1

Place of Meta-analysis among Other Methods of Research Synthesis

Julia Koricheva and Jessica Gurevitch

In the most general terms, meta-analysis is one method of research synthesis. Research synthesis may be defined as a review of primary research on a given topic with the purpose of integrating the findings (e.g., for creating generalizations or resolving conflicts). Research synthesis is central to the scientific enterprise. Without it, the evidence for various alternative hypotheses cannot be properly evaluated and generalizations cannot be reached, thus the advance of the scientific field as well as any potential practical applications are inhibited. Research synthesis can be performed either qualitatively, in the form of a narrative review, or quantitatively, by employing various statistical methods for the integration of results from individual studies.

Research reviews in ecology and evolutionary biology have traditionally been carried out either in the form of narrative reviews, or by "vote counting," where the number of statistically significant results for and against a hypothesis are counted and weighed against each other (see below). In other fields, it has been widely recognized that neither of the above methods is adequate to address current needs for quantitative research synthesis, and we believe that this is also true in ecology, evolution, and conservation biology. Narrative reviewing offers expert interpretation and perspective, but it is inherently subjective and nonquantitative. Vote counting has very poor properties as a statistical procedure, as discussed below. Neither procedure is able to provide critically important information on the magnitude of the effects or the sources of variation in outcomes among studies.

The lack of training in formal, rigorous protocols for research synthesis in ecology and evolutionary biology contrasts markedly with the standard quantitative training that scientists in these fields receive for primary research, including training in experimental design and analysis. Fortunately, this asymmetry in the tools employed for primary versus synthetic investigations began to change dramatically in the past two decades as meta-analytic methods have started to be introduced and incorporated into standard practice. The summarization and integration of research results across studies is increasingly viewed as a scientific research process in its own right in ecology and evolutionary biology, as it has come to be recognized in other disciplines. Like primary research, research synthesis follows the scientific method and is held to scientific standards. Meta-analysis employs specialized techniques for data gathering and analysis developed specifically for the purposes of research synthesis. Below, we briefly review the variety of approaches to research synthesis and discuss their relative merits, advantages, limitations, and drawbacks.

"TEXTBOOK EXAMPLES"

The simplest method of presenting the results of research on a particular topic is the "textbook example." While not a formal method of research synthesis, it is one with which every biologist is familiar, because that is the way we learn and subsequently teach our fields of study. This ubiquitous method picks out particular case studies which seem to best illustrate the evidence for a particular phenomenon, and uses these examples to explain the phenomenon, to provide the evidence that it exists, and to summarize the findings on that phenomenon. The practice of using "textbook examples" or case studies as a summary of scientific findings in a research field is based on a widespread but erroneous belief that a single primary study, particularly a well-designed experiment, is able to provide the ultimate test for a hypothesis and hence resolve an issue. This is a mistaken idea. Research results are probabilistic and subject to artifacts such as sampling error and measurement error; thus, the findings of any single study may have occurred simply by chance and may be refuted by subsequent research (Taveggia 1974, Schmidt 1992, Ioannidis 2008). Therefore, in the majority of current methods for research synthesis, and in meta-analysis in particular, any individual primary study is considered to be one of a population of studies on a given question, rather than a single definitive and conclusive piece of evidence capable of fully resolving the question.

NARRATIVE REVIEWS

Traditionally, research synthesis in ecology and evolutionary biology has been carried out in the form of narrative reviews; numerous examples can be found in journals such as *Annual Review of Ecology, Evolution and Systematics*, *Trends in Ecology & Evolution*, *Biological Reviews*, or *The Quarterly Review of Biology*. These narrative reviews are generally invited contributions written by senior scientists who are recognized experts in their fields. Some of these reviews are narrower in their scope while others are more comprehensive, and their bibliographies may contain dozens to hundreds of primary studies. The structure of these reviews differs profoundly from that of primary research papers. There is usually no "material and methods" section where the author explains search methods or selection criteria for the inclusion of primary studies into the review. An analysis of 73 review articles on conservation and environmental management topics published during 2003–2005 found that 63% of the reviews were narrative. Of these, only 30% reported details of the sources used to acquire studies for review, 34% defined inclusion criteria for identification of relevant studies, and fewer than 20% reported the search terms used (Roberts et al. 2006). This lack of methodological and reporting rigor can increase the degree of subjectivity in which papers are chosen for review and how their results are interpreted. For example, two authors could review the literature on the same topic and, by including or excluding different subsets of studies, arrive at opposite conclusions. Narrative reviews are opaque in the sense that the methods used for searching the literature are seldom laid out explicitly, and this means that the review cannot be repeated or updated systematically as more information becomes available.

Another very practical limitation of narrative reviews is that they are inefficient in handling a large number of studies and ill equipped for dealing with variation in outcomes among studies. At best, the results of the studies may be presented in the form of large tables, which can be difficult to interpret. At worst, large proportions of the studies are ignored, while the results of a small number of "exemplars" are emphasized. This places narrative reviewing uncomfortably close to "textbook examples" or case studies as a method of summarizing research. If a large amount of research has been done on a particular topic, chance alone dictates that studies exist that report inconsistent and contradictory findings, particularly when sample sizes and the statistical power to detect significant outcomes differ greatly among studies. When the number of reviewed studies is relatively small, potential explanatory variables could be suggested by comparing various study attributes in a table. As the number of studies increases, however, the task of sifting through many potential explanatory variables becomes unmanageable. Consequently, the results of studies in narrative reviews may appear to be "inconsistent," "inconclusive," or "conflicting"; these types of results are uninformative for resolving intellectual or methodological conflicts. They are also insufficient to address needs for making practical decisions, which is becoming an increasingly important task in applied ecology and conservation biology.

Narrative reviews can serve important functions by presenting perspectives, historical development of ideas, and other conceptual contributions. But their inherently high subjectivity and low repeatability, and their poor suitability for dealing with variation in results among studies greatly limit their utility for research syntheses, including evaluation of whether hypotheses are supported by the existing data. This can lead to a paradoxical situation that occurs when many hypotheses exist to explain some phenomenon or process, and all of them have been repeatedly tested, but none has ever been rejected, and they all coexist by virtue of supporting evidence in one system or another (e.g., Berenbaum 1995).

Ecologists and evolutionary biologists have often been criticized for having long-term debates which seemingly persist for generations (e.g., on the role of competition in structuring communities, or on the relationship between the diversity and stability of communities). We argue that the reliance of ecologists and evolutionary biologists on "textbook examples" and narrative reviews as ways of informing conclusions and synthesizing the results of published literature is at least partly responsible for this pattern.

VOTE COUNTING

Vote counting is a quantitative method for research synthesis that has been discredited and largely abandoned in other scientific fields but is unfortunately still all too commonly employed in ecology and evolutionary biology. In its simplest form, the available studies are sorted into three categories: those that report significant results in the predicted direction, those that yield significant results in the opposite direction, and those that yield nonsignificant results. Sometimes the latter two categories are combined because both types of studies are considered as evidence not in support of the hypothesis (e.g., Watt 1994). The studies are counted up and the category into which most studies fall is declared the "winner," and is considered to provide the best evidence about the direction of the effect. The relative number of studies "voting" for or against the effect may be considered evidence for its magnitude, strength, and consistency. Alternatively, outcomes of studies could be classified into more than three categories. For example, in a review on the relationship between productivity and diversity (Mittelbach et al. 2001), the shape of the relationship in each of the studies being reviewed was classified into five categories: linear positive relationships, linear negative, humped (maximum diversity at intermediate productivity values), U-shaped, and no significant relationship. The percentages of studies displaying the above shapes were then compared. Authors sometimes use goodness of fit or other statistical tests to compare the expected and observed frequencies of positive and negative outcomes (Waring and Cobb 1992, Daehler 2003, Colautti et al. 2006), with the null hypothesis being that the number of positive and negative outcomes is equal.

The advantages of vote counting are simplicity and seemingly broad applicability; regardless of design and the statistical approach used in the primary study and the way the results are reported, the outcomes are likely to be broadly classifiable into one of the vote-counting categories. Although counting up significant results seems to be logically straightforward, it is unfortunately seriously flawed as a statistical technique for research synthesis. The results of a vote count are statistically biased and often misleading. More importantly, they do not provide the information most necessary and relevant in synthesizing the results of different studies. There are several reasons for this. First, the vote-counting procedure gives one vote to each study regardless of its sample size and the statistical precision with which the outcome was tested. Thus, a study testing an effect with three replicates, for example, receives the same weight as a study with 100 or 1000 replicates. A second limitation of vote counting as a method for quantitative research synthesis is that the number of studies reporting statistically significant versus nonsignificant outcomes tells us nothing about the magnitude of the effect of interest, which we would argue is more likely to be biologically meaningful. One could partially deal with the problem of accounting for differences among studies in the precision with which an effect is tested by comparing the number of positive and negative outcomes regardless of their statistical significance. However, this still does not tell us anything about the magnitude of the effect in question. Also, the direction of the outcome may not be reported for nonsignificant results.

Vote counting may also lead to a wrong conclusion about the overall outcome across studies. Because vote counting is based on the statistical significance of the research findings, it has low power for effects of relatively small magnitude (which might still be biologically or otherwise significant); this is due to the statistical power of small studies possibly being too low to detect an effect. Effect sizes and sample sizes in ecology and evolution are often rather small. For example, Møller and Jennions (2002) have shown that across 43 published meta-analyses on various ecological and evolutionary topics the average effect size reported for Pearson's correlation coefficient was between 0.180 and 0.193. The sample size required to detect such a small effect with a power of 80% is over 200 replicates, which is considerably larger that the sample size of most studies in ecology and evolutionary biology. Moreover, Hedges and Olkin (1980) demonstrated that the power of the vote-counting procedure decreases as the number of studies integrated increases—a strange and undesirable property for a statistical test.

Surprisingly, despite the serious statistical drawbacks outlined above, which have been detailed in the ecological literature for well over a decade, vote counting is still used, published, and accepted as a "valid" method of research synthesis in ecology and evolution. For example, in a recent paper in *Science* on the shape of the productivity/plant species richness relationship, Adler et al. (2011) counted the number of study outcomes in five categories (nonsignificant, positive linear, negative linear, concave up, or concave down) at local, regional, and global scales. The most common relationship was nonsignificant, and the authors concluded that there was no clear relationship between productivity and plant species richness. While their results may very well be correct, a vote count is not the best way to investigate this question, and the authors did not attempt to justify this approach. We return to the reasons for the persistence of vote counting in the ecological literature below (see "Choosing the Method for Research Synthesis").

Vote counting can be more pervasive and subtle than is sometimes realized. Narrative reviews may rely upon a kind of "virtual vote count" in that they report on findings based on the statistical significance of their outcomes, and may then base conclusions on the preponderance—that is, the number—of statistically significant findings. Informally, one's assessment of the literature read on a question may likewise be based on the number of studies that report statistically significant findings. Formal systematic reviews (below) and meta-analyses bring a more rigorous, scientific approach to understanding the available evidence.

COMBINING PROBABILITIES AND RELATED APPROACHES

Another approach to quantitative research synthesis commonly used in the social sciences, but quite rarely in ecology and evolution, is based on combining probability values (significance levels) from statistical tests in individual studies that are testing the same scientific hypothesis. Methods for combining probabilities across studies have a long history dating at least from procedures suggested by R. A. Fisher in his book *Statistical Methods for Research Workers* (1925). As many as 18 different methods of combining probabilities currently exist (Becker 1994) depending on the statistical distribution employed. The method most commonly applied in the social sciences is the sum of Z's method based on normal distribution (Becker 1994, Cooper 1998). Another common approach to combining probabilities is Fisher's sum of logs methods, as described by Sokal and Rohlf (1995), which uses the inverse of the chi-square distribution to determine significance.

Similar to vote counting, the advantage of the combining-probabilities method is its broad applicability (any kind of statistical test can be used). Combining probabilities is also less problematic than vote counting because it uses exact probabilities, hence there is no deep gap between studies reporting $P = 0.04$ and $P = 0.06$ (note that in vote counting the above two studies would have been placed in different categories, nonsignificant and significant). However, similar to vote counting, combining probabilities is not very informative. The null hypothesis for the test of combined significance is that the effects of interest are not present in any of the studies (Becker 1994). If the null hypothesis is rejected, all the synthesist can tell is that in at least one of the studies the effect is not zero. This approach also provides no information about the sign or magnitude of the effect. Small P values may arise because of large effects or because of large sample sizes. The alternative hypothesis can be made more specific; for example, one-tailed tests are often used to estimate the probability that effects of interest are positive, negative, or larger than a specified value in at least one study (Becker 1994).

A practical problem with the application of the method of combining probabilities is that exact probability values are often not reported in primary studies. Commonly, significance values of statistical tests are reported as $P > 0.05$, $P < 0.05$, $P < 0.01$, or $P < 0.001$. In addition, Cooper (1998) pointed out that the method of combining probabilities is too liberal; if many tests exist and at least one of them provides a very low P value, the null hypothesis is practically always rejected.

The combining-probabilities method is seldom used for research synthesis in ecology and evolution. However, Koricheva et al. (2004) used the sum of Z's method in a meta-analysis testing for the potential trade-offs between constitutive and induced plant defenses against herbivores. If induced response decreases with increasing constitutive defense, as predicted by the trade-off hypothesis, the slope of the regression line of induced defense on constitutive defense will be < 1. The authors combined the one-tailed probabilities from individual t-tests of slopes of the linear regression of induced defense on constitutive defense. The combined Z value for 14 tests yielded 3.9394, which corresponds to the overall $P < 0.00005$, and thus results in rejecting the null hypothesis that the slope of the linear regression was < 1 in none of the 14 tests. A weighted meta-analysis of the same studies using Pearson's correlation coefficient between induction ratios and constitutive defenses provided further support for the existence of the trade-offs between plant allocations to constitutive and induced defenses against herbivores (Koricheva et al. 2004). Both approaches were used because correlations between induction ratios and constitutive defenses may be spurious, as the ratio is correlated with its own denominator. The combination of two approaches thus reinforced the conclusions of the meta-analysis. Combining probabilities is often used in combination with weighted meta-analysis in the social sciences (Kraus 1995).

A related approach, unfamiliar in ecology, is the combination of t-statistics. This approach has the advantage of considering the magnitude of effects in the individual studies. It also takes the sign of the effect into account, unlike the combination of probabilities. However, it suffers from other statistical drawbacks (described by Becker and Wu 2007, among others), particularly in terms of the inferences that it is possible to make from the outcome.

META-ANALYSIS

The term "meta-analysis" was coined by Glass (1976) in reference to "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings." The above definition is broad and covers all the techniques used in quantitative research synthesis including vote counting and combining probabilities, as described earlier. In this book, we define meta-analysis more narrowly, as a set of statistical methods for combining the magnitudes of the outcomes (effect sizes) across different data sets addressing the same research question. The methods of meta-analysis were originally developed in medicine and various social sciences (Glass et al. 1981, Hedges and Olkin 1985); they were introduced in ecology and evolutionary biology in the early 1990s (Järvinen 1991, Gurevitch et al. 1992, Arnqvist and Wooster 1995a).

Meta-analysis provides a powerful, informative, and unbiased set of tools for summarizing the results of studies on the same topic. It offers a number of advantages over narrative review, vote counting, and combining probabilities (Table 1.1). Meta-analysis is based on expressing the outcome of each study on a common scale. This measure of outcome, an "effect size," includes information on the sign and magnitude of an effect of interest from each study. In many cases the variance of this effect size can also be calculated (see Chapters 6 and 7). These effect size measures can then be combined across studies to estimate the grand mean effect size and

TABLE 1.1. Comparison of methods of research synthesis.

Characteristics of the review type	Narrative review	Vote counting	Combining probabilities	Meta-analysis
Imposes restrictions on the type of studies that can be used in review	No	No	No	Yes
Interprets study outcome based on its statistical significance	Yes	Yes	Yes	No
Takes into account sample size and statistical power of the individual studies being combined	No	No	Yes	Yes
Assesses statistical significance of the mean (overall) effect (i.e., whether it is significantly different than zero)	No	No	Yes	Yes
Assesses the magnitude of the mean effect	No	No	No	Yes
Allows analysis of sources of variation among studies	No	No	No	Yes

its confidence interval, and to test whether this overall effect differs significantly from zero. In many cases in ecological meta-analysis, it is of interest to examine heterogeneity among outcomes and model the relative contribution of different factors to the magnitude of the effect sizes. Hence, unlike other methods of research synthesis described above, meta-analysis allows the research synthesist to estimate the magnitude and sign of the grand mean effect across studies, assess whether the confidence interval around the effect includes zero, and examine sources of variation in that effect among studies.

A key aspect of modern approaches to meta-analysis is accounting for unequal precision in the magnitude of the effect among studies by weighting each study's effect size by the inverse of its variance. Meta-analysis offers an improved control of type II error rates (Arnqvist and Wooster 1995a), because the low power of individual studies to detect an effect is "corrected" by the accumulation of evidence across many studies. This is particularly important in areas where failure to reject false null hypothesis may have large detrimental impacts, as in conservation biology (Fernandez-Duque and Vallegia 1994) and medicine. Meta-analysis can potentially allow the detection of an effect even in situations when none of the studies included in the analysis show statistically significant results because of low statistical power (Arnqvist and Wooster 1995a). An important contribution of meta-analyses can be to identify gaps in the literature where more research is needed, and also to identify areas where the answer is definitive and no new studies of the same type are necessary. When the data are already sufficient to resolve a question, it is a waste of time and money to keep accumulating more of the same kind of information; meta-analysis can be an invaluable tool in this regard, and has been used in this way in ecology and other fields.

The first meta-analysis in ecology was published in 1991 (Järvinen 1991) and the number of meta-analyses in ecology increased greatly over the next two decades, reaching 119 publications by 2000 (Gurevitch et al. 2001), and exceeding 500 publications per year by 2010 (see Fig. 25.1C in Chapter 25). The first general review of meta-analysis in ecology and evolution, published in 1995 (Arnqvist and Wooster 1995a), stated: "Meta-analysis is still rare in our domain, but the first applications show that it can successfully help address a variety of questions." The second review was published in 2001 (Gurevitch et al. 2001) and concluded that at the turn of the century, meta-analysis had started to have a substantial impact on the way data were summarized in ecology and evolutionary biology. The authors predicted that incorporation of meta-analysis as a routine and familiar approach in ecology and evolution will fundamentally change the nature of these scientific disciplines. It is too early yet to assess the accuracy of their prediction.

More widespread adoption of meta-analysis may have a number of subtle but profound effects on ecologists' perspective of research. Most importantly, individual studies contributing to meta-analysis are seen as members of a population of studies that all provide information on a given effect, rather than as isolated and presumably definitive examples. More controversially, meta-analysis can influence one's interpretation of results by reducing the emphasis on the statistical significance of results and moving the focus to the magnitude, direction, and variance in effects (Chapter 23). Variability of results among studies may become more a source of hypothesis generation regarding the nature of the sources of variation rather than being seen as a hindrance to understanding and interpreting a given phenomenon. While scientists have always emphasized biological importance over statistical significance, meta-analysis makes this view more compelling, and shifts the interpretation of biological importance from case studies and textbook examples to "the weight of evidence" across all of the literature on a particular question. Meta-analysis helps us to advance from seeking complete answers based on individual experiments, however creative and elegant, to combining evidence across existing studies, each of them perhaps imperfect in some way, but which together provide a wealth of information. Meta-analysis is thus much more than just a new method of research synthesis; it is a new approach to scientific research which requires major changes in our views of the role of individual studies and the acceptance of the cumulative nature of scientific knowledge (Schmidt 1992).

SYSTEMATIC REVIEW

An important method of research synthesis that has become ubiquitous in the medical literature (Khan et al. 2003), and is beginning to be used in conservation and environmental management (Davies et al. 2008, Newton et al. 2009, Stewart et al. 2009), is the "systematic review." Systematic review is research synthesis on a precisely defined topic using explicit methods to identify, select, critically appraise, and analyze relevant research. The crucial element of the systematic review that distinguishes it from an ordinary narrative review is an a priori protocol, which describes the methodology, including detailed search strategy and inclusion criteria. This review protocol makes the review process rigorous, transparent, and repeatable. Details and examples of review protocols are discussed in Chapter 3.

Systematic reviews may or may not include meta-analysis or other methods of quantitative research synthesis. Systematic reviews without meta-analyses are used to identify the current state of knowledge, including gaps when insufficient data exists to conduct meta-analysis. On the other hand, meta-analyses are also produced without fully defined systematic reviews, but this practice may lead to biased or erroneous results if systematic methods have not been used to obtain and synthesize the data. Throughout this book, we shall retain the formal distinction between systematic review as the systematic collation and analysis of data and meta-analysis as the statistical methods for combining effect sizes.

In medicine, systematic reviews in combination with meta-analyses provide critical information needed for evidence-based medicine (Khan et al. 2003; Chapter 25). Use of systematic reviews for supporting decision making in conservation and environmental management has been promoted by the Centre for Evidence-Based Conservation at Bangor University, UK (http://www.cebc.bangor.ac.uk/). Guidelines for systematic reviews in conservation and environmental management have been developed (Pullin and Stewart 2006) and completed systematic reviews on these topics can be found at http://www.environmentalevidence.org/Reviews.html. For published examples of systematic reviews in conservation and management see Davies et al. (2008), Newton et al. (2009), and Stewart et al. (2009).

CHOOSING THE METHOD OF RESEARCH SYNTHESIS: IS META-ANALYSIS ALWAYS THE BEST CHOICE?

Meta-analysis has many advantages over other methods of research synthesis (Box 1.1). The rest of this handbook focuses specifically on meta-analysis and aims to promote its correct and thoughtful use as a part of systematic reviews in ecological and evolutionary research. Does this mean that meta-analysis is always the preferred method of research synthesis, and that narrative reviews, combining probabilities, and vote-counting procedures have to be abandoned altogether? Are there situations when it is not advisable or feasible to conduct a meta-analysis? Below we evaluate these questions, examine some common objections to the use of meta-analysis in ecology and evolutionary biology, and suggest situations where the use of other methods of research synthesis in combination with meta-analysis can be advantageous.

BOX 1.1.
Why should ecologists and evolutionary biologists learn about meta-analysis?

(1) Meta-analysis provides a more objective, informative and powerful means of summarizing the results from individual studies as compared to narrative/qualitative reviews and vote counting.
(2) Applications of meta-analysis in ecology are becoming increasingly more common, and thus even if you are not planning to conduct your own meta-analyses, you need to understand the method to follow and evaluate the literature in your field.
(3) Application of meta-analysis to applied fields (e.g., conservation and environmental management) can make results more valuable for policy makers.
(4) Learning the basics of meta-analysis can dramatically improve standards for data reporting in primary studies so that the results can be included in subsequent research synthesis on the topic.
(5) Conducting meta-analysis changes the way you read and evaluate primary studies; it makes you acutely aware that the statistical significance of the results depends on statistical power, and in general improves your abilities to critically evaluate evidence.

Meta-analysis is more demanding than other research synthesis methods in terms of the format of the data required for analysis (Table 1.1). In ecology and evolutionary biology this problem is further exacerbated by uneven and poor reporting standards. Many primary studies in ecology and evolution do not report essential data needed to calculate effect sizes, or needed for critical appraisal of the results (e.g., standard deviations or other measures of variation, and sample sizes). As a result, those studies have to be excluded from conventional weighted analyses, often resulting in a dramatic reduction in the number of available studies and a consequent loss of information. This is sometimes used as an argument to justify the use of vote counting instead of meta-analysis in ecology and evolution (Heck et al. 2003). However, the problem of poor reporting of primary data can sometimes be solved in more satisfactory ways than by resorting to the statistically flawed vote-counting procedure. Two approaches have been suggested by Gurevitch and Hedges (1999) for situations where lack of variance or sample size information prevents the use of weighted meta-analysis; these are randomization tests (Adams et al. 1997), or unweighted standard parametric statistical tests, such as ANOVA or least-squares regression. Examples of the use of these approaches in ecology are provided by Johnson and Curtis (2001) and Coleman et al. (2006). The advantage of randomization tests and unweighted parametric methods over vote counting is that they provide an estimate of the magnitude of the effect, which vote counting lacks. In addition, resampling methods are free from normality assumptions of the parametric statistical tests. Note, however, that both randomization tests and unweighted parametric methods assume homogeneity of variances; this assumption is likely to be seriously violated when combining studies, because different studies typically have vastly different sample sizes and vary greatly in precision (Gurevitch and Hedges 1999).

Another possible way to avoid the loss of information when only a subset of studies provide data needed for meta-analysis is to conduct a proper weighted meta-analysis on studies that provide enough information to calculate effect sizes, with a vote count on the remaining studies. This approach is common in ecology (e.g., Borowicz 2001, Huberty and Denno 2004, Liu and Stiling 2006, Attwood et al. 2008); it allows all studies to be used, and the results of the meta-analysis can be compared with those of vote counting. Good agreement between the methods is observed in some cases, thus reinforcing the conclusions of the research synthesis, but discrepancies are also common (Liu and Stiling 2006). Bushman and Wang (1996, 2009) have proposed a procedure that combines estimates based on effect size and vote-counting procedures to obtain an overall estimate of the population effect size. To our knowledge, this approach has not been used yet in ecological and evolutionary meta-analyses. In general, we concur with the recommendation by Bushman and Wang (2009) that vote-counting procedures should never be used as a substitute for effect size procedures and should never be used alone. Combining probabilities may also be used in combination with weighted meta-analysis (Kraus 1995). Additional quantitative and qualitative sources of information like expert opinions can be incorporated into research synthesis by using a Bayesian approach (Kuhnert et al. 2005, Newton et al. 2007, Choy et al. 2009).

Some ecologists argue that meta-analysis is ill advised in situations when the data are very heterogeneous (e.g., in terms of research methods, experimental design, response variable measured, and organisms studied) and suggest that vote counting offers a more cautious and conservative approach to research synthesis in these circumstances (Daehler 2003, Heck et al. 2003, Ives and Carpenter 2007, Tylianakis et al. 2008). This argument refers to the "apples and oranges dilemma" and essentially repeats the argument of Eysenck (1994) that "meta-analysis is only properly applicable if the data summarized are homogeneous—that is, treatment, patients, and end points must be similar or at least comparable." While it is true that sometimes collected studies are too heterogeneous to allow meaningful synthesis (Markow and Clarke 1997), using this argument to justify a vote count instead is problematic. Vote counts of a truly heterogeneous body of studies offer no advantages, and considerable disadvantages, over meta-analysis. If the studies are considered to be too heterogeneous to be sensibly combined by meta-analysis, why is it justifiable then to lump them together in a vote count, or even a narrative review? Meta-analysis at least provides a way to explicitly analyze the extent of variation in effect sizes among studies and to reveal the causes of this heterogeneity whereas both vote counting and narrative reviews are unsuited to doing so. The scope of the review has to be carefully considered before any research synthesis is undertaken, and generalizations have to be sought on a biologically meaningful level. If the original question is too broad and results in a set of primary studies that is too heterogeneous, the question has to be redefined more narrowly to allow meaningful synthesis, regardless of the synthesis method used. Sometimes, combining heterogeneous data can be valid, depending on the level of generalization one wishes to make. For example, meta-analysis has a very important role in generalizing across species, ecosystems, and other larger-scale entities, beyond the scope of individual studies (Chapter 23).

SUMMARY AND CONCLUSIONS

Meta-analysis alone or in combination with other methods of research synthesis should be used whenever the estimate of the magnitude of an effect and an understanding of sources of variation in that effect is of interest, and when at least some of the primary studies gathered provide sufficient data to carry out the analysis. However, if few studies on the question exist, and the aim is largely to make the reader aware of an emerging research field or a new direction in the established field (as in *Current Opinion in . . .* and *Trends in . . .* journal series), meta-analysis combined with a systematic review of the topic may be unnecessary and a short narrative review may suffice.

If ecologists and evolutionary biologists wish to bring scientific methodology to bear on using the "weight of evidence" to inform policy making in conservation and environmental management, moving toward more scientific methods of synthesizing available data will become increasingly necessary. This is already widely accepted practice in medicine and in social policy. As the need for ecologists to be heard by policy makers and the public in addressing critical issues, such as biodiversity loss and climate change, becomes ever more urgent, mastery and implementation of the "evidence-based" tools of systematic review and meta-analysis becomes all the more compelling.