A statistical view of synthesizing patterns of species richness along productivity gradients: devils, forests, and trees

JESSICA GUREVITCH1,3 AND KERRIE MENGersen2

1Department of Ecology and Evolution, Stony Brook University, Stony Brook, New York 11794-5245 USA
2School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia

INTRODUCTION

Robert Whittaker (2010) offers a critique of quantitative research syntheses attempting to generalize species richness patterns along gradients of productivity. As he says, the results of such syntheses have been controversial and disagreed in their conclusions. Beginning with a large research synthesis by Mittelbach et al. (2001), a number of authors (Gillman and Wright 2006, Pärtel et al. 2007, Laanisto et al. 2008) have attempted to classify patterns from individual studies by the shape of the responses; Whittaker and Heegaard (2003) criticized what they felt were methodological flaws in the Mittelbach et al. (2001) paper and the critique was rebutted by Mittelbach et al. (2003). Due to what he feels are persistent methodological flaws in the papers attempting quantitative syntheses of this literature, and because he now believes that it is impossible to carry out meaningful meta-analyses on this relationship, Whittaker (2010) recommends an end to meta-analyses on this topic, expresses concern over whether quantitative data synthesis is a legitimate and repeatable approach to making sense of the data on this question, and suggests a profound change in the way meta-analyses are conducted and reviewed in ecology.

We offer a short discussion of why we feel that Whittaker’s dismissal of meta-analysis is inappropriate, add a brief critique of this literature of our own from a statistical perspective, and most importantly, point the way to improved statistical approaches. While the devil is certainly in the details, we also don’t want to lose sight of the forest for the trees.

While Whittaker is correct in demanding unambiguous and transparent criteria for selecting studies and carrying out the meta-analysis, this is not a legitimate argument against the use of meta-analysis. In fact, because there are established criteria for carrying out systematic reviews and quantitative research syntheses—a broader field in which meta-analysis is one component—meta-analyses that follow contemporary established protocols are more likely to be repeatable than other forms of literature reviews (e.g., Borenstein et al. 2009). Whittaker is among many authors who have suggested specific criteria for study inclusion in ecological research synthesis; for example, Hillebrand and Cardinale (2010) argue for using different criteria. Just as the past four decades of debate about the process of systematic review in medicine have seen the establishment of widely accepted standards and protocols by the Cochrane Collaboration, so this engaged discussion can contribute positively to the development of the application of scientific principles and methodology to research synthesis in ecology (description of the Cochrane Collaborations available online). More broadly, various formal techniques have been applied to account for variation in study quality in meta-analysis, including weighting by both inverse variance and study quality, or including moderators to account for methodological flaws in the statistical models employed (e.g., Cooper et al. 2009). Methods for dealing with variation in selection criteria among meta-analyses have been developed focusing on levels of generalizability of the outcomes (Sutton et al. 2000, Wolpert and Mengersen 2004; B. J. Becker and A. Aloe, unpublished data). Similar problems are endemic to research review and synthesis regardless of the method of synthesis that is adopted. For example, they are just as much a problem in qualitative evaluation as they are in quantitative pooling of effect sizes. A model-based approach to meta-analysis allows for the possibility of formal, transparent adjustment for selection bias and similar problems, whereas this is much more difficult with narrative methods or vote counts.

A second reason that we do not find Whittaker’s dismissal of meta-analysis convincing is that flaws in particular research syntheses, real or perceived, do not invalidate meta-analysis itself. By carrying out a systematic review or quantitative research synthesis using flawed aims, assumptions or data, one may indeed undermine the validity of its conclusions. However, this is true for any scientific enterprise or statistical technique.
We would not conclude that we should dismiss all future attempts at ecological lab or field work because of individual flawed studies, and likewise, while regression and other statistical techniques in ecology are often applied incorrectly and inappropriately, we would be reluctant to discard their use on that basis. Otherwise, we would be back to narrative case studies and individual natural history observations, as were standard fare until the 1960s when modern statistical techniques first became widely adopted in this field (Simpson et al. 1960).

Similar discussions occurred in medicine over a decade ago (see review and discussion, e.g., by Borenstein et al. 2009:384–386). Different issues regarding the adoption of meta-analysis methodology have been raised in different disciplines. For instance, in medicine, there were fierce debates about the use of fixed vs. random effects models in meta-analysis; this has not been an issue at all in the ecological literature. On the other hand, a number of key similarities exist in the arguments for and against the incorporation of scientific principles for research review (including meta-analysis) across disciplines. Issues such as how to define and account for variation in study quality in meta-analysis methodology, for instance, are common in all disciplines. Lau et al. (in press) explore the comparisons between the history of the adoption of meta-analysis in medicine, social scientific research, and ecology, and the parallels and contrasts are enlightening.

Whittaker (2010) suggests that more appropriate solutions than meta-analysis to understanding the diversity-productivity relationship are to devise rigorous field studies that will make meaningful contributions to the question, or to undertake a narrative review. Devising more field studies may be useful for filling in knowledge gaps, improving the quality of the available data and for many other reasons, but it will not help synthesize the results of the existing studies. Moreover, no one study can replace all of the existing studies or be complete enough to resolve this question across all systems, organisms, and scales. Interestingly, this parallels a debate in medicine, where the relative value of research syntheses and very large clinical trials with tens of thousands of subjects and sometimes lasting many years has been discussed at length; e.g., Lau et al. (1992), LeLorier et al. (1997), and Ioannidis et al. (1998). Nor will narrative reviews provide closure on this question. If a narrative review seeks generalizations, it will face most of the same problems that Whittaker (2010) finds with quantitative reviews (and others, besides, including reviewer bias and vote counting; e.g., Lipsey and Wilson 1993, Sharpe 1997). On the other hand, we stand to gain little from a narrative review that discusses each study as a unique example that is not comparable to others and where there is no generalization possible. In fact, if it is utterly impossible to generalize from the results of a study, or to compare studies to reach generalizations, we would argue that there is little value in the individual studies as well, because the exact circumstances of any one ecological study are unlikely to ever occur again.

Whittaker’s final recommendation is to use “best evidence synthesis” (Slavin 1995). Slavin’s ideas have been developed considerably since the publication of that and an earlier paper (Slavin 1986). Slavin’s recommendations for clear problem statement, formalized literature search, critical literature review, evidence tabulation, and qualitative synthesis are now part of (albeit with some controversy) the larger body of work on systematic reviewing, which we cannot discuss at any length here, and many of these elements are standard practice in high quality meta-analyses. With the exception of qualitative synthesis, which is prone to biases and may indeed be less transparent than the quantitative approaches, these activities should be part of any synthesis, whether it is nominally “best evidence” or “meta-analysis.” The adoption of these methods and their incorporation into current practice is an example of substantive progress in the establishment of systematic criteria for literature review.

As pointed out by Hillebrand and Cardinale (2010) and by many others, meta-analysis does not necessarily require that the aims of the original studies be the same, that there are no modifying factors that differ among studies, or that sampling schemes and study designs be identical among studies, as claimed by Whittaker (2010). These are neither conceptual nor statistical requirements for quantitative data synthesis, and rigorous statistical methodology has been developed to deal with all of these issues. Modifying factors that influence the outcomes and that vary among studies can in many cases be modeled. Of course, if there is true confounding among these factors, this can limit the inferences possible. While none of these issues are uniquely problematic for meta-analysis, unlike meta-analysis, other methods (e.g., narrative reviews) have not developed methods for addressing them. It is, of course, essential that studies be synthesized in a meaningful manner, and this can be challenging.

LIMITATIONS TO VOTE COUNT APPROACHES

Each of the quantitative research syntheses on the productivity–diversity relationship from Mittelbach et al. (2001) on proceed by determining the shape of the curve in each of the primary studies being synthesized according to various criteria, and then tallying the numbers in each shape category and comparing them. Due to different criteria (and other factors) they arrive at different numbers of unimodal, U-shaped, and positive and negative monotonic curves across the literature. A fundamental problem with this approach is that the results are obtained using a statistical technique known as vote counting (Hedges and Olkin 1980; Gurevitch and Hedges 1999, Borenstein et al. 2009), in which studies are judged by their significance.
levels to “cast a vote” in favor of or against a particular outcome. Vote counting involves simple estimation of the proportion of studies that show a “significant” effect (where the definition of statistical significance may vary from one reviewer to another) in response to a specified hypothesis. In the species richness and productivity assessment, the primary test is whether the quadratic coefficient is not equal to zero (i.e., there is a curvilinear relationship).

The disagreement and confusion found among these studies is precisely what one would predict from a set of vote counts, because it is a flawed statistical technique that results in biased and inconsistent outcomes when used as a research synthesis method. Unfortunately vote counting is not only a weak form of inference, it is potentially misleading and may not provide the answers to the questions authors are generally most interested in addressing when synthesizing results across studies: the overall magnitude and direction of a parameter or an effect (such as a slope) and explanations for variation in that effect. In the case of the diversity–productivity relationship, there are various potentially important covariates and confounders, including the issue of scale. Because of the statistical problems with vote counts, the practice of applying simple or elaborate statistical tests to vote counts are likely to be misleading as well. Vote counting has essentially been abandoned in other disciplines but continues to be relied upon in ecology. Vote counting is not meta-analysis although it is sometimes misidentified as such in the ecological literature. At least some of the arguments made by Whittaker (2010) are valid critiques of vote counting, rather than meta-analysis.

Some of the reasons given by ecologists for adopting vote counting are that the data are not available for doing a meta-analysis, that the results are too heterogeneous to warrant a formal meta-analysis, and that vote counts are somehow more conservative or reliable than meta-analysis because they appear to have fewer assumptions (e.g., Ives and Carpenter 2007, Tylianakis et al. 2008).

In fact, like any statistical technique, vote counts also are based on assumptions, although these are often not examined. Missing data (e.g., means, sample sizes, and errors) can indeed pose major challenges for meta-analysis. Methods have been proposed to deal with partial missing data (e.g., Fahrbach 2001, Pigott 2009; Lajeunesse and Schmid, in press) but if too much data are unreported it may be impossible to accurately synthesize the data quantitatively; vote counts will not provide more precise or accurate syntheses in this case. If an effect is consistently reported over different scales, locations, time periods, and study designs, a vote count may provide some support for a true association. However, the analyst should clearly state the reasons for the adoption of vote counting despite its limitations and the limited inferences that can be made on the basis of such an analysis. Vote counts may also prove to be of some use in a first-pass exploratory data analysis, to gauge the overall patterns of responses. Another alternative where the analyst feels that data are too limited for formal meta-analysis is combining t statistics. Like vote counting and combining P values, the combination of t statistics can accommodate heterogeneous study results, but unlike vote counting or combining P values, it has the advantage of taking into account the magnitude of the study-specific effect estimates. This approach is an improvement on the combination of P values, since the latter does not discriminate between positive and negative values, but still suffers from other major drawbacks (e.g., see Becker and Wu 2007).

In the case of the productivity–diversity relationship, the results are not consistent across studies, but vote counting is not a good tool for analyzing the sources of this heterogeneity. If the results of a group of studies are strongly heterogeneous or cannot be statistically combined for other reasons, one option might be to conduct a descriptive narrative review that does not depend on the P values of the outcomes of the studies being combined. A review that categorizes studies by characteristics other than those based on the statistical significance of the outcomes may be informative and meaningful; for instance, a review that finds that 70% of studies on invasive species concern only plants is an interesting and useful finding and is not subject to the limitations of vote counts based on significance tests. If a researcher deems that a group of studies is irreducibly heterogeneous at all levels of generality, it may be worthwhile to reframe the research question to something more meaningful and tractable.

The charge of misleading inferences arising from vote counting follows in part from two major drawbacks of this approach: it takes no account of the magnitude of the effect or of the uncertainty in the estimate of that effect (i.e., the confidence interval around the effect estimate). As a very simple example, if vote counting of “statistically significant” studies is used to determine if an effect is substantiated across studies, and if one study shows a very strong quadratic relationship and two studies show a “nonsignificant” quadratic effect, vote counting would, possibly erroneously, lead to a conclusion of “inconsistent effects” or of “no overall effect.” Similar problems exist if counts of reported (or computed) “significant” U-shaped or hump-shaped relationships are compared. Moreover, a change in the threshold for tests of “significance” can subtly or dramatically change the “votes” and thus the resultant inferences. Mittelbach et al. (2003:3393) acknowledged this issue, stating that “the strength of the quadratic terms is a legitimate issue separate from its existence and this is not something we attempted to address...” More formally, the vote count estimate does not meet any of the criteria for a good estimator—it is not unbiased, consistent, or sufficient—so its usefulness in
providing meaningful quantitative syntheses is quite limited.

**Statistical Approaches to Quantitative Synthesis of the Diversity–Productivity Relationship**

What are the alternatives to vote counting? Contemporary meta-analysis practice relies upon on a model-based combination of the study-specific data. In brief, meta-analysis involves obtaining a measure of the effect from each study, called an “effect size” (such as a standardized mean difference, response ratio, correlation coefficient, or a regression coefficient), weighting the effect sizes by the inverse of their sampling variances, and then modeling these weighted mean outcomes across studies. We note that there are substantial benefits to weighting in this way, which is why it has become standard practice in meta-analysis. These include being able to model the within- and between-study heterogeneity, and accounting for differences among studies in the precision of the effect size estimate. Weighting effect estimates by their respective (inverse) variances has mathematical properties that may be lost if the weights are formed using some other, arbitrary criteria.

Ecologists have occasionally objected to using variances in meta-analysis either as a basis for effect size calculations (e.g., as used in standardized mean differences) or as weights, with the rationale that there may be systematic differences between field and lab studies in the magnitude of variances, creating a statistical bias in favor of lab-based studies. Curiously this hypothesis (that lab-based studies have smaller variances, or larger effect sizes on average than field-based studies), while not unreasonable, has never been demonstrated to be true. Moreover, the overall effect estimates that are obtained with arbitrary weights must be auditable (i.e., based on a robust mathematical justification for selection of the weights) and must be interpretable. This is still a matter of great debate in the statistical literature. In any case, if there are systematic differences such that bias is introduced by combining two very different types of data, the synesthetist certainly has the option of analyzing those two groups of studies separately rather than combining them. This is a better alternative than discarding a valuable statistical tool for which there is no obvious substitute.

The techniques for modeling the outcomes (i.e., effect sizes) across studies have experienced considerable development over the past four decades, and can range from the very simple—e.g., finding a weighted grand mean across studies and its confidence limits—to modeling variation in the effects across studies, including both frequentist and Bayesian approaches (e.g., Hedges and Olkin 1985, Borenstein et al. 2009, Cooper et al. 2009). These techniques offer statistically unbiased and robust means for asking questions about the overall magnitude and direction of the effect and about heterogeneity among studies, i.e., variation in the magnitude, confidence limits (or credible intervals), and statistical significance of that effect.

Let us assume that the meta-analyst has compiled a set of studies that are biologically relevant, satisfy conditions of comparability of scale, meet study quality and design thresholds, report consistently on important covariates (that is, factors influencing the nature of the relationship between productivity and diversity), and provide either primary or summary data about the relationship of interest. In the present case, this relationship is the association between species richness and productivity; the primary data available from each study would comprise a set of pairwise values of the two variables over a defined range This is one of the approaches taken by Mittelbach et al. (2001), who obtained relevant data from the 171 studies published and then reanalyzed the available study-specific data using well-defined linear and quadratic regression models in a formal, if limited, meta-analysis. Alternatively, the summary data from each paper would be the regression parameter estimates (intercept, linear coefficient, quadratic coefficient, and so on), the corresponding standard errors of these estimates and/or the t or P values resulting from the hypothesis that the parameter estimates are equal to zero. An interesting recent hybrid between a very large primary study and the use of study-specific data synthesis on the productivity–diversity relationship in marine systems was recently published by Witman et al. (2008).

The compilation of primary data from all studies is often argued to be a “gold standard” approach in meta-analysis. One important benefit is that synthesis based on primary data allows for consistent analysis of data within studies and thus provides a directly comparable set of study-specific estimates for input into a meta-analysis. There are also drawbacks to obtaining and analyzing study-specific data: the process is extremely time-consuming and it is often not possible to obtain the required data from all identified studies. It may be possible to extract data from published figures if the original data are not available, although this may potentially introduce inaccuracies (Whittaker and Hegardt 2003). Although the ecological community is beginning to redress the problems of poor data reporting with increasing pressure on authors to make complete data sets available online, they will persist for the foreseeable future since meta-analysis relies on a historical profile of published papers. A global problem is the possibility of incorrectly analyzing data due to ignorance of important characteristics of the study design and conduct, adjustment for biases and confounders, data collection and management, treatment of missing outcomes and covariates, and so on.

A strong advantage of having access to the primary data is the ability to build more comprehensive meta-analysis models. In the meta-analysis of quadratic (or other order) regressions, a number of options are possible. For example, the meta-analyst can build a
hierarchical (multilevel) model that allows for separate study-specific regression models at the local level, and then combines the regression parameters using fixed- or random-effects assumptions at a global level. Intermediate levels can be included to describe subgroups of studies with common parameter values. Texts describing these methods include Raudenbush and Bryk (2002), Goldstein (2003), and West et al. (2007), and papers on their application in ecology include Helser and Lai (2004) and Thompson and Hobbs (2006). Alternatively, a meta-analysis analog to an ANOVA approach can be taken, whereby individual regressions are fitted to each study, then a test is conducted to assess whether a common quadratic coefficient can be fitted, which if passed is followed by a test to assess whether a common linear coefficient can be fitted, and finally whether a common intercept can be fitted (Tweedie and Mengersen 1995). In both of the above cases, the analysis can include an assessment of the contribution of the higher order polynomial terms; that is, the increased goodness of fit achieved by including a quadratic term at all is formally tested, which if failed is followed by a test of the linear term.

If the meta-analyst uses the primary data to obtain summary estimates of the regression parameters, the resultant data set is equivalent to that obtained by extracting these summary estimates from the published papers themselves (assuming that the same models have been fitted and that the required information is published or otherwise available). The meta-analyst now has a choice between a simple combination of $t$ statistics or a more complete approach involving combining parameter estimates according to an explicit statistical model. In the study of the association between species richness and productivity, these parameter estimates might include correlations, linear and quadratic coefficients, and goodness of fit measures.

This brings us, then, to the more complete statistical approach to combining the parameter estimates themselves. In the context of a quadratic regression, this involves fitting a (typically) random effects model to the multiple estimates of intercept, linear coefficient and quadratic coefficient. A random effects model is generally preferable because we would usually assume in ecology that the regression coefficients differ among studies, in addition to which there is sampling error between the estimates of the coefficients among studies.

Thus, assume that the fitted model for the $j$th study is

$$\hat{y}_j = b_0 + b_L X_j + b_Q X_j^2$$

where $\hat{y}_j$ is the vector of expected responses and $X_j$ is the vector of covariates. This model can be fitted to the original or transformed values for the study outcomes (e.g., log transformation) and under different assumptions may be represented and estimated using ordinary least squares or a variant of this approach (iterative least squares, generalized least squares, etc.), or as a generalized linear model. Assuming that the regression coefficients are normally distributed, a simple random effects model that then combines these study-specific regression estimates is as follows:

$$\begin{pmatrix} \hat{b}_0 \\ \hat{b}_L \\ \hat{b}_Q \end{pmatrix} \sim \mathcal{N}(\mu, \Sigma + C_J)$$

$$\mu = \begin{pmatrix} \beta_0 \\ \beta_L \\ \beta_Q \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \sigma_0^2 & \sigma_{0L} & \sigma_{0Q} \\ \sigma_{0L} & \sigma_L^2 & \sigma_{LQ} \\ \sigma_{0Q} & \sigma_{LQ} & \sigma_Q^2 \end{pmatrix}$$

where $\mu$ is the global mean vector of regression estimates, $\Sigma$ is the between-study variance–covariance matrix, and $C_J$ is the study-specific variance–covariance matrix for the parameter estimates (the intercept, linear, and quadratic coefficients). In the context of estimating the shapes of the productivity–diversity curves across studies, the estimate of the overall quadratic regression estimate, $\hat{b}_Q$, is perhaps of primary interest. This model is described by Becker and Wu (2007) in a general statistical context and by Arends (2006) as one of a wide range of methods for multivariate meta-analysis. On the other hand, if one is only interested in the existence (and magnitude) of a quadratic effect and not the whole shape of the productivity–species richness relationship, then it might suffice to undertake a univariate (random effects) meta-analysis on the quadratic coefficients, rather than the full multivariate model (comprising the combination of intercept and linear and quadratic coefficients).

Note that the above model explicitly accommodates both within-study variation and between-study heterogeneity, which is possible because the individual study parameters are weighted by the inverse sampling variances. Specific sources of heterogeneity can be included as covariates by directly extending the above model to a meta-regression, or through a multi-level model in which different subsets of the studies are combined within and across the different levels. Heterogeneity is dealt with as in any other meta-analysis: if it can be described using covariates (moderators), one can extend the model to a meta-regression; additional variation can be included either as simple between-study variation using a random-effects model, or more elaborately in a hierarchical model. This proposed statistical approach may not be applicable if the study designs are very different, if insufficient data are available to calculate the parameters, or if the responses are on very different scales, among other limitations.

Mittelbach et al. (2001) and others discuss the problem of determining whether a putative maximum
or minimum is indeed a turning point within the observed range of productivities, that is, whether the regression warrants a quadratic rather than linear form. In a quantitative meta-analysis such as the model described above, all of the quadratic coefficients, whether “U shaped” or “hump shaped” or “no maximum or minimum within the range studied,” can be combined to provide an overall estimate of the species richness–productivity relationship. This is the same as combining positive and negative estimates of any measure in a meta-analysis and then testing for heterogeneity among the studies, and depends of course on the ecological validity and interpretation of the result. If desired, tests such as the Mitchell-Olds and Shaw (1987) test applied by Mittelbach et al. (2001) can then be applied to assess whether this overall estimate is a turning point within the observed range of productivities. If the results are heterogeneous, a hierarchical statistical model could be explored to test for differences in the shape among groups of studies, and for heterogeneity within groups. Alternatively, in a Bayesian framework, we suggest that the assessment of its use for the meta-analysis of 42 published experiments of mitochondrial electron transport; here, nonlinear regression was used to estimate the relationship of interest in each study and the results of the regression analyses were synthesized by a random effects model. Van Houwelingen et al. (2002) provide a general description of the bivariate version of this model (i.e., intercept and linear term only), and Paul et al. (2006) applied this in an ecological context, with the aim of analyzing the relationship between fusarium head blight and deoxynivalenol content of wheat among 126 field studies. A Bayesian analogue of the model has also been described by Riley et al. (2007). The same model framework can be used to combine other parameters of interest such as correlations; see Paul et al. (2006) for an example and discussion.

More flexible regression models, such as fractional polynomial regression and spline regression, may also be considered as alternatives to quadratic (and higher-order polynomial) descriptions of a nonlinear relationship. These models can be applied to the study-specific data and then combined using an inverse-variance-weighted random-effects model; Bagnardi et al. (2004) describe this approach in an epidemiological context. To our knowledge, this has not been applied to any ecological problems; it may also be interesting to assess whether this would assist with the problem of asserting that a maximum/minimum has occurred in a specified range.

We note that Mittelbach et al. (2001) did utilize the multivariate meta-analysis model described above and obtained a negative parameter estimate for the overall quadratic coefficient, $b_0$, with an associated 95% confidence interval that did not include zero. However, the authors embedded this result in the body of the paper, preferring to use it as a supplementary rather than primary analysis, and did not elaborate on it more than obtaining the overall mean, confidence limits and heterogeneity. Not surprisingly, the quadratic coefficient was highly heterogeneous across studies, but this was not explored further quantitatively.

A drawback of the model described above is the need for estimates of the covariances of the study-specific regression parameters. These are rarely reported in the published papers and are typically difficult or impossible to estimate using surrogate information. Thus adoption of this method usually relies on access to the primary data. However, alternatives are being devised. For example, Riley et al. (2008) suggest a slight reparameterization of the random effects meta-analysis model that does not involve the covariance terms, at the expense of modified inferences. Although the synthesis of regression parameters was not explicitly discussed, it may be possible to transfer these results to this context.

We recognize that the model expressed in the form above does not explicitly address many of the substantive issues identified by Mittelbach et al. (2001), Whittaker and Heegaard (2003), Gillman and Wright (2010), and others. For example, if the gradients in the individual studies are not of the same length, it is probably meaningless to combine parameters across all studies; instead, the above model could be applied to ecologically more homogeneous categories of studies, such as those identified by Mittelbach et al. (2001) based on local and global scale ranges. Moreover, instead of maintaining these as separate analyses, an additional hierarchy could be added to the model that allows combination of the outputs from the different categories; again, although this is valid statistically, the resultant estimates and inferences would need to be ecologically interpretable and supportable. If plot size was considered to be the most ecologically meaningful covariate (Whittaker 2010), similar approaches could be taken using plot size or other important qualifying features of the studies. Other issues, such as what measures are appropriate surrogates for productivity, are scientific rather than statistical matters, and we do not comment on those here.
Based on this discussion, it is obvious that different meta-analysis methods are applicable in different situations. The informed meta-analyst, then, has at his or her disposal a progression of increasingly comprehensive models and methods. It is incumbent not to stop at simple vote counting, but equally not to use more quantitative methods without suitable data or satisfaction of assumptions. We therefore recommend that “best statistical practice” is included as part of the evolving “best practice” of meta-analysis. This embeds an exploratory phase that allows for qualitative discussion of comparable estimates from different studies (which may include vote counting) followed by an inferential phase that allows for different types of statistical modeling and analysis, based on what is supported by the data and by scientific understanding. In both phases, the principles that are now recognized as underpinning all of the “best practice” components of meta-analysis will be expected; that is, the methods that are adopted must be stated clearly, underlying assumptions defended and caveats about the limitations of the methods acknowledged.

Ecological relationships are invariably complex, so it is not unexpected that meta-analysis is difficult in this discipline area; however, this is also why meta-analysis can be a powerful tool for providing an overall, informed opinion about the collected body of literature. Similar issues to the ones in this Forum have been discussed in the context of meta-analysis and systematic review in other disciplines (e.g., see reviews by Mullen and Ramirez 2006, Quintana and Minami 2006). As is evident from recent literature, the identification of problems with the application in ecology of existing statistical techniques for meta-analysis motivates the development of new techniques, which in turn motivates statistical techniques for meta-analysis. This embeds an exploratory phase that allows for qualitative discussion of comparable estimates from different studies (which may include vote counting) followed by an inferential phase that allows for different types of statistical modeling and analysis, based on what is supported by the data and by scientific understanding. In both phases, the principles that are now recognized as underpinning all of the “best practice” components of meta-analysis will be expected; that is, the methods that are adopted must be stated clearly, underlying assumptions defended and caveats about the limitations of the methods acknowledged.

Ecological relationships are invariably complex, so it is not unexpected that meta-analysis is difficult in this discipline area; however, this is also why meta-analysis can be a powerful tool for providing an overall, informed opinion about the collected body of literature. Similar issues to the ones in this Forum have been discussed in the context of meta-analysis and systematic review in other disciplines (e.g., see reviews by Mullen and Ramirez 2006, Quintana and Minami 2006). As is evident from recent literature, the identification of problems with the application in ecology of existing statistical techniques for meta-analysis motivates the development of new techniques, which in turn motivates statistical techniques for meta-analysis. This embeds an exploratory phase that allows for qualitative discussion of comparable estimates from different studies (which may include vote counting) followed by an inferential phase that allows for different types of statistical modeling and analysis, based on what is supported by the data and by scientific understanding. In both phases, the principles that are now recognized as underpinning all of the “best practice” components of meta-analysis will be expected; that is, the methods that are adopted must be stated clearly, underlying assumptions defended and caveats about the limitations of the methods acknowledged.

Ecological relationships are invariably complex, so it is not unexpected that meta-analysis is difficult in this discipline area; however, this is also why meta-analysis can be a powerful tool for providing an overall, informed opinion about the collected body of literature. Similar issues to the ones in this Forum have been discussed in the context of meta-analysis and systematic review in other disciplines (e.g., see reviews by Mullen and Ramirez 2006, Quintana and Minami 2006). As is evident from recent literature, the identification of problems with the application in ecology of existing statistical techniques for meta-analysis motivates the development of new techniques, which in turn motivates statistical techniques for meta-analysis. This embeds an exploratory phase that allows for qualitative discussion of comparable estimates from different studies (which may include vote counting) followed by an inferential phase that allows for different types of statistical modeling and analysis, based on what is supported by the data and by scientific understanding. In both phases, the principles that are now recognized as underpinning all of the “best practice” components of meta-analysis will be expected; that is, the methods that are adopted must be stated clearly, underlying assumptions defended and caveats about the limitations of the methods acknowledged.

Ecological relationships are invariably complex, so it is not unexpected that meta-analysis is difficult in this discipline area; however, this is also why meta-analysis can be a powerful tool for providing an overall, informed opinion about the collected body of literature. Similar issues to the ones in this Forum have been discussed in the context of meta-analysis and systematic review in other disciplines (e.g., see reviews by Mullen and Ramirez 2006, Quintana and Minami 2006). As is evident from recent literature, the identification of problems with the application in ecology of existing statistical techniques for meta-analysis motivates the development of new techniques, which in turn motivates statistical techniques for meta-analysis. This embeds an exploratory phase that allows for qualitative discussion of comparable estimates from different studies (which may include vote counting) followed by an inferential phase that allows for different types of statistical modeling and analysis, based on what is supported by the data and by scientific understanding. In both phases, the principles that are now recognized as underpinning all of the “best practice” components of meta-analysis will be expected; that is, the methods that are adopted must be stated clearly, underlying assumptions defended and caveats about the limitations of the methods acknowledged.

**Literature Cited**


