Research Synthesis Methods

(www.interscience.wiley.com) DOI: 10.1002/jrsm.8

Published online in Wiley Interscience

BOOK REVIEW

Ingram Olkin and Therese D. Pigott

Evidence-based medicine, cost effectiveness, decision-making coupled with research synthesis and meta-analysis have generated a near tsunami of publications in the medical, social, and behavioral sciences. There is, however, generally a lag between the publication of research papers and the publication of books that collate the new results and make them available for use by the practitioner. Some isolated papers in meta-analysis go back many years, but more began to appear in the late 1970s' and early 1980s.

The first book '*Meta-Analysis in Social Research*' by G. V. Glass, B. McGraw and M. G. Smith [1] appeared in 1981. Subsequently, 9 books appeared by 1990, 16 from 1991 to 2000, and 16 since 2000. A review of the book '*Statistical Methods for Meta-Analysis*' by L. V. Hedges and I. Olkin [2] published in 1985 appeared in the *Journal of Educational Statistics* (1988), authored by Stanley Wasserman [3]. A second set of book reviews examining many of the books published prior to 1994 appeared in the same journal (1998). This collection was authored by Betsy Becker and the Synthesis Research Group [4] at Michigan State University.

The current plan is to periodically review books on research synthesis and meta-analysis designed to serve the readership of this journal.

References

1. Glass GV, McGaw B, Smith ML. *Meta-Analysis in Social Research*. Sage Publications: Thousand Oaks, 1981.
2. Hedges LV, Olkin I. *Statistical Methods for Meta-Analysis*. Academic Press: New York, 1985.
3. Wasserman S. *Journal of Educational Statistics* 1988 **13**:75.
4. Becker BJ, Syn RG. *Journal of Educational Statistics* 1998 **23**:77.

A review of three introductory texts for meta-analysis

Therese D. Pigott*†

Over the past decade, interest in comprehensive research reviews and meta-analysis has grown, as reflected in a number of texts on methods and statistical techniques published during this period. This review will discuss three of these texts that are aimed at practitioners who are learning meta-analysis for the first time. These books are Lipsey & Wilson's (LW) *Practical meta-analysis* [1], Sutton, Abrams, Jones, Sheldon & Song's (SAJSS) *Methods for meta-analysis in medical research* [2], and the newest addition to these introductory resources, Borenstein, Hedges, Higgins, & Rothstein's [BHHR] *Introduction to meta-analysis* [3]. The review will focus on the statistical techniques of meta-analysis that vex students new to this area of study, namely the computation and analysis of effect sizes.

While the three texts share common features and approaches to statistical methods for meta-analysis, there are areas of difference between the practice of meta-analysis in medicine and in the social and behavioral sciences. LW draws on the authors' extensive experience in the social and behavioral sciences in the authors' presentation of techniques for comprehensive research

Research Methodology, School of Education, Loyola University Chicago, Chicago, IL, U.S.A.
**Correspondence to: Therese D. Pigott, Research Methodology, School of Education, Loyola University Chicago, 820 N. Michigan Ave., Chicago, IL 60611, U.S.A.*
†E-mail: tpigott@luc.edu

reviews. SAJSS's discussion orients toward the research designs used in medicine, namely randomized controlled trials, and the usually limited number of studies included in medical meta-analyses. BHHR aims to bridge the gap between the medical and social sciences by presenting material relevant to both contexts. As a result, each text differs in the terminology and techniques used to address particular issues in meta-analysis.

This review has two aims: (1) to provide an overview of the content of these three texts for novice meta-analysts, and (2) to assist instructors of meta-analysis in choosing appropriate resources for training. The focus here is on issues that students find problematic in a first course on meta-analysis. These issues are the computation of effect sizes, the differences between random and fixed effects assumptions in meta-analysis, and the use of models to explore heterogeneity among effect sizes.

Computation of effect sizes. One of the first issues raised by students is the meaning of the effect size of a study. In the social sciences, the most commonly used effect sizes are the standardized mean difference and the correlation coefficient. In medicine, some form of the odds ratio or rate ratio is used most frequently. All three texts take a slightly different approach to covering the computation and interpretation of effect sizes although the basic formulas and terminology remain consistent across the three presentations. The texts also differ in the amount of guidance given for computing effect sizes from a range of summary statistics. Many novice meta-analysts struggle with whether an effect size can be computed when the standard statistics are not provided in a study report.

Computing overall effect size, and considering sources of heterogeneity. Most students studying meta-analysis have difficulty grasping the possible sources of variance among effect sizes. Since the practice of meta-analysis has shifted toward the assumption of random sources of variance, meta-analysts need to understand the differences between the interpretation and computation of effect size means under the assumptions of fixed and random sources of variance. The greatest difference between the three texts lies in their treatment of how to compute the overall mean effect size and to assess homogeneity in the context of fixed and random sources of variance. Some students also struggle with the idea that the analysis of effect sizes requires weighting by the inverse of the variance of the effect size. The three texts provide a range of explanations for the use of weighting in meta-analysis.

Effect size models: *Examining heterogeneity.* All three texts motivate the use of effect size models as a way to explore the heterogeneity among effect sizes. Each discusses both categorical models (analogues to ANOVA) and regression techniques for fixed and random effects. They differ in the emphasis on each modeling strategy as well as on the support provided for computing the models. The largest difference among the texts is reflected in their treatments of fixed, random, and mixed effects models. The following sections examine how each text in turn treats these three issues.

Lipsey & Wilson's (*2001*) *Practical Meta-Analysis* [*LW*]

LW's text, one of the earliest written for teaching meta-analysis, includes a discussion of all stages of a comprehensive research review, including specifying a research question, establishing study eligibility, developing coding strategies, managing data, and using statistical techniques for meta-analysis. The emphasis in this text is on providing practical advice and guidance for conducting a meta-analysis. As such, this text is well-suited for an introductory overview of the process of research reviews with an emphasis on basic statistical methods of meta-analysis.

Computation of effect sizes. LW introduces the effect size by defining a research finding as a 'statistical representation of one empirical relationship involving the variable(s) of interest' (p. 35). The effect size is the encoding of the relationship into a standardized statistical form. This text divides the types of effect sizes into those for one- and two-variable relationships. These effect sizes include single-group pre–post effects, correlations, differences in both means and proportions, and differences in rates or odds ratios. Each discussion is accompanied by the formulas for computing the effect and a narrative example of a study that would use the given effect size.

In keeping with their goal to provide practical advice, LW use a unique notation for effect sizes and for effect size models that is geared toward non-statisticians. For example, the standardized mean difference effect size is designated as

$$ES_{\text{sm}} = \frac{\overline{X}_{G1} - \overline{X}_{G2}}{s_{\text{p}}} \tag{1}$$

where the subscript sm refers to the standardized mean difference effect size, the $\overline{X}_{Gi}$ are the group means for groups $i = 1$ and 2, and s_{p} is the pooled standard deviation. LW refer to each type of effect size, *ES*, with different subscripts. For example, the odds-ratio is given as ES_{OR}, and the correlation is written as ES_r. A similar notation is used for the standard error of the effect size, with SE_{sm} indicating the standard error of the standardized mean difference.

LW also provides a number of tables and appendices geared toward reviewers conducting a meta-analysis for the first time. For example, the authors include flowcharts for deciding on an appropriate effect size with group contrast data and with data involving correlations and associations. Another useful contribution in the effect size chapter is a summary of all the formulas for each type of effect size presented along with its associated standard error. Unlike the other two texts, LW also presents methods for computing an effect size from a range of statistical information in the appendix. For example, methods for computing a standardized mean difference from *t*- and *F*-tests, from frequency data, and from analysis of variance results are given.

They provide a parallel discussion for computing the odds ratio and the correlation from alternative data sources. In the text, LW also explicitly discuss computing effect sizes when the two associated variables are on different measurement scales (such as the relationship between a dichotomous variable and a continuous one).

Computing overall effect size, and considering sources of heterogeneity. LW introduce the analysis of effect sizes by contrasting the nature of individual person data with data from a meta-analysis. Whereas social and behavioral scientists are familiar with issues related to individual subjects, meta-analysis has as its unit of analysis a primary study. They describe data from a meta-analysis as 'lumpy', comprised of studies with different sample sizes, and thus differing levels of sampling error. Since effect sizes are based on sample statistics, these effect sizes will also differ in their precision with larger studies producing more precise estimates of effect size than smaller studies. Thus, all data analyses involving effect sizes require weighting to account for differences in precision.

LW discuss the analysis of effect sizes in two separate chapters in their text. Chapter 6 of the text outlines the stages of the analysis of effect sizes, including the computation of the fixed effects weighted mean, its standard error, and the homogeneity test, *Q*. The next part of Chapter 6 moves to a discussion of the assumptions a meta-analyst might make about the variation that is beyond subject-level sampling error, an introduction to the notion of fixed and random effects models. The text does not present the formulas for computing the weighted mean under the random effects model in this section of the book, reserving this discussion for Chapter 7. The beginning of Chapter 7 provides computational strategies for obtaining the fixed effects and random effects weighted mean, their confidence intervals, and the test of homogeneity.

Effect size models: *Examining heterogeneity.* As mentioned above, Chapter 6 of LW introduces the distinctions among fixed, random, and mixed models by describing three assumptions a meta-analyst could make about the variation among effect sizes. Variation among effect sizes may be due to (1) random differences among studies whose sources cannot be identified, (2) systematic and identifiable differences among studies, and (3) both systematic and random sources. These three assumptions correspond to the random, fixed, and mixed effect models.

In general, the authors limit their discussion of the differences among these three assumptions, focusing more on a conceptual treatment rather than on the estimation of these models. LW explain that in a random effects model, the variance associated with each effect size has two components, one associated with subject-level sampling error within each study, and one associated with random effects variance. LW give the total variance, v_i^* of an effect size under random effects as

$$v_i^* = v_\theta + v_i, \tag{2}$$

where v_i is the standard error of the target effect size (i.e., SE_{sm} for the standardized mean difference), and v_θ is the random effects or between-studies variance component. LW also acknowledge the difficulty with computing v_θ, referencing the iterative method using maximum likelihood, and providing the method of moments estimator as

$$v_\theta = \frac{Q-(k-1)}{\sum w_i - (\sum w_i^2 / \sum w_i)}, \tag{3}$$

where Q is the value of the homogeneity statistic, k is the number of effect sizes in the sample, and w_i is the inverse of the standard error of the effect size, or $1/v_i$. While the method of moments estimator in Equation (3) is easily computed, LW acknowledge that it results in less accurate estimates than other methods.

LW's Chapter 6 presents a conceptual discussion of mixed models and includes an example of the results from a mixed model in Chapter 7. LW define mixed models as assuming that 'the effects of between-study variables, such as treatment type, are systematic, but that there is a remaining unmeasured … random effect in the effect size distribution in addition to sampling error' (p.124). In Chapter 7, LW expand this treatment in their presentation of results from a mixed effects model using weighted regression. In order to estimate a mixed effects regression, LW compute the random effects variance component, v_θ after accounting for moderator variables. In other words, the variance component is computed based on the residual variance from the regression model that includes a set of fixed moderators. LW do not provide formulas for this model, instead providing an illustration of the output from their SPSS macros for mixed effects models.

Chapter 7's discussion of modeling strategies is integrated with illustrations using both EXCEL and SPSS to compute the models. They begin with examples of how to compute the weighted mean, its standard error, and the homogeneity test in fixed effects models. They provide the formulas for the within-group, and between-group Q statistics using the output from Wilson's SPSS macros (available on Wilson's website, http://mason.gmu.edu/~ dwilsonb/ma.html). Weighted regression models are then presented using SPSS, with the caveat that the F-test for regression, the t-tests for the coefficients, and the standard errors of regression coefficients are not correct, a point not always recognized by novice meta-analysts. Wilson's website also provides macros for SAS, SPSS, and STATA that compute the corrected regression coefficient standard errors.

In summary, LW's text offers a comprehensive introduction for researchers new to the field of meta-analysis. A researcher beginning a research synthesis in the social sciences could use the advice and macros in the book to complete a basic meta-analysis. Wilson's macros continue to be available to aid meta-analysts in completing the statistical analyses in SPSS, SAS, and STATA. The appendices included are particularly useful in providing methods for computing effect sizes from various summary statistics given in studies, and resources for computation and for coding studies. Given that LW is one of the earlier instructional texts on meta-analysis, recent advances are not included. Meta-analysts wishing to conduct more complex analyses will need to consult recent texts for guidance.

Sutton, Abrams, Jones, Sheldon, & Song's (2000) Methods for Meta-Analysis in Medical Research [SAJSS]

SAJSS's text is aimed at those conducting meta-analyses in medicine. While SAJSS naturally focus on the issues faced in medical meta-analyses, they also include many techniques that have potential applications in the social and behavioral sciences. SAJSS provide introductions to a number of meta-analytic techniques that are likely unfamiliar to social and behavioral scientists such as Bayesian meta-analysis, individual patient meta-analysis, techniques for observational studies and survival data, and cumulative meta-analysis. The second half of SAJSS covers many different topics including ways to detect and correct for publication bias, the assessment and adjustment of meta-analysis results for study quality, and a treatment of formal methods for sensitivity analysis. The technical expertise assumed in SAJSS's text is higher than in the other two texts, so that a novice meta-analyst will need the guidance of a statistician to conduct the analysis, or a strong background in statistics. Nevertheless, SAJSS provide a comprehensive overview of the range of methods used in medicine, and an introduction to techniques that have yet to be widely applied to the social and behavioral sciences.

Computation of effect sizes. In medicine, the most common effect sizes are methods of encoding incidence rates or odds. SAJSS discuss the major forms of effect sizes as either binary or continuous, and provide a comprehensive treatment of the various ways to represent binary outcomes. SAJSS make a distinction between binary outcomes for descriptive purposes in a single sample such as incidence rates, and outcomes for comparative purposes such as the odds ratio and relative risk. They also present effect sizes not common in the social sciences such as the relative risk and the number needed to treat. The continuous data effect sizes covered are the raw mean difference and the standardized mean difference. The discussion of each effect size is followed by a worked example. SAJSS use notation for the effect size that is commonly used in the meta-analytic literature and is based on standard statistical notation. For example, the effect size is generically referred to as T_i with its associated variance, v_i.

Although SAJSS do not discuss the computation of effect sizes when typical summary statistics are unavailable as in LW, they do provide a chapter on meta-analysis of different types of data. In Chapter 14, SAJSS provide strategies for dealing with ordinal outcomes, combining binary outcomes reported on different scales, and non-parametric methods for combining effects. This chapter also includes an introduction to diagnostic test accuracy meta-analysis, a topic not covered in either of the other texts. This chapter concludes with a discussion of vote-counting methods and the combining of *p*-values across studies.

Computing overall effect size, and considering sources of heterogeneity. In presenting methods for computing the mean effect size, SAJSS take an approach consistent with the smaller numbers of studies typical in a meta-analysis in medicine. After introducing the types of effect sizes, they begin their treatment of the analysis of effect sizes by discussing the assessment of heterogeneity. They present the homogeneity test called *Q* by LW as the standard chi-square test [4], but emphasize that the chi-square test is usually underpowered due to the small sample sizes in medical meta-analyses. Given the low power of the homogeneity test in medical meta-analysis, SAJSS suggest the use of graphical techniques for assessing heterogeneity including the forest, radial, and L'Abbe plots.

Chapters 4 and 5 in SAJSS discuss fixed and random effects models, respectively. The fixed effect model chapter opens with the explanation that in the fixed effects method for combining study effects, studies are assumed to be estimating a single true underlying effect size. The chapter then continues with a presentation of the inverse variance-weighted method attributed to Birge [5] and Cochran [6] along with worked examples for combining the log-odds ratio and the standardized mean difference. SAJSS do not provide an elaboration of why the inverse variance-weighted method is necessary. This chapter also presents two alternative methods for directly combining odds ratios much more common in medicine, the Mantel–Haenszel method and Peto's method, describing both the benefits and difficulties in applying these methods.

In Chapter 5, SAJSS describe the random effects methods for combining study estimates as reflecting the random variation both within and among trials in a meta-analysis. SAJSS state that the random effects method 'assumes that the study-specific effect sizes come from a random distribution of effect sizes with a fixed mean and variance' (p. 73). SAJSS draw from random effects models in the general statistical literature in their discussion. In the random effects method, estimates of effect sizes, T_i, can be conceptualized as the sum of the true effect size, θ_i, plus the error of the estimate, e_i, or, algebraically,

$$T_i = \theta_i + e_i. \tag{4}$$

The variance of the effect size, T_i, has two components, the random effects variance, τ_θ^2, and the sampling variance for the *i*th effect size, v_i, given by

$$\text{var}(T_i) = \tau_\theta^2 + v_i. \tag{5}$$

SAJSS present a computation of the random effects variance, τ_θ^2, that differs slightly from LW's method of moments computation of v_θ (LW's notation for the random effects variance). SAJSS define a quantity *U*, as

$$U = (k-1)\left(\overline{w} - \frac{s_w^2}{k\overline{w}}\right), \tag{6}$$

where $\overline{w}$ is the mean of the weights, $w_i = 1/v_i$, from the k studies, and s_w^2 is the variance of the weights, w_i. Given this quantity, the estimate of τ_θ^2 is

$$\begin{aligned} \tau_\theta^2 &= 0 \quad \text{if } Q \leqslant k-1 \\ \tau_\theta^2 &= (Q-(k-1))/U \quad \text{if } Q > k-1. \end{aligned} \tag{7}$$

SAJSS provide a short discussion of the derivation of this estimate of τ_θ^2, and briefly cite alternative estimation methods using maximum likelihood and restricted maximum likelihood. Chapter 5 concludes with what SAJSS call extensions to the random effects model, including other methods for estimating τ_θ^2 under special conditions.

Effect size models: *Examining heterogeneity*. The medical meta-analysis literature refers to effect size models in terms slightly different from the social sciences. In the medical meta-analysis literature, the term subgroup analysis refers to the use of separate meta-analyses conducted on particular subsets of studies. In the first example given in Chapter 6, SAJSS present results from a meta-analysis of 34 randomized controlled trials for reducing cholesterol. One subgroup analysis divides the trials into those that used either a dietary intervention or a drug treatment. These subgroup analyses consist of separate meta-analyses for each group, and in this example, each study contributes an effect size to only one group. The results presented include forest plots of each subgroup of effect sizes, and reporting of the random effects mean and confidence interval. No formal statistical comparisons are made between the means of the two subgroups.

The analysis presented in SAJSS differs from the categorical ANOVA models discussed by LW. In the ANOVA models, effect sizes are divided into groups on the basis of a factor such as type of intervention, and statistics analogous to the between- and within-group *F*-tests are computed. The ANOVA models presented in LW include a formal test of within-group heterogeneity (as do the subgroup analyses presented by SAJSS), and also a test of the differences among the subgroup mean effect sizes, a test not formally presented in the subgroup analyses in SAJSS.

SAJSS also make a distinction between two types of subgroup analyses. The first involves subsets of effect sizes defined by study or patient characteristics as discussed above. The second is the analyses defined by subsets of patients within the studies being pooled. The latter set of techniques refers to methods that require individual-level patient data. Individual-level patient meta-analyses are more commonly conducted in medicine than in the social sciences since individual-level data can be more accessible in medical trials than in the social sciences.

The bulk of Chapter 6 on exploring between-study homogeneity focuses on regression methods. SAJSS distinguishes between two types of regression analyses. They use the term meta-regression to refer to regression models using fixed effects. SAJSS present meta-regression using the following notation. As above, SAJSS assume k independent effect sizes, $T_1,\ldots,T_k$ with estimated sampling variances of $v_1,\ldots,v_k$. The underlying effect size parameters are given by $\theta_1,\ldots,\theta_k$. The meta-regression model is based on the assumption that there are p predictor variables, $X_1,\ldots,X_p$ that are related to the effect size parameters by a linear model given as

$$\theta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} \tag{8}$$

As SAJSS point out, this model can be computed using weighted regression techniques included in most commercially available statistical analysis programs. SAJSS also highlight the need for adjusting the standard errors of the coefficients when using weighted regression in most commercial statistics packages.

The term mixed effect model, or random-effects regression, is reserved for regression models that have both fixed and random effects. To describe the mixed effect model, SAJSS extend Equation (4), the random effects model for a given effect size, to the model for true effects, θ_i. The true effect sizes for each study depend on a set of predictors plus error, or

$$\theta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + u_i \tag{9}$$

where u_i is the random effect of study i, the deviation of the true effect in study i from the value predicted on the basis of the linear model. The random effect, u_i, is the deviation from the true effect after accounting for the predictors in the linear model. Substituting Equation (4) into Equation (9), SAJSS show that the estimate of the effect size in the mixed effect model is

$$T_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + u_i + e_i. \tag{10}$$

Given Equation (10), the variance of the effect sizes in the mixed effects model, T_i, controlling for the p predictor variables, $X_1,\ldots,X_p$, is

$$v_i^* = \text{var}(u_i + e_i) = \tau_\theta^2 + v_i, \tag{11}$$

where τ_θ^2 is the random effects between-study variance, and v_i, the variance of the ith effect size. The variance component, τ_θ^2, given in Equation (11), depends on knowing the regression coefficients in Equation (10), and thus cannot be estimated using the method of moments given earlier. The regression coefficients, in turn, depend on the value of τ_θ^2. SAJSS provide a description of how to use the method of moments in an iterative process, and also state that maximum likelihood can be used to estimate both the regression coefficients and the variance component, τ_θ^2, simultaneously. The advanced content at the end of Chapter 6 provides more details on methods for obtaining the mixed model solution.

The treatment of mixed effect models in SAJSS is consistent conceptually to that given in LW. Both texts describe the mixed effect model as containing a set of predictors that are associated with some of the heterogeneity among effect sizes with the remaining variance attributed to random sources of variance. Both also describe estimating the random effects variance component based on the variance not accounted for by the predictors in the model. However, SAJSS do not provide a discussion of strategies to estimate the mixed effect regression model.

In summary, SAJSS's text is grounded in the issues faced by meta-analysts in medical research. The emphasis on effect sizes for incidence rates and proportions, on ways to examine heterogeneity other than statistical tests, and on subgroup analyses, all reflect the nature of data in medical meta-analyses, namely, small samples of randomized controlled trials. SAJSS also assume a level of statistical knowledge equivalent to training in advanced statistical techniques. Given that the text does not provide extensive treatment of computational strategies, a more novice meta-analyst using SAJSS's text may need additional resources to complete a meta-analysis. SAJSS, however, also give an introduction to strategies not common in the social sciences, an important contribution to the meta-analytic literature.

Borenstein, Hedges, Higgins, & Rothstein's (2009) Introduction to meta-analysis (BHHR)

BHHR, the newest of the three texts, provides a hybrid treatment of meta-analytic techniques drawing from both the social sciences and medicine. It includes chapters on methods typically used in medicine such as cumulative meta-analysis, and direct methods for combining odds ratios. It also includes discussions of issues faced by researchers but not covered in the other texts such as conditions for conducting a meta-analysis, and power analysis, an area not developed at the time of the other two texts' publication. The real strength of BHHR's text is its attention to conceptual explanations of the methods of meta-analysis.

Computation of effect sizes. BHHR begin their treatment of effect sizes by distinguishing between a *treatment effect* and an *effect size.* The difference here invokes the nature of the study design; a treatment effect refers to a difference between two experimental groups, whereas the effect size is a more generic term for any group difference including those from observational studies such as gender differences. BHHR divide the effect size types based on the kind of summary statistic used. The book includes chapters on effect sizes based on means (raw and standardized mean differences including pre–post designs), those based on binary data (risk and odds ratios, risk difference), and those using correlational data. In each chapter, formulas for computing the effect size are presented, followed immediately by a worked, numerical example. BHHR also include a short chapter discussing the factors that affect the precision of effect sizes. In BHHR's text, effect sizes are denoted by symbols commonly used in the social sciences. For example, d is used for the standardized mean difference, with its variance given by V_d. BHHR use Y_i to refer to the generic effect size, with V_{Y_i} as the effect size variance.

The authors do not provide a discussion of alternative ways to compute effect sizes since they have a text in preparation devoted to effect size computation that serves as a companion to the text discussed here [7]. In the section on effect size computation, BHHR present formulas for converting among effect sizes. In addition, they discuss vote-counting, combining p-values, and techniques of psychometric meta-analysis.

Computing overall effect size, and considering sources of heterogeneity. After discussing how to compute various effect sizes, and how to convert them into different scales, BHHR begin their treatment of the analysis of effect sizes by discussing fixed and random effect models separately. Each chapter focuses on the computation of the weighted mean effect size, its standard error, and confidence interval. BHHR point out that in the fixed effects model, the meta-analyst assumes that the studies are all measuring a single, common effect size. The fixed effects chapter proceeds with a graphical representation of the common effect size in the fixed effects model, and the impact of sampling error on a study's estimate of that true effect. The chapter concludes with the presentation of formulas for computing the fixed-effects weighted mean, its standard error, and the Z-test for the weighted mean effect size.

In contrast to fixed effects models, BHHR describe the random effects model as assuming that the true effect size varies from study to study; there may be different effect sizes underlying each estimated effect. BHHR use a graphical representation to distinguish between the distribution of the true effect size that underlies each study, and the distribution of the mean of those true effect sizes. The graphical representation illustrates that the difference between the observed effect size and the mean of the true effects consists of two components: the variance of the observed effect size from its true study effect size, and the distance of the true study effect size from the overall mean of all true effect sizes. These two components correspond to the within-study and between-study variances.

BHHR provide the steps needed to compute the random effects mean. They refer to the parameter τ^2 as the between-studies variance. (SAJSS use the notation τ_θ^2, and LW use v_θ to indicate the between-studies variance, or random-effects variance). The first step is to estimate τ^2. BHHR use the notation T^2 to designate the estimate of τ^2, and provide the method of moments estimator as

$$T^2 = \frac{Q - df}{C}, \tag{12}$$

where Q is the value of the homogeneity test, df are the degrees of freedom for Q (k, the number of studies minus 1), and C is given by

$$C = \sum W_i - \frac{\sum W_i^2}{\sum W_i}. \quad (13)$$

The W_i are the inverse of the within-study variances, denoted as V_{Y_i} by BHHR. This formula corresponds to that given in LW.

Following the fixed and random effects model chapters, BHHR provide an elaborate discussion of the differences between fixed and random effects models in meta-analysis. They compare and contrast how the two models would behave under certain conditions, such as the influence of small or large studies on the computation of the weighted mean effect, and the width of confidence intervals for the overall effect size. BHHR also give explicit advice about how to choose between the models, and warn against using the test of homogeneity test as the reason to choose random effects.

After the random and fixed effects discussion, BHHR provide a chapter containing worked examples for computing the weighted mean, standard error, and Z-test for continuous, binary, and correlational effect sizes under both the fixed and random effects models. The chapter on identifying and quantifying heterogeneity provides the most comprehensive treatment of the three texts, incorporating a discussion of I-squared, a measure that post-dates the other two texts. The chapter begins with a graphical representation of homogeneity, the dispersion across studies relative to the within-study error. BHHR first present the Q test of homogeneity (SAJSS's chi-square test). They emphasize that Q is a test of the null hypothesis that all studies share a common effect size, and does not provide an estimate of the amount of true variance. BHHR then continue with a discussion of the method of moments estimator for τ^2, making explicit the notion that the estimate of the random effects variance is a measure of homogeneity. The last section of the chapter on homogeneity presents the I-squared statistic [8], a measure of the proportion of observed variance that reflects real differences in the effect size. BHHR point out that like Q, I-squared does not allow a substantive interpretation of the true variance among effect sizes, only a measure of the relative amount of between- to within-study variation. The chapter concludes with formulas for confidence intervals of the random effects variance component and for I-squared. The next chapter in BHHR presents the idea of prediction intervals for random effects models, the interval that incorporates the dispersion of the true effect size. The section on homogeneity concludes with a chapter containing worked examples of the measures of homogeneity for continuous, binary, and correlational data.

Effect size models: *Examining heterogeneity*. Of the three books, BHHR's text provides the most elaborated treatment of effect size models, particularly with regard to categorical models (the analogue to ANOVA in meta-analysis) with random effects. Three separate chapters discuss subgroup analyses, meta-regression, and special issues related to subgroup analyses and meta-regression. BHHR use the term subgroup analyses to refer to ANOVA models where the goal in the analysis is to compute the mean effect and variance for each subgroup, and then to compare the mean effect across subgroups. This use of the term subgroup analysis differs from that used in SAJSS, where separate meta-analyses are conducted within each subgroup, but the means of those subgroups are not formally tested. The chapter on subgroup analyses centers on three different models: fixed effects, random effects with separate subgroup estimates of the variance component, and random effects with a pooled estimate of the variance component. Once the subgroup mean effects and variances are estimated, BHHR provide three methods for comparing subgroup means. With only two subgroups, BHHR illustrate the use of the Z-test. The second method BHHR present is the analogue to ANOVA models, where the overall homogeneity test Q is partitioned into within-group and between-group components, analogous to the within-group and between-group F-tests in ANOVA. This method is equivalent to that given in LW. The third method presented in BHHR is to treat the subgroup means as individual effect sizes, and compute the homogeneity test, Q, using the subgroup means.

While BHHR's presentation of fixed effects subgroup analysis shares features of LW's and to a lesser extent SAJSS's texts, their discussion of subgroup analyses with random effects provides an explicit rationale for the different ways to estimate the variance component, and represents a unique contribution to the meta-analysis training literature. BHHR see the decision to use a pooled estimate of the variance component as dependent on whether the meta-analyst assumes the same between-study dispersion within subgroups. In some cases, the researcher may believe that studies within each subgroup have a between-study variance component that differs from another subgroup. As they also point out, computing separate variance components for each subgroup requires sufficient numbers of studies within each subgroup. BHHR summarize their recommendations in a flowchart for selecting a computational model for subgroup analyses.

At the end of the chapter, BHHR clarify their use of the terms fixed, random, and mixed models in a conceptual way. A fixed-effects subgroup analysis utilizes a fixed-effects model both within and across subgroups. BHHR use the example of a subgroup analysis based on gender to illustrate the fixed-effects subgroup analysis. The meta-analyst using a fixed-effects subgroup model would assume that the studies within each subgroup are estimating a common effect. In addition, since there are only two possible categories for gender, the factor gender is considered fixed. A mixed-effects subgroup model, which they advocate, uses a random effects model within subgroups. The comparison among subgroup means would treat the grouping variable as fixed, as opposed to a random factor, which would indicate that the subgroups are sampled from some population of subgroups. For example, a primary study might randomly sample a set of counties to be included in an analysis so that the comparison among county means would need to include a variance component for variation among counties. A fully random-effects model would treat the subgroups as randomly sampled from a population of subgroups when comparing the subgroup means. However, BHHR indicate that a discussion of the fully random effects model is beyond the scope of the text.

Table I. Comparison of the three introductory texts.

Topic	LW	SAJSS	BHHR
Effect size computation and interpretation	Computation of a range of effect sizes	Focus on effect sizes for incidence rates	Computation of a range of effect size
	Guidance for computing effect sizes from various statistics given in a study	No guidance for computation of effect sizes from study reports	Guidance for computation of effect sizes to be published in a companion volume
	Illustrations of computations	Illustration of computations	Illustration of computations
Combining effect sizes and assessing overall homogeneity	Thorough discussion of computing fixed effects mean and homogeneity statistics	Focus on assessing homogeneity using both statistical and graphical methods	Thorough discussion of assumptions and computation of fixed and random effects means
	Conceptual discussion of random effects mean	Thorough discussion of assumptions and computation of fixed and random effects means	Comprehensive presentation of statistical methods for assessing homogeneity including Q, I^2, and test of τ^2
	Macros provided for computing on author website	Illustrations of computations but no guidance for use of commercially available packages	Illustrations of computations and EXCEL spreadsheets on author website
Examining homogeneity	Thorough discussion of fixed effects models using ANOVA	Discussion of subgroup analyses as defined in the medical literature	Comprehensive treatment of effect size models, particularly ANOVA models with random effects
	Conceptual discussion of random and mixed models using ANOVA and regression	Thorough discussion of meta-regression (fixed effects regression), and mixed effects regression model	Thorough discussion of regression models under fixed and random effects
	Illustrations of analyses and macros provided on author website	Illustrations of analyses but no guidance for computations	Illustrations of analyses but limited guidance for computations of regression models
Unique features	Notation suited for non-statisticians	Inclusion of techniques not common in social sciences	Inclusion of more recent meta-analysis methods such as power analysis
	Guidance for all stages of a research review so that text could be used to complete a simple meta-analysis	Advanced topics including publication bias and Bayesian meta-analysis	Explicit guidance for complex data within studies such as multiple outcomes
Limitations	Publication date limits inclusion of more recent techniques	Requires knowledge of advanced statistical techniques	Limited references for techniques discussed
	Complex data analyses not covered	Little guidance given for computation of models	Limited guidance for computing effect size models such as meta-regression

The next two chapters in BHHR examine meta-regression and provide further reflections on subgroup analyses and regression. The chapter on meta-regression presents and interprets the results from examples of both fixed and random effects meta-regression. BHHR uses the term meta-regression in a more generic way than SAJSS's text; it uses meta-regression to refer to any linear model of effect sizes, whether it utilizes fixed or random effects. The final chapter in this section discusses in more detail the issues involved in choosing between fixed and random effects models, and provides a short note on software. BHHR cover software in the final chapter, but do not provide the meta-analyst explicit guidance on how to obtain the results of meta-regression.

In summary, BHHR's text provides the most comprehensive treatment of meta-analysis methods of the three texts. Like LW, it focuses on a conceptual treatment of statistical techniques, and does not require readers to have a pre-existing and extensive knowledge of statistical theory. One limitation of the text related to this approach, however, is that few references are given for readers who want to understand the underlying theory and background for the advice given. For example, while the treatment of random effects models is an important contribution of this text, authors do not provide any citations for the approaches they

advocate. BHHR's text also has strengths and limitations related to the computation of meta-analytic results. While extensive examples are given for computing effect size means and variances, and for subgroup models based on ANOVA, there is little guidance given on how to compute regression models of effect size.

Summary

All three texts provide a comprehensive treatment of the fundamental statistical techniques of meta-analysis. As a general overview, LW provides practical guidance on all stages of a research review. In addition, the macros and appendices included give enough detail for a meta-analyst with basic statistical knowledge to carry out a simple review. BHHR focuses exclusively on statistical techniques for meta-analysis, providing detailed, conceptual treatments of the techniques commonly used in both medicine and the social and behavioral sciences. Excel spreadsheets for all computations in the text are provided on the text's web site so that a meta-analyst who is at the analysis stage could easily follow the advice given and produce simple analysis of effect sizes. However, BHHR does not provide examples of how to use commercially available statistical packages to obtain estimates of effect size models. SAJSS's text requires a higher level of statistical sophistication, but also introduces a number of issues central to medical meta-analyses that also have potential application in the social sciences. However, social scientists without a strong statistical background may not be able to conduct the analyses described in SAJSS, although this text could serve as a reference for instructors of meta-analysis who want to introduce new methods to social scientists. Table I presents a summary of the strengths and limitations of the texts discussed here.

Reviewing these three texts also highlights areas that need clarification across the fields of social science and medicine. For example, the term subgroup analysis has slightly different meanings across the two contexts. In addition, meta-regression may also imply different types of analyses across the three texts. The three texts also provide differing guidance on the use and interpretation of random effects models. These differences reflect the ongoing discussions in the methodological literature about the proper use of random effects models. Unfortunately, the methodological discussions have left many meta-analysts without a clear understanding of how best to employ and estimate random effects models.

In many ways, these three texts reflect the current state of meta-analysis as a broad field. This state is partially due to the nature of the data and the research questions and designs used in the two fields. However, even this small exercise in reviewing these texts illustrates the need for more communication between those who use meta-analysis in diverse contexts in order to develop methods for assessing the evidence relevant to important social problems. Using a combination of these texts as a basis for a course in meta-analysis in any field would advance the practice of comprehensive research reviewing.

References

1. Lipsey MW, Wilson DB. *Practical Meta-Analysis*. Sage Publications: Thousand Oaks, 2001.
2. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. *Methods for Meta-Analysis in Medical Research*. New York, 2000.
3. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. Wiley: Chichester, 2009.
4. Cochran WG. *Biometrics* 1954; **10**:101.
5. Birge RT. *Physical Review* 1932; **16**:1.
6. Cochran WG. *Journal of the Royal Statistical Society* 1937; **4**(Suppl):102.
7. Borenstein M. *Meta-Analysis*: *Computing Treatment Effects and Effect Sizes*. Wiley: Chichester, in press.
8. Higgins JPT, Thompson SG. *Statistics in Medicine* 2002; **21**:1539.