THE META-ANALYSIS OF RESPONSE RATIOS IN EXPERIMENTAL ECOLOGY

Larry V. Hedges,[1] Jessica Gurevitch,[2] and Peter S. Curtis[3]

[1] *Department of Education, University of Chicago, 5835 South Kimbark Avenue Chicago, Illinois 60637 USA*
[2] *Department of Ecology and Evolution, State University of New York, Stony Brook, New York 11794 USA*
[3] *Department of Plant Biology, Ohio State University, 1735 Neil Avenue, Columbus, Ohio 43210 USA*

Abstract. Meta-analysis provides formal statistical techniques for summarizing the results of independent experiments and is increasingly being used in ecology. The response ratio (the ratio of mean outcome in the experimental group to that in the control group) and closely related measures of proportionate change are often used as measures of effect magnitude in ecology. Using these metrics for meta-analysis requires knowledge of their statistical properties, but these have not been previously derived. We give the approximate sampling distribution of the log response ratio, discuss why it is a particularly useful metric for many applications in ecology, and demonstrate how to use it in meta-analysis. The meta-analysis of response-ratio data is illustrated using experimental data on the effects of increased atmospheric CO_2 on plant biomass responses.

Key words: effect size; elevated CO_2, plant biomass responses; meta-analysis in experimental ecology; meta-analysis of response ratios; research synthesis; response ratio; statistical techniques.

Introduction

The increasingly large body of studies in experimental ecology has led to interest in methods for summarizing this evidence to reach general conclusions. Meta-analysis, the formal application of quantitative methods to summarize evidence across studies, has recently been introduced to ecology (Gurevitch et al. 1992, Arnqvist and Wooster 1995). In a meta-analysis, the result of each independent experiment is usually expressed as an index of effect; these effect estimates are then combined across studies to produce a summary of the findings. Subgroupings of experiments may be examined separately to determine whether experimental results differ across biologically meaningful groupings of experiments (e.g., whether the effects of a manipulation such as elevated atmospheric CO_2 on total biomass differ across plant taxa).

Meta-analyses can only provide meaningful summaries if the effect-size index used is a meaningful summary of any one experiment. Several recent meta-analyses in ecology have used the standardized difference between means (the difference between the mean of the treated group and a control group, divided by the within-group standard deviation), also called a "*d* index," as an index of effect. Although its statistical properties are well understood, and a substantial set of meta-analytic procedures using the *d* index are available (see Hedges and Olkin 1985), it is not always a meaningful way to summarize experiments in ecology (Osenberg et al. 1997).

The *response ratio,* the ratio of some measured quantity in experimental and control groups, is commonly used as a measure of experimental effect because it quantifies the proportionate change that results from an experimental manipulation. Examples of such ratios include relative competition intensity, relative yield, and relative crowding coefficient. The response ratio is also closely related to metrics that quantify the effects of treatments on per capita rates of change (Osenberg et al. 1997, 1999, Laska and Wootton 1998, Downing et al. 1999). The response ratio has been used informally in reviews to describe the effects of predation in streams (Cooper et al. 1990, Englund et al. 1999), the effects of grazers on algal biomass (Sarnelle 1992), and the response of plant biomass to increased CO_2 levels (Kimball 1983, Poorter 1993, Gunderson and Wullschleger 1994). To take full advantage of the statistical techniques available for statistical inference in meta-analysis, it is necessary to know the sampling properties of the effect-size index used. To fully utilize meta-analytic techniques, however, it is necessary to understand the sampling properties of the response ratio. The purpose of this article is to provide the information necessary to carry out meta-analytic procedures using response ratios and demonstrate their use with an important set of ecological data. We use data from recent experiments on the effects of increased ambient CO_2 on growth of woody plants (Curtis and Wang 1998) to illustrate the use of response ratios in meta-analysis.

The Response Ratio as an Effect Index

Consider a set of experiments, each of which compares dry mass gain in a replicated control group in which plants are grown with ambient levels of CO_2

with an experimental group in which plants are grown with approximately double the ambient level of CO_2. We denote the mean and standard deviation of the outcome in the experimental group by $\bar{X}_E$ and SD_E, the sample size (number of replicates) by n_E, and the mean, standard deviation and the sample size of the outcome in the control group by $\bar{X}_C$, SD_C and n_C, respectively. Then, the (sample) response ratio, $R = \bar{X}_E/\bar{X}_C$, is an estimate of the ratio, ρ, of the population means. However, it is desirable to perform statistical analyses in the metric of the natural logarithm of the response ratio, $L = \ln(R) = \ln(\bar{X}_E) - \ln(\bar{X}_C)$, for two reasons. The first is that the logarithm linearizes the metric, treating deviations in the numerator the same as deviations in the denominator. That is, while the ratio is affected more by changes in the denominator (especially when the denominator is small), the log ratio is affected equally by changes in either numerator or denominator. The second reason is that the sampling distribution of R is skewed, and the distribution of L is much more normal in small samples than that of R (see Appendix A).

If $\bar{X}_E$ and $\bar{X}_C$ are normally distributed and $\bar{X}_C$ is unlikely to be negative, then L is approximately normally distributed with mean approximately equal to the true log response ratio and variance, ν, approximately equal to

$$\frac{(SD_E)^2}{n_E\bar{X}_E^2} + \frac{(SD_C)^2}{n_C\bar{X}_C^2}. \tag{1}$$

If the variances in the treatment and control groups are approximately equal then the variance can be computed using the pooled within-group standard deviation SD_P as

$$(SD_P)^2\left(\frac{1}{n_E\bar{X}_E^2} + \frac{1}{n_C\bar{X}_C^2}\right).$$

An approximate $100(1 - \alpha)\%$ confidence interval for the individual log response ratio parameter λ is given by

$$L - z_{\alpha/2}\sqrt{\nu} \le \lambda \le L + z_{\alpha/2}\sqrt{\nu} \tag{2}$$

where $z_{\alpha/2}$ is the $100(1 - \alpha/2)\%$ point of the standard normal distribution, and the corresponding confidence interval for the (unlogged) response ratio ρ is obtained by taking the antilogs of the confidence limits for the log response ratio. While L has a slight small-sample bias and its sampling distribution is slightly skewed, both bias and skew disappear rapidly as the smaller of $\sqrt{n_C}\bar{X}_C/SD_C$ or $\sqrt{n_E}\bar{X}_E/SD_E$ (which is usually $\sqrt{n_C}\bar{X}_C/SD_C$) becomes large. This approximation is quite accurate whenever the smaller of $\sqrt{n_C}\bar{X}_C/SD_C$ or $\sqrt{n_E}\bar{X}_E/SD_E$ is larger than ~ 3.

META-ANALYSIS OF RESPONSE RATIOS

Let k denote the number of studies and let λ_i, L_i, and ν_i be the log response-ratio parameter (the true log response ratio), the sample (estimated) log response ratio, and the variance (squared standard error) of the sample log response ratio, respectively, from the ith study, computed from Eq. 1. Thus the data from the entire collection of studies is the set of response ratios $L_1, \ldots, L_k$ and their variances $\nu_1, \ldots, \nu_k$. The goals of meta-analyses are typically to describe the distribution of the effect sizes associated with the set of experiments by estimating the mean and variance of that distribution, although details of analyses may differ depending on whether it is assumed that all experiments share the same true value of that effect size (fixed-effects model) or not (random- or mixed-effects models). Special analytic methods are needed because the observations $L_1, \ldots, L_k$ in meta-analysis are not expected to be identically distributed (e.g., Hedges and Olkin 1985, and see Gurevitch and Hedges 1999). That is, the variances of the observations $\nu_1, \ldots, \nu_k$ are assumed to be unequal and are typically markedly so.

Recent discussions of meta-analysis emphasize that analogous meta-analytic methods can be applied to many indices of effect size, and many methods that were developed independently for different metrics actually can be subsumed under the same global statistical framework (see Hedges 1994). Thus the methods discussed below can be used with any effect-size index that is approximately normally distributed and for which a consistent estimate of the variance is available.

Examining variation

There are two components of variation in the sample log response ratios. One component is due to sampling variation in the estimate for each experiment, that is variation of the estimate L_i about the parameter $\lambda_i = \ln(\rho_i)$. This component is quantified by ν_i, the square of the standard error of estimation, and would decline to zero if the sample size of the experiment were very large. The second component of variation is due to between-experiment variation in the experiment-specific parameters $\lambda_1, \ldots, \lambda_k$. This component represents the variation between experimental results that would remain even if the estimates from all of the experiments had negligible internal (standard) errors. This between-experiment variance is often of scientific interest because it quantifies the degree of true (nonsampling) variation in results across experiments.

A simple estimate of the between-experiment variance component can be derived from the statistic Q, which is used to test the statistical significance of this second variance component. This estimate, denoted $\hat{\sigma}_\lambda^2$ is given by

$$\hat{\sigma}_\lambda^2 = \frac{Q - (k - 1)}{\sum_{i=1}^{k} w_i - \frac{\sum_{i=1}^{k} w_i^2}{\sum_{i=1}^{k} w_i}} \tag{3}$$

where

$$Q = \sum_{i=1}^{k} w_i (L_i)^2 - \frac{\left(\sum_{i=1}^{k} w_i L_i \right)^2}{\sum_{i=1}^{k} w_i} \qquad (4)$$

and $w_i = 1/v_i$.

A test (at significance level α) of the hypothesis that the between-experiment variance component is zero, that is, a test of

$$H_0: \sigma_\lambda^2 = 0$$

is based on the Q statistic used in computing the variance component estimate. The test procedure consists of rejecting the null hypothesis whenever Q exceeds the $100(1 - \alpha)$ percentage point of the chi-squared distribution with $k - 1$ degrees of freedom.

The square root of the variance component estimate is a useful descriptive statistic to quantify between-study variation in results, and it is often useful to compare the variance component to the average of the v_i as an index of the amount of variance between experiments as compared to the average amount of variance due to sampling error within experiments. We believe that careful evaluation of the scientific aspects of the set of experiments should be the predominant consideration in any decision to combine experimental results into a summary. However, examination of between-study variation may be an important secondary consideration influencing the decision whether to combine a given set of estimates. If the between-experiment variation is too large (e.g., many times the average within-experiment sampling-error variation), one might question whether the experimental results are similar enough to warrant combination.

Estimation of the mean effect size

The most obvious summary of a set of effect estimates is the mean, or some other measure of central tendency. However the effect estimates from different experiments will typically differ in precision (standard error). Therefore a weighting of the individual study estimates giving greater weight to experiments whose estimates have greater statistical precision (smaller standard error) will increase the precision of the combined estimate. Consequently, a weighted mean is typically used in meta-analysis (e.g., Hedges 1983, DerSimonian and Laird 1986). The weighted mean of the log response ratio that produces greatest precision (minimum variance) is

$$\overline{L^*} = \frac{\sum_{i=1}^{k} w_i^* L_i}{\sum_{i=1}^{k} w_i^*} \qquad (5)$$

where $w_i^* = 1/(v_i + \hat{\sigma}_\lambda^2)$, the reciprocal of the total (unconditional) variance of L_i. When the number of studies is large, (say ≥ 50) the standard error of this weighted mean is given by

$$\mathrm{SE}(\overline{L^*}) = \sqrt{\frac{1}{\sum_{i=1}^{k} w_i^*}} . \qquad (6)$$

However if the number of studies is not large, statistical inference based on the variance estimate given in Eq. 6 may be seriously misleading, particularly if the within-study sample sizes are small (e.g., $n < 5$ per group). A more accurate estimate of the standard error of the weighted mean is

$$\mathrm{SE}(\overline{L^*}) \qquad (7)$$

$$= \sqrt{\left[\frac{1}{\sum_{j=1}^{k} w_j^*} \right] \left\{ 1 + 4 \sum_{i=1}^{k} \frac{1}{df_i} \left(\frac{w_i^*}{w_i} \right)^2 \frac{w_i^* \left[\left(\sum_{j=1}^{k} w_j^* \right) - w_i^* \right]}{\left(\sum_{j=1}^{k} w_j^* \right)^2} \right\}}$$

where df_i is the number of degrees of freedom in the ith study ($n_{iE} + n_{iC} - 2$).

Because each of the individual effect estimates (log response ratios) is approximately normally distributed, the weighted mean would be normally distributed if the weights were constants. However, here—and in most meta-analyses—the weights are not truly constants because they depend upon the data. However, the asymptotic distribution of the weighted mean is still normal with the same variance as if the weights were constants, and the normal approximate distribution based on that asymptotic distribution has proven to be quite accurate in other meta-analytic situations (see Rao 1973: 389–391, Hedges 1983, Hedges and Olkin 1985, Cooper and Hedges 1994: 36–38). Thus the weighted mean is approximately normally distributed and a $100(1 - \alpha)\%$ confidence interval for the average log response ratio μ_λ is given by ($\mathrm{CL_L}$, $\mathrm{CL_U}$):

$$\mathrm{CL_L} = \overline{L^*} - z_{\alpha/2}\mathrm{SE}(\overline{L^*}) \leq \mu_\lambda \leq \overline{L^*} + z_{\alpha/2}\mathrm{SE}(\overline{L^*})$$

$$= \mathrm{CL_U} \qquad (8)$$

where $z_{1-\alpha}$ is the $100(1 - \alpha/2)\%$ point of the standard normal distribution, and the corresponding confidence interval for the mean (unlogged) response ratio, μ_ρ, is obtained by taking antilogs. Thus the confidence interval for μ_ρ is

$$\exp(\mathrm{CL_L}) \leq \mu_\rho \leq \exp(\mathrm{CL_U}). \qquad (9)$$

Note that the confidence interval for the log response ratio is symmetric, but the back-transformed confidence interval for μ_ρ, the unlogged mean response ratio, will not be symmetric about the point estimate, $\exp(\overline{L^*})$, of the response ratio. Back-transforming the mean of the logs introduces a bias into the estimate of the mean response ratio due to the convexity of the log transform (Jensen's inequality; e.g., see Feller 1966, Rao 1973: 58). This bias also arises, for example, in

the averaging of correlation coefficients by backtransforming the average of several Fisher z-transforms, or in the averaging of odds ratios by backtransforming the average of several log-odds ratios. However, since the magnitude of the bias depends upon the variance of the weighted mean (which will typically be small if the number of studies and the precision of their estimates are not both small), the bias will usually be slight.

Sample size requirements

Simulation studies suggest that there are roughly three situations that need to be distinguished. If the number of studies is large (e.g., $k \geq 50$), 95% confidence intervals based on the standard error given in Eq. 6 are reasonably accurate (e.g., the actual probability content of the confidence intervals is at least 93%) as long as no more than ~20% of the studies have sample sizes of $n_E = n_C = 2$. If the number of studies is large but there are many studies with very small sample sizes, or if the number of studies is intermediate (e.g., $20 \leq k \leq 50$), then confidence intervals based on the standard error given in Eq. 7 are reasonably accurate, while those based on Eq. 6 are too narrow. Finally, if the number of studies is small (e.g., $k \leq 20$), confidence intervals based on either Eq. 6 or Eq. 7 may be too narrow, with the actual probability content of confidence intervals based on Eq. 6 being lower than 89% and that based on Eq. 7 being as low as 91%.

Example: effects of elevated CO_2 on woody-plant growth

Curtis and Wang (1998) collected 102 response-ratio estimates on the effects of elevated CO_2 on total biomass (above- plus below-ground) of woody plants. In these experiments, plants were grown under controlled growth or field conditions within chambers allowing the manipulation of atmospheric CO_2 partial pressure. Control treatments received air at current ambient CO_2 levels (<40 Pa) while experimental treatments received air at approximately twice current ambient CO_2 levels (60–80 Pa). The data set (see Appendix B) is rather typical in that it exhibits much of the diversity that is common in research of this type. Synthesizing the responses of plants to elevated CO_2 across studies is important for model parameterization and ultimately for the implications it has for both ecology and policy. Meta-analysis offers a uniquely valuable tool for this purpose.

We first note that the values of the standardized means of the control group, $\sqrt{n_C}\bar{X}_C/SD_C$ are relatively large. Only two of the values are <3 and 84% exceed 6. Therefore the L's will have little bias and the normal approximation to their sampling distribution should be quite good. The within-study sample sizes ranged from $n_E = n_C = 2$ (in 24% of the studies) to $n_E = n_C = 48$ (in 4% of the studies), with about 50% of the studies having sample sizes greater than $n_E = n_C = 5$. Because

the number of studies is large and the preponderance of within-study sample sizes is larger than $n_E = n_C = 2$, confidence intervals based on the standard error given in Eq. 6 might be expected to be reasonably accurate.

In addition, the between-studies variance component, given in Eq. 3 is $\hat{\sigma}_\lambda^2 = 0.021$, approximately the same size as the average within-experiment variance, $\Sigma v_i/k = 0.023$ (where k = number of studies), suggesting that this collection of studies is similar enough to warrant combination. However, we also note that $\hat{\sigma}_\lambda^2$ is highly significant ($Q = 768.9$, $P < 0.001$), indicating statistically reliable (nonsampling) variation across studies.

The weighted mean log response ratio computed from Eq. 5 is $\overline{L^*} = 0.253$, and its standard error computed from Eq. 6 is SE $(\overline{L^*}) = 0.0182$. Using this standard error in Eq. 8 to compute the 95% confidence interval for $\overline{L^*}$ yields

$$0.217 = 0.253 - 1.96(0.0182) \leq \mu_\lambda$$
$$\leq 0.253 + 1.96(0.0182) = 0.288.$$

Taking antilogs to examine the (unlogged) response ratios yields an estimate of μ_ρ of $1.29 = \exp(0.253)$ and a 95% confidence interval for μ_ρ of

$$1.24 = \exp(0.217) \leq \mu_\rho \leq \exp(0.288) = 1.33.$$

The standard error computed from the more accurate formula given in Eq. 7 yields $SE\overline{L^*} = 0.0188$. Using this standard error in Eq. 8 to compute the 95% confidence interval for $\overline{L^*}$ yields

$$0.216 = 0.253 - 1.96(0.0188) \leq \mu_\lambda$$
$$\leq 0.253 + 1.96(0.0188) = 0.290.$$

Taking antilogs to examine the (unlogged) response ratios yields an estimate of μ_ρ of $1.29 = \exp(0.253)$ and a 95% confidence interval for μ_ρ of

$$1.24 = \exp(0.216) \leq \mu_\rho \leq \exp(0.290) = 1.34.$$

Therefore the use of the more accurate (expression Eq. 8) for the standard error would have produced almost identical results in this study.

By comparison, the unweighted average $\overline{L}$ (i.e., $\Sigma L_i/k$) for this data set = 0.284, 12% greater than $\overline{L^*}$ but within the 95% confidence interval of the weighted average. Although these two estimates are comparable in magnitude, the confidence interval computed from $\overline{L^*}$ is ~20% narrower than that based on elementary procedures for computing a confidence interval for the mean of identically distributed observations using the t distribution. However, the confidence interval based on the assumption of identically distributed observations (i.e., the unweighted L's) would not necessarily be valid for these data because the variances of the observations differ by a factor of as much as 1100 to 1.

The significant between-studies variance component suggests that partitioning the data set could reveal

groups of studies whose L's differ from one another. Curtis and Wang (1998) tested several predictions regarding the magnitude of the plants' responses, coding each observation according to whether plants were exposed to other treatments in factorial combination with elevated CO_2 (e.g., low soil nutrients, high ozone levels, drought, etc.). Their meta-analysis revealed, among other findings, that there were statistically significant differences in biomass response to CO_2 between unstressed plants (+31%) and plants stressed by low soil nutrients (+16%), although there was a significant response to CO_2 even under the nutrient-stressed conditions. That is, elevated atmospheric CO_2 will stimulate net biomass gain in trees even when trees are stressed by low soil-nutrient availability. Thus, the use of L worked well in the meta-analysis of this large data set, revealing information that would not otherwise have been available.

Example: effects of competition on plant biomass

Not every collection of ecological experiments will be suitable for meta-analysis by response ratios using the methods given here. Gurevitch, et al. (1992) collected data from field experiments on the effects of competition on biomass, which they analyzed using the d index. We examined a subset of that data, the results of the 77 experiments on plant biomass, and found that they were not suitable for analysis via response ratios. Nine of the studies had zero biomass in the control groups, making computation of a response ratio impossible. Of the 68 remaining experiments with non-zero values of $\bar{X}_C$, 16 (23.5%) had values of $\sqrt{n_C}\bar{X}_C/$ SD_C that were <2.0, and 25 (36.8%) had values <3.0. Therefore more than one third had values of the denominator that were too small for the normal approximation to be adequate. Thus we would not advise meta-analysis of these data using response ratios via the methods presented in this paper.

EXTENSION TO OTHER ANALYSES

Other more complex meta-analytic procedures (such as the mixed-model approach to comparing groups of experiments discussed by Gurevitch and Hedges [1993, 1999], or any of the procedures in Cooper and Hedges [1994]) can also be applied to response ratios. In general all that is required is to substitute the log response ratio for the effect-size index and the variance given in this paper (Eq. 1) for the variance of the effect-size index in the formulas for estimates and test statistics. Resampling methods provide a nonparametric approach to calculating confidence limits on response ratios and testing for homogeneity among groups (see, e.g., Rosenberg et al. 1997). If the sample variances are not available, resampling methods may still be able to provide homogeneity tests and confidence intervals around average effect sizes, but cannot provide estimates of between- or within-experiment variance components (Adams et al. 1997).

Finally note that other indices of effect size are simple functions of the response ratio. For example, Osenberg et al. (1997) proposed using indices of the form

$$t^{-1}\log(\bar{X}_E/\bar{X}_C),$$

where t is a measure of the duration of the experiment, while other ecologists employ $(\bar{X}_E - \bar{X}_C)/\bar{X}_C$. The methods given in this paper for meta-analysis of response ratios can also be used (with slight modification) to conduct meta-analyses using these and related indices.

The index $t^{-1}\log(\bar{X}_E/\bar{X}_C)$ is simply equal to a constant (t^{-1}) times the log response ratio. Multiplication by a constant does not affect the normality of the log response ratio, but makes the variance t^{-2} times as big as the variance of the log response ratio. Consequently the analysis of the index $t^{-1}\log(\bar{X}_E/\bar{X}_C)$ is exactly like that for the log response ratio except that the variance for this effect size is t^{-2} times as big.

The relative competition index (RCI) commonly used by plant ecologists, $(\bar{X}_{NC} - \bar{X}_C)/\bar{X}_{NC}$, is just one minus the response ratio (that is, $(\bar{X}_{NC} - \bar{X}_C)/\bar{X}_{NC} = 1 - \bar{X}_C/\bar{X}_{NC}$), where $\bar{X}_{NC}$ indicates the mean performance of plants grown with no competitors, and $\bar{X}_C$ those with competition (cf. Goldberg et al. 1999). Note that here the group with no competitors is treated, for the purposes of computing the index, as the control group. Therefore it would be appropriate to transform the effects (by subtracting from 1) and conduct the analysis in the metric of log response ratios, back-transforming the results of the analysis back to the original metric by subtracting 1 and changing the sign.

CONCLUSIONS

Meta-analytic methods have been adapted to summarize experimental evidence in many areas of the behavioral, medical, and social sciences. One aspect of that adaptation has been the development of analytic procedures so that indices of effect can be used that are most meaningful for that research. The response ratio is an appealing index of effect size for many ecological experiments. When the standardized mean of the denominator is not too small, the bias of the log response ratio is small and its sampling distribution is approximately normal. In these cases, the log response ratio is a suitable index for meta-analysis. This paper demonstrates how to use response ratios to conduct a simple meta-analysis. The approach suggested here has the advantage over naive summaries in that it provides statistical confidence statements (confidence intervals) for the summary of effects across experiments. It also provides a quantification of between-experiment variation that may be useful in the interpretation of the results.

SPECIAL FEATURE

Barbara, and the State of California. Development of the elevated CO_2 database was assisted by Xianzhong Wang and supported by the interagency program on Terrestrial Ecosystems and Global Change, USDA grant no. 9504800 to P. S. Curtis. We received valuable comments from Allan Stewart-Oaten, Philip Dixon, and an anonymous reviewer.

LITERATURE CITED

Adams, D. C., J. Gurevitch, and M. S. Rosenberg. 1997. Resampling tests for meta-analysis of ecological data. Ecology **78**:1277–1283.

Arnqvist, F. R., and D. Wooster. 1995. Meta-analysis—synthesizing research findings in ecology and evolution. Trends in Ecology and Evolution **10**:236–240.

Cooper, H. M., and L. V. Hedges,editors. 1994. The handbook of research synthesis. Russell Sage Foundation, New York, New York, USA.

Cooper, S. D., S. J. Walde, and B. L. Peckarsky. 1990. Prey exchange rates and the impact of predators on prey populations in streams. Ecology **71**:1503–1514.

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.

DerSimonian, R., and N. Laird. 1986. Meta-analysis in clinical trials. Controlled Clinical Trials **7**:177–188.

Downing, J. A., C. W. Osenberg, and O. Sarnelle. 1999. Meta-analysis marine nutrient-enrichment experiments: variation in the magnitude of nutrient limitation. Ecology **80**:1157–1167.

Englund, G., Sarnelle, O., and Carpenter, S. 1999. The importance of data-selection criteria: Meta-analyses of stream predation experiments. Ecology **80**:1132–1141.

Feller, W. 1966. Introduction to probability theory and its applications. Volume II. John Wiley & Sons, New York, New York, USA.

Fieller, E. C. 1932. The distribution of the index in a normal bivariate population. Biometrika **24**:428–440.

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

Goldberg, D. E., T. K. Rajaniemi, J. Gurevitch, and A. Stewart-Oaten. 1999. Empirical approaches to quantifying interaction intensity: competition and facilitation along productivity gradients. Ecology **80**:1118–1131.

Gunderson, C. A., and S. D. Wullschleger. 1994. Photosynthetic acclimation in trees to rising atmospheric CO_2: a broader perspective. Photosynthesis Research **39**:369–388.

Gurevitch, J., and L. V. Hedges. 1993. Meta-analysis: combining the results of independent studies in experimental ecology. Pages 378–398 *in* S. Scheiner and J. Gurevitch,

editors. The design and analysis of ecological experiments. Chapman & Hall, London, UK.

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

Gurevitch, J., L. L. Morrow, A. Wallace, and J. S. Walsh. 1992. A meta-analysis of field experiments on competition. American Naturalist **140**:539–572.

Hedges, L. V. 1983. Combining independent estimators in research synthesis. British Journal of Mathematical and Statistical Psychology **36**:121–131.

———. 1994. Statistical considerations in research synthesis. Pages 29–38 *in* H. Cooper and L. V. Hedges, editors. The handbook of research synthesis. Russell Sage Foundation, New York, New York, USA.

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

Hinkley, D. V. 1969. On the ratio of two correlated normal random variables. Biometrika **56**:635–639.

———. 1970. Correction. Biometrika **57**:683.

Kimball, R. A. 1983. Carbon dioxide and agricultural yield: an assemblage and analysis of 430 prior observations. Agronomy Journal **75**:779–788.

Laska, M. S., and J. T. Wootton. 1998. Theoretical concepts and empirical approaches to measuring interaction strength. Ecology **79**:461–476.

Marsaglia, G. 1965. Ratios of normal variables and ratios of sums of uniform variables. Journal of the American Statistical Association **60**:193–204.

Meier, P. (1953). Variance of a weighted mean. Biometrics **9**:59–73.

Osenberg, C. W., O. Sarnelle, and S. D. Cooper. 1997. Effect size in ecological experiments: the application of biological models in meta-analysis. American Naturalist **150**:798–812.

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.

Poorter, H. 1993. Interspecific variation in growth response of plants to an elevated ambient CO_2 concentration. Vegetatio **104/105**:77–97.

Rao, C. R. 1973. Linear statistical inference. John Wiley & Sons, New York, New York, USA.

Rosenberg, M. S., D. C. Adams, and J. Gurevitch. 1997. MetaWin. Statistical software for conducting meta-analysis: fixed effect models, mixed effect models, and resampling tests. Version 1.0. Sinauer Associates, Sunderland, Massachusetts, USA.

Sarnelle, O. 1992. Nutrient enrichment and grazer effects on phytoplankton in lakes. Ecology **74**:551–560.

APPENDIX A

TECHNICAL DETAILS

Let $X_1, \ldots, X_n$ be the observations in the treatment group and let $Y_1, \ldots, Y_m$ in the control group. Assume that the data collected in the two groups are independently normally distributed within each group. That is, assume that

$$X_i \sim N(\mu, \sigma_X^2) \qquad i = 1, \ldots, n$$

$$Y_i \sim N(\eta, \sigma_Y^2) \qquad i = 1, \ldots, m.$$

Let $\bar{X}$ and SD_X be the mean and standard deviation of $X_1, \ldots, X_n$ and $\bar{Y}$ and SD_Y be the mean and standard deviation of $Y_1, \ldots, Y_m$, respectively. Then $\bar{X}$ and $\bar{Y}$ are independently normally distributed with means μ and η and variances σ_X^2/n and σ_Y^2/m, respectively. The sampling distribution of the ratio $R = \bar{X}/\bar{Y}$ has been studied by Geary (1930), Fieller (1932), Marsaglia (1965), and Hinkley (1969, 1970). If $\sqrt{m}\eta/\sigma_Y$ is sufficiently large that there is negligible probability that $\bar{Y}$ is

negative (or if this is impossible for practical reasons), then the sampling distribution of R (the probability density function) is given by

$$f(R) = \frac{1}{\sqrt{2\pi}} \frac{\sqrt{mn}(nr\mu\sigma_Y^2 + m\eta\sigma_X^2)}{(nR^2\sigma_Y^2 + m\sigma_X^2)^{3/2}}$$

$$\times \exp\left[-0.5\left(\frac{mn(\mu, R\eta)^2}{nR^2\sigma_Y^2 + m\sigma_X^2}\right)\right]. \qquad (A.1)$$

When $\sigma_X^2 = \sigma_Y^2$, the distribution of R depends only on the two standardized means $\sqrt{n}\mu/\sigma$ and $\sqrt{m}\eta/\sigma$ or, equivalently, their inverses, which are the coefficients of variation. Therefore in evaluating the adequacy of approximations to the distribution of R or transformations of R, we only need to consider the range of situations defined by these two parameters. When-

ever $\eta \neq 0$, the large sample approximation to the distribution of R is normal, with mean μ/η and variance

$$\left[\frac{\mu}{\eta}\right]^2 \left[\frac{\sigma_X^2}{n\mu^2} + \frac{\sigma_Y^2}{m\eta^2}\right].$$

Comparing the exact distribution of R with the normal approximation reveals that they are very similar if $\sqrt{m}\eta/\sigma_Y$ is large, but the exact distribution is positively skewed. This skewness can be marked if $\sqrt{m}\eta/\sigma_Y$ is small but decreases rapidly as this quantity increases.

Transforming by taking the natural logarithm of R reduces the skewness substantially. The exact probability density function of $L = \ln(R)$ is given by

$$g(L) = \exp(L)f(\exp(L))$$

where $f(R/)$ is the probability density function of R given in Eq. A.1. The large-sample approximation to the distribution of L is that L is normal with mean $\lambda = \ln(\mu/\eta)$ and variance

$$\frac{\sigma_X^2}{n\mu^2} + \frac{\sigma_Y^2}{m\eta^2}.$$

Comparing the exact distribution of L with the normal approximation shows that they are very similar in most cases.

One question that might arise is the bias of L as an estimate of λ, the true log response ratio. The fact that distribution of L depends only on $\sqrt{n}\mu/\sigma$ and $\sqrt{m}\eta/\sigma$ implies that the bias depends only on $\delta = \sqrt{m}\eta/\sigma$ and λ. Integrating the exact distribution of L to obtain the bias shows that the bias of L can be substantial if δ is small, but it decreases rapidly as δ

increases. For example when $\rho = 1.5$ ($\lambda = 0.40$), the relative bias of L is ~15% for $\delta = 2$, but decreases to <5% for $\delta > 4$ and to <1% for $\delta > 9$. Since L will usually be used when δ is relatively large, bias would not appear to be large enough to be problematic.

Therefore the large-sample, normal approximation to the distribution of $L = \ln(R)$ will be adequate in most meta-analytic situations, particularly since the exact normality of each effect size estimate is less crucial than the approximate form, which is used to compute the sampling standard error for weighting purposes. Indeed it appears that L is actually better behaved than some indexes (such as the odds ratio from 2×2 tables) routinely used in meta-analysis in medicine and epidemiology.

The distribution of L is used by computing an estimate of the variance of L, and substituting the sample mean and standard deviation for the population parameters in the expression for the variance to yield

$$v = \frac{(SD_X)^2}{n\bar{X}^2} + \frac{(SD_Y)^2}{m\bar{Y}^2}$$

which is the weighted sum of the squared coefficients of variation.

The variance estimate underlying Eq. 6 is derived from an asymptotic model that assumes large within-study sample sizes so that the weights are essentially known. The expression given in Eq. 7 is derived by applying a theorem given in Meier (1953) to take account of the uncertainty of the (estimated) weights themselves in computing the variance of the weighted mean.

APPENDIX B

A data set of 102 response-ratio estimates collected by Curtis and Wang (1990) is available in ESA's Electronic Data Archive: *Ecological Archives* E080-008.