
The Procedure of Meta-analysis in a Nutshell

Isabelle M. Côté and Michael D. Jennions

IT IS SAID THAT a picture is worth a thousand words. Taking this at face value, we offer two figures to summarize the entire meta-analytic process. In Figure 2.1 we cover Part I, the initial stage of formulating a question and systematically searching the primary literature for suitable studies. In Figure 2.2 we cover Part II, the stage at which you extract data from publications, run statistical tests, and present and interpret your results. The two figures explicitly link each step to the relevant chapters (indicated by a circled number). These indicate where Chapters 3 to 27 fit into the process of meta-analysis. You will notice teardrop-shaped symbols next to each step. These are not intended to indicate the tears of frustration that will be shed as you proceed (although they might!). Instead, they are meant to reflect the beads of sweat, or hard work, that should be expected. Our reason for including them is to reassure the novice that the hardest part of a meta-analysis is not mastering complex statistics (statistical procedures for meta-analysis, as we will see, may range from very simple to advanced) or software, but the labor involved in gathering, evaluating, and assimilating papers. In particular, meta-analysis requires the additional work of extracting data, which is often frustrating and challenging, and is not typically part of an old-fashioned narrative review. The key information to extract is an estimate of the strength of a relationship in the form of an effect size. Effect sizes are a common currency into which the outcomes of all papers are translated so that they can be combined and compared (e.g., Hedges' d is the difference between a control and treatment group expressed as the number of standard deviations by which they differ, or Pearson's r is the correlation between two variables of interest). For now you can think of an effect size as a P -value corrected for sample size with the direction of the relationship also provided (for details, see Chapter 6). If two studies report the relationship between two variables as $P = 0.05$, $n = 20$, and $P = 0.05$, $n = 50$, the relationship must be stronger—hence Pearson's r larger—in the first study. Why? With less data the confidence interval on the estimate of r is wider, so the estimated mean value of r must be greater for the lower limit of the 95% confidence interval to be above the null value of 0 (recall that $P = 0.05$ when the 95% confidence interval for a test statistic just touches the null value).

Performing a meta-analysis is in fact conceptually no different from an empirical study in the sense that sometimes statistical problems bog you down. But researchers usually design a study with their own statistical abilities in mind, or they follow an established design that allows them to replicate a standard analytic approach. The difference between a good and a bad empirical study often boils down to whether an interesting question is being asked, and to the quality and quantity of the data collected using an unbiased sampling technique. The same is true of meta-analysis. Statistical elegance is wonderful, but it is subservient to the quality/quantity of data collected. (After all, you can always reanalyze data.) If you are a novice you will be reassured to know that anyone who uses a basic Windows-driven statistical package, such as Minitab, SPSS,

or JMP, has the ability to do the statistics necessary to produce a publishable meta-analysis. Even tricky issues often have simple solutions if you are prepared to use less powerful statistical approaches (e.g., use the mean of a set of nonindependent measures rather than a repeated-measures model; see Chapter 16). Another alternative is to collaborate with someone with advanced statistical skills when you wish to delve into more complex models. If like many biologists, you consider yourself “mathematically challenged,” do not be deterred; you likely already have most of the skills, training, and insights needed to conduct a top-notch meta-analysis.

SOME CAVEATS ABOUT THE ROAD MAP

The two figures offer a road map depicting the usual sequence of events that make up a meta-analysis. You should, however, be aware that the exact content of the activities in each box can vary considerably among meta-analyses.

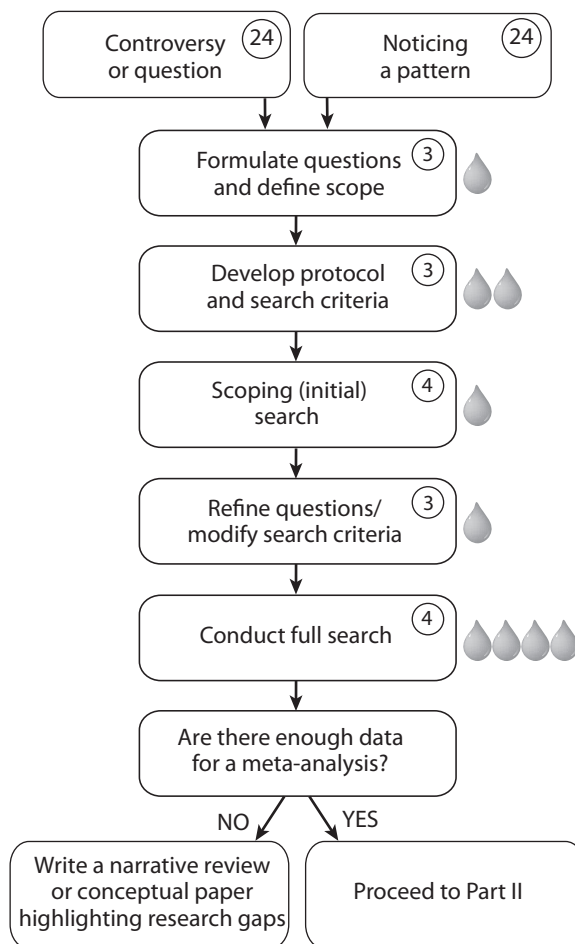


Figure 2.1. Part I of conducting a meta-analysis. The teardrop shapes represent (in our experience) the relative effort (drops of sweat) associated with each step of the process. Circled numbers refer to relevant chapters in this book.

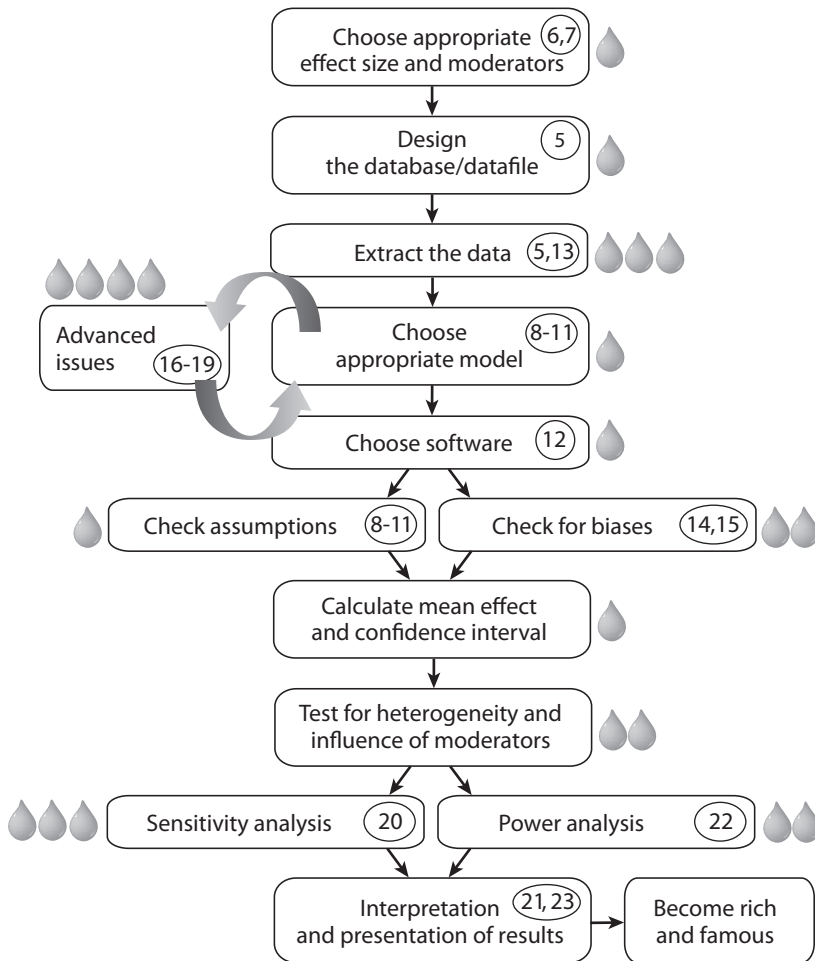


Figure 2.2. Part II of conducting a meta-analysis. The teardrop shapes represent (in our experience) the relative effort (drops of sweat) associated with each step of the process. Circled numbers refer to relevant chapters in this book.

First, the extent of the initial (scoping) search and the rigor of the final full search vary among published meta-analyses. The ideal scenario is to search systematically for all available data from all possible sources (Chapter 3), but very few ecologists and evolutionary biologists carry out this type of full-scale systematic review. In fact, current practice leans rather strongly toward trying to find all published (and some unpublished) papers on the question of interest, with less attention given to gray literature (government reports, conference proceedings, and so on). The extent to which the gray literature needs to be searched is likely to vary widely depending on the type of meta-analysis. It is expected to be important when tackling a question about conservation or environmental policy (e.g., the effect of traffic on frog populations), but unlikely to provide useful data when conducting a meta-analysis on a highly theoretical evolutionary topic (e.g., the effect of sperm competition on mutation rates).

Second, while our outline makes the process seem linear, there is often uncertainty that leads to some steps being repeated. In particular, in Part I you might have to repeat several cycles of scoping searches, adjusting study questions and modifying the protocol and search criteria until you are confident that a full search will generate the appropriate amount of data (i.e., enough to produce meaningful results, but not so much as to make it logistically impossible to extract data from all relevant sources before your retirement).

Third, you will usually find that once you start to extract data from papers, your database has to be redesigned. This is usually because you have overlooked a factor that is a potential source of variation in study outcomes (i.e., effect sizes). For example, perhaps only after reading 20 papers on photosynthesis do you realize that there could be important differences in the results because researchers use two different methods to estimate average daily light intensity. This means that you can end up reading some papers several times to fully encode them into your database, and then you will also have to update your protocol.

Fourth, although the best practice is to fully test all model assumptions, look for publication bias, perform a sensitivity analysis, and calculate the power of every test you conduct, few published ecology and evolution meta-analyses do this. Almost none, for example, report statistical power. In our view, such shortcomings do not invalidate a meta-analysis any more than, for example, failure to analyze subsets of the data invalidates a primary empirical study. In short, while you should strive to be as thorough as possible, do not be put off by fears that you will be tackling some elements of the process less thoroughly than others (due to lack of expertise and/or time). The bottom line is that if you have a clearly defined protocol you will produce a data set whose limitations can be interrogated further by others, or yourself, long after you have published your meta-analysis.

A FEW WORDS ON EACH STEP

As the two figures show, each step in the process is covered in detail in one or more relevant chapters. Here we simply provide a short description of what each step involves. We also offer tips based on our personal experiences. Someone once said, “advice is just a set of anecdotes about what worked for the giver,” so you can take what we say at face value, or be more circumspect and see where our views fall relative to the expert advice offered by our colleagues in the rest of this book.

Formulate question and define scope

This is the fun bit. What do you want to know?

- (1) Do you want to evaluate a theoretical hypothesis? If so, you will need studies that use experiments to test for causality. You should be careful not to mix up studies that report an observed relationship and relationships identified through experimental manipulation (i.e., you should encode the study type). (Note that in evolutionary biology, “natural experiments” and “phylogenetic correlations” are sometimes considered to provide insights on causality, especially with respect to identifying selective forces; see Shadish et al. 2002).
- (2) Do you simply want to test whether there is a consistent relationship between two or more variables (i.e., a biological rule)? If so, observational and experimental data can probably be combined (although it is still advisable to encode study type and test for its influence).

- (3) When formulating a question, think carefully about how widely you wish to generalize and what size data set you can manage. For example, it would be great to know the general relationship between genetic heterozygosity and fitness, but do you have the resources to examine all studies published on this subject? Or should you confine yourself to a specific aspect of fitness (e.g., mating success or longevity) and/or taxonomic group (e.g., only birds or all vertebrates)? The answer will partly depend on what your initial search of the literature reveals; if there are too many studies, you will have to be less ambitious in the scope of your questions (or get more funding for library assistance). Our anecdotal tip is that the closer the questions you ask are to your own area of expertise, the easier your life will be. When a research field is distantly related to your own, it becomes more difficult to read papers and extract data, and the likelihood grows that you will overlook studies (e.g., due to use of unsuitable search terms, and unfamiliarity with key narrative reviews and the identity of researchers) or miss some important potential moderators of effect sizes.
- (4) In most cases, the key question in ecological and evolutionary meta-analyses (as in many other fields) can be distilled to the following: What is the mean effect? Does it differ from the null expectation? Can we explain variation (heterogeneity) in the outcome of different studies? The last question requires that you think about factors (moderators) that differ among studies that might affect the effect size estimate. There is much opportunity here for creative insights (see Griffin et al. 2005 for a great example from evolutionary biology). Our practical tip to get you started is to encode studies based on the *population* studied, with the most obvious groupings of taxonomy, geography, habitat or ecotype, functional type (e.g., predator/herbivore, tree/shrub), and organism size; the *methodology* used (e.g., lab or field study, duration of study); how the *outcome* is measured (e.g., sperm count/ejaculate volume, leaf area/leaf mass); and the population to which a treatment is *compared*, that is, the baseline level of the measured outcome (e.g., rate of disease in control plants). You should note that in some cases the comparison will simply be to a null value (e.g., $r = 0$). Remember that moderators of effect sizes can be discrete factors or continuous variables. A final tip is that there is no point having a very large number of moderators, or many levels of a discrete moderator. If the number of papers per group is small you will have very low statistical power. Be prepared to pool groups before conducting your final statistical analyses (e.g., you might code “bird,” “mammal,” “insect,” and “spider,” but end up using the codes “vertebrate” and “invertebrate” instead). You should be keenly aware of confounded moderators, and decide how you want to handle this. For example, if all studies on woody plants are longer than one year and all studies on herbaceous plants are shorter than four months, then one cannot test for study duration without also comparing functional groups and vice versa.

Develop protocol and search criteria

Once you have formulated your questions, you need to write up a protocol that will (1) formally specify the questions you intend to ask, (2) specify an objective search strategy, and (3) establish study inclusion criteria. Specifying the questions you intend to ask (1) will include being specific about potential sources of heterogeneity in effect sizes.

Specifying an objective search strategy (2) entails not biasing data collection toward relevant papers that you are already familiar with; otherwise, this could affect the outcome as it is often easier to remember papers with significant results (Chapter 14). Primarily, this involves

making a list of which electronic databases you will search and what search terms you will use. Chapter 4 includes excellent practical tips on how to search. Secondly, an objective search strategy will involve a decision about how much effort to expend searching the “gray literature.” For example, will you write to colleagues asking for unpublished data, and if so, who and why?

Once you have conducted a search and compiled a list of potential papers, you have to establish study inclusion criteria (3). These criteria are often fairly obvious, and include the following: Does the study fit the scope of your questions? (For example, is it on a marine ecosystem?) Does the methodology fit with how your question is defined? (For example, was there an experimental manipulation?) If so, was it of sufficient magnitude or duration? Does the study contain extractable data, that is, is there sufficient information to extract an effect size, its variance, and the sample size used? Your inclusion criteria will sometimes have to take into account study quality. This is far more difficult to assess than the criteria we have listed above, but it can be just as important (see Chapter 5).

To date, most published ecological and evolutionary meta-analyses have not presented a detailed protocol. This is not a good reason to do the same; quality standards for meta-analysis are rising in ecology and evolutionary biology. It is worth noting, however, that as in primary research, your protocol for searching literature and extracting effect sizes will almost certainly be modified as you proceed. The reality is therefore that, in many respects, your final protocol will end up describing what you did rather than what you ideally wanted to do. Our best advice is to keep your eye focused on two goals. First, you must tell the reader how you collected your data. So, just as in primary research, you provide the reader with enough information on data collection and analysis to allow your review to be repeated and updated in the future. Second, you must have a protocol that forces you to evaluate continually whether your sampling is biased. A protocol increases the objectivity with which you compile data, but it should not blind you to the reality that the process of meta-analysis involves numerous subjective decisions; these are most apparent when trying to decipher the results of a given paper and deciding whether you can extract the necessary data for your synthesis. If more than one person is collecting the data, a well-described and tested protocol is very important to ensure uniformity in data extraction and coding decisions about moderators (Chapter 4).

Scoping (initial search), refining questions, modifying search criteria, and deciding on an effect size measure

Sometimes, if you are confident that most studies will be confined to a few key sources, you might only search a limited set of journals. This was how almost all research syntheses were done originally, before online databases were available (e.g., Gurevitch et al. 1992). No one uses this approach any more because of the explosion of data accessibility (at least for those with access to scientific journals). However, the question of whether to delve into unpublished or “gray” sources remains an important one. In ecology and evolution, using only easily accessed published literature is arguably justifiable when dealing with a fundamental topic with no obvious applications. This work is almost always published in peer-reviewed journals. For example, studies on strategic investment in sperm by males in response to female mating status are unlikely to be featured in government reports and, due to the specificity of the topic, it is fairly easy for an experienced researcher to know which papers will be irrelevant based purely on the journal of publication. In other cases, however, such as questions about the loss of rainforest or coral reefs, it is likely that suitable and sometimes crucial data can be extracted from many sources (e.g., journals, conservation reports, ecological impact assessments). If so, you

need to give some consideration to how many effect sizes each source is likely to provide. This can be difficult to estimate in the absence of prior experience. It matters though, because these and other issues, such as whether non-English language literature is included, can have a real bearing on how unbiased your analysis will be (Chapter 14).

Remember that during a scoping search you are *not* trying to find every study or obtain a preliminary estimate of the mean effect. Your main goals are to (a) estimate how much data there is so that you can decide whether to broaden or narrow your study questions; (b) work out what factors vary among studies that you might encode as potential moderators; (c) decide what criteria mark a study as obviously irrelevant (e.g., if your search identifies 2000 papers to read in full, you will definitely have to make some exclusion decisions based on the title, abstract, and/or place of publication); (d) work out what criteria each potentially relevant study must fulfill before you try to extract an effect size; (e) establish the format of your data extraction form/spreadsheet; and (f) decide upon the most suitable measures of outcome (effect sizes). This last decision will often depend on whether data are reported as a relationship between two continuous variables, in which case the effect size r is the most popular choice in ecology and evolution. Alternatively, the decision may involve a comparison between two groups, in which case there is a range of options depending on whether the response variable is discrete (e.g., alive/dead) or continuous (e.g., body size) (Chapters 6 and 7). Our quick tip is that it is sometimes easiest to conduct separate meta-analyses dividing studies on the basis of the most appropriate effect size. For example, in a review of the effect of a treatment on animal mortality, you might want to first do a meta-analysis using the log odds ratio for studies that measured whether animals were alive or dead at the end of the study, and then do a second meta-analysis using Hedges' d for those that measured lifespan.

Conduct a full search

Once you have established your protocol you will proceed to do a full search. This will generate a large number of studies, but many will be discarded as obviously irrelevant using criteria based on the study's title, abstract, or place of publication. The remaining "potentially relevant" studies will need to be read more closely and divided into relevant and irrelevant. This process can lead to a large reduction in the number of papers at each step (Fig. 4.2). Be prepared for a large number (often the majority) of studies that you initially identify as relevant turning out to be unsuitable for the meta-analysis. This reflects in part the low publication standards in ecology and evolutionary biology (e.g., variance estimates or sample sizes are often missing) and, in part, the diversity of approaches used by biologists to address specific questions. The final step is to try and extract the information you need (effect sizes and moderators) from relevant papers.

A practical tip we offer is that there is a trade-off between building up a pile of relevant papers and then returning to them to extract effect size once you have a finalized data spreadsheet versus extracting data from a paper as you read it. The advantage of the former is that you can be more confident that, having read all the relevant papers once, your spreadsheet contains all the information that you want to extract. The advantage of the latter is that you do not have to read a paper in depth twice. It can be surprisingly hard to establish exactly how a study was designed and which are the relevant data to extract an effect size. Unless you keep good notes (which we strongly recommend), it is often no easier on a second reading. If you are confident that you have a good understanding of the main features of the relevant studies, then you might consider designing a database and extracting data as soon as you classify a paper as relevant. The caveat, of course, is that you might still have to return to these papers if you later discover that you need to encode an additional moderator term or adjust your study inclusion criteria.

Extracting information on the initial reading is most feasible when dealing with studies that closely follow a specific and commonplace experimental design.

Designing the database/datafile and extracting data

Designing a database is an art, and Chapter 5 offers concise and practical advice. We recommend you read it. The basic rules are the same as those for an empirical study; you should ensure the datasheet contains all the information you need and is set out logically so that it is easy to complete and difficult to enter data in the wrong place. A trial run can help to tweak the format. Our practical tip is to use the longest, most complicated looking papers you have in order to trial test your datasheet. If it can handle them, you are off to a good start.

The main difference between an empirical study and a meta-analysis is that you need to record which subjects you did and did not collect data from. In short, keep a bibliographic library of studies and explain why some were excluded (e.g., irrelevant, missing key information needed to satisfy inclusion criteria, not possible to extract an effect size and/or variance estimate). Again, Chapter 5 has sound practical advice on organizing a bibliographic library.

With a good protocol it is fairly easy to encode information on study moderators. In contrast, extracting effect sizes is among the most challenging parts of meta-analysis. It can lead to self-doubt, especially during your first meta-analysis. To extract effect sizes you often have to make subjective decisions (albeit based on procedures outlined in your protocol) about such seemingly easy questions as the relevant sample size (often used to calculate the effect size variance). You will be shocked to discover how poorly results are often presented in the ecological and evolutionary literature (this is a favorite topic among meta-analysts meeting over beer!). In some cases you might simply give up on a paper and place it in the “not possible” category. Do not be put off by this “failure”; everyone in this field who conducts a meta-analysis has the same experience. Take it as a sign that your meta-analysis will be a valuable contribution. You are helping others identify papers with results that are too confusing to be appropriately cited as showing a given effect. These papers were presumably previously cited based on qualitative statements made in the abstract or discussion, which you now know to be difficult to substantiate based on the results presented.

Finally, have a protocol in place to deal with studies that report multiple effect sizes. Specifically, if treatment effects are measured repeatedly over time, you need criteria to determine which comparisons you will use. Likewise, if there is a single control and several treatments varying in magnitude, will you extract all treatment-control comparisons, or only the one between the control and the most extreme treatment? The statistical issues surrounding nonindependence of effect sizes are dealt with in Chapters 16 and 17. If you want to keep life simple, the easiest solution is to design a protocol that only extracts one effect size per study. Be aware, however, that you are sacrificing information and statistical power for ease of analysis. For an excellent basic introduction to complex data structures within a study and the similarities and differences among multiple independent subgroups, multiple outcomes (including time points), and multiple comparisons, we can highly recommend Borenstein et al. (2009, 215–45). They provide very simple, fully worked examples of how to calculate effect sizes and variance in each case. These should help you to decide on the best protocol for your meta-analysis.

One complicating issue in ecological and evolutionary meta-analyses is that there might be nonindependence between studies due to a shared phylogeny history. This can occur when studies of the same species or closely related species produce more similar results (effect sizes) than those from distantly related taxa. Statistical techniques to incorporate analysis of correlations among studies due to shared phylogeny are covered in Chapter 17. However, to be

pragmatic we should note that most published meta-analyses do not fully control for this form of nonindependence. At best, they treat studies of the same species as the equivalent of multiple independent subgroups in a study, such that the final analysis is based on the mean effect size for each species. This may be a serious limitation for some meta-analyses (e.g., those in evolutionary morphology), and little or no problem for other meta-analyses (e.g., those investigating many ecological traits). Note, however, that this relaxed approach will undoubtedly change soon because phylogenies are becoming more readily available and a user-friendly software including phylogenetic data in meta-analyses has recently been developed (Lajeunesse 2011a). For now, however, we think that it is reasonable to publish meta-analyses that do not fully control for phylogeny (as long as this limitation is acknowledged), because when the data are presented in full they can always be reanalyzed in the future by other researchers.

Statistical models, software, calculating means, and looking for moderators

There is no easy way to summarize the many different statistical models for meta-analyses in a few sentences (for details, see Chapter 8). However, we can provide at least some reassurances for the novice meta-analyst. First, in the same way that you use statistics to test empirical data, even though their formal mathematical basis may be a distant memory from your undergraduate years, you can use meta-analysis software without having to recall every detail of the underlying mathematics. However, you do need to be aware of the underlying assumptions and understand at least the gist of the mathematics involved to ensure that you choose the right model for the task at hand. As with conventional statistics, interpreting your results correctly (which requires a good conceptual understanding of what has actually been tested) is vital. (Again, we can highly recommend Borenstein et al. (2009) who provide a very clear explanation of the main statistics used in meta-analysis. The authors take the reader through worked examples in sufficient detail so that each step is easy to follow. By “copying” their approach it is easy to ensure that you appropriately analyze your own data set and know which approaches are needed to answer standard questions and resolve generic problems common to many data sets.)

A second reassurance is that the main tests you will conduct have the same basic logic as those used in conventional statistical analyses. For example, in a primary empirical study you might want to know whether the difference between two treatments is greater than zero (the null value) (i.e., a t -test that is equivalent to an ANOVA with two levels), and whether there is a confounding variable that explains why the difference is bigger in some cases than others (i.e., a two-way ANOVA if the confounding variable is discrete, or an ANCOVA if it is continuous). Likewise, in a meta-analysis you usually want to calculate the mean effect size and see whether it differs from the null value. For example, for the effect size r , does the 95% confidence interval overlap zero? You also usually want to test for more variation among studies in effect sizes than expected by chance. In other words, you are testing for heterogeneity, so that you know whether there is a reason to look for confounding variables. You then test whether the effect size covaries with a continuous moderator variable, or whether the mean effect varies significantly among groups defined by a discrete moderator variable.

Third, as with analyses of empirical data, the real issue is whether your data on effect sizes conform to the assumptions of the statistical model you use, and whether you can interpret the output. For example, it matters that you recognize how a major outlier can change the P -value of a test, or whether nonsignificance is hard to interpret when sample sizes are small and statistical power is low.

Finally, as with conventional statistics, the models become more challenging when there are more predictor variables, when response variables are not statistically independent, and when there are nested, split, or crossed designs. Comparable solutions to those used with empirical data can be applied in a meta-analysis. They can range from the path of least resistance (e.g., use of the mean value when you have repeated measures from the same subject), to the more sophisticated (e.g., use of the most powerful possible statistical design and collaboration with a colleague who has greater statistical expertise).

Chapter 12 offers help in deciding which software to use. It is also useful to consult published meta-analyses by colleagues in your own or allied research areas to see what packages they have used. As a practical tip, we have mainly used MetaWin (Rosenberg et al. 2000). It has limitations (the most obvious being that you can only look at one moderator variable at a time), but the advantage for the novice is that it has a “point and click” format that is easy to use. You should be able to calculate mean effect sizes and test for differences among groups (or for an effect of a continuous moderator variable) within the first hour of using it.

Rigor: Check assumptions, publication bias, and power and sensitivity analysis

As with any scientific study it is always advisable to be self-critical. In the case of a meta-analysis, this can be distilled down to (a) checking that your data fit the assumptions of the statistical model you use, (b) testing for publication bias, (c) calculating the statistical power of key tests, and (d) testing how sensitive your results are to the inclusion/exclusion of data (e.g., whether the removal of effect sizes from unpublished studies or studies you classify as “lower-quality” changes your main conclusions). In general, it is possible to follow simple recipes to conduct these analyses (Chapters 14, 22, and 24), but the more you understand your data and the more you have thought about them, the higher the quality of your meta-analysis is likely to be. Also, as already noted, while it is best to be as thorough as possible, failure to, for example, report statistical power is no more likely to invalidate your meta-analysis than it would a primary study.

Presentation and interpretation of results

Once you have compiled your data set and obtained your results, you need to write up your paper and publish it. Chapter 20 contains a comprehensive checklist that you can work through to maximize the quality of your meta-analysis and publication. In our view, it is important to remember that, like a primary study, there is much benefit in presenting “raw data” showing the effect sizes associated with each study (note that this is now required by some funding sources), and to also offer a full list of studies that were included and excluded from your final data set. These can be lengthy documents, so they are often better presented as electronic appendixes than as tables in the main text. These raw data can considerably facilitate updates of a review as new studies become available and can provide an explanation for conflicting results of two reviews on the same topic. The other point we highlight is that it pays to provide a detailed account of your methods (i.e., your protocol). In some cases this could lead to difficulties with reviewers unfamiliar with systematic reviews because such descriptions may, in their eyes, take up too much space in the Methods section. Again, a simple solution is to be concise in the main text and use an electronic appendix to publish your protocol in its entirety. Finally, while it can be useful to consult published meta-analyses to gain a feel for how to present your results, you should remember that meta-analysis is still a relatively new technique in ecology and

evolutionary biology. Many published meta-analyses are the first conducted by their authors, and the reviewers of the publication might have had no experience conducting meta-analyses and even limited experience reading them. In short, standards are certain to rise so you should not assume a meta-analysis is well presented simply because it is published. (That is, unless it was done by one of us, of course, but we will not say which one!)

CONCLUSIONS

So there you have it. Ultimately, like riding a bicycle, the only way to become comfortable with conducting a meta-analysis is to give it a try. We strongly encourage you to do so because there are many benefits (Chapter 1). You will find that thinking in terms of effect sizes makes you a better scientist who is more aware of the disparity between the outcome of individual studies and the sweeping generalizations that can ensue (Chapter 23), and of the perils of relying solely on *P*-values to classify studies (Chapters 1 and 23). You will also find that if you persevere and publish your meta-analysis it will be a far longer-lasting (and we hope more widely cited) publication than an equivalent narrative review. If nothing else, we can assure you that simply attempting to extract effect sizes from seemingly relevant papers will teach you some valuable lessons. You will discover that many researchers do an appalling job of presenting their statistical results. (Where are the degrees of freedom? What is the sample size? What is the direction of the difference between the two treatments? How did this paper ever get past peer reviewers and editors?) This will help you improve both the presentation of your own primary research and how you review manuscripts. You will realize that many studies do not show what they are cited as demonstrating. This can be a sobering experience. Meta-analysis forces you to examine the primary literature more critically. Finally, you will find that there are often far fewer papers with usable data than you originally assumed. This can be a useful insight allowing you to decide which primary studies are worthwhile and/or to convince a reviewer that, in fact, your own primary research is not testing a repeatedly confirmed hypothesis. So think of a question and start formulating your search protocol. Good luck!