16

Statistical Models for the Meta-analysis of Nonindependent Data

Kerrie Mengersen, Michael D. Jennions, and Christopher H. Schmid

In previous chapters we considered meta-analysis in which each primary study contributes one estimate of an effect size to the analysis. Each study and its corresponding estimate were treated as statistically independent. In many meta-analyses, however, independence is questionable because there are several effect estimates per study and/or some of the individual studies included in the meta-analysis might not provide independent estimates of the effect (e.g., if the studies have been conducted at the same site). Within-study nonindependence can arise due to multiple measures of the same effect on the same experimental units being made over time, multiple treatments being compared to the same set of control individuals, or different measures being taken (e.g., plant height, dry weight, and photosynthesis rate) from the same experimental units to generate several different effect size estimates. Reasons for likely nonindependence among studies include multiple studies being made at the same location or on the same animals or plants. Sometimes there is a judgment call as to whether studies will be non-independent. For example, a research team might differ slightly in their approach to measurement, the available equipment, or the study system being examined. Different studies by the same team are then likely to produce results that are more similar to each other than to those from another team (i.e., the team is a thus moderator that is a source of heterogeneity in effect sizes). Effect size estimates from the same team are then nonindependent. On the other hand, different research groups might use very standardized techniques on generic study systems so that there is little reason to anticipate team identity influencing effect size estimates. It is also important to remember that several different studies are often published from a single experimental setup in ecology (e.g., FACE, or free air carbon enrichment) experiments are expensive, long term, and complex to set up) so that separate publications are not necessarily independent of each other. All of these examples are concerned with nonindependence among response variables, including effect size estimates. Another common type of nonindependence encountered in meta-analyses is that among explanatory moderator variables; this is dealt with separately in Chapter 20. Finally, a special, but ubiquitous, case of nonindependence among response variables (and effect sizes) in biology occurs due to phylogenetic relationships among study species. This topic is covered in Chapter 17, which we recommend to the reader, even if they are not specifically interested in accounting for phylogeny. It provides an excellent worked example of a general approach to dealing with nonindependence that can be adapted to contexts beyond a shared phylogenetic history (e.g., spatial or temporal correlation).

TABLE 16.1. Examples of situations in which nonindependence arises in meta-analysis.

Type of nonindependence	Aim of the meta-analysis
Studies are independent but can be grouped by site, year, research group, and so on, which might be factors that affect the effect size estimates.	Learn about the group and overall mean effects (and sources of heterogeneity).
A study reports repeated estimates of an effect. Examples include repeated measures, estimates over time, or experimental studies at different doses (i.e., in a dose-response study).	Learn about the mean effect at individual time points or set dosages, compare effects at different time points or doses, and/or assess the overall effect or trend.
A study reports estimates of more than one effect relating to a common ecological/evolutionary phenomenon (e.g., offspring size at birth, survival, immune function as measures of evolutionary fitness or performance).	Create a composite effect for each study, and combine these composite effects to estimate the overall mean effect.
A study reports estimates of more than one effect (e.g., different estimates of plant or animal performance such as biomass, mortality, etc.).	Learn about each effect, while taking the other estimated effects into account.

Here, we discuss nonindependence among effect sizes both within and among studies. We confine ourselves to describing four commonplace situations where nonindependence can occur in ecology and evolution meta-analyses (Table 16.1). Each of these four situations is illustrated with a single case study (data sets in Appendix 16.1). In the real world, a data set will often contain several types of nonindependence. For example, each study may provide several different effect sizes obtained from the same individuals, and there may be some sites that contribute several studies to the data set. In such cases the hierarchical approaches developed in this chapter might need to be extended.

The effects of not accounting for nonindependence in meta-analysis are now well documented in the literature. For example, Riley (2009) argues against the practice of meta-analyzing each of a set of outcomes independently. He shows that ignoring within-study correlation in this way can, on average, increase the mean squared error and standard deviation of pooled estimates. Similarly, Jones et al. (2009) argue against the practice of ignoring correlation between time periods and analyzing a set of time points independently. These authors also show that this practice can result in different pooled estimates and corresponding standard errors. Note that these problems are not resolved by another common practice of conducting the independent analyses and performing some form of post hoc adjustment for nonindependent tests, such as the Bonferroni test.

As a simple example of the effect of nonindependence, consider the problem of computing the difference between two effects, say T_a and T_b. The estimated difference is $D_{ab} = T_a - T_b$. If the effects are independent the variance of D_{ab} is simply the sum of the variances, say ${\sigma_a}^2 + {\sigma_b}^2$. If the outcomes are dependent, however, the variance is:

$$\sigma_a^2 + \sigma_b^2 - 2\rho_{ab}\sigma_a\sigma_b \tag{16.1}$$

where ρ_{ab} is the correlation between the effects. If the effects are positively correlated, then the variance of the difference is smaller after taking this correlation into account. For example, if ${\sigma_a}^2 = 0.02$, ${\sigma_b}^2 = 0.018$, and $\rho_{ab} = 0.2$, then the variance of D_{ab} is equal to 0.0304, compared to 0.038 if the effects were assumed to be independent.

How can nonindependence be addressed in a meta-analysis? Three options are available. First, we could exclude multiple nonindependent estimates and/or only focus on a single

response variable. This reduces the problem to a simple meta-analysis that can be modeled using the approaches described in Chapters 8 to 11. However, in doing this we are ignoring relevant data and losing potential information. Second, we can (wrongly) assume that all the effect size estimates are independent and press ahead with the approaches described in Chapters 8 to 11. Unfortunately, ignoring dependence between effects might increase the likelihood of a type I error (false positive), bias the parameter estimates, and generate incorrect estimates of the corresponding variances. This is because correlated observations contribute a different amount of information about an effect than independent observations. The key idea is that correlated data are not as informative as independent data; the true sample size is smaller if the observations are positively correlated. This affects not only the weighting, but also the variances, and is clearly shown in the example above. There, the incorrect weighting (inverse of variance) will be given to the effect D_{ab} if we erroneously assume that ρ_{ab} is always 0 in all the studies we wish to combine to estimate the mean value of D_{ab}. Third, we can extend the models introduced in Chapter 8 to accommodate nonindependence. We will now describe how to do this for each of the four situations described in Table 16.1. For each situation, we discuss the relevant multivariate hierarchical (sometimes called a multilevel or nested) model. The multivariate component captures the nature of the correlation between effects, and the hierarchical component captures how these effects are nested and combined. We also consider some extensions and alternatives to the general multivariate approaches.

In practice, it is not always possible to fit the desired model due to a lack of information or incomplete reporting within and across studies compiled for a meta-analysis. The problems this creates, and how to deal with them, are considered in the final section of the chapter.

SITUATION 1: GROUPED ESTIMATES OF A SINGLE EFFECT

Effect estimates can often be grouped based on an external factor that might lead to their being nonindependent (e.g., when effect sizes are calculated from data collected at the same study site or on the same species). Data set 1 (Table A16.1) contains 25 effect size estimates and corresponding variances (i.e., 25 studies) that can be assigned to five groups (labeled A to E). Each group contains 3 to 8 studies. We have kept this example deliberately vague since it is intended to represent a wide range of common scenarios. Readers would do well, however, to visualize a plausible example relevant to their own field of study.

The first question to ask is whether effect estimates are statistically independent *within* a grouping, so that they each provide unique data about the effect. This can require some biological insights. For example, if the grouping is by population but each study collected data from different animals, we might be happy to treat each study within a given population as independent. If, however, we have the additional information that studies were collected at different times of the year, we might suspect that there is greater similarity between those studies conducted in autumn than those conducted in spring; this would lead to the correlation among studies varying within each population. Second, we need to decide whether the groups themselves are independent. Again, this can require biological insights, and will often depend on the study question. For example, consider a study measuring growth rates of island-dwelling birds, where the grouping variable is "island." We might be willing to treat data collected on islands two km apart as independent if we are asking a question about growth rates on islands in a Finnish lake. We would be less happy to do so if the data set included islands from lakes across Europe, because the similarity in food availability among islands on a Finnish lake is now high relative to the variation in food availability among islands in the full data set. The correlation among effect sizes is thus likely to be far stronger for some pairs of islands than for others in the European data set.

The general meta-analysis model assumes that each study reports an estimate of a true effect for that study, the true effects within groups are distributed around an overall group effect, and the group effects are distributed around a global effect. In general, suppose that there are I groups, and that the ith group reports P_i estimates of the effect of interest. In the ith group, let T_{ij} and S_{ij}^2 denote the jth estimate and associated variance of that estimate. Each T_{ij} is an estimate of a true effect θ_{ij}. Within the ith group, the θ_{ij} are assumed to be distributed around an overall group effect μ_i, with variance γ_i^2. These overall effects, $\mu_{i.}$ are then assumed to be distributed around a global effect μ_0 with variance ω^2 representing the global effect for all I groups. As stated above, independence is assumed among effects within groups, and among group effects.

In Data set 1 there are five groups. Hence $I = 5$, $P_1 = 4$, $P_2 = 3$, . . . , $P_5 = 8$, $T_{11} = 0.2$, $T_{12} = 0.6$, $T_{21} = 0.8$; the estimated variance of the first effect in the first group is given by $S_{11}^2 = 0.10$, likewise $S_{23}^2 = 0.04$, and so on. This hierarchical model is depicted in Figure 16.1.

Assume that all of the groups are independent, and that the studies within groups are independent. Further, assume that the effect estimates (T_{ij}) and true effects (θ_{ij}) are normally distributed with true variances σ_{ij}^2 (which are estimated by S_{ij}^2 described above). Then the meta-analysis model can be written:

$$\begin{aligned} T_{ij} &\sim \mathrm{N}(\theta_{ij}, \sigma_{ij}^2) \\ \theta_{ij} &\sim \mathrm{N}(\mu_i, \gamma_i^2) \\ \mu_i &\sim \mathrm{N}(\mu_0, \omega^2). \end{aligned} \tag{16.2}$$

Note that it is not necessary to stipulate normal distributions in Equation 16.2. Other distributions can be used if these are more applicable. This is discussed elsewhere in this book.

A fixed-effects model within groups is obtained by setting $\gamma_i^2 = 0$. A fixed-effects model among groups is obtained by assuming that for all studies, $\mu = \mu_0$ and hence, $\omega^2 = 0$. If necessary, covariates can be added at any level of the hierarchy of the model (Chapter 7). A variety of software is available to fit this model (e.g., SAS [PROC MIXED], Stata, MLwiN, WinBUGS,

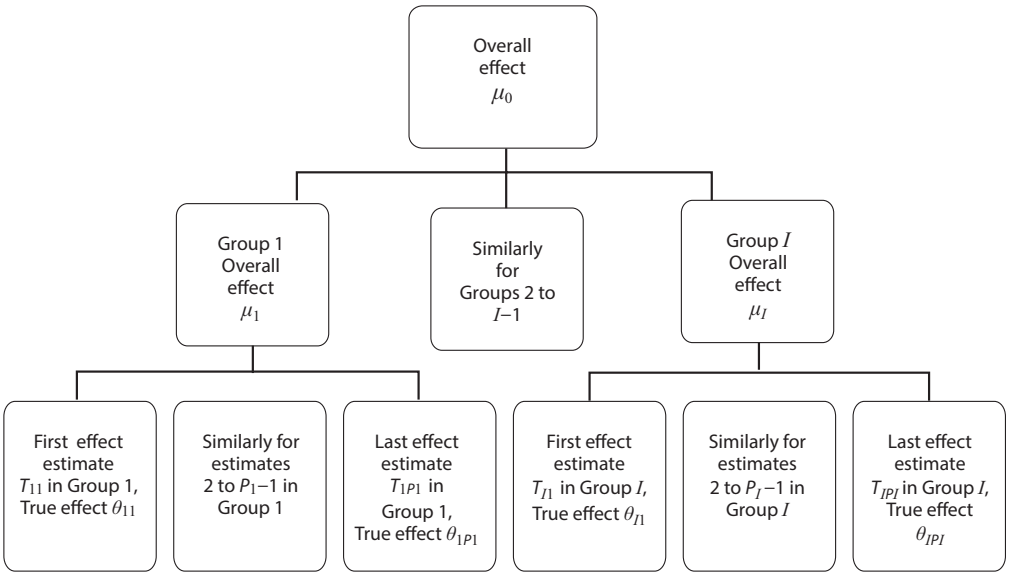

Figure 16.1. Situation 1: Schematic representation of general meta-analysis structure.

TABLE 16.2. The estimates of the mean effect size (and its variance) for meta-analyses of Data set 1 (Table A16.1) that make different assumptions about the independence of effect size estimates.

Research Group:	A	B	C	D	E	Mean
Analysis 1: Ignoring grouping information and assuming all estimates are independent						
FE estimate (SE)	-	-	-	-	-	0.65 (0.040)
RE estimate (SE)	-	-	-	-	-	0.63 (0.076)
Analysis 2: Include group effect: MM approach						
FE estimate (SE)	0.331 (0.137)	0.704 (0.132)	0.647 (0.098)	0.578 (0.067)	0.806 (0.072)	0.65 (0.040)
RE between-study variance	0	0.025	0.162	0.285	0.141	
RE estimate (SE)	0.331 (0.137)	0.701 (0.162)	0.679 (0.194)	0.785 (0.277)	0.855 (0.154)	0.63 (0.113)

Notes:

FE = fixed-effects analysis; RE = random-effects analysis; MM = method of moments.

Analysis 1: In the RE analysis, the between-study variance was approximated as $\hat{\gamma}^2 = \max(0, (X^2 - (k-1))/(\sum W_i - \sum W_i^2/\sum W_i))$ where $X^2 = \sum W_i(T_i - \hat{\mu}_{0FE})^2$ and k = number of estimates. Hence $W_i' = 1/(S_i^2 + \hat{\gamma}^2)$. See DerSimonian and Laird (1986). Note that a similar equation can be used to estimate variance between-study effects within-groups or between-group effects.

Analysis 2: Method of moments estimates were obtained using a two-step approach, with separate meta-analyses of studies for each group to obtain group effect estimates, and the group effect estimates were then combined in a further (separate) meta-analysis. Between-group variance = 0.026.

R [lme, lmer, mcmcglmm]) (Chapter 12). Alternatively, a simple two-step procedure can be used. A meta-analysis is fitted for each group separately by direct calculation using, say, an Excel spreadsheet or statistical software. The calculated group means and variances are then used in a second meta-analysis to calculate the global mean effect.

Returning to Data set 1, we first ran a meta-analysis in which group effects were ignored. We fitted both a simple fixed- and random-effects model (Chapter 9). We then consider a meta-analysis in which groups were included, as in Equation 16.2. The results based on method of moments estimates (Chapter 9) are given in Table 16.2. As you would expect, the inclusion of dependence via the group effect does not change the estimated global mean and its variance when looking at fixed-effects models (since there is no allowance for between-study or between-group variation). In contrast, there was a difference between the random-effects models. The 50% increase in the variance of the global mean effect estimated when fitting a model that includes groups reflects the fact that there is substantial variation in effect sizes among groups. Maximum likelihood and Bayesian estimates of the global mean effect size can also be obtained that take group into account. The WinBUGS code to do so is provided in Appendix 16.2.

An important modification

If there is dependence between studies within groups, Equation 16.2 can be modified so that within each group the study effects θ_{ij} have a multivariate normal distribution. Thus the second

line of Equation 16.2 is replaced by $\theta_i \sim \text{MVN}(\mu_i, \Gamma_i)$, where θ_i is a vector of the true study effects θ_{ij} in the ith group, and Γ_i is a variance-covariance matrix that contains the variances, γ_i^2, of these study effects along the diagonal and their covariances (describing the relationship between the true study effects) in the off-diagonal elements.

Similarly, if there is dependence among groups, the third line of Equation 16.2 can again be modified so that the group effects (μ_i) have a multivariate normal distribution with a variance-covariance matrix containing the (conditional) variances of the group effects (γ_i^2) along the diagonal and covariances among the group effects in the off-diagonal elements.

In short, Situation 1 is a restricted example (all covariances = 0) of more general situations.

SITUATION 2: REPEATED MEASURES ESTIMATES OF A SINGLE EFFECT

Repeated measures studies, including time series and dosage response studies, are common in ecology. The repeated measurements might be taken on the same individual, or they might represent unique sets of individuals measured at different times or at different dosages within a study. In the former case, there is obvious dependence between measurements for each individual. In the latter case, while there is no dependence within individuals, there is still dependence over time within studies. If the aim is simply to obtain an overall estimate of the effect of interest, then the meta-analysis can be collapsed to Situation 1. For example, each time period or dosage can be treated as a separate "study" and the original studies as "groups." We can then modify Equation 16.2 by replacing γ_i^2 with a variance-covariance matrix where the strength of covariance either decreases as the time interval between "studies" increases, or is only dependent on the previous study (autocorrelation). In many cases, however, our aim is to evaluate time or dose trends within the original studies and estimate the overall trend based on the combined information.

We illustrate models that describe trends by considering Data set 2 (Table A16.2), which contains information on annual estimates of shrimp biomass from nine North Atlantic sites; this is part of a large database analyzed by MacKenzie et al. (2003). The data, consisting of biomass (log scale) each year for each study, are provided in Appendix 16.1. The least squares (LS) regression slope coefficient (and associated standard error of this estimate) for biomass per year, fitted for each study independently, is also provided in Appendix 16.1.

When we have access to the primary data (in our example this is the estimated biomass at each time point for each study), our meta-analysis model can have a two level hierarchy. First, we can fit a linear regression over time for each study, and then we can combine the study-specific intercepts and slope coefficients. If, on the other hand, we only have access to the estimated regression coefficients and corresponding standard errors from each study, these estimates can be combined directly across studies using a fixed- or random-effects model.

We will first consider the case in which we have access to the primary data from each study. At the first level, we fit a linear trend assuming that the vector of responses in the ith study has a linear trend over time and that the residuals are normally distributed. Thus

$$\begin{aligned} T_{ij} &= \beta_{0i} + \beta_{1i} X_{ij} + e_{ij} \\ e_{ij} &\sim \text{N}(0, \sigma_T^2) \end{aligned} \tag{16.3}$$

where T_{ij} is the observed response in the ith study at the jth time (X_{ij}), β_{0i} and β_{1i} are the study-specific intercept and linear trend parameters, respectively, and e_{ij} is the residual between the observed and fitted responses in the ith study, $i = 1, \ldots, I$.

Note that Equation 16.3 assumes a common variance for residuals. The model does not necessarily have to be this restrictive; a different variance could be assumed for each study by replacing σ_T^2 with σ_i^2 in the second line of the equation.

At the second level, a random-effects model used to combine these study-specific regression estimates is:

$$\begin{aligned}\begin{pmatrix}\beta_{0i}\\ \beta_{1i}\end{pmatrix} &\sim \text{MVN}\left(\begin{pmatrix}\mu_0\\ \mu_1\end{pmatrix},\Gamma\right)\\ \Gamma &= \begin{pmatrix}\gamma_0^2 & \gamma_{01}\\ \gamma_{10} & \gamma_1^2\end{pmatrix}\end{aligned} \tag{16.4}$$

where $(\mu_0, \mu_1)'$ is the vector of global regression parameters and Γ is the between-study variance-covariance matrix for these parameters (i.e., to take into account that intercept and regression coefficient estimates might covary across studies, rather than assuming that they do not). Thus, μ_0 and μ_1 are the overall intercept and trend values obtained when we synthesize the data from all I studies.

Second, we consider the case in which we only have estimates b_{0i} and b_{1i} of the intercept and slope parameters β_{0i} and β_{1i} for each study, and estimates S_{0i}^2 and S_{1i}^2 of the corresponding variances σ_{0i}^2 and σ_{1i}^2. (Note that although we have used the same notation S and σ, here they are not directly comparable to the terms in Equation 16.3, since we do not have the primary data.) In an ideal world, we would also have estimates of the corresponding covariances, but since these are rarely provided, we will ignore them here. In this case, we can again use Equation 16.4 after first defining

$$b_{0i} \sim \text{N}(\beta_{0i},\sigma_{0i}^2); \qquad b_{1i} \sim \text{N}(\beta_{1i},\sigma_{1i}^2). \tag{16.5}$$

If the focus of the meta-analysis is on a single regression parameter, such as the linear trend, then the above model reduces to a simple univariate fixed- or random-effects model involving only b_{1i}, β_{1i} and μ_1.

Sometimes, we may want to relate the regression slope or intercept to other variables with a regression model called a meta-regression; this is used to explain their heterogeneity. It is tempting to do separate analyses in each primary study and use the resulting estimates in the meta-regression. This type of two-stage analysis must account for the uncertainty in the estimates from the primary studies in the first stage. Without such weighting, the meta-regression estimates will have standard errors that are too small because they will have assumed the first-stage estimates to be true values.

The above models explicitly accommodate both within-study variation (i.e., whenever the regression in the first line of Equation 16.3 has $R^2 < 1$ or the variances in Equation 16.5 are not equal to zero), and between-study heterogeneity. Note, however, that if there is high between-study heterogeneity in slopes, it might be misleading to combine them and discuss a global trend. Again, if necessary, other potential sources of heterogeneity can be included as covariates in the model by directly extending it to a meta-regression or, for categorical moderators, through a multilevel model in which different subsets of studies are first combined within, and then across, the next level. In our example we have tested for a linear trend, but other forms of temporal changes could readily be investigated. Nonlinear trends can be represented by quadratic or higher order polynomial functions or nonparametric splines. The main point is that if each study generates one or more estimated parameters (and associated variances) that describe the trend, these can be subject to meta-analysis using the above model. If normality cannot be assumed, it is sometimes possible to use transformations (e.g., log transformation for count data or logit transformation for binary data), or generalized linear modeling.

If residuals are assumed to be normally distributed, models can be estimated using ordinary least squares (OLS) or a variant of this approach; an example of a variant is the iterative least squares or generalized least squares (GLS). The OLS estimator assumes that the residual term (e_i) has a constant variance of σ^2. In contrast, the GLS estimator assumes that the variance is not constant, which is more typical in ecology. A transformation is then applied by dividing all terms in the regression by the square root of the nonconstant term in the variance of e_i, resulting in a constant variance for the error term that satisfies OLS assumptions. Multivariate meta-analysis methods that utilize GLS have been proposed by Hedges and Olkin (1985), Raudenbush et al. (1988), and Becker (1992). The GLS approaches to meta-analysis are discussed by Dear (1994), Kalaian and Raudenbush (1996), and Timm (1999). The use of GLS models to correct for phylogeny is also covered in Chapter 17. The models can also be estimated using maximum likelihood (ML) or Bayesian approaches (Chapters 10 to 11).

The general model we have described is extensively discussed by Becker and Wu (2007) in a more general statistical context, and by Arends (2006) as one of many methods for multivariate meta-analysis. McIntosh (1996) describes both ML via the EM (expectation-maximization) algorithm, and Bayesian estimation via Markov chain Monte Carlo (MCMC); these descriptions are for analysis of a bivariate two-level hierarchical regression model for treatment effects, when using the control group event rates as covariates. A worked example of its application in epidemiology is provided by Bagnardi et al. (2004). Gelman and Hill (2007) give an excellent description of these models in the wider context of general multilevel models and ML estimation using the software package R.

The choice of a statistical model for meta-analysis of repeated measures data depends on the available data and the aim of the analysis. We briefly consider some examples in the context of time series data, noting that they also apply to dosage-response and general repeated measures studies.

As a first example, a single time point might be of particular biological interest in a time series experiment (e.g., when animals mature, or the endpoint of the study). In these cases, we can extract a single effect estimate from each primary study, and apply a standard meta-analysis model (Chapter 8). The major limitation of this approach is the tremendous loss of information. Moreover, while this approach might make studies with differences in sampling dates more consistent, it is important to critically consider the comparability of the estimates being combined. (For example, is it appropriate to use the end point of studies that differ widely in duration?) It might be possible to include the time interval since the application of a treatment (or some key biological date such as time since hatching) as a covariate in a meta-regression. It is also important to recognize that it is easy to introduce bias into an analysis by only defining the relevant time point after the data have been assessed (Higgins and Green 2011).

As a second example, a common aim in many ecological meta-analyses is to compare effect sizes at two time points. For example (case A), if each study reports effect estimates at the two time points, and/or reports (or we can calculate) a measure of the difference between the two effect estimates, the aim is then usually to estimate the overall mean difference in effect across studies. Alternatively (case B), if independent sets of studies report effect estimates at only one of the two relevant time points, the aim might be first to combine the reported effect estimates for studies looking at each time point, and then to estimate the difference in the mean effect estimates between the two types of studies. (Case A and case B can be considered to be equivalent to a paired and unpaired experimental design, where time periods are paired by study in the first example). The way in which time points are defined can influence the analysis. For example, they might be relative to an event (e.g., weights before and after a treatment, or treatment-control differences before and after an intervention), depict an epoch (e.g., different stages of development), or be chronological (e.g., effect of hunting on population growth in 1970 and 2000).

Cases A and B can both be addressed through a hierarchical meta-analysis model. For case A, at the first level, a single estimate of the measure of interest, such as a standardized difference in the effect estimates between the two time points is obtained for each study (taking into account the covariance between the data at these time points by using Equation 16.1. At the second level, these study-specific measures are combined using a random- or fixed-effects model. (Of course, if the studies directly report the measure of interest, then only the second level of this hierarchy is required.) In case B, at the first level, a random- or fixed-effects model is used to combine the estimated effects within each time point; at the second level, the mean effects are compared by calculating the difference in the means and using Equation 16.1 to calculate the 95% CI for this difference. In both cases, the two levels can be performed as separate steps or, more preferably, a mixed-effects model can be fit to perform both steps in a single analysis. Depending on the nature of the two time points being considered, it might be prudent to include the interval between the two measures in each study as a covariate. Worked examples of assorted approaches to the problem of variation in the interval between estimates are provided by Guo and Gifford (2002), Côté et al. (2005), and Stewart, Pullin, and Coles. (2007). An obvious extension of the previous aim is to compare multiple time points. Again it is possible that either case A or B applies. When comparing, however, it is necessary to correct for multiple testing using, say the Bonferroni correction (Quinn and Keogh 2002, 49) or the false discovery rate (Benjamini and Hochberg 1995).

As a third example, there is the option of combining effect estimates from all time periods across all studies. This approach is useful if there is no well-defined way to choose a particular time point from each study, and if temporal trends are of minor interest. Depending on the data and research question, "time" can be treated as a replicate or as a continuous or categorical covariate. For example, effect estimates can be averaged, ideally taking into account the temporal dependence in the data (Equation 16.1), since observations at close time points are likely to be more correlated than those that are further apart. Alternatively, time can also be treated as a fixed factor in a mixed model setting if the variation among time points is of specific interest to the experimenter, or as a random effect if the time points simply represent a sample of times in a time frame. Rustad et al. (2001) provide a useful worked example from a meta-analysis of the effect of experimental warming. They estimated effects for each site, year, and outcome of interest for three ecosystem soil variables (soil respiration, net nitrogen mineralization, and above-ground plant growth), and then combined the effect sizes across experimental sites at each time period.

We now use Data set 2 to illustrate some of the above approaches. Here, we treat the problem as a meta-analysis of nine individual time series studies (Fig. 16.2).

To illustrate the full hierarchical model, we fit a linear regression for each site at the first level of the hierarchy, and then combined the site-specific intercept and slope parameters (weighted by their respective inverse variances and covariance) in a bivariate model at the second level of the hierarchy. We did this in a one-step process, taking a Bayesian approach using WinBUGS (Chapter 11; code is provided in Appendix 16.2), with normal priors for the trend parameters, a uniform prior for the standard deviation of the residual of the within-site regressions, and Wishart priors for the precision matrices (the inverse of the variance-covariance matrices). The key results are presented in Table 16.3 (posterior estimates for intercepts were obtained but for brevity are excluded).

The estimated overall linear slope is 0.0199 with a 95% credible interval of (–0.046, 0.50). We might therefore conclude there is no evidence for a directional change in shrimp biomass over the study period. Note that, as expected, the posterior estimates of the slope at each site have "shrunk" from the observed LS estimates to the posterior mean. This is particularly evident for Site 2J, where the observed strongly positive trend (LS = 0.16) was an evident outlier.

Figure 16.2. Plot of trends in shrimp biomass at 9 study sites between 1967 and 2000 (Data set 2, Table A16.2). Each line represents a different site.

TABLE 16.3. Posterior estimates of the trend for each site in the Bayesian analysis of Data Set 2 (Table A16.2) and the corresponding 95% credible intervals for these values. Also shown are the mean intercept and slope and the corresponding precision matrix for the mean parameter estimates. The precision is equal to the inverse of the variance (i.e., smaller variance equals greater precision). The ordinary least squares (LS) estimates of slopes are included for comparison. The WinBUGS code for this model is in Appendix 16.2.

Site	Parameter	LS Estimate	Posterior Estimate	95% CI lower	95% CI upper
2J	Slope	0.16	0.02	0.006	0.03
3K	Slope	0.06	0.05	0.03	0.08
3M	Slope	0.05	0.05	0.02	0.08
NGSL	Slope	0.03	0.03	0.01	0.04
ESS	Slope	0.02	0.02	0.01	0.03
5Y2	Slope	0.000	0.001	–0.02	0.01
ICE	Slope	0.01	0.01	0.01	0.02
BS	Slope	–0.002	0.00	–0.02	0.02
SKAG	Slope	0.008	0.01	–0.01	0.03
Mean	Intercept		1.03	0.36	1.71
	Slope		0.02	–0.46	0.50
$1/\gamma_1^2$	1/Variance		1.56	0.102	5.35
$1/\gamma_{12}$	1/Variance		–0.025	–2.75	2.65
$1/\gamma_2^2$	1/Covariance		2.92	0.211	9.03

There was, however, substantial heterogeneity in the overall estimates, as indicated by the precision matrix. This cautions us about how we should interpret the mean slope.

Data set 2 was also analyzed using a linear mixed model estimated by restricted ML (REML) in the statistical software package R. We arranged the shrimp biomass data into a file "shrimp" with three columns: Study (1, 2, . . . , 9), Year (rescaled by "Year-1972," so that 1 = 1973), and Biomass (log). The meta-analysis model was then fit using the command

lme(Biomass∼Year,random=∼1|Study/Year,data=shrimp, method=REML)

This command describes the fixed-effects (within-study) component of the model (i.e., the linear regression of Biomass on Year) and the random-effects (between-study) component (an overall effect, denoted by "∼1," which allows the mean to vary within each study and year). If the variation between studies is considered to be a fixed effect, the above equation is changed to "Study*Year." The output includes the log-restricted likelihood for the model, the overall regression estimates (intercept and slope), and the standard deviation corresponding to the random effects (the between-study variation in trend). Note that in this meta-analysis the studies are automatically weighted by the uncertainty of their respective estimates, since the regression analyses, and consequently the variation in the regression estimates, are included as part of the model.

For additional worked examples of meta-analyses of repeated measures data we refer the reader to Marsh (2001), Harley et al. (2001), Harley and Myers (2001), Barrowman et al. (2003), and Blenckner et al. (2007).

SITUATION 3: COMPOSITE EFFECTS

We next consider the problem of combining effect sizes for different response variables obtained from the same individuals. One option is to perform a separate meta-analysis for each response variable, which means that we must take into account that we will then need to correct for multiple testing, as there is nonindependence between effect sizes estimated from the same study. Sometimes, however, we are interested in obtaining an overall estimate of a general phenomenon (a "latent" or "composite" measure) that is described by a synthesis of these variables (Borenstein et al. 2009, 225).

The hypothetical example we use is inspired by experimental studies that assess the benefits of polyandry by comparing the offspring of females assigned a single male they can mate with x times (monogamous), with the offspring of females assigned x males that they can each mate with once (polyandrous) (Simmons 2005). Data set 3 (Table A16.3) contains data from six studies, each of which reported an effect size (Fisher's z transformed correlation between number of mates and outcome) for up to three different traits: offspring size at birth, survival, and immune function. Each outcome can be considered an indirect measure of "offspring fitness" or "offspring performance." Two points should be noted. First, the variances are very similar for the different traits in a given study. This is because the effect is a correlation, so sample size determines the variance (i.e., $1/[n - 3]$ for z-transformed Pearson's correlation coefficient) and sample sizes tend to be similar within studies. There is, however, no general expectation for variance to be similar for different response traits within a study. Different traits can have different variances, and differences in sample sizes among traits will also influence the variance of the estimates. This is a potential problem (see below). Second, certain traits were not measured in some studies; in studies C and F, the researcher did not measure immune function. Our analysis does not require that every study measures every response trait.

Our first step is to combine the outcomes for a given study in order to calculate a composite mean effect size for the effect of polyandry on offspring performance. There is no information

about which response variables (size, survival, immune function) are more important indicators of performance, so we given them equal status. A composite effect estimate for a study can then be calculated as the variance-weighted average of the available estimated effects, X_i say, for that study (i.e., $\hat{\mu}_{0FE} = (\sum W_i X_i)/(\sum W_i)$, with $W_i = 1/V_i$, assuming a fixed-effects model, and m = number of estimated effects) . If the variances are equal for each response, as in study E, this simplifies to the arithmetic mean:

$$[0.07/0.01 + 0.08/0.01 + 0.08/0.01]/[1/0.01 + 1/0.01 + 1/0.01] = [0.07 + 0.08 + 0.08]/3 = 0.0766.$$

The corresponding variance estimate for the mean effect must take into account the nonindependence (i.e., covariances) between the response traits. In general, if variables are correlated, the variance of their sum is the sum of their covariances. Thus the variance of the arithmetic (unweighted) mean of $X_1, \ldots, X_m$, say, is obtained by dividing this value by m^2, so that (departing from the above notation for a little while):

$$V_{mean} = (\sum_{i=1}^{m} \sum_{j=1}^{m} \mathrm{Cov}(X_i X_j))/m^2. \tag{16.6}$$

Note that the term in the outside brackets includes all combinations of i and j, (i.e., both X_1X_2 and X_2X_1) and the variances, since $\mathrm{Cov}(X_1X_1) = \mathrm{Var}(X_1)$. Let V_i represent the variance of X_i, V_{ij} represent the covariance between X_i and X_j, and r_{ij} represent the correlation between X_i and X_j. Then the above equation can be written as follows:

$$V_{mean} = \left(\sum_{i=1}^{m} V_i + 2\sum_{i,j} V_{ij}\right) \Big/ m^2 = \left(\sum_{i=1}^{m} V_i + 2\sum_{i,j} (r_{ij}\sqrt{V_i}\sqrt{V_j})\right) \Big/ m^2 \tag{16.7}$$

where r_{ij} is the correlation between outcomes i and j, the summation over i, j is for distinct pairs of i and j with $i \neq j$ (i.e., only one combination of a pairing is required, hence the multiplication by 2 of the second term).

As an illustration, for study B the variance of the composite mean is given by

$$V_{mean} = [0.01 + 0.095 + 0.09 + (2 \times r_{size\ survival} \times \sqrt{0.1} \times \sqrt{0.09}) + (2 \times r_{size\ immune} \times \sqrt{0.1} \times \sqrt{0.095}) + (2 \times r_{survival\ immune} \times \sqrt{0.09} \times \sqrt{0.095})]/3^2.$$

Note that the above variance equations are appropriate for an unweighted mean (or, equivalently, when the weighting is identical for all effects so that it can be effectively ignored when calculating the mean). If all the study variances are equal (to V, say) and all the correlations are also equal (to r, say), these equations reduce to

$$V_{mean} = (V/m)[1 + (m - 1)r] \tag{16.8}$$

Hence, when the m effect sizes are uncorrelated ($r = 0$), then the variance of the composite effect is V/m so the variance halves each time the number of estimates is doubled. When $r = 1$, however, there is no extra information provided so that variance of the composite effect is the same as that of the individual effects. The multiplier term $(1 + [m - 1]r)$ is known as a variance inflation factor (Borenstein et al. 2009).

Once the composite mean and variance for each study are calculated, they can be combined using a study level meta-analysis, based on either a fixed- or random-effects model. Note that both the estimated mean effect and its variance will be affected by correlations between traits, and although the mean effect may not change much if the variances are very similar within studies, its associated variance may still change considerably. Strictly speaking, we cannot use Equation 16.7 with Data set 3 because the variances are not identical within each study. For

TABLE 16.4. The effect of different levels of correlation between traits when calculating the composite mean for Data set 3 (Table A16.3). Note that the composite mean for each study was calculated giving each response trait equal weighting (i.e., ignoring slight differences in variance within studies). Similarly, so that the original response traits all had the same variance, we used the arithmetic mean variance in Equation 16.7. For simplicity, we assume the same correlation between all three original traits. For ease of interpretation we present the 95% CI for the overall mean. The variance can be calculated as: Upper value − mean = 2.57 × sqrt(Variance) [$t_{5,\ 0.05(2\text{-tailed})}$ 2.57].

Study	No traits	Composite Mean	Variance of original traits	Variance of Composite mean when r =			
				0	0.3	0.6	1
A	3	0.3	0.019	0.0063	0.0101	0.0139	0.019
B	3	0.4	0.095	0.0317	0.0507	0.0697	0.095
C	2	0.2	0.05	0.025	0.0325	0.04	0.05
D	3	0.5167	0.03	0.01	0.016	0.022	0.03
E	3	0.7667	0.01	0.0033	0.0053	0.0073	0.01
F	2	0.325	0.065	0.0325	0.0423	0.052	0.065
			FE Mean	0.545	0.538	0.534	0.530
			95% CI	0.445–0.645	0.413–0.663	0.388–0.679	0.361–0.699
			RE Mean	0.436	0.442	0.448	0.456
			95% CI	0.154–0.719	0.158–0.726	0.163–0.733	0.172–0.741

illustrative purposes, however, given that they differ only slightly, we will use the mean variance to illustrate the effect of changes in the correlation between traits on our overall estimate of the mean composite effect and its variance. The results are shown in Table 16.4.

If there had been several independent estimates of effect sizes for any of the response traits in a study, then we would need to take a hierarchical approach. First, for each study we would separately calculate the weighted mean and its variance using the standard fixed- or random-effects model for each trait. Once we had one effect size per trait per study, we could then calculate the weighted mean composite effect size and its associated variance using Equation 16.7.

Composite based on differences in effect sizes

If the researcher is interested in comparing the difference between two nonindependent effect sizes, such as offspring size and survival function, then Equation 16.1 can be used to obtain the variance of the estimated difference for each study. The estimated differences can then be combined using a fixed- or random-effects model. For example, if $V_{immune} = 0.02$, $V_{size} = 0.018$, and $r_{immune\text{-}size} = 0.2$, then $V_{immune\text{-}size} = 0.0304$. As the correlation becomes larger, the shared variation increases and $V_{immune\text{-}size}$ decreases. For example, if $r = 0.9$ in the above calculation, $V_{immune\text{-}size}$ declines to 0.004. Combining the estimated differences between offspring size and survival for the six studies under a random-effects model when $r = 0.2$, the overall method of moments estimate given by MetaWin is –0.097 (95% CI; –0.301 to 0.117)—which in this case is equivalent to a fixed-effects model since the between-study variance collapses to zero. When $r = 0.9$, the overall estimate is –0.072 (95% CI: –0.214 to 0.069). There is therefore insufficient evidence that the effect size for polyandry differs for offspring size and survival (although there is a trend for it to be smaller for size).

It is apparent from closer inspection of Equations 16.1 and 16.7 that a larger positive correlation induces a *larger* variance for the sum (or mean) of two effect sizes (see Table 16.4), but a *smaller* variance for the difference between two effect sizes, compared with independence

($r = 0$). This makes sense biologically. Consider a case where we want to measure offspring performance, using a number of proxy measures. Imagine that the breeding season is long and that environmental conditions vary widely. When conditions are good then offspring growth, survival, and escape speed are high; when conditions are poor they are low. Consequently, the more measures we take, the more we will over- or underestimate the offspring performance relative to their average performance during the breeding season. A combined effect therefore increases variance among individuals. In contrast, when we calculate the difference between two measures (say, growth and survival), we are removing differences that arise due to environmental noise (i.e., $r > 0$). Hence, the mean difference is estimated with greater precision and the variance is lower when we account for environmental noise (Borenstein et al. 2009).

In the same way that we can compare the effect sizes of two original traits, we can also compare two composite effect sizes based on different outcomes, or even one composite effect with that for a specific trait. Of course, if multiple comparisons are made using subsets of the same data, we have to adjust for multiple testing. One issue that obviously arises when calculating composite effects is how to obtain estimates of the correlation between effect sizes (*rij*). We will return to this problem at greater length after discussing Situation 4. There is, however, one situation that is a special case, but that is fairly commonplace, where it is easy to calculate correlations. It arises when we want to combine or compare the effect of several treatments, when the treatment subjects are independent, but the control always involves the same set of subjects. For example, many studies investigating the benefit of polyandry, mate females with either a single male four times (= control, monandry), two males twice each (treatment 2, polyandry), and four males once each (treatment 4, polyandry) (Tregenza and Wedell 1998). We can then ask two questions. First, is there a positive effect of polyandry (regardless of whether it involves two or four males) on offspring performance? Second, is the effect greater for four males than two males?

To answer the first question, before calculating the mean effect across studies we have to calculate the combined effect size for control versus treatment 2, and control versus treatment 4 for each study using Equation 16.7. In this case, we know the correlations between the effect sizes: treatment 2 and treatment 4 females have no relationship ($r = 0$) while control females do ($r = 1$). If the treatment and the control have equal sample sizes, then the correlation between the effect size for treatment 2 and that for treatment 4 is $r = 0.50$ (the average of the two). If the sample sizes differ, this needs to be weighted accordingly. For example, if the same sizes are $n_{control} = 20$, $n_{treatment\ 2} = 10$, and $n_{treatment\ 4} = 5$, then $r = 1 \times (20/35) + 0 \times (15/35) = 0.57$. To answer the second question, one option is to calculate the difference between the treatment 2 versus control effect, and the treatment 4 versus control effect (using Equation 16.1). It is, however, better to take advantage of the fact that the control effectively "cancels out" when we do so. In short, if the original study provides the data (e.g., means and a measure of variance for each treatment), we simply calculate the effect size for the difference between the original outcomes in the two treatments (Borenstein et al. 2009, 241).

SITUATION 4: MULTIPLE EFFECTS

Suppose that I studies each report estimates of P effects, and our interest is in the analysis of each of these effects, allowing for the possibility that they could be correlated. Note that this is different from Situation 3, in which we had the same type of data, but the aim was to create a composite effect. Here, we wish to estimate each effect.

For example, experimental studies of the effects of elevated CO_2 report a number of effects on plants such as CO_2 assimilation rate, growth rate, and stomatal conductance. We might be

interested in finding the overall estimates of these effects. To do this, we could conduct separate meta-analyses for each response variable, but this would ignore any possible correlation between these effects so that any broader inferences or estimates might be misleading. Consequently, the use of a single meta-analysis is preferable because it can take into account correlations between different outcome measures. We pursue this example below.

We can denote the effect estimates for the ith study by the vector $\mathbf{T}_i = (T_{i1}, \ldots, T_{iP})$, and the true effects for this study by the vector $\theta_i = (\theta_{i1}, \ldots, \theta_{iP})$. The overall effects (combined across all studies) will be denoted by $\mu = (\mu_1, \ldots, \mu_P)$. As noted in the Introduction, the models for the various situations described in this chapter are all variations of the same theme, introduced in Equation 16.2. It is no surprise then, that the following model is analogous to that described at the end of the discussion of Situation 1, where a single outcome is replaced by multiple outcomes, and the estimate of variance at a given level of a hierarchy is replaced with a variance-covariance matrix that contains the variances of the effects along the diagonal, and covariances between the effects in the off-diagonal elements. The differences in the models are simply due to the nature of the correlation structure in the data and between effects, and to the questions we wish to answer.

For the ith study, then, let C_i be the (observed) variance-covariance matrix for the effect estimates $\mathbf{T}_i$ and let Σ_i be the (unknown) variance-covariance matrix for the true effects θ_i. Thus the observed and "true" variance-covariance matrices, C_i and Σ_i, are given by

$$C_i = \begin{bmatrix} S_{i1}^2 & S_{i12} & \cdots & S_{i1P} \\ \vdots & S_{i2}^2 & & \vdots \\ & & \cdots & \\ S_{iP1} & \cdots & & S_{iP}^2 \end{bmatrix} \quad \Sigma_i = \begin{bmatrix} \sigma_{i1}^2 & \sigma_{i12} & \cdots & \sigma_{i1P} \\ \vdots & \sigma_{i2}^2 & & \vdots \\ & & \cdots & \\ \sigma_{iP1} & \cdots & & \sigma_{iP}^2 \end{bmatrix}.$$

Here, for example, for the first study, S_{11}^2 denotes the estimated variance of the effect estimate T_{11} (the first effect in the first study), $S_{11}^2 S_{123}$ denotes the estimated covariance between effect estimates T_2 and T_3 (in this study, where σ_{11}^2 is the variance of the first (true) effect θ_1), and σ_{123} is the covariance between the true effects θ_2 and θ_3 in the first study.

As before (Situation 2), the variance-covariance matrix Γ describes the relationship between the estimated overall parameters (here, $\mu_1, \ldots, \mu_P$):

$$\Gamma = \begin{bmatrix} \gamma_1^2 & \gamma_{12} & \cdots & \gamma_{1P} \\ \vdots & \gamma_2^2 & & \vdots \\ & & \cdots & \\ \gamma_{P1} & \cdots & & \gamma_P^2 \end{bmatrix}.$$

Commonly (but not necessarily) the study-specific effect estimates are assumed to be multivariate normal. This is often justified based on the assumed distribution of the corresponding population effect (or an appropriate transformation), or on the basis of a large sample size and appeal to the central limit theorem, so that the sample means (standardized differences, mean log response ratios, etc.) are approximately normally distributed. Under this assumption, the model can be written as

$$\mathbf{T}_i \sim \text{MVN}(\theta_i, \Sigma_i)$$
$$\theta_i \sim \text{MVN}(\mu, \Gamma).$$

Note that we do not attempt to combine these to obtain a single effect, μ_0, as we did in Equation 16.2, since here the effects are measuring phenomena that we wish to examine separately (e.g., CO_2 assimilation and stomatal conductance).

If the meta-analysis contains two effects, for example, the model expands as follows:

$$\begin{pmatrix} T_{i1} \\ T_{i2} \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix}, \Sigma_i\right), \qquad \Sigma_i = \begin{pmatrix} \sigma_{i1}^2 & \sigma_{i12} \\ \sigma_{i12} & \sigma_{i2}^2 \end{pmatrix}$$
$$\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Gamma\right), \qquad \Gamma = \begin{pmatrix} \gamma_1^2 & \gamma_{12} \\ \gamma_{12} & \gamma_2^2 \end{pmatrix}$$

and Σ_i is estimated by C_i.

Note also that it is sometimes easier to describe the model in terms of correlations. For example, if r_{i12} is the observed correlation between the first and second effect estimates, then $S_{i12} = r_{i12}S_{i1}S_{i2}$ and similarly, if ρ_{12} is the correlation between the overall effects μ_1 and μ_2, then $\gamma_{12} = \rho_{12}\gamma_1\gamma_2$.

By merging the levels of the model, the observed effects are related to the overall effects as follows:

$$\begin{pmatrix} T_{i1} \\ T_{i2} \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, V_i\right), \qquad V_i = \begin{pmatrix} \sigma_{i1}^2 + \gamma_1^2 & \sigma_{i12} + \gamma_{12} \\ \sigma_{i12} + \gamma_{12} & \sigma_{i2}^2 + \gamma_2^2 \end{pmatrix}.$$

Thus the observed effects are centered on the overall effects (μ_1, μ_2) and, as for a simple random-effects model (Chapters 7 and 9), the combined variance is the sum of the within-study variances (given by the S^2's as our best estimates of σ^2) and the between-study variances (given by the γ's).

The addition of covariates to the above model has been described by Van Houwelingen et al. (2002), and has been developed further by Arends et al. (2003), Riley, Abrams, et al. (2007), Riley, Simmonds, et al. (2007), and Riley et al. (2008).

Sometimes it is of interest to conduct univariate meta-analyses even if multiple effects are reported in each study. For example, if one of the effects measured in the primary studies is of key interest, the other effects can either be ignored, in which case the model becomes a simple univariate random- or fixed-effects meta-analysis; or, if the primary effect is highly correlated with the other effects, one might use the latter as covariates predicting the primary effect, in which case the model reduces to a univariate regression meta-analysis. To illustrate the latter case, if $T_1, \ldots, T_P$ effects are reported in each study and T_P is the principal effect of interest, the meta-analysis model becomes

$$T_{iP} = \beta_{0i} + \beta_{1i}T_{i1} + \ldots + \beta_{P-1,i}T_{i,p-1} + \varepsilon_i$$

where ε_i is the residual. Other covariates can be added as appropriate.

The meta-regression approach has the advantage of accounting for all of the effects while retaining the simplicity of a univariate meta-analysis. Moreover, the relationship between T_{iP} and another effect T_{il}, say, given the presence of all the effects, is indicated by the regression coefficient β_{1i}. Although the effects can be included as independent variables in the above model, dependency between them can also be included via a study-level variance-covariance matrix for both β and ε. The disadvantage of this approach is that relationships between other effects are not evaluated, so the inferential capacity of the meta-analysis under this formulation is constrained.

An alternative to the above approaches, where effects are considered separately, is to calculate a composite measure of the effects for each study and then combine these using standard univariate meta-analysis models (Situation 3). The most simplistic way of dealing with multiple effects is to assume that the effects are independent. The meta-analysis then reduces to a series of separate univariate meta-analyses for each effect (Rosenthal and Rubin, 1978), using the models and methods described in Chapters 8 to 11. Rosenthal and Rubin (1982) and Hedges and Olkin (1985) argue that this approach is useful if outcomes are independent or weakly

correlated. If, however, outcomes are highly related this approach can lead to biased parameter estimates, inaccurate variance estimates, and misleading inferences (Walsh 1947; Holmes and Matthews 1984; Hedges and Olkin 1985; Sohn, 2000; Ades 2003). The type and extent of this bias will depend on the type and degree of dependency between outcomes. In general, most evolutionary and ecological meta-analyses to date have taken a univariate approach when dealing with multiple effect sizes per study.

We illustrate the meta-analysis of multiple effects using Data set 4 (Table A16.4). This is a real data set showing the response of plants to increased CO_2 concentration. Peter Curtis compiled and meta-analyzed a comprehensive database of the effect of increased CO_2 concentration on plants (Curtis and Wang 1998).

In Data set 4 we only consider two response variables: net CO_2 assimilation (PN) and stomatal conductance (GS) under "No Stress" condition. The reported effect size estimates are the standardized mean difference (Hedges' *d*) and corresponding variance based on the mean response of plants grown under ambient and elevated CO_2 conditions. The data consist of 34 "studies" published in 17 papers. The scatterplot of the pairs of effect estimates (GS and PN) across studies suggests there is a positive correlation between the pairs of effects (Fig. 16.3). This motivates a meta-analysis that takes into account potential nonindependence.

The multivariate model described above is extended to include an additional hierarchy that first combines study estimates derived from the same paper (i.e., "Paper" is a grouping factor; Situation 1). Let T_{ijk} denote the estimated effect for the *k*th outcome (GS, PN) from the *j*th study in the *i*th paper. The meta-analysis model proceeds as follows: (i) estimate the true effect θ_{ijk} for each T_{ijk}—since we don't have any information about the correlations between estimates we use a univariate distribution; (ii) based on the vectors $(\theta_{ij1}, \theta_{ij2})$, estimate a combined paper effect (μ_{i1}, μ_{i2}), and here we use a bivariate distribution to allow for correlation between the effects; and (iii) based on the vectors (μ_{i1}, μ_{i2}), estimate an overall effect (μ_{01}, μ_{02}) for GS and

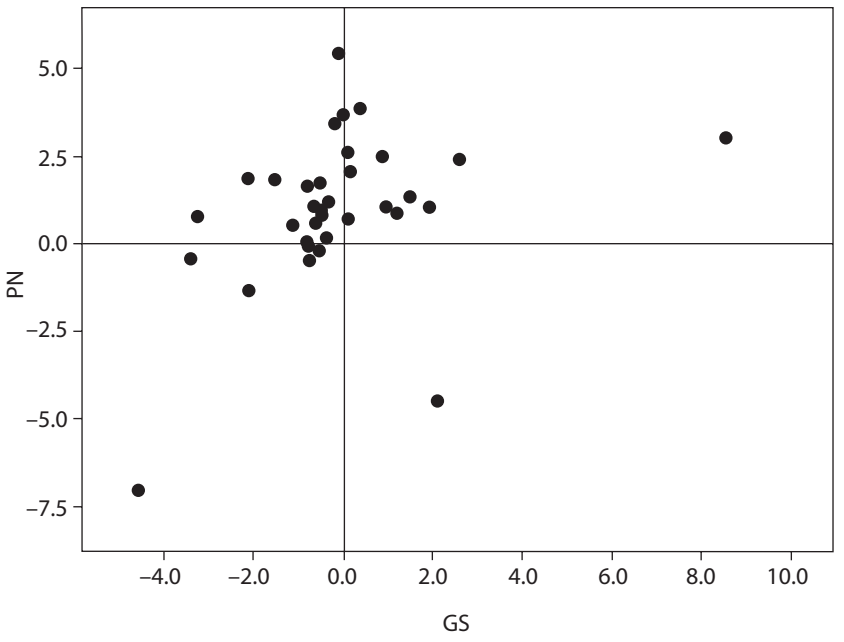

Figure 16.3. The relationship between effect sizes for stomatal conductance (GS) and net CO_2 assimilation (PN) from 34 studies in Data set 4 (Table A16.4).

PN—again we use a bivariate distribution to allow for correlated effects. Formally, the model is as follows:

$$\begin{aligned} T_{ijk} &\sim \mathrm{N}(\theta_{ijk}, \sigma^2_{ijk}); \quad \theta_{ij} = (\theta_{ij1}, \theta_{ij2})' \\ \theta_{ij} &\sim \mathrm{MVN}(\mu_i, \Gamma_i); \quad \mu_i = (\mu_{i1}, \mu_{i2})' \\ \mu_i &\sim \mathrm{MVN}(\mu_o, \Omega); \quad \mu_0 = (\mu_{01}, \mu_{02})' \end{aligned} \tag{16.9}$$

where σ_{ijk}^2 is estimated by the reported variance S_{ijk}^2, Γ_i is the variance-covariance matrix for the effects within papers, and Ω is the variance-covariance matrix of the overall effect.

For publications with a small number of reported estimates ("studies"), the terms of the matrix Γ_i may be poorly estimated, particularly the covariance. An alternative model is to maintain a univariate representation for θ_{i1} and θ_{i2} for such a publication, then combine these at the global level, at the last line of Equation 16.8. The method of Riley et al. (2008) for cases in which covariances are unreported and are not easily estimated is also applicable here.

We adopted a Bayesian approach to this meta-analysis, with a normal prior for μ_0 and Wishart priors for Γ_i and Ω. An equivalent univariate analysis, in which each response is analyzed separately, can be fit by changing this model slightly. The models were fit in WinBUGS (the code is in Appendix 16.2 as are the changes to the DATA list needed for the univariate analysis).

Table 16.5 shows the posterior estimates of the overall mean for GS and PN, and the three terms in the overall variance-covariance matrix for the multivariate and univariate models. It can be seen that there is some difference in posterior global variance–covariance terms, and the univariate model undergoes much greater shrinkage of the effects θ_{ijk}, θ_{ij}, which in turn drives the slight differences between multivariate and univariate estimates of the overall means for each response.

DEALING WITH MISSING CORRELATIONS

As noted earlier, a problem that often arises in meta-analysis of dependent outcomes is that the correlations between study effect estimates within groups, and/or between group effect estimates are not reported. The question that then arises is whether to ignore or estimate these correlations. The extreme positions are to assume that the outcomes are independent or that they are completely related (i.e., measuring both outcomes is an exercise in redundancy, aside from the fact that doing so reduces sampling error because the mean of two estimates is closer to the true mean). The same assumptions can be made at the group level. If independence is

TABLE 16.5. The estimates of the mean effect size (and its standard deviation) for net CO_2 assimilation (PN) and stomatal conductance (GC) from Bayesian meta-analyses of Data set 4 (Table A16.4), based on two different modeling assumptions. The WinBUGS code for these models is provided in Appendix 16.2.

Global	Multivariate		Univariate	
Parameter	Mean	SD	Mean	SD
Mean 1 (GS)	–0.3509	0.6104	–0.2336	0.5078
Mean 2 (PN)	1.236	0.6005	1.069	0.4164
Precision 1	0.3703	0.1917	0.2593	0.2037
Precision 2	0.5372	0.254	0.8079	2.931
Precision12	–0.179	0.1547		

assumed, covariance terms in matrices are equal to zero and the full model reduces to a simpler model—that is, of the basic form shown in Equation 16.2. Assuming independence will overestimate, and assuming complete dependence will underestimate, the precision with which the mean effect is known. It is strongly recommended that the researcher undertake a sensitivity analysis to assess the effect of these extreme assumptions by generating study level means and variances using both $r = 0$ and $r = 1$ (Table 16.4).

Alternatively, the researcher can estimate the missing correlations. Consider the problem of combining multiple outcomes within each study (Situation 3). If the outcomes are different measures of the same effect, then they will almost certainly be positively correlated. As a starting point, the researcher should ensure that there is a common predicted direction of the effects. For example, if improved offspring performance is associated with being larger but having a shorter development time, then the sign of the treatment versus control difference for size should be opposite that for development time (hence signs should be reversed for one of the measures). In the case of repeated measures over time, however, it is possible that there is a negative correlation. For example, a large effect of a treatment at time x might lead to a trade-off with a crash at time $x + 1$, so that the difference in outcome between the treatment and control reverses between the two time periods. As an illustration, sexual selection models show that it is possible for males in good condition to "overinvest" in a courtship signal in the present, because the short-term gains more than compensate for any detrimental effect on signaling in the future (Kokko 1997, Hunt et al. 2004).

The researcher might be able to make an assumption about the magnitude of missing correlations by a comparison with available estimates from other studies of the same or closely related outcomes (Berkey, Anderson, et al. 1996). A number of imputation schemes are available (Little and Rubin 2000). If a correlation matrix is estimated or imputed, it is important to check that it is positive definite. Missing within-study correlations can sometimes be estimated from the observed data; Gleser and Olkin (1994) provide relevant formulas for multiple treatment and multiple endpoint studies respectively. For example, if in a polyandry study (Situation 3) the correlation between offspring size and offspring immune function is given separately for the monandry and polyandry treatment, the mean of these correlations may be used as an estimate of the correlation between the effect size for polyandry on size and immunity. A correlation based on the whole data set from a study would be trickier to interpret because it includes any treatment effect of polyandry that changes the values of both the traits of interest. Although it might seem risky to assume correlation from a single study will be similar to those in other studies, Berkey, Anderson, et al. (1996) has noted that correlations from one study are often closer to unreported correlations for the remaining studies than is zero; this last is the assumed correlation when treating outcomes as independent.

When imputing, it is most important to bear in mind that the imputed value is only an estimate of the unknown value. The uncertainty introduced by the imputation must be accounted for in the model to get an unbiased estimate of the truth with a proper estimate of the standard error. Pretending as if the imputed value were the real missing value will lead to underestimation of the variability of the parameter estimate.

A final issue that has to be considered is that when true correlations differ among studies, this should affect the relative weight assigned to each study because of its effect on the variance estimate for that study (e.g., Equation 16.7). This could lead to very different estimates of the overall mean effect (Borenstein et al. 2009, 233). When correlations are unknown and a hypothetical value is assigned, it might be prudent to also conduct a sensitivity analysis to calculate the mean effect after judiciously assigning different extreme values ($r = 0$ and 1) to different studies in order to estimate the mean effect.

DEALING WITH MISSING STUDIES

A problem arises in the analysis of multiple effects if studies report on different subsets of the effects. If there are only a few missing studies it might be possible to estimate them using techniques such as randomization or multiple imputation. Alternatively, if there is sufficient overlap in the subsets, enabling reasonable estimates of the covariances between effects to be obtained, then the multivariate model described above can be altered to allow for different effects in different studies (for an example, see Nam et al. 2003). Difficulties arise, however, if there is insufficient overlap of effects reported by the studies. These include increased variation in effect estimates that are positively correlated due to the inclusion of uncertainty about covariance estimates. This can lead to substantially biased estimates (Riley, Abrams, et al. 2007). Analysts must also beware of reporting bias when encountering studies that report different subsets of effects. Reporting bias arises when authors selectively report some effects and not others, perhaps because of statistical significance or other reasons related to the message they are trying to communicate. Analysis of available effects, even if appropriately done, will lead to bias if the effects are not missing at random (Little and Rubin 2002)

DISCUSSION

We have focused on several possible approaches to the meta-analysis of dependent data, where dependence arises for a number of different reasons. We have shown that the meta-analysis model that is adopted partly depends on the question being answered, the type of data available (i.e., primary or aggregate), the format of these data (e.g., means or regression coefficients), the hierarchical or multilevel nature of the data, the correlation structure in the data and between the effects, and the extent to which we are willing to estimate or ignore covariance. Our intention has been to consider a restricted number of situations in which dependence is commonly exhibited in ecology and evolution meta-analyses, and to describe the meta-analysis models in terms of a general multivariate hierarchical structure that describes the dependencies relevant to that situation. We hope that the reader will now see that the models are simple variations on the same structure, and realize that this structure can be generalized to other situations of interest. We have not gone into details about other, more complex, approaches to the meta-analysis of dependent data, but we provide a brief resume in Appendix 16.3 for the interested reader.

Many of the models we have described can be applied using either primary or aggregated data from the relevant studies. Access to primary data has the strong advantage of ensuring that the same model is fit for each study. It has the disadvantages of necessarily excluding studies for which appropriate data cannot be obtained, and requires that the meta-analyst, who has typically not been involved in the primary studies, has sufficient background knowledge to adequately undertake the study-specific analyses.

One question that often arises is how to decide between a univariate and a multivariate approach. One consideration is on the basis of the available data. If the study-specific effect estimates, variances and within-study variances, *and* covariances are known (the latter usually obtained from the correlation between traits) or estimated (at least partially), then a multivariate analysis can be a profitable approach. However, if the within-study covariances are unknown or cannot be estimated, or if the covariance between outcomes across studies can only be poorly estimated, then a univariate approach is more defensible. Of course, there is no reason not to present analyses based on both approaches as a form of sensitivity analysis to differing model assumptions.

We end by noting that in some published meta-analyses, the authors use several types of meta-analysis approaches to the question being answered (e.g., Rustad et al. 2001, McCarthy

et al. 2006). There is rarely a single, perfect solution. Issues of data availability and format influence the choice of approach. Ultimately, however, there is little use running a meta-analysis if it does not address the question it was designed to answer. Ensuring this is the case should always be the priority. Put slightly differently, a sophisticated model that is poorly estimated is often no better than a simple approach in which missing information is acknowledged (e.g., lack of information on correlations between effect sizes at various levels) and then simplifying assumptions are made (e.g., test when $r = 0$ and $r = 1$, or include and ignore grouping factors that might be a source of nonindependence). If these weaknesses are clearly stated the reader can then, at least, interpret the results in a straightforward manner. Equally, it is important to acknowledge the possible biases that can be introduced when using simpler but less appropriate models, so care should be taken in any inferences that are made based on these analyses.

APPENDIX 16.1: DATA SETS

Note: All of the material in section 16.1 and 16.2 of this appendix can also be found on the web site for this book, http://press.princeton.edu/titles/10045.html

TABLE A16.1. Data set 1 ("grouped data set"): A hypothetical data set used to illustrate the analysis of Situation 1. There are 25 studies that can be assigned to five groups (a group could be a study site, research time, year, the technique used to collect the data, and so on). The mean effect size and estimate of its variance from each study are shown.

Group	Effect	Variance
A	0.2	0.10
A	0.6	0.15
A	0.5	0.05
A	0.1	0.06
B	0.8	0.08
B	0.4	0.05
B	0.9	0.04
C	0.2	0.09
C	0.7	0.11
C	0.5	0.03
C	0.2	0.05
C	1.1	0.06
C	1.4	0.07
D	0.1	0.01
D	0.8	0.02
D	0.9	0.02
D	1.4	0.04
E	0.2	0.03
E	0.3	0.05
E	0.8	0.06
E	0.9	0.02
E	1.4	0.07
E	1.2	0.08
E	1.3	0.05
E	0.9	0.04

TABLE A16.2. Data set 2 ("shrimp data set"): Shrimp biomass at nine sites over successive years (ranging from 13 to 34 per site). These data are used to illustrate the analysis of Situation 2.

	Study								
Year	2J	3K	3M	NGSL	ESS	5Y2	ICE	BS	SKAG
1967							1.50		
1968						0.99	1.64		
1969						0.92	1.66		
1970						1.05	1.63		
1971						0.96	1.48		
1972						1.01	1.37		
1973						0.76	1.43		
1974						0.36	1.44		
1975						0.46	1.45		
1976						0.25	1.56		
1977	2.47				1.09	–0.34	1.68		
1978	2.43				1.06	0.10	1.68		
1979	2.30				1.22	0.19	1.79		
1980	2.30				1.10	0.21	1.73		
1981	2.31				1.10	0.66	1.84		
1982	2.27			1.86	1.09	–0.04	1.86	2.51	
1983	2.16			1.83	1.08	0.38	1.80	2.63	
1984	2.22			1.80	1.02	0.76	1.65	2.67	
1985	2.11			1.81	0.86	0.69	1.69	2.39	2.04
1986	2.40			1.83	0.92	0.88	1.62	2.22	2.09
1987	2.37			1.87	0.94	0.70	1.73	2.16	2.02
1988	2.33	1.56	0.34	1.92	0.91	0.56	1.90	2.26	2.04
1989	2.35	1.44	0.28	1.98	1.10	0.74	1.92	2.33	1.98
1990	2.39	1.48	0.33	2.02	1.11	0.80	2.04	2.42	1.98
1991	2.32	1.58	0.91	2.01	1.28	0.46	1.96	2.51	1.97
1992	2.29	1.73	1.22	1.87	1.22	0.27	1.96	2.38	2.01
1993	2.36	1.79	0.97	1.95	1.26	0.75	1.92	2.37	2.00
1994	2.41	1.86	0.52	2.02	1.28	0.84	1.83	2.21	2.03
1995	2.48	1.98	0.73	2.11	1.35	0.75	1.84	2.29	2.08
1996	2.57	2.01	0.81	2.18	1.37	0.54	2.00	2.44	2.06
1997	2.68	2.11	0.71	2.23	1.37	0.62	1.86	2.48	2.13
1998	2.66	1.99	1.23	2.30	1.51	0.84	1.68	2.53	2.20
1999	2.69	2.00	1.09	2.26	1.59		1.56	2.50	2.11
2000	2.77	2.13	0.99	2.30	1.61		1.77		
No. Yrs	24	13	13	19	24	31	34	18	15
Trend T	0.160	0.058	0.054	0.028	0.022	0.000	0.012	–0.002	0.008
SE(T)	0.004	0.006	0.020	0.003	0.004	0.007	0.002	0.007	0.003

TABLE A16.3. Data set 3 ("polyandry data set"): A hypothetical data set used to illustrate the analysis of Situation 3. There are 16 effect sizes from six separate studies. Each study provides an estimate of the effect of polyandry on two or three offspring traits (offspring size, survival rate, and immune function).

Study	Response	Effect Estimate	Variance
A	size	0.3	0.02
A	survival	0.4	0.018
A	immune	0.2	0.019
B	size	0.5	0.1
B	survival	0.3	0.09
B	immune	0.4	0.095
C	size	0.2	0.05
C	survival	0.2	0.05
D	size	0.3	0.03
D	survival	0.6	0.03
D	immune	0.65	0.03
E	size	0.7	0.01
E	survival	0.8	0.01
E	immune	0.8	0.01
F	size	0.35	0.07
F	survival	0.3	0.06

TABLE A16.4. Data set 4 ("CO_2 data set"): The effect of elevated CO_2 levels on two plant traits (GS and PN) that is used to illustrate the analysis of Situation 4 (Curtis and Wang 1998). There are 68 effect sizes from 17 studies. Each study has at least one pair of effect sizes for PN and GS (range: 1–4).

	GS		PN	
Paper No[1]	Hedges' *d*	Var (*d*)	Hedges' *d*	Var (*d*)
121	–0.4862	0.3432	0.9817	0.3735
121	0.1535	0.3343	2.0668	0.5113
121	0.0965	0.3337	2.6101	0.6172
121	0	0.2857	3.6586	0.7638
168	–1.5271	0.4305	1.8355	0.4737
222	–3.2447	1.544	0.7747	0.7167
222	–3.4039	1.6322	–0.4267	0.6818
233	–4.57	1.4442	–7.0561	2.8895
290	–2.1284	0.3916	1.8644	0.3586
290	–0.6377	0.2627	1.0313	0.2832
290	0.1082	0.2504	0.7069	0.2656
340	0.9597	0.3717	1.0391	0.3783
340	–0.7788	0.3586	-0.0805	0.3336
506	1.9111	0.5826	1.0516	0.4553
506	2.6041	0.7391	2.39	0.6856
550	8.5458	2.0258	3.0211	0.8563
582	–0.3432	0.5074	1.1915	0.5887
582	–0.6008	0.5226	0.5762	0.5208
582	–0.5389	0.5182	–0.1952	0.5024
582	–2.1085	0.7779	–1.3417	0.6125
582	–0.7445	0.5346	–0.476	0.5142
745	–0.7991	0.18	1.6405	0.2227
745	0.8741	0.1826	2.4785	0.2946
2035	0.3736	0.2261	3.8708	0.7222
2035	–0.5231	0.188	1.7512	0.2515
2045	–0.3751	0.0848	0.1703	0.0836
2045	–0.8026	0.09	0.0557	0.0834
2047	2.1102	0.5189	–4.4809	2.3398
2065	–0.1068	0.3338	5.4279	1.5609
2121	1.4902	0.1661	1.3346	0.8151
2121	1.201	0.1432	0.878	0.7309
2122	–0.1738	0.4015	3.4233	0.986
2129	–0.4823	0.5145	0.8128	0.5413
2131	–1.1225	0.4797	0.532	0.4308

[1] Refers to paper identification numbers in Curtis and Wang (1998)

APPENDIX 16.2: CODE FOR ANALYSES

1. *WinBUGS* code for Bayesian analysis of Data set 1

```
model{
    for (i in 1:No.Ests){                          # No.Ests is number of estimates
        prec.Y[i] <- 1/Var[i]                      # prec.Y is precision of Y = 1/var(Y)
        Y[i] ~ dnorm(th[i], prec.Y[i])             # th[i] is true effect theta[i]
        th[i] ~ dnorm(mu[T[i]], prec.th[T[i]])     # mu[T[i]] is true effect for group T[i]
        }                                          # T[i] is the group for the ith estimate
    for (j in 1:No.Groups){                        # No.Groups is the number of groups
        prec.th[j] <- 1/(sig.th[j]*sig.th[j])      # sig.th[i] is the SE of the theta[i]
        sig.th[j] ~ dunif(0,100)                   # noninformative uniform prior
        mu[j] ~ dnorm(mu.0, prec.mu)               # mu.0 is the overall effect
        }
    mu.0 ~ dnorm(0.0, 1.0E-2)                      # noninformative normal prior
    prec.mu <- 1/(sig.mu*sig.mu)                   # sig.mu is the SE of the mu[i]
    sig.mu ~ dunif(0,100)                          # noninformative uniform prior
    }

Data
list( No.Ests=25, No.Groups=5,
Y=c(0.2, 0.6, 0.5, 0.1, 0.8, 0.4, 0.9, 0.2, 0.7, 0.5, 0.2, 1.1, 1.4, 0.1, 0.8, 0.9, 1.4, 0.2, 0.3,
0.8, 0.9, 1.4, 1.2, 1.3, 0.9),
Var=c(0.10, 0.15, 0.05, 0.06, 0.08, 0.05, 0.04, 0.09, 0.11, 0.03, 0.05, 0.06, 0.07, 0.01, 0.02,
0.02, 0.04, 0.03, 0.05, 0.06, 0.02, 0.07, 0.08, 0.05, 0.04),
T=c(1, 1, 1, 1, 2, 2, 2, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5)
)

Initial values
list( sig.mu=1, sig.th=c(1,1,1,1,1))
```

2. *WinBUGS* code for Bayesian analysis of Data set 2

```
# full hierarchical model

model {
    for( i in 1 : N ) {
        beta[i , 1:2] ~ dmnorm(mu.beta[], R[i, , ])
        R[i,1:2 , 1:2] ~ dwish(Omega[ , ], 2)
        for( j in 1 : T[i] ) {
            Y[i , j] ~ dnorm(mu[i , j], tauC)
            mu[i , j] <- beta[i , 1] + beta[i , 2] * x[i,j]}}
    mu.beta[1:2] ~ dmnorm(mean[],prec[ , ])
    tauC <- 1/pow(sigma,2)
    sigma ~ dunif(0.0,10.0)
}
```

```
# DATA
list( N = 9,
T=c(24,13,13,19,24,31,34,18,15),
      Omega = structure(.Data = c(1.0, 0.0, 0.0, 1.0), .Dim = c(2, 2)),
      mean = c(0,0),
      prec = structure(.Data = c(1.0E-6, 0, 0, 1.0E-6), .Dim = c(2, 2)),
Y=structure(.Data=c(
2.47,2.43,2.3,2.3,2.31,2.27,2.16,2.22,2.11,2.4,2.37,2.33,2.35,2.39,2.32,2.29,2.36,2.41,
2.48,2.57,2.68,2.66,2.69,2.77,0,0,0,0,0,0,0,0,0,1.56,1.44,1.48,1.58,1.73,1.79,1.86,
1.98,2.01,2.11,1.99,2,2.13,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.34,0.28,0.33,
0.91,1.22,0.97,0.52,0.73,0.81,0.71,1.23,1.09,0.99,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,1.86,1.83,1.8,1.81,1.83,1.87,1.92,1.98,2.02,2.01,1.87,1.95,2.02,2.11,2.18,2.23,
2.3,2.26,2.3,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.09,1.06,1.22,1.1,1.1,1.09,1.08,1.02,0.86,
0.92,0.94,0.91,1.1,1.11,1.28,1.22,1.26,1.28,1.35,1.37,1.37,1.51,1.59,1.61,0,0,0,0,0,
0,0,0,0,0,0.99,0.92,1.05,0.96,1.01,0.76,0.36,0.46,0.25,-0.34,0.1,0.19,0.21,0.66,-0.04,
0.38,0.76,0.69,0.88,0.7,0.56,0.74,0.8,0.46,0.27,0.75,0.84,0.75,0.54,0.62,0.84,0,0,0,
1.5,1.64,1.66,1.63,1.48,1.37,1.43,1.44,1.45,1.56,1.68,1.68,1.79,1.73,1.84,1.86,1.8,
1.65,1.69,1.62,1.73,1.9,1.92,2.04,1.96,1.96,1.92,1.83,1.84,2,1.86,1.68,1.56,1.77,2.51,
2.63,2.67,2.39,2.22,2.16,2.26,2.33,2.42,2.51,2.38,2.37,2.21,2.29,2.44,2.48,2.53,2.5,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2.04,2.09,2.02,2.04,1.98,1.98,1.97,2.01,2,2.03,2.08,2.06,
2.13,2.2,2.11,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),.Dim=c(9,34)),
x=structure(.Data=c(
11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,0,0,0,0,0,0,0,0,0,0,
22,23,24,25,26,27,28,29,30,31,32,33,34,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
22,23,24,25,26,27,28,29,30,31,32,33,34,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,0,0,0,0,0,0,0,0,0,0,
2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,0,0,0,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,
16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),
.Dim=c(9,34)))
```

3. *WinBUGS* code for Bayesian analysis of Data set 4

```
model {
      for(n in 1:Nobs) {
            for (j in 1:2){
            tau.d[n,j] <- 1/vard[n,j]
            d[n,j] ~ dnorm(mu.d[n,j], tau.d[n,j])}
            mu.d[n,1:2] ~ dmnorm(mu.i[StudyNo[n],],tau.i[StudyNo[n],,])}
      for (i in 1:NStudy) {
            mu.i[i,1:2] ~ dmnorm(mu.o[], tau.o[,])
            tau.i[i,1:2,1:2] ~ dwish(Gamma.i[,],2)}
      mu.o [1:2] ~ dmnorm(mu.c[], tau.c[,])
      tau.o[1:2,1:2] ~ dwish(Omega[,],2)}
```

```
# Data
list(Nobs=34, NStudy=18,
StudyNo=c(1, 1, 1, 1, 2, 3, 3, 4, 5, 5, 5,6, 6,7, 7,8, 9, 9, 9, 9, 9,
10, 10, 11, 11, 12, 12, 13, 14, 15, 15, 16, 17, 18),
d=structure(.Data=c(-0.4862, 0.9817, 0.1535, 2.0668, 0.0965, 2.6101,
0, 3.6586, -1.5271, 1.8355, -3.2447, 0.7747, -3.4039, -0.4267,
-4.57, -7.0561, -2.1284, 1.8644, -0.6377, 1.0313, 0.1082, 0.7069,
0.9597, 1.0391, -0.7788, -0.0805, 1.9111, 1.0516, 2.6041, 2.39,
8.5458, 3.0211, -0.3432, 1.1915, -0.6008, 0.5762, -0.5389, -0.1952,
-2.1085, -1.3417, -0.7445, -0.476, -0.7991, 1.6405, 0.8741, 2.4785,
0.3736, 3.8708, -0.5231, 1.7512, -0.3751, 0.1703, -0.8026, 0.0557,
-4.4809, 2.1102, -0.1068, 5.4279, 1.4902, 1.3346, 1.201, 0.878,
-0.1738, 3.4233, -0.4823, 0.8128, -1.1225, 0.532),.Dim=c(34,2)),
vard=structure(.Data=c(0.3432, 0.3735, 0.3343, 0.5113, 0.3337,
0.6172, 0.2857, 0.7638, 0.4305, 0.4737, 1.544, 0.7167, 1.6322,
0.6818, 1.4442, 2.8895, 0.3916, 0.3586, 0.2627, 0.2832, 0.2504,
0.2656, 0.3717, 0.3783, 0.3586, 0.3336, 0.5826, 0.4553, 0.7391,
0.6856, 2.0258, 0.8563, 0.5074, 0.5887, 0.5226, 0.5208, 0.5182,
0.5024, 0.7779, 0.6125, 0.5346, 0.5142, 0.18, 0.2227, 0.1826,
0.2946, 0.2261, 0.7222, 0.188, 0.2515, 0.0848, 0.0836, 0.09,
0.0834, 2.3398, 0.5189, 0.3338, 1.5609, 0.1661, 0.8151, 0.1432,
0.7309, 0.4015, 0.986, 0.5145, 0.5413, 0.4797, 0.4308),.Dim=c(34,2)),
mu.c = c(0,0),
tau.c = structure(.Data=c(1.0E-1,1.0E-6,1.0E-6,1.0E-1), .Dim=c(2,2)),
Gamma.i = structure(.Data=c(10,0.1,0.1,10.0), .Dim=c(2,2)),
Omega = structure(.Data=c(10,0.1,0.1,10.0), .Dim=c(2,2))
)

# Univariate model – different variances for each study
model {
    for(n in 1:Nobs) {
        mu.d[n,1] ~ dnorm(mu.i[StudyNo[n],1],tau.i[StudyNo[n],1])
        mu.d[n,2] ~ dnorm(mu.i[StudyNo[n],2],tau.i[StudyNo[n],2])
      for (j in 1:2){
        tau.d[n,j] <- 1/vard[n,j]
        d[n,j] ~ dnorm(mu.d[n,j], tau.d[n,j]) }}
    for (i in 1:NStudy) {
      mu.i[i,1] ~ dnorm(mu1.o, tau1.o)
      mu.i[i,2] ~ dnorm(mu2.o, tau2.o)
        tau.i[i,1] <- 1/pow(sigma1.i[i],2)
        tau.i[i,2] <- 1/pow(sigma2.i[i],2)
        sigma1.i[i] ~ dunif(0,100)
        sigma2.i[i] ~ dunif(0,100) }
    mu1.o ~ dnorm(0.0,1.0E-4)
    mu2.o ~ dnorm(0.0,1.0E-4)
    tau1.o <- 1/pow(sigma1.o,2)
    tau2.o <- 1/pow(sigma2.o,2)
    sigma1.o ~ dunif(0,100)
    sigma2.o ~ dunif(0,100) }
```

```
# Data – univariate model
list(Nobs=34, NStudy=18,
StudyNo=c(1, 1, 1, 1, 2, 3, 3, 4, 5, 5, 5,6, 6,7, 7,8, 9, 9, 9, 9, 9,
10, 10, 11, 11, 12, 12, 13, 14, 15, 15, 16, 17, 18),
d=structure(.Data=c(-0.4862, 0.9817, 0.1535, 2.0668, 0.0965, 2.6101,
0, 3.6586, -1.5271, 1.8355, -3.2447, 0.7747, -3.4039, -0.4267,
-4.57, -7.0561, -2.1284, 1.8644, -0.6377, 1.0313, 0.1082, 0.7069,
0.9597, 1.0391, -0.7788, -0.0805, 1.9111, 1.0516, 2.6041, 2.39,
8.5458, 3.0211, -0.3432, 1.1915, -0.6008, 0.5762, -0.5389, -0.1952,
-2.1085, -1.3417, -0.7445, -0.476, -0.7991, 1.6405, 0.8741, 2.4785,
0.3736, 3.8708, -0.5231, 1.7512, -0.3751, 0.1703, -0.8026, 0.0557,
-4.4809, 2.1102, -0.1068, 5.4279, 1.4902, 1.3346, 1.201, 0.878,
-0.1738, 3.4233, -0.4823, 0.8128, -1.1225, 0.532),.Dim=c(34,2)),
vard=structure(.Data=c(0.3432, 0.3735, 0.3343, 0.5113, 0.3337,
0.6172, 0.2857, 0.7638, 0.4305, 0.4737, 1.544, 0.7167, 1.6322,
0.6818, 1.4442, 2.8895, 0.3916, 0.3586, 0.2627, 0.2832, 0.2504,
0.2656, 0.3717, 0.3783, 0.3586, 0.3336, 0.5826, 0.4553, 0.7391,
0.6856, 2.0258, 0.8563, 0.5074, 0.5887, 0.5226, 0.5208, 0.5182,
0.5024, 0.7779, 0.6125, 0.5346, 0.5142, 0.18, 0.2227, 0.1826,
0.2946, 0.2261, 0.7222, 0.188, 0.2515, 0.0848, 0.0836, 0.09,
0.0834, 2.3398, 0.5189, 0.3338, 1.5609, 0.1661, 0.8151, 0.1432,
0.7309, 0.4015, 0.986, 0.5145, 0.5413, 0.4797, 0.4308),.Dim=c(34,2)))
```

APPENDIX 16.3: LITERATURE GUIDE TO ADVANCED MODELS

Mixed models

A general mixed model was developed by DuMouchel (1995) for meta-analysis of dose response estimation. This paper and later ones by DuMouchel allow for the combination of studies representing multiple outcomes or multiple treatments, accommodating both binary and continuous outcomes (although a common scale is required for all the outcomes in any particular model). This method does not require that correlations between outcomes within studies are known. Instead they are estimated by the model. This model extends the standard random-effects model of DerSimonian and Laird (1986), allowing heterogeneous study designs to be accounted for and combined. Unlike the methods of Gleser and Olkin (1994) for multiple treatment designs, no common treatment or control group is required to be present in all studies. Different treatment groups can consist of results from separate (sub)groups of subjects, or from groups that cross over and are subject to multiple treatments (e.g., repeated measures of same individuals). A key difference from univariate random-effects models is that each group of subjects is modeled separately, so the model does not treat them as truly multivariate. Van Houwelingen et al. (2002) also adopted a mixed model extension to his 1993 approach to investigate if covariates at the study level are associated with baseline risk via multivariate regression. Finally, Von Ende (2001) discusses how to analyze a repeated measures response over levels of a treatment.

Composite models

There are many variations in the calculation of composite effect sizes. For example, Gleser and Olkin (1994) illustrate the calculation of composite measures for combining effect sizes in multiple treatment and multiple endpoint situations. Marín-Martínez and Sánchez-Meca (1999)

compare three methods of estimating standardized mean differences, each with different assumptions about the correlation between estimates within studies; the authors conclude that the methods can yield different overall estimates, depending on the degree of variation and covariation between effects. An alternative that has received increasing attention is the use of factor analysis to construct a composite summary measure. This approach is being developed and applied in such diverse disciplines as psychology (e.g., Bushman et al. 1991 and later publications) and marketing (e.g., Peterson 2000); see also Hunter and Schmidt (2004) for a brief discussion of this and related approaches. Of course, the interpretation of a summary effect is crucial. If such an effect is adopted, the ability to examine differences in the individual effects is largely lost. On the other hand, if the effects are very highly correlated, the use of a summary effect is only slightly superior to the selection of any single estimate (Hedges and Olkin 1985, 221).

Full multivariate models

A number of Bayesian approaches to the full multivariate model (Situation 4) have been proposed. General Bayesian mixed model approaches for meta-analysis have been promoted and illustrated by DuMouchel and coauthors (DuMouchel and Harris 1983; DuMouchel 1990, 1995), with particular focus on combining studies of interest to the pharmaceutical sciences (DuMouchel 1990). Dominici et al. (1999) proposed a hierarchical Bayesian group random-effects model and presented a complex meta-analysis of both multiple treatments and multiple outcomes. This analysis evaluates 18 different treatments for the relief of migraine headaches that were grouped into three classes. Multiple heterogeneous reported outcomes (including continuous effect sizes, differences between pairs of continuous and dichotomous outcomes) and multiple treatments were incorporated via relationships between different classes of treatments.

Three Bayesian multivariate models were proposed and compared by Nam et al. (2003). The preferred model was the full hierarchical representation that we described. Nam et al. (2003) also detail a Bayesian approach to the same problem considered by Raudenbush et al. (1988) and Berkey, Anderson, et al. (1996, 1998), focusing on multiple outcomes after adjusting for study level covariates. Turner et al. (2006) provide details of a Bayesian hierarchical model for multivariate outcome data that is applicable to randomized cluster trials. These are studies in which entire clusters of subjects are randomized across treatments, but outcomes are measured on individuals (e.g., fields are given different fertilizer treatments and then growth of individual plants is then measured). Here, correlation between observations in the same cluster, as well as correlation between outcomes, creates dependence between the data that must be accounted for. The authors also consider a parameterization of the model that allows for the description of a common intervention effect using data measured on different scales. The models proposed by the authors are applicable to a multivariate meta-analysis with clustering among studies.

Other examples

There are many other examples of meta-analysis related to dependent variables in the literature. Two examples are given here, as encouragement for the reader to search for situations that are closer to their needs. First, Becker (1992, 1995) describes methods for combining correlations. Inter-relationships between variables in the correlation matrix can be analyzed using fixed- or random-effects models. A set of linear models is derived from the correlation matrix that describes both direct and indirect relationships between outcomes of interest and explanatory variables. Inferences are made on the coefficients of the resultant standardized regression equations, including tests for partial relationships.

Second, a fixed-effects meta-regression model was proposed by Berkey, Mosteller, et al. (1996) to combine studies that report multiple continuous outcomes and/or multiple treatments groups, allowing for study and treatment level covariates corresponding to one or more outcomes. As with the above methods, each study can report all or only a subset of outcomes, and consider all or a subset of treatments on the outcome of interest. This model also allows comparisons without a common control group and, similar to Dear (1994), inclusion of trials that consider only one treatment or control group (single-arm trials), hence allowing better use of the available data. It differs from the models previously described in that outcomes and differences between treatments are measured in the original units and not in terms of effect sizes. This can simplify interpretation and allow for a wider range of applications.

Missing correlations

Missing correlations can be estimated in a Bayesian framework by giving them prior distributions and including them as variables to be estimated in the analysis. An advantage of this approach is that the uncertainty induced by estimating missing correlations is both quantified (in terms of variances of the estimates) and is propagated through the remainder of the analyses. As is intuitively reasonable, this results in larger variances for other estimates. This is arguably better than inserting a point estimate for missing data and pretending that it is "known" (i.e., equivalent in value to observed data).

Riley et al. (2008) address the problem of estimating unknown within-study correlations in the context of a bivariate random-effects meta-analysis. By assuming a single overall correlation parameter ρ, and using the fact that a covariance σ_{ij} is equal to $\rho\sigma_i\sigma_j$, they develop a model that uses only the data required for univariate random-effects meta-analysis of each effect/outcome. They show that, unless the estimate ρ is very close to 1 or –1, the pooled estimates derived from their model have little bias, and better statistical properties than those obtained from separate univariate meta-analyses of each effect.

Models with mixed summary and primary data

Riley et al. (2010) discuss models for situations when primary data are available for some studies and only summary data for others. Such models may have two types of risk factors, those that apply to the individual and those that apply to the study. For example, survival of fish populations may depend on the weights of individual fish captured, but also on the average temperature of the water measured in the study. While the weights will differ by fish within a study, the temperature is the same for each fish in the study. Riley's models permit the primary data to inform the estimates of the individual factors, while both primary and summary data inform estimation of the study-level factors.