CLINICAL TRIALS *ARTICLE* *Clinical Trials* 2007; **4:** 525–539

Meta-analysis of longitudinal studies

K Jack Ishak[a,b], Robert W Platt[a,c], Lawrence Joseph[a,d], James A Hanley[a,e] and J Jaime Caro[a,f]

Background Longitudinal studies typically report estimates of the effect of a treatment or exposure at various times during the course of follow-up. Meta-analyses of these studies must account for correlations between effect estimates from the same study.
Purpose To describe and contrast alternative approaches to handling correlations inherent to longitudinal effect estimates in meta-analyses.
Methods. Linear mixed-effects models can account for correlations in a number of ways. We considered three alternatives: including study-specific random-effects, correlated time-specific random-effects or a general multivariate specification that also allows correlated within-study residuals. Data from a review of studies of the effect of deep-brain stimulation (DBS) in patients with Parkinson's disease are used to illustrate the application of these models. Results are contrasted with those from a naïve meta-analysis in which the correlations are ignored.
Results The data included 46 studies that yielded 82 estimates of the effect of DBS measured at 3, 6, 12 months or later after implantation of the stimulator. Models that accounted for correlations, particularly the full multivariate specification, provided better fit (lower AIC) and yielded slightly more precise effect estimates. This was in part due to a relatively extreme observation from a study that provided similar estimates at other times, which in the naïve approach exerts greater influence since it is treated as an independent observation.
Limitations Since the true values of the parameters are not known, it is impossible to confirm that estimates from the multivariate approach are necessarily more accurate.
Conclusion Standard meta-analytic models can be readily extended to account for correlations between effects in longitudinal studies. These models may provide better fit and possibly more precise summary effect estimates. *Clinical Trials* 2007; **4**: 525–539. http://ctj.sagepub.com

Introduction

Longitudinal studies are commonly used in clinical and epidemiological research to assess the effect of a treatment or exposure over time. These studies typically involve a series of measurements of the response at pre-determined intervals. Treatment effects can then be described by estimates calculated at various times (corresponding to the measurement times in the study). Alternatively, longitudinal data are sometimes analyzed using summary measures [1] (e.g., mean or slope of response values for each participant), in which case the treatment effect is expressed in terms of (relative or absolute) differences in the summary measure (e.g., difference in mean slope of response comparing treated to controls).

Meta-analyses of longitudinal studies reporting treatment effects in terms of a single summary measure can be handled with standard approaches [2–4]. Special consideration is required, however, when effect estimates are reported at different times or in terms of multiple parameters,

[a]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada, [b]Caro Research Institute, Montreal, Canada, [c]The Montreal Children's Hospital Research Institute, McGill University, Montreal, Canada, [d]Division of Clinical Epidemiology, Montreal General Hospital, Department of Medicine, Montreal, Canada, [e]Division of Clinical Epidemiology, Royal Victoria Hospital, Department of Medicine, Montreal, Canada, [f]Caro Research Institute, Concord, MA, USA
Author for correspondence: K. Jack Ishak, Caro Research Institute, 185 Dorval Ave, Suite 500, Dorval, Quebec, H9S 5J9, Canada. E-mail: jishak@caroresearch.com

SAGE Publications, Los Angeles, London, New Delhi and Singapore 10.1177/1740774507083567

(e.g., polynomial functions) since these are inherently correlated. The correlations may be biological (e.g., course of the disease), structural (e.g., effects calculated at different times relative to a common baseline) or statistical (e.g., measurements subject to similar 'errors' or derived from same equation – e.g., intercept and slope).

A few examples of meta-analyses of longitudinal data have appeared recently [5–7]. Lopes *et al*. [5] describe a Bayesian model for the meta-analysis of longitudinal studies with patient-level data using mixtures of multivariate normal distributions. Farlow *et al*. [6] also present a meta-analysis of longitudinal studies based on patient-level data. Maas *et al*. [7] describe a mixed model for the meta-analysis longitudinal effect estimates; they handle the correlation between observations by allowing random intercepts and linear time effects.

In this article, we discuss a general linear mixed-effects model for the meta-analysis of longitudinal effect estimates and discuss alternative specifications to account for correlations between the observations. This is a new application of the multiple-outcome models that have been used for meta-analyses of two or more related outcomes [8–18]. We compare findings to those from independent meta-analyses at each time (i.e., ignoring correlations).

The use and relative performance of the models are illustrated with data from a meta-analysis of the effect of deep-brain stimulation (DBS) on motor skills of patients with Parkinson's disease at 3, 6, and 12 months and beyond after implantation of the stimulator.

Models for longitudinal meta-analytic data

General linear mixed model for longitudinal data

Suppose K measurements are taken over time on N units (e.g., studies in a meta-analysis); we denote by y_i, the $K \times 1$ vector of observed values from the ith unit and by y_{ij} the jth observation from this unit. A general linear mixed model that can account for the correlations between the observations is given by [19,20]:

$$y_i = X_i\theta + Z_i\delta_i + \varepsilon_i,$$

where X_i is a $K \times p$ matrix of possibly time dependent covariates, θ is a $p \times 1$ vector of fixed effects, Z_i is a $K \times q$ matrix of covariates which are usually a subset of those included in X_i, δ_i is a $q \times 1$ vector of random-effects and ε_i is a $K \times 1$ vector of residuals. Observations from different studies are assumed to be independent, so that $\mathrm{cov}(\varepsilon_{ij}, \varepsilon_{ml}) = 0$ when $i \neq m$ and for any observations j, $l = 1 \ldots K$. It is also assumed that residuals and random-effects are independent: $\mathrm{cov}(\varepsilon_i, \delta_i) = 0$.

The joint distribution of the random-effects is assumed to be a q-dimensional multivariate normal (MVN) distribution with mean 0 and $(q \times q)$ covariance matrix D: $\delta_i \sim MVN_q(0, D)$. The residuals are also assumed to have a joint MVN distribution: $\varepsilon_i \sim MVN_k(0, S_i)$, where S_i is a $K \times K$ covariance matrix. The marginal distribution of y_i is then given by $MVN_k(X_i\theta, V_i)$, where $V_i = \mathrm{var}(y_i) = Z_i' D Z_i + S_i$.

The structure imposed on D and S_i determines how the between- and within-unit (e.g., study) correlations are handled. This may be done in a number of ways [19]: for instance, one may choose to not include any random-effects and set S_i to be a general unstructured positive definite matrix that is constant for all units: i.e., $S_i = S$. Another approach is to allow a random intercept (i.e., set Z_i to be a column of 1s) with (scalar) variance D and set $S_i = \sigma^2 I_K$, where I_K is the $K \times K$ identity matrix. This leads to a compound symmetric correlation structure, whereby $\mathrm{corr}(y_{ij}, y_{il}) = D/D + \sigma^2$ when $j \neq l$. When multiple random-effects are involved, D must be given a structure to describe the relationship between the random-effects within each unit. Some of the most commonly used structures are [21]: compound-symmetry or constant correlations; first-order auto-regressive (AR(1)), where correlations weaken by powers of the lag between observations; Toeplitz, or banded, where observations at different lags are allowed to be differently correlated; and, unstructured, where no patterns are assumed aside from positive-definiteness.

Mixed models for meta-analysis of longitudinal data

The general model described above can be adapted to meta-analyze longitudinal data with slight modifications. In fact, the univariate random-effects meta-analysis model commonly used in practice [2–4] is a special case with $K = 1$ and $Z_i = 1$ (i.e., random-intercept). The within-study variance, which is a scalar in the univariate case, is set to the variance of the observed estimate reported in each study and assumed known without error. This has the effect of weighting each observation by the reciprocal of its variance, so that more precise estimates are more influential.

In meta-analyses of longitudinal data, the *units* of observation are studies from which estimates of the treatment effect are collected at K times. These times may be chosen based on the available data and the objectives of the analysis. Studies may vary with respect to the timing of measurements, and results at some measurement

times may not be reported. This can be a form of publication bias if reporting is related to the magnitude of the effect, similar to selection bias that may occur in a study when loss to follow-up or failure to return for scheduled visits is related to treatment and prognosis.

Longitudinal models for meta-analytic data also differ from the general case in the specification of covariance matrices; in fact, some of the approaches described above would not be appropriate in this context. For instance, random-effects models are increasingly considered to be better suited for meta-analysis [22–24]. Thus, omitting random-effects and accounting for correlations through S_i alone (i.e., a fixed-effects approach) would not be ideal. Furthermore, studies are likely to vary with respect to sample size, measurement methods and other aspects that will cause estimates from one study to be more or less precise than another. The variance may also vary over time within studies because of loss to follow-up, for instance. Thus, setting the within-study covariance matrix S_i to $\sigma^2 I_K$ would not be adequate since it assumes a constant variance for all effect estimates in all studies. S_i should be set to be at least a diagonal matrix with components that can have different values. And, to ensure proper weighting of the data, we set the diagonal elements of S_i to values reported in the studies, as in univariate meta-analyses.

Model specification begins by setting the fixed-effects, which will guide the choice of random-effects. Since the goal of the analysis is to summarize the effect of treatment over time, X_i will include covariates that relate the response variable to time. This might consist of some parameterization of time (e.g., linear, quadratic, logarithmic, etc.). Here, we consider the case where effects are to be summarized at each of the K measurement occasions. This may be done by omitting the intercept and including K time indicators, or by including $K-1$ indicators with an intercept. With $K=4$, we could set:

$$X_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ or } X_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

In the first case, a summary estimate is obtained for each time while in the second, the covariates measure differences in summary effect estimates at each time relative to the first occasion, which is estimated by the intercept. Study-level covariates (e.g., proportion of males, average age) may also be included in X_i. These typically have fixed-effects and, therefore, do not impact the specification of the correlation structure.

Specification of correlation structure

In what follows, unless otherwise stated, we assume S_i is a diagonal matrix with values set to the variances of estimates reported in the studies (we denote the jth diagonal element by S_{ij}^2 for study i). To facilitate notation, we assume X_i only includes time indicators (i.e., no intercept or other covariates); including other fixed-effects to the models described below would not change the correlation structures.

Random study-effect model

The simplest way to account for the correlation between observations is to allow a random-effect that is common to all observations from a given study. This can be thought of as a random-intercept model. If, for example $K=4$, we would set $Z_i' = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$, so that δ_i is a scalar; the model can then be written as:

$$y_{ij} = \sum_{j=1}^{K} X_{ij}\theta_j + \delta_i + \varepsilon_{ij},$$

where $\text{var}(\delta_i) = D$, the covariance of the vector of residuals is $\text{cov}(\varepsilon_i) = S_i$ and $\text{cov}(\delta_i, \varepsilon_{ij}) = 0$ for all i and j. The variance of the marginal distribution of y_{ij} is $D + S_{ij}^2$. And, the covariance between two observations collected on occasions j and l in study i is given by $\text{cov}(y_{ij}, y_{il}) = D$; observations from different studies are assumed independent. Thus, the correlation between observations is $\text{corr}(y_{ij}, y_{il}) = D/\sqrt{\left(D + S_{ij}^2\right)\left(D + S_{il}^2\right)}$, which allows different correlations between observations since within-study variances can differ at each time.

Random time-effects model

The previous model is somewhat restrictive since it assumes that between-study heterogeneity affects observations at each measurement occasion in a given study the same way. This may not be true, however. For instance, attrition may occur differently in each study, causing greater heterogeneity in effects estimated at later times. We can extend the model by allowing a random-effect at each measurement occasion. That is, we set $Z_i = X_i$; the model is then given by:

$$y_{ij} = \sum_{j=1}^{K} X_{ij}\left(\theta_j + \delta_{ij}\right) + \varepsilon_{ij},$$

or

$$y_i = X_i(\theta + \delta_i) + \varepsilon_i.$$

Thus, δ_i is now a K-vector of random-effects with covariance matrix D. Assuming these are independent (i.e., D is diagonal) is equivalent to assuming the observations themselves are independent, since

$$\text{cov}(y_{ij}, y_{il}) = \text{cov}(\delta_{ij} + \varepsilon_{ij}, \delta_{il} + \varepsilon_{il}) = \text{cov}(\delta_{ij}, \delta_{il}) + \text{cov}(\varepsilon_{ij}, \varepsilon_{il}) = 0,$$

because both D and S_i are diagonal and residuals are independent of random-effects. This is equivalent to treating each observation as coming from a different study, or meta-analyzing the data at each time separately.

A structure must, therefore, be imposed on D to account for the correlations between observations. Ideally, D would be left unstructured to allow a different correlation between pairs of random-effects. This adds $K(K-1)/2$ parameters to the model, however. Considering that data are often limited in meta-analysis, this can make it impossible to reliably estimate the parameters of the model. More parsimonious structures like compound-symmetry or AR(1) are more practical, as they involve only one additional parameter. This comes at the cost of more constrained relationships between random-effects, however.

Multivariate model: within- and between study correlations

The models described so far have assumed within-study variation to occur independently at different measurement occasions (diagonal S_i) and accounted for correlations between observations through random-effects. Correlations exist at both within- and between-studies, however. For instance, sampling variability and measurement error might affect estimates within a study at different times in a similar way. In general, factors causing variability between studies (e.g., blinding methods, randomization, etc) may cause correlations within-studies as the impact of these factors is shared by all observations. On the other hand, background factors that cause a given study to be prone to overestimate the effect at one occasion might also have a similar impact at other times.

The approaches considered so far capture the *total* correlation in the marginal distribution of the observations. To consider these separately, we extend the model in the last section by relaxing the assumption that the within-study covariance matrix is diagonal. Two approaches are possible: if studies report the covariance between estimates, these values may be used to fix the off-diagonal elements of S_i, as with the variances. This is rarely the case, however. Alternatively, a hybrid approach may be adopted whereby the diagonal elements of S_i are held fixed and the off-diagonal elements estimated from the data. This requires specification of a structure for S_i, however, which introduces more parameters to the model.

Since data are usually limited in meta-analyses, some simplifications are in order to limit the parameters added to the model. First, we assume within-study correlations are constant across studies (i.e., $\text{corr}(y_{ij}, y_{il}) = \text{corr}(y_{mj}, y_{ml}) = \rho_{jl}$ for studies i and m), still allowing the variances to differ, however. Furthermore, we adopt a single-parameter correlation structure, like compound-symmetry or AR(1).

Implementation of the model

Mixed models for meta-analysis are commonly estimated by likelihood maximization or generalized least squares. While generalized estimating equations are sometimes used with repeated measures patient data, they are rarely applied in meta-analyses as these do not allow estimation of study-level random-effects. These can be useful, for instance, to derive empirical Bayes confidence intervals for effects in each of the studies [16].

The models in our example were estimated by maximum likelihood, using SAS/STAT PROC MIXED software. Van Houwelingen *et al*. [16] describe how the procedure can be adapted for general multivariate (multiple-outcome) meta-analyses with fixed within-study covariances. Additional manipulation is required, however, if only the within-study variances are fixed and covariance components are to be estimated from the data - these are described in Appendix A. Nam *et al*. [12] implemented a similar model in a Bayesian context with BUGS [25].

Case study

Deep-brain stimulation

Data from a meta-analysis of the effect of DBS on the relief of symptoms of Parkinson's disease [26] was used to explore the application of the models described above. Parkinson's disease is a chronic progressive disease characterized by declining motor function and, eventually, severe disability [27]. Although pharmacological treatments are available, medication effects can

become increasingly unpredictable and short-lived, leaving patients with little or no relief of symptoms for much of the day [28]. To prolong effective ('on') periods, physicians increase the medication dose at the risk of side-effects [29]. DBS, which is delivered through thin surgically implanted wires in specific areas of the brain and controlled by the patient, is meant to provide relief with lower doses of medication. DBS of the subthalamic nucleus (STN) [30] has been found to be more beneficial than stimulation of other sites in the brain for patients with Parkinson's disease [31].

We reviewed studies published between 1980 and 2004 reporting on the effects of DBS of the STN on changes in medication dose (and other outcomes of interest, not considered here) [26]. Studies were included if they reported estimates of effect for the outcomes of interest with a standard error or a confidence interval. We were interested in changes in the effect of DBS during the course of the first year of use and beyond. Motor function was measured with the Unified Parkinson's Disease Rating Scale (UPDRS-part III) [32] and generally reported at 3, 6, 12 months and long-term after implantation of the stimulator. Scores can range between 0 and 108, lower values indicating better motor function. The effect of DBS was measured as the difference in scores while the stimulator is active and baseline scores (before implantation of the stimulator).

We recorded effect estimates (y_{ij}) reported at each of the $K=4$ occasions as well as the variances of each estimate (S_{ij}^2). As is common, the covariances between estimates at different times in each study (S_{ijl}) were not reported. We recorded at baseline the mean age of patients, the proportion of females, mean duration of disease and the mean baseline UPDRS scores. These were included in the analyses as predictors of the effect of DBS over time to explain between-study heterogeneity in effects. The data are shown in Appendix B.

Models compared

The data were meta-analyzed using the following model specifications: assuming random study-effects; correlated random time-effects; a full multivariate model that allows correlations within and between studies; assuming complete independence between random-effects and residuals. The latter is equivalent to meta-analyzing the data at each time separately.

We compared the relative magnitude and precision of summary effect estimates from the four models. We acknowledge, however, the relative accuracy of the estimates is difficult to gauge since the true treatment effects are not known; furthermore, gains in precision are not necessarily a sign of improvement as this may lead to narrower confidence intervals that have coverage probabilities below the nominal values. Thus, comparisons aimed to identify differences in results due to the way correlations were handled in each model. For instance, we were interested in how results from the independence model differ from those that account for correlations, and how those from models that capture the total correlation (random study- and time-effects) differ from the full multivariate model.

We also compared goodness-of-fit using the likelihood-ratio statistics of nested models and Akaike's information criterion (AIC), which penalizes the likelihood by twice the number of parameters in the model. All the models included the same fixed effects but differed in terms of the random-effects and covariance parameters.

Specification of covariance matrices

The correlated random time-effects and multivariate models require specifying covariance structures for the random-effects (D) and/or residuals (S_i). In applications with relatively small data sets, like the current and likely most meta-analyses, multi-parameter covariance structures (e.g., unstructured or Toeplitz) can complicate or even inhibit estimation. In fact, we were unable to fit the models in our example with either of these structures; the estimation procedure diverged or produced nonpositive definite covariance matrix estimates with only point estimates.

Our choice was, therefore, limited to the AR(1) and compound symmetric structures. AR(1) seems well-suited for longitudinal data since background factors that induce heterogeneity may vary over time and have a more similar impact on effects measured closer in time than further apart. Alternatively, the compound symmetric structure assumes the correlation is constant. Results in our case study were similar with either formulation; for brevity, we only report findings based on models using the AR(1) structure for both between- and within-study covariance matrices.

In the full multivariate model, we parameterized the covariance components of S_i in terms of correlation parameters which we assume to be common to all studies. That is, for any study i and observations j and l, $S_{ijl} = \rho_{jl}^S \times \sqrt{S_{ij}^2 \times S_{il}^2}$, where ρ_{jl}^S is the correlation between estimates at times j and l in all studies. As before, we assumed an AR(1) structure for the correlations.

Results

Forty-six studies reporting estimates of the effect of DBS of the STN in the absence of medication in at least one of the time intervals of interest were included in the meta-analysis. Half of these studies reported effects for a single time interval and only three reported effects at all four times of interest. Eighty-two observations were extracted: 24 were measured at three months after implantation of the DBS, 22 at 6 months, 25 at 1 year and 11 in the second year or later.

Figure 1(a) and (b) display the observed estimates with corresponding 95% confidence intervals. Considering the invasive nature of the intervention, many studies had small samples, and, thus, yielded estimates with wide confidence intervals that sometimes included the null value. The general pattern of point estimates reveals a considerable improvement in motor skills, however.

Summary estimates from the models are presented in Table 1. Time-specific meta-analyses (i.e., independence model) suggest a reduction of 24.9 points (95% CI: −27.3, −22.4) on the UPDRS-Motor scale at month 3 and slightly stronger effects at 6 and 12 months. The long-term effect appears to decline, however, as the summary estimate suggests improvements similar to those observed at month 3. The random study-effects and correlated random time-effects models yielded similar point estimates and confidence intervals for summary effects. While the pattern was generally consistent with those from the independence model, these suggest a slightly stronger effect at month 3 and long-term. This is evident from both the point estimate and upper-bounds of the confidence intervals, which were also narrower than those from the independence model. The long-term effect was stronger still in the multivariate model.

Estimates of the between-study variances (Table 1) suggest substantial heterogeneity in effects across studies. While point estimates of between-study variances were similar in all of the models, confidence intervals were narrower when correlations were taken into account, particularly in the long-term interval, which included the fewest observations. The upper bound of the 95% confidence interval was reduced from 186.6 in the independence model to 108.4 in the multivariate model.

Models that accounted for correlations also had better fit than the uncorrelated random-time-effects model, which had noticeably higher −2log-likelihood and AIC values. In fact, these statistics were lowered by 39 points in the correlated random-time-effects model which differs from the independence model by a single correlation parameter. The random study-effects model had comparable fit to the former but only includes a single random-effect. The multivariate model had the best fit; allowing within-study correlations reduced the −2log-likelihood by 11.1 points compared to the correlated random time-effects model and 14.1 points compared to the random study-effects model. The last reduction was achieved with five additional parameters (within-study correlation + three additional random-effects + 1 between-study correlation), however. In both cases, improvements in fit with the multivariate model are statistically significant based on chi-square test (with 1 and 5 degrees of freedom) at 5% level of significance, respectively.

We attempted to explain some of the heterogeneity between studies by including the mean age of patients, the proportion of females, mean duration of disease and the mean baseline UPDRS scores as predictors of effects over time. The first two did not appear to be predictive of the size of effects and so were not retained. Mean disease duration and mean baseline score were centered at their mean values (14 years and 52 points) and two studies were excluded because the mean duration of disease of the study population was not reported. Table 2 summarizes the findings from these analyses. The fits of all models were significantly improved with the inclusion of the covariates; the −2log-likelihood and AICs were reduced by 37.5 points or more. The change was largest for the independence and multivariate models, for which the reductions exceeded 50 points. Thus, accounting for within-study correlations appears to lead to an even greater improvement in fit relative to the random study or time effects models in this situation as the gap in fit statistics is even greater than in the unadjusted analyses.

Summary effect estimates were very similar to those obtained in the unadjusted models (Table 1), but confidence intervals were slightly narrower. We attribute this, at least partly, to the substantial reduction in between-study variances by the inclusion of mean disease duration and mean baseline score. Thus, these factors account for an important part of the observed heterogeneity. The effects of mean disease duration and mean baseline score were slightly stronger in the independence model, which also yielded narrower confidence intervals for these factors than the other models. This is because these variables are common to all observations in the study; the uncertainty in their effects is underestimated in the independence model, as this ignores the correlation between repeated observations within studies. The same does not occur with the time indicators, which have different values for each observation.

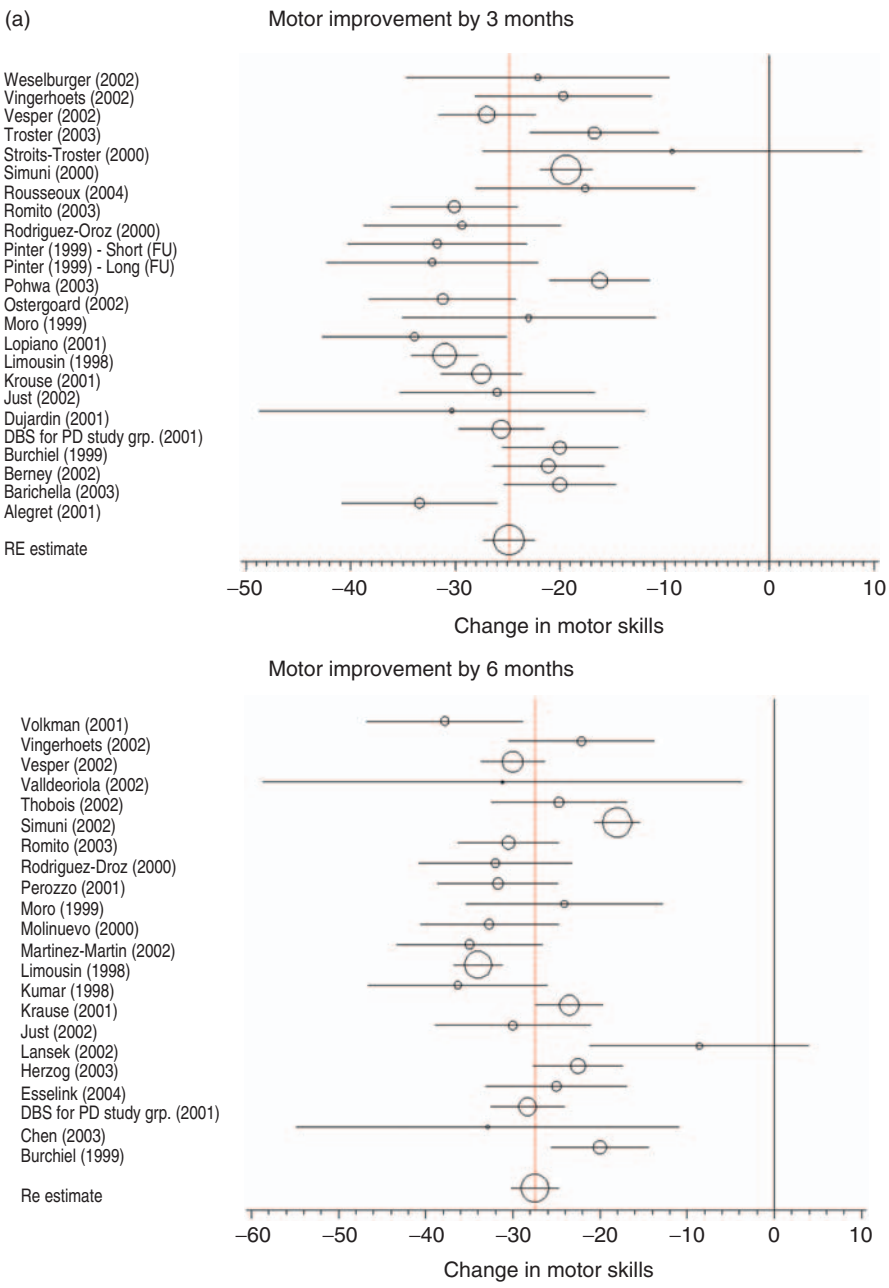

Figure 1 (a) Effect estimates (negative = improvement) by months 3 and 6 with corresponding 95% confidence intervals, as well as summary estimates from univariate random-effects models. *Note*: The two studies by Pinter MM in month 3 are based on two different populations. (b) Effect estimates by month 12 and long-term with corresponding 95% confidence intervals, as well as summary estimates from univariate random-effects models. The size of circles reflects the precision of estimates.

In addition to effect estimates, the correlated random time-effects and multivariate models also quantified within- and between-study correlations. In the former case, the covariance matrix of the random-effects was not positive-definite, so that correlation estimates were not returned from the estimation procedure. In contrast, the within- and between-study correlations were estimated successfully in the multivariate models with and without other covariates (Tables 1 and 2). Both models suggested strong within- and between-study correlations; for instance, in the model without other covariates, the within-study correlation estimate was 0.97 (95% CI: 0.88, 1.00) (The confidence interval was truncated to the maximum of the scale;

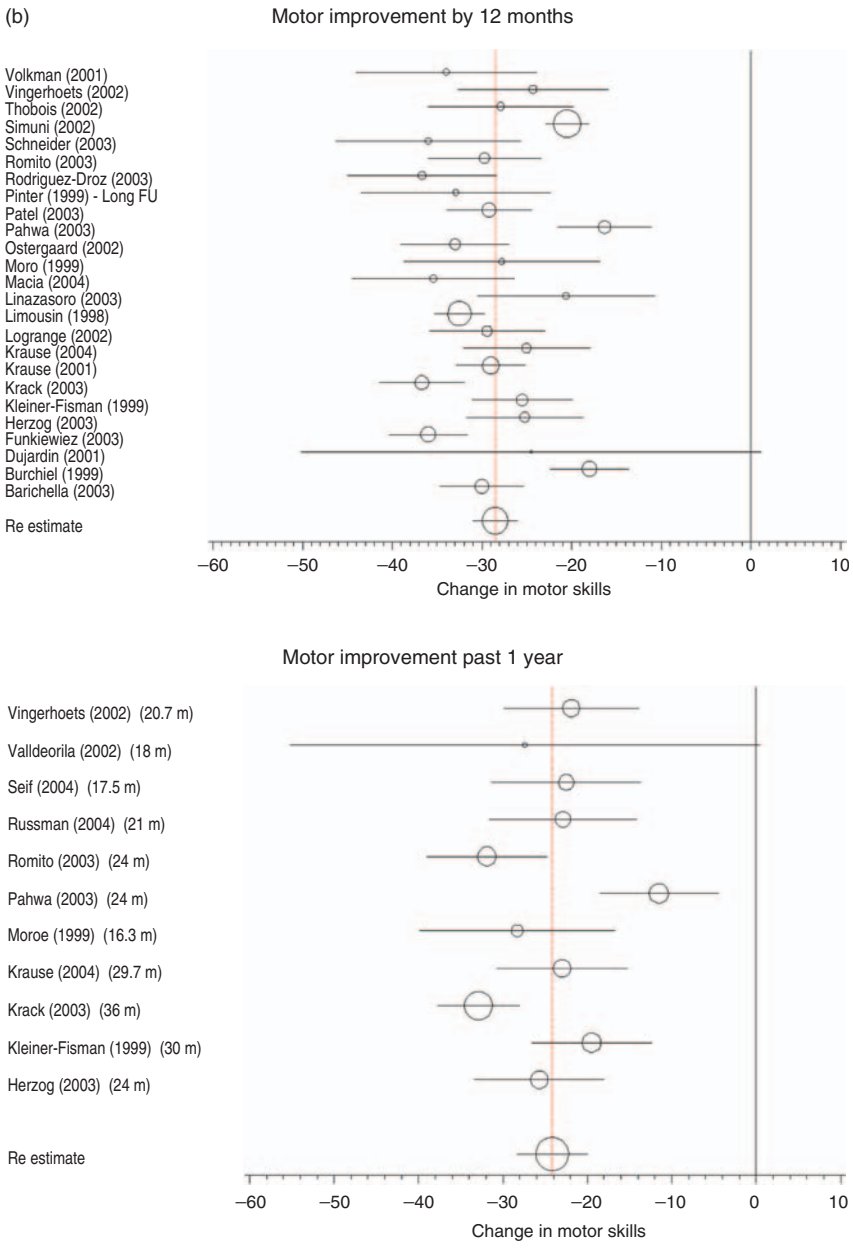

Figure 1 Continued

the estimation procedure returned an upper bound of 1.06), and 0.88 (95% CI: 0.79, 0.98) between studies. We attribute this to the relative stability of effects over time.

Comparison of models

Models that capture the correlations between observations appeared to have significantly better fit than the independence model and produced summary effect and between-study variance estimates with narrower confidence intervals. This suggests a *borrowing of strength* across times in these models, which cannot occur in the independence model which treats the observations as though each arose from a different study. Furthermore, the multivariate model yielded the narrowest interval estimates for the long-term data and had better fit than the other longitudinal models. Thus, it seems that modeling within- and between-study correlations separately was beneficial.

Table 1 Summary effect and between-study variance estimates and 95% confidence intervals from various meta-analysis models

	Time-specific (Independence)	Random study effects	Correlated random time effects[b]	Multivariate model[c]
Summary effect estimates				
3 Months	−24.9 (−27.3, −22.4)	−26.2 (−28.3, −24.1)	−26.0 (−27.9, −23.9)	−25.9 (−27.9, −23.9)
6 Months	−27.5 (−30.2, −24.7)	−27.2 (−29.3, −25.1)	−27.5 (−29.7, −24.9)	−27.5 (−29.7, −25.2)
12 Months	−28.5 (−31.0, −26.0)	−28.5 (−30.6, −26.5)	−28.6 (−30.6, −26.5)	−28.7 (−30.7, −26.6)
Long-term	−24.1 (−28.3, −20.0)	−25.6 (−28.9, −22.5)	−25.8 (−29.0, −22.6)	−26.5 (−29.2, −23.8)
Estimates of between-study variances				
3 Months	23.1 (11.2, 71.6)	26.7 (16.1, 52.4)[a]	20.4 (11.0, 49.7)	22.6 (11.6, 61.3)
6 Months	27.8 (13.4, 88.9)		36.0 (18.7, 85.1)	33.7 (18.7, 78.4)
12 Months	27.7 (14.7, 69.6)		26.4 (15.1, 58.0)	26.1 (14.7, 58.7)
Long-term	29.9 (11.5, 186.6)		30.1 (13.1, 124.7)	31.1 (14.5, 108.4)
Estimates of correlations				
Within-Study	NA	NA	NE	0.97 (0.88, 1.06)[d]
Between-study	NA	NA	NE	0.88 (0.79, 0.98)
Model fit				
−2LogL	524.7	488.7	485.7	474.6
AIC	532.7	490.7	493.7	486.6

[a]Estimate and 95% confidence interval of variance of the random intercept.
[b]Implemented with AR(1) covariance structure for the joint distribution of random-effects.
[c]Implemented with an AR(1) covariance structure for within- and between-study covariance matrices.
[d]The confidence interval was truncated to the maximum of the scale; the estimation procedure returned an upper bound of 1.06.
NA: Model does not include correlation parameters.
NE: Correlation parameters not estimated, since covariance matrices were not positive definite.

Table 2 Summary effect and between-study variance and 95% confidence intervals of between-study variance parameters from meta-regressions including mean disease duration and mean baseline UPDRS-motor scores of patients in each study

	Time-specific (Independence)	Random study effects	Correlated random time effects[b]	Multivariate model[c]
Summary effect estimates				
3 Months	−25.0 (−26.4, −23.7)	−26.2 (−28.0, −24.4)	−25.7(−27.2, −24.3)	−25.7 (−26.9, −24.5)
6 Months	−27.0 (−29.2, −24.8)	−27.1 (−28.9, −25.3)	−27.5 (−29.6, −25.4)	−27.7 (−29.4, −25.9)
12 Months	−28.1 (−30.3, −26.0)	−28.4 (−30.2, −26.7)	−28.5 (−30.3, −26.7)	−28.8 (−30.4, −27.1)
Long-term	−24.0 (−27.1, −20.0)	−25.1 (−28.0, −22.2)	−25.2 (−28.1, −22.3)	−26.1 (−28.1, −24.1)
Disease duration	−0.58 (−1.18, 0.02)	−0.81 (−1.67, 0.04)	−0.78 (−1.53, −0.04)	−0.63 (−1.29, 0.02)
Baseline score	−0.60 (−0.74, −0.46)	−0.54 (−0.76, −0.33)	−0.57 (−0.76, −0.38)	−0.69 (−0.86, −0.52)
Estimates of between−study variances				
3 Months	1.8 (0.35, 3491.2)	13.3(7.4, 30.4)[a]	5.6 (2.1, 38.1)	1.8 (0.44, 152.0)
6 Months	14.1 (6.1, 60.2)		20.9 (10.3, 62.3)	12.7 (5.9, 43.4)
12 Months	16.3 (7.6, 55.6)		14.8 (7.6, 40.8)	12.0 (6.1, 33.0)
Long-term	9.9(2.7, 407.5)		12.9 (4.4, 137.5)	6.9 (2.0, 161.3)
Estimates of correlations				
Within-study	NA	NA	NE	0.95 (0.88, 1.00)[d]
Between-study	NA	NA	NE	0.77 (0.52, 1.00)[d]
Model fit				
−2LogL	472.0	451.2	445.6	423.2
AIC	480.0	453.2	453.6	435.2

[a]Estimate and 95% confidence interval of variance of the random intercept.
[b]Implemented with AR(1) covariance structure for the joint distribution of random-effects. Final estimate of covariance matrix was not positive-definite.
[c]Implemented with an AR(1) covariance structure for within- and between-study covariance matrices.
[d]The confidence interval was truncated to the maximum of the scale; the estimation procedure returned an upper bound of 1.06.
NA: Model does not include correlation parameters.
NE: Correlation parameters not estimated, since covariance matrices were not positive definite.

Models that accounted for correlations yielded stronger long-term summary estimates. This may be due to a relatively extreme effect estimate (identified by an arrow in Figure 2) in this time interval. This observation came from a study that also reported relatively low effects at other times. In fact, patients in this study had better motor skills (lower UPDRS score) at baseline compared to other studies, as a result there was lower potential for improvement. We removed the long-term observation and refitted the models. Figure 3 illustrates the new results (dashed lines) and the original (solid lines) effect estimates for the independence and multivariate models, which had previously yielded the weakest and strongest estimates, respectively. A greater change can be seen in results from the independence model compared to the multivariate model (and other two longitudinal models – not shown), which was essentially unaffected. This is because the extreme point is modeled as an independent observation in the independence model. However, similar effects were observed from the same study at previous times, which implies strong within-study correlations. Thus, the *effective* information provided by the observed effects is less than if had they been observed in different studies (i.e., if they were truly independent). This is taken into consideration in models that account for the correlations and, hence, the inherent loss of information this implies. The extreme observation is, therefore, less influential in these models and exerts far less *pull* compared to the independence model.

To explore this further, we also removed a relatively extreme observation at month six – the only observation from that study – and found similar changes in summary estimates from all models. Therefore, we infer that accounting for the correlations reduced the potential influence of the extreme observation. We must consider, however, that correlations between observations were very strong in our data; it is not clear whether the extreme observation would have been more influential otherwise.

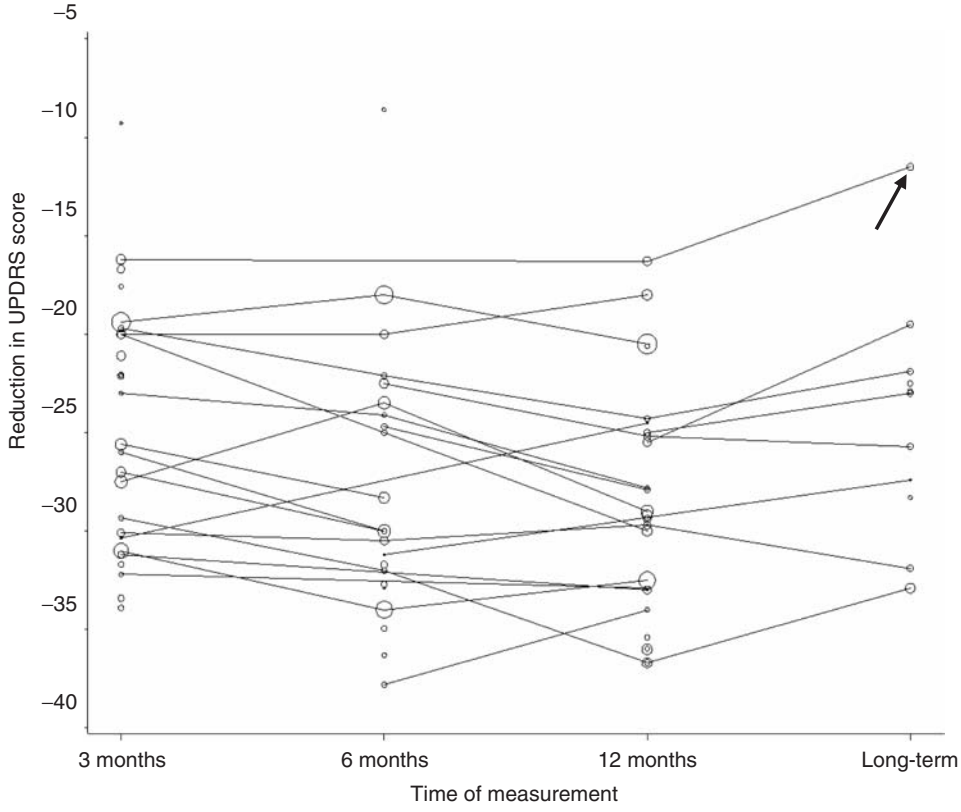

Figure 2 Observed effects of deep brain stimulation of the subthalamic nucleus by months 3, 6, 12, and long-term and summary effect estimates. Connected points arise from the same study; unconnected points are from studies reporting an effect estimate at a single time. The size of circles reflects the precision of estimates. The arrow points to an apparently extreme observation, which was removed to assess its impact on summary estimates (Figure 3)

Discussion

In our case study, models that accounted for correlations had better fit and produced slightly more precise summary effect and heterogeneity estimates, particularly in intervals where data were limited. In fact, confidence intervals from the multivariate model were slightly narrower than those from random study- and time-effects models, which also had poorer fit. Furthermore, we found that the multivariate model (and to some extent the random-time-effects models) was less affected by a relatively extreme estimate in the last interval.

The gain in precision with models that account for correlations seems surprising, since accounting for correlations would be expected to lead to increases in variances [33]. It is possible that factors that induce this variability likely affect the measurements at different times in a similar way; that is, studies that tend to find a strong effect may do so at all times. Thus, allowing independent random-effects for each outcome (as in univariate meta-analyses) perhaps exaggerates the heterogeneity. Allowing correlations between random-effects in the multivariate model, on the other hand, may partially account for the overlap in heterogeneity between time-intervals, leading to lower between-study variance estimates and, hence, increased precision. In our example, this was apparent only when we accounted for mean baseline score and mean disease duration; in this case, the multivariate model suggested lower heterogeneity than the independence model. In the unadjusted models, we suspect the greater precision in the multivariate model was in large part due to the reduced impact of the extreme observation noted above. This is evidenced by the fact that lower bounds of the confidence intervals were almost identical, while upper bounds (and point-estimates) changed.

Although the correlated random-time-effects and multivariate models generally allow more flexibility in the specification of covariance structures, their estimation may be hampered by the relatively small size of meta-analyses. In our example, we restricted the covariance matrices to have AR(1) or compound-symmetry structures since estimation algorithms did not converge otherwise. Despite these simplifications, we encountered difficulties in estimating the correlation parameters. The situation can be improved if within-study covariances are reported in publications, as these can be used to fix the parameters in the models (thus reducing the number of parameters to estimate). Otherwise, within-study covariances can also be fixed using approximations of the correlation without much risk of error in summary effect estimates [34].

Most studies reported effects at some but not all of the times; 'unreported' effects are treated as missing in the analyses. To be valid, time-specific meta-analyses require that missingness occurs completely at random (MCAR). By including data from multiple times and accounting for the correlations between them, the multivariate

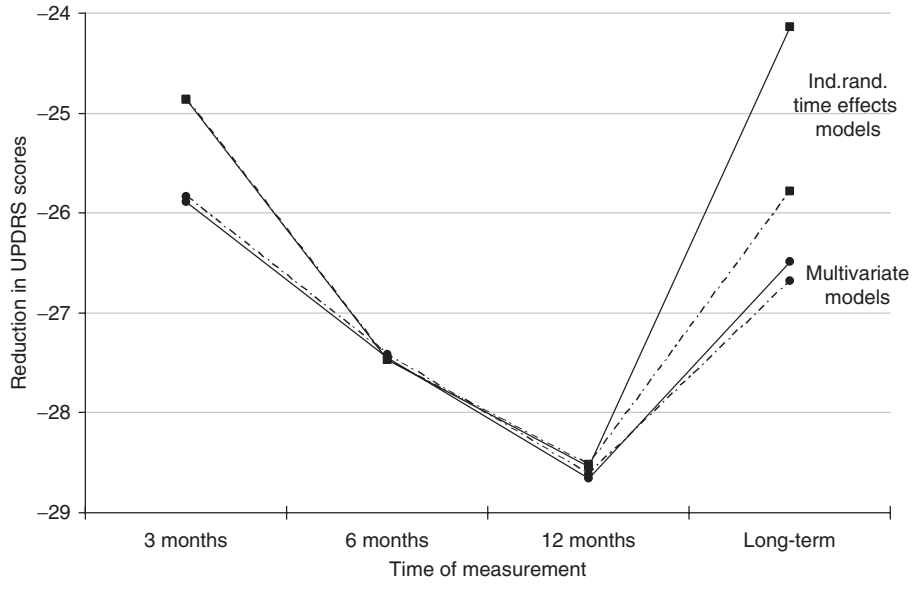

Figure 3 Summary estimates of effect with all observations (solid lines) and when a relatively extreme observation in the long-term interval is removed (dashed line)

model can borrow information requires and only that the data be missing at random (MAR).

While our example was based on effects measured as mean changes from baseline in a single arm, the models we described can be used for meta-analyses of randomized control trials without complications. They can also be applied to different effect measures (e.g., log-odds-ratios calculated at various times), as long as these have (at least approximately) normal sampling distributions. Furthermore, although we used categorical time intervals, continuous formulations of the time variable could be implemented, assuming the shape of the relationship can be specified reasonably well. In this case, studies included in the meta-analysis need not be restricted to those publishing specific time points of interest.

Applications of multivariate meta-analytic models [8–17] have typically shown little difference from multivariate and outcome-specific meta-analyses. Riley *et al*. [18] have shown slight gains in precision in scenarios with potential for borrowing strength. Our case study also suggests there may be a potential for gaining precision with a multivariate analysis. Since the true values of the parameters of the model were not known in our example, we cannot confirm that the observed differences were favorable. Further simulation studies are required to verify our findings.

Acknowledgements

This work was supported by a doctoral fellowship from the Natural Sciences and Engineering Research Council of Canada to KI, a Chercheur-boursier award from the Fonds de la Recherche en Santé du Québec to RP, a Senior Scientist award from the Canadian Institutes of Health Research to LJ and grants from the Natural Sciences and Engineering Research Council of Canada and the Fonds Québécois de la Recherche sur la Nature et les Technologies to JH. The authors are grateful to the three reviewers, editor and associate editor for their helpful comments, which greatly improved the manuscript.

References

1. **Senn S, Stevens L, Chaturvedi N.** Repeated measures in clinical trials: simple strategies for analysis using summary measures. *Stat. Med.* 2000; **19**: 861–77.
2. **DerSimonian R, Laird N.** Meta-analysis in clinical trials. *Controlled Clin. Trials* 1986; **7**: 177–88.
3. **Normand ST.** Tutorial in biostatistics: Meta-analysis: formulating, evaluating, combining and reporting. *Stat. Med.* 1999; **18**: 321–59.
4. **Hardy RJ, Thompson SG.** A likelihood approach to meta-analysis with random effects. *Stat Med* 1996; **15**: 619–29.
5. **Lopes HF, Muller P, Rosner GL.** Bayesian meta-analysis for longitudinal data models using multivariate mixture priors. *Biometrics* 2003; **59**: 66–75.
6. **Farlow MR, Small GW, Quarg P, Krause A.** Efficacy of Rivastigmine in Alzheimer's disease patients with rapid disease progression: results of a meta-analysis. *Dement Geriatr Cogn Disord* 2005; **20**: 192–7.
7. **Maas CJM, Hox JJ, Lensvelt-Mulders GJLM.** Longitudinal meta-analysis. *Quality & Quantity* 2004; **38**: 381–9.
8. **Raudenbush SW, Becker BJ, Kalaian H.** Modeling multivariate effect sizes. *Psychological Bulletin* 1988; **103**: 111–20.
9. **Berlin JA, Longnecker MP, Greenland S.** Meta-analysis of epidemiologic dose-response data. *Epidemiology* 1993; **4**: 218–28.
10. **Berkey CS, Anderson JJ, Hoaglin DC.** Multiple outcome meta-analysis of clinical trials. *Stat Med* 1996; **15**: 537–57.
11. **Berkey CS, Hoaglin DC, Antczak-Bouckoms A, Mosteller F, Colditz GA.** Meta-analysis of multiple outcomes by regression with random effects. *Stat Med* 1998; **17**: 2537–50.
12. **Nam I, Mengersen K, Garthwaite P.** Multivariate meta-analysis. *Stat Med* 2003; **22**: 2309–33.
13. **Dear KBG.** Iterative generalized least squares for meta-analysis of survival data at multiple times. *Biometrics,* 1994; **50**: 989–1002.
14. **Arends LR, Voko Z, Stijnen T.** Combining multiple outcome measures in a meta-analysis: an application. *Stat Med* 2003; **22**: 1335–53.
15. **Van Houwelingen HC, Zwinderman K, Stijnen T.** A bivariate approach to meta-analysis. *Stat Med* 1993; **12**: 2272–84.
16. **Van Houwelingen HC, Arends LR, Stijnen T.** Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* 2002; **21**: 589–624.
17. **McIntosh MW.** The population risk as an explanatory variable in research synthesis of clinical trials. *Stat Med* 1996; **15**: 1713–28.
18. **Riley RD, Abrams KR, Lambert PC, Sutton AJ, Thompson JR.** An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Stat Med,* 2006; **26**: 78–97.
19. **Demidenko E.** *Mixed Models. Theory and Applications*. Wiley Series in Probability and Statistics, Hoboken, NJ, 2004.
20. **Cnaan A, Laird NM, Slasor P.** Using the general linear mixed model to analyze unbalanced repeated measures and longitudinal data. *Stat Med* 1997; **16**: 2349–80.
21. **Littell RC, Pendergast J, Natarajan R.** Modeling covariance structures in the analysis of repeated measures data. *Stat Med* 2000; **19**: 1793–819.
22. **Brockwell SE, Gordon IR.** A comparison of statistical methods for meta-analysis. *Stat Med* 2001; **20**: 825–40.
23. **Thompson SG, Sharp SJ.** Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med* 1999; **18**: 2693–708.
24. **Mosteller F, Colditz GA.** Understanding research synthesis (meta-analysis). *Annu Rev Public Health,* 1996; **17**: 1–23.
25. **Spiegelhalter D, Thomas A, Best N, Gilks W, Lunn D.** BUGS: Bayesian inference using Gibbs sampling. MRC Biostatistics unit, Cambridge, England. www.mrc-bsu.cam.ac.uk/bugs/
26. **Ishak KJ, Proskorovsky I, Caro I, Caro JJ.** A meta-analysis of the effect of deep brain stimulation of the subthalamic nucleus in Parkinson's disease. Submitted for publication.

27. **Dodel R, Berger K, Oertel W.** Health-related quality of life and healthcare utilization in patients with Parkinson's disease. Impact of motor fluctuations and dyskinesias. *Pharmacoeconomics* 2001; **19**: 1013–38.
28. **Dewey RB.** Management of motor complications in Parkinson's disease. *J Neurology* 2004; **62**: S3–7.
29. **Krack P, Pollak P, Limousin P, Benazzouz A, Deuschl G, Benabid AL.** From off-period dystonia to peak-dose chorea. The clinical spectrum of varying subthalamic nucleus activity. *Brain* 1999; **122**: 1133–46.
30. **Benabid A, Koudsié A, Benazzouz A *et al*.** Subthalamic stimulation for Parkinson's disease. *Arch. Med. Res.* 2000; **31**: 282–9.
31. **Kumar R, Lozano AM, Montgomery E, Lang AE.** Pallidotomy and deep brain stimulation of the pallidum and subthalamic nucleus in advanced Parkinson's disease. *Movement Disorders* 1998; **13**: 73–82.
32. **Fahn S, Elton RL.** The UPDRS Development Committee. Unified Parkinson's disease rating scale. In **Fahn S, Marsden CD, Calne D, Goldstein M.** (Eds). *Recent developments in Parkinson's disease* Vol 2 Macmillan Healthcare Information, Florham Park, NJ, 153–63, 1987.
33. **Liang KY, Zeger SL.** Regression analysis for correlated data. *Annual Review of Public Health* 1993; **14**: 43–68.
34. **Ishak KJ, Platt RW, Joseph L, Hanley J.** Impact of approximating within-study covariances in multivariate meta-analyses. *Stat Med* (in press).

Appendix

A. SAS code for multivariate meta-analysis

We describe below generic SAS code to implement a multivariate meta-analysis model for longitudinal effect estimates in which both within- and between-study correlations are estimated. The challenge lies in specifying the SAS code to estimate all parameters of the between-study covariance matrix (D) and the within-study correlation ρ_S, while holding within-study variances fixed to their observed values.

We assume the data for the meta-analysis are stored in a data set called **mult_ma_data**, in which data pertaining to each effect estimate reported in each study are recorded as a separate row. The data set should include the following variables:

Study_id	Unique identifier for each study; may be an abbreviated reference (Author, year) or other unique numeric code assigned to each study.
Meas_time	Time of measurement of effect; should be stored as numeric (3, 6, 12 . . .) if time will be considered as continuous variable; otherwise, can use descriptive names ('3 Months', '6 Months' . . .).
Eff_est	Estimate of the effect.
Var_est	Variance of the estimate.
Covariates	Variables describing the study characteristics, like mean age of the population, etc.

For each study, one row should appear for each measurement occasion, with missing values for effects estimates (and other data elements) at occasions where no data were reported. Records from the first five studies in our example data set were as follows:

study_id	meas_time	eff_est	var_est
Alegret (2001)	3 Months	−33.41	14.339
Alegret (2001)	6 Months	.	.
Alegret (2001)	12 Monhts	.	.
Alegret (2001)	Long−term	.	.
Berney (2002)	3 Months	−21.10	7.257
Berney (2002)	6 Months	.	.
Berney (2002)	12 Monhts	.	.
Berney (2002)	Long−term	.	.
Chen (2003)	3 Months	.	.
Chen (2003)	6 Months	−32.90	124.986
Chen (2003)	12 Monhts	.	.
Chen (2003)	Long−term	.	.
Dujardin (2001)	3 Months	−30.34	88.203
Dujardin (2001)	6 Months	.	.
Dujardin (2001)	12 Monhts	−24.50	170.683
Dujardin (2001)	Long−term	.	.

For instance, only month 3 results were reported by Alegret (2001), so that data only appears in the first record; observations in subsequent records are missing.

The multivariate model is fitted using PROC MIXED, in which the structure of within- and between-study covariance matrices (AR(1) in our example) must be specifed. The between-study covariance matrix is common to all studies, and all of its parameters are to be estimated from the data. The specification of the within-study covariance matrices requires special consideration, on the other hand, since the variance components must be allowed to vary across studies and their values held fixed to the observed variances of the estimates, while the correlation parameter is common to all studies (to ensure identifiability) and is to be estimated from the data. We can represent the structure of the within-study covariance matrices as follows:

$$S_i = W_i^{-1/2} \times C \times W_i^{-1/2}$$

$$= \begin{pmatrix} s_{i11}^{-2} & 0 & 0 & 0 \\ 0 & s_{i22}^{-2} & 0 & 0 \\ 0 & 0 & s_{i33}^{-2} & 0 \\ 0 & 0 & 0 & s_{i44}^{-2} \end{pmatrix}^{-1/2}$$

$$\times \begin{pmatrix} 1 & \rho_S & \rho_S^2 & \rho_S^3 \\ \rho_S & 1 & \rho_S & \rho_S^2 \\ \rho_S^2 & \rho_S & 1 & \rho_S \\ \rho_S^3 & \rho_S^2 & \rho_S & 1 \end{pmatrix}$$

$$\times \begin{pmatrix} s_{i11}^{-2} & 0 & 0 & 0 \\ 0 & s_{i22}^{-2} & 0 & 0 \\ 0 & 0 & s_{i33}^{-2} & 0 \\ 0 & 0 & 0 & s_{i44}^{-2} \end{pmatrix}^{-1/2}$$

where C is common to all studies with an unknown the correlation parameter, while W_i is specific to each study and must be held fixed to their observed values. This can be accomplished with the parms and weight commands in PROC MIXED. The parms statement accepts starting values for the covariance matrix parameters (stored in a data set), and allows specific parameters to be held fixed to the starting values (with the hold option). We pass initial values for D and C, and hold the variance parameters in C to 1 throughout estimation. The inverse of within study variances are specified as weights in PROC MIXED, which by definition are constants, and are used as multipliers to the within-study correlation matrix as described in the formula above. The SAS code to fit a multivariate model is described below.

data mult_ma_data; set mult_ma_data; w = **1**/var_est; **run**;	Create a variable called w, calculated as the reciprocal of the variance of the effect estimate, corresponding to the diagonal components of S_i.
data whithin_study_cov; param='Var−1'; est=1; output; param='Var−2'; est=1; output; param='Var−3'; est=1; output; param='Var−4'; est=1; output; param='Corr'; est=0.5; output; **run**;	Create starting values for C and store in a dataset called within_study_cov. One row appears for each parameter, and a variable called est with values set to 1 for variances and 0.5 for the correlation parameter, which should appear last.
data btw_study_cov; param='Var−1'; est=23.1; output; param='Var−2'; est=27.7; output; param='Var−3'; est=27.8; output; param='Var−4'; est=29.9; output; param='Corr'; est=0.5; output; **run**;	Similarly, starting values are specified for D and stored in a dataset called btw_study_cov. Starting values can be obtained from the data; for instance, we used the between−study variance estimates from univariate analyses. We set the correlation parameter starting value to 0.5.
data initial_values; set btw_study_cov within_study_cov; keep param est; **run**;	Initial values for both within− and between−study covariance matrices are appended into a single dataset, called initial_values. It is important that starting values of between−study covariance matrix.
proc mixed method=REML cl data=mult_ma_data; class study_id meas_time; model eff_est = meas_time /noint s cl ddf=1000,1000, 1000,1000;	The parameters are estimated by REML maximization. The cl option requests confidence interval estimates for the parameters. Class specifies variables that are to be treated as classification or categorical variables. The model includes the measurement times and no intercept (noint) so that an estimate is obtained for each occasion. The ddf = option specifies the degrees of freedom for the t−tests; they are set to 1000 to obtain standard normal tests as in Van Houwelingen [16].
random meas_time /subject=study_id type=arh(1) s; repeated meas_time /subject= study_id type=arh(1);	The *random* and *repeated* statements specifiy the structure of between and within−study covariance matrices, respectively. The subject = option specifies the units within which observations are correlated. The s option requests that the parameter estimates be displayed.
parms/parmsdata=initial_values hold=6 to 9; weight w; **run**;	The parms statement specifies the starting values of the covariance parameters. Hold = specifies the rows of parameters in the initial_values dataset that are to be held at the provided values (in the est variable). The variable holding the weights for each observation is specified with the weight command.

Appenndix B: Data for deep-brain stimulation meta-analysis

The table below shows the data for the example discussed in the paper. The mean age and proportion of males at baseline are not shown as these variables were not retained in the meta-regression analyses.

	Month 3		Month 6		Month 12		Long−term			
Reference	Effect estimate	Variance	Effect estimate	Variance	Effect estimate	Variance	Effect estimate	Variance	Mean disease duration	Mean baseline score
Alegret (2001)	−33.4	14.3							16.1	53.6
Barichella (2003)	−20.0	7.3			−30.0	5.7			13.5	45.3
Berney (2002)	−21.1	7.3							13.6	45.6
Burchiel (1999)	−20.0	8.0	−20.0	8.0	−18.0	5.0			13.6	48.0
Chen (2003)			−32.9	125.0					12.1	65.7
DBS for PD Study Grp. (2001)	−25.6	4.2	−28.3	4.6					14.4	54.0
Dujardin (2001)	−30.3	88.2			−24.5	170.7			13.1	65.0
Esselink (2004)			−25.0	17.0					12.0	51.5
Funkiewiez (2003)					−36.0	5.0			14.0	56.0
Herzog (2003)			−22.5	6.8	−25.2	11.0	−25.7	15.4	15.0	44.9
Iansek (2002)			−8.6	41.0					13.0	27.6
Just (2002)	−26.0	22.4	−30.0	20.6					14.0	44.0
Kleiner−Fisman (1999)					−25.5	8.2	−19.5	13.0	13.4	50.1
Krack (2003)					−36.7	5.8	−32.9	6.1	14.6	55.7
Krause (2001)	−27.5	3.8	−23.5	3.8	−29.0	3.8			13.7	59.0
Krause (2004)					−25.0	13.0	−23.0	15.4	14.4	60.0
Kumar (1998)			−36.3	27.3					14.3	55.7
Lagrange (2002)					−29.4	10.7			14.0	53.7
Limousin (1998)	−31.0	2.6	−34.0	2.0	−32.5	2.0			14.0	57.0
Linazasoro (2003)					−20.6	25.3			13.7	47.7
Lopiano (2001)	−33.9	20.1							15.4	59.8
Macia (2004)					−35.4	21.2			15.0	55.2
Martinez−Martin (2002)			−34.9	18.0					16.4	55.7
Molinuevo (2000)			−32.7	16.3					15.8	49.6
Moro (1999)	−23.0	38.1	−24.1	32.9	−27.8	31.0	−28.3	34.6	15.4	67.6
Ostergaard (2002)	−31.2	12.7			−33.0	9.5			15.0	51.3
Pahwa (2003)	−16.2	5.9			−16.3	7.0	−11.5	12.7	12.1	41.3
Patel (2003)					−29.2	5.8			10.0	47.8
Perozzo (2001)			−31.7	12.4					15.4	59.7
Pinter (1999) − Long FU	−32.2	26.5			−32.9	29.0			11.3	60.0
Pinter (1999) − Short FU	−31.7	19.1							11.5	59.7
Rodriguez−Oroz (2000)	−29.3	22.9	−32.0	20.0	−36.7	17.8			16.5	51.5
Romito (2003)	−30.1	9.4	−30.5	8.7	−29.7	10.4	−31.9	13.3	13.8	63.9
Rousseaux (2004)	−17.6	28.4							12.0	52.3
Russman(2004) (21 m)							−22.9	20.0	15.9	47.1
Schneider (2003)					−36.0	27.7			17.0	51.3
Seif(2004) (17.5 m)							−22.5	20.3	15.0	44.2
Simuni (2002)	−19.4	1.6	−18.0	1.7	−20.5	1.5			16.7	43.5
Straits−Troster (2000)	−9.3	85.2							8.0	47.4
Thobois (2002)			−24.7	15.5	−27.9	17.1			13.5	44.9
Troster (2003)	−16.7	9.8							9.5	41.6
Valldeoriola (2002)			−31.2	196.0			−27.4	201.6	15.6	49.0
Vesper (2002)	−27.0	5.5	−30.0	3.5					14.0	53.0
Vingerhoets (2002)	−19.7	18.5	−22.1	18.1	−24.3	18.2	−21.9	16.7	16.0	48.8
Volkman (2001)			−37.8	20.9	−34.0	26.4			13.1	56.4
Weselburger (2002)	−22.1	40.8							14.0	50.3