models. Ecology 80:1105–1117.

Parker, J. D., D. E. Burkepile, and M. E. Hay. 2006. Opposing effects of native and exotic herbivores on plant invasions. Science 311:1459–1461.

Peters, J. L., and K. L. Mengersen. 2008. Meta-analysis of repeated measures study designs. Journal of Evaluation in Clinical Practice 14:941–950.

Raudenbush, S. W., B. J. Becker, and H. Kalaian. 1988. Modeling multivariate effect sizes. Psychological Bulletin 103:111–120.

Rosenthal, R. 1984. Meta-analytic procedures for social research. Sage Publications, Beverly Hills, California, USA.

Schmitz, O. J., P. A. Hambäck, and A. P. Beckerman. 2000. Trophic cascades in terrestrial systems: a review of the effects of carnivore removals on plants. American Naturalist 155:141–153.

Sokol-Hessner, L., and O. J. Schmitz. 2002. Aggregate effects of multiple predator species on a shared prey. Ecology 83:2367–2372.

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

Walsh, J. E. 1947. Concerning the effect of the intraclass correlation on certain significance tests. Annals of Mathematical Statistics 18:88–96.

APPENDIX

Discussion on large-sample theory, derivation of variances of response ratios from factorial designs, raw data used in examples, and variance–covariance matrices used to estimate the covariance between two response ratios (*Ecological Archives* E092-178-A1).