

Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials

Ashley P Jones^a, Richard D Riley^b, Paula R Williamson^a and Anne Whitehead^c

Background In clinical trials following individuals over a period of time, the same assessment may be made at a number of time points during the course of the trial. Our review of current practice for handling longitudinal data in Cochrane systematic reviews shows that the most frequently used approach is to ignore the correlation between repeated observations and to conduct separate meta-analyses at each of a number of time points.

Purpose The purpose of this paper is to show the link between repeated measurement models used with aggregate data and those used when individual patient data (IPD) are available, and provide guidance on the methods that practitioners might use for aggregate data meta-analyses, depending on the type of data available.

Methods We discuss models for the meta-analysis of longitudinal continuous outcome data when IPD are available. In these models time is included either as a factor or as a continuous variable, and account is taken of the correlation between repeated observations. The meta-analysis of IPD can be conducted using either a one-step or a two-step approach: the latter involves analysing the IPD separately in each study and then combining the study estimates taking into account their covariance structure. We discuss the link between models for use with aggregate data and the two-step IPD approach, and the problems which arise when only aggregate data are available. The methods are applied to IPD from 5 trials in Alzheimer's disease.

Results Two major issues for the meta-analysis of aggregate data are the lack of information about correlation coefficients and the effect of missing data at the patient-level. Application to the Alzheimer's disease data set shows that ignoring correlation can lead to different pooled estimates of the treatment difference and their standard errors. Furthermore, the amount of missing data at the patient level can affect these estimates.

Limitations The models assume fixed treatment effects across studies, and that any missing data is missing at random, both at the patient-level and the study level.

Conclusions It is preferable to obtain IPD from all studies to correctly account for the correlation between repeated observations. When IPD are not available, the ideal aggregate data are model-based estimates of treatment difference and their variance and covariance estimates. If covariance estimates are not available, sensitivity analyses should be undertaken to investigate the robustness of the results to different amounts of correlation. *Clinical Trials* 2009; **6**: 16–27. <http://ctj.sagepub.com>

^aCentre for Medical Statistics and Health Evaluation, School of Health Sciences, University of Liverpool, Brownlow Street, Liverpool, L69 3GS, UK, ^bPublic Health, Epidemiology & Biostatistics, The Public Health Building, The University of Birmingham, Birmingham B15 2TT, ^cCentre for Medical Statistics and Health Evaluation, School of Health Sciences, University of Liverpool, Brownlow Street, Liverpool, L69 3GS, UK

Author for correspondence: Ashley P Jones, Centre for Medical Statistics and Health Evaluation, School of Health Sciences, University of Liverpool, Brownlow Street, Liverpool, L69 3GS, UK. Tel: +44(0)151 794 4750. E-mail: apjones@liverpool.ac.uk

Introduction

The assessment of a patient who has a chronic condition, for example cystic fibrosis or a neurodegenerative disorder, often involves taking measurements of an outcome at a number of time points throughout the course of a trial. Such data are referred to as longitudinal data or repeated measures data, and allow one to assess the change in an outcome over a period of time [1]. When multiple trials of the same longitudinal outcome exist, it is clearly important to synthesize the data across trials to facilitate evidence-based conclusions. However, the longitudinal nature of the data creates added complexity for those performing meta-analysis. For example, it is common to find that the timing of the repeated measurements varies across trials. In this situation, a standard meta-analysis conducted at a specific time point will usually be based on only a subset of the trials. Further, meta-analyzing each time-point independently may be inappropriate, as correlation exists between time-points. Thus a more complex meta-analysis that synthesizes all trials and all time-points simultaneously, whilst accounting for correlation, is potentially superior.

Several authors have considered the meta-analysis of longitudinal data [2–4]. Most recently, Ishak *et al.* consider when only aggregate data are available from each study [4], such as the mean treatment difference and its standard error at each time-point. They propose models that account for the correlation between the multiple time-points, and suggest the approach ‘may provide better fit and possibly more precise summary effect estimates’ than approaches that ignore correlation. However, the authors note that they ‘encountered difficulties in estimating the correlation parameters’, even when very simple correlation structures were assumed, and could not confirm the benefits of the proposed models since in their example ‘the true values of the parameters of the model were not known’. This highlights some of the disadvantages of having only aggregate data, with the unavailability of within-study correlations a major concern and limitation [5,6].

In this article we consider the meta-analysis of longitudinal data using individual patient data (IPD) [7], where the original data from all participants in each trial are obtained and then synthesized. We show that availability of IPD allows the correlation between multiple time-points to be calculated directly in each study, thus overcoming the aforementioned problem shown in Ishak *et al.* [4] when the correlations are unknown. We describe how to appropriately

meta-analyze IPD by: (i) using models that directly synthesize the IPD in a one-step approach, or (ii) first reducing the IPD to aggregate data in each study, and then synthesizing the aggregate data across studies using a multivariate meta-analysis model. This allows us to show the relationship between IPD and aggregate data models, and thus recommend how best to proceed with aggregate data models when IPD are not available.

The outline of the article is as follows. In section ‘A review of current practice’ we summarize a review of practice for meta-analyzing longitudinal studies in the Cochrane Library [8]. In section ‘Methods for meta-analysis of longitudinal data’ we describe methods for IPD and aggregate meta-analysis of longitudinal data, and in section ‘Illustration of the meta-analysis methods’ these are applied to an example where IPD are available for five Alzheimer trials, which enables us to compare IPD and aggregate data approaches directly. Finally, in section ‘Discussion’ we critically discuss our work and make recommendations for practice and further research.

A review of current practice

To investigate current practice for meta-analyzing longitudinal data we searched Issue 3 2005 of the Cochrane Library [8], which contained 2435 completed reviews. We searched using the three terms ‘longitudinal*’, ‘repeated NEXT measure*’ and ‘serial NEXT measure*’, and reviews needed to contain one or all of these search terms to be identified. The term ‘longitudinal*’ was found in 289 reviews (12%), the term ‘repeated NEXT measure*’ in 70 (3%) and ‘serial NEXT measure*’ in 6 (0.25%). These results related to 345 independent reviews, but after further investigation only 113 reviews related to a longitudinal data meta-analysis.

In 85 of these 113 reviews the Methods section contained no information regarding the analysis of longitudinal data; however in 22 of these the Results section indicated that a separate meta-analysis had been carried out at each of a number of time points, with the correlation between time points not taken into account. Indeed, the issue of correlated data was not discussed in the Results or Discussion section of any of these reviews. The remaining 28 reviews did discuss the analysis of longitudinal data in the Methods section, as follows: (i) in 15 reviews it was stated that an analysis at separate time points would be conducted, although this was ultimately only possible in 10 due to the lack of data; (ii) in

six reviews authors stated that they would choose data from one of the time points from each of the trials (four of these reviews pre-specified that they would analyze data from the final time point reported in the trials and two stated that they would analyze the data reported at the time point nearest to the one pre-specified in their protocol); (iii) in five reviews it was mentioned that the length of follow up was a possible source of heterogeneity across trials (two of these studies ultimately used single time point analysis and three did not have any data to perform a meta-analysis); (iv) in one review the authors stated that the statistical package used within Cochrane reviews was not appropriate for analyzing longitudinal data and so they just reported the results from the original trials in a table; and (v) in the remaining review incidence rates based on person time were reported.

It is thus clear from this review that practitioners are undecided on how to appropriately meta-analyze longitudinal studies, and it seems that only simple meta-analysis methods are used. In particular, ignoring correlation between time points and undertaking a meta-analysis at each of the several time points separately was the most frequently used method in practice.

Since this review of the literature was conducted, there has been limited research into the analysis of data of this form [4] and we therefore feel that the situation has not changed a great deal and review authors are still not analyzing this form of data using the most appropriate methods. Our article gives a number of methods that could be used depending on the information that is available to the review author.

Methods for meta-analysis of longitudinal data

In this section, appropriate methods for an IPD analysis of longitudinal continuous outcome data are described, first for a single study and then for a meta-analysis of several studies. Models are presented both for time as a factor and as a continuous variable, and they assume fixed treatment effects across studies. We then consider aggregate data meta-analysis of longitudinal data, and describe the aggregate data required to produce similar results to the IPD meta-analysis. Extension to random treatment effects is considered in the discussion.

Across the studies in the meta-analysis, observations may have been recorded at different time points post-treatment. Our notation below regarding time points assumes that all of the time points which have occurred in any of the

studies have been ordered and are referred to as t_k ($k = 1, \dots, b$). Thus any one trial may only provide observations for some of the b time points. Also, for a particular study an assessment may have been performed prior to any study treatment (baseline). Our models below assume that there are only post-treatment assessments. However, modifications to the models to include baseline measurements in the response variable as an additional assessment time will be discussed briefly where relevant.

IPD from a single study

Treating time as a factor, a model for a single study (study i) which includes treatment, time and treatment by time interaction terms is given by

$$y_{hijk} = \alpha_{ik} + \beta_{hik} + \varepsilon_{hijk}, \quad (1)$$

where y_{hijk} is the observation in study i from patient j ($j = 1, \dots, n_{hi}$) on treatment h ($h = 1, \dots, a$) at time t_k , and ε_{hijk} is the residual error. In this article we take treatment a to be the reference treatment, so that β_{hik} represents the effect of treatment h minus the effect of treatment a at time t_k , and $\beta_{aik} = 0$. The parameter α_{ik} is the effect of treatment a at time t_k . If required, baseline observations can be included as observations y_{hij0} and model (1) defined for $k = 0, 1, \dots, b$. Usually in randomized controlled trials the parameters β_{hi0} , $h = 1, \dots, a$, would be set equal to zero as no differences due to treatment are expected at baseline.

Alternatively, treating time as a continuous variable with a linear effect the model can be written as

$$y_{hijk} = \lambda_i + v_{hi} + \alpha_i t_k + \beta_{hi} t_k + \varepsilon_{hijk}, \quad (2)$$

where λ_i is the intercept for treatment a , v_{hi} is the difference in the intercepts between treatment h and treatment a ($v_{ai} = 0$), α_i is the time slope for treatment a , and β_{hi} is the difference in the time slopes between treatment h and treatment a ($\beta_{ai} = 0$).

Model (2) incorporates separate intercepts for each treatment, thus only making the assumption of a linear trend with time over the time points included in the model. If the baseline observations are to be included or a linear trend from time = 0 is to be fitted, then v_{hi} would be removed from the model.

In both models (1) and (2), the ε_{hijk} are assumed to be normally distributed with zero mean and a covariance structure that allows for the correlation between repeated observations on the

same patient. There are various possible structures one could use [9]: for example, a fully unstructured covariance matrix would allow a separate variance parameter for each time point within each treatment group, and a separate covariance parameter for each pair of time points (k and m , say) in each treatment group, so that

$$V(\varepsilon_{hijk}) = \sigma_{hik}^2 \quad (3)$$

and

$$\text{Cov}(\varepsilon_{hijk}, \varepsilon_{hijm}) = \rho_{hikm} \sigma_{hik} \sigma_{him} \quad \text{for } k \neq m, \quad (4)$$

where ρ_{hikm} is the correlation between observations from the same patient in treatment group h at time points k and m . Simpler specifications are also possible; for instance, we could assume common variances and correlation coefficients across all times and treatment groups. However, in this article we use the fully unstructured covariance matrix above in all of our analyses. The major advantage of this is that it allows the correlation between time points close together to be larger than the correlation between time points which are farther apart. The disadvantage is that it requires a large number of parameters to be estimated [10]. In some situations, the autoregressive covariance structure might be used as an acceptable compromise. Equations (1) and (2), and all subsequent models in this article, can be fitted using software for repeated measurements analysis, for example SAS PROC MIXED [11] (model codes are available on request).

IPD meta-analysis of several studies

When IPD are available from each of the studies, the meta-analysis can proceed in a one-step or a two-step framework [7]. The one-step approach simultaneously models the IPD from all of the studies. The two-step approach first fits a model to the IPD from each study separately, and then the study parameter estimates are combined in a meta-analysis. We discuss both approaches in this section: it is the second approach which links directly with an aggregate data approach when IPD are unavailable.

One-step IPD meta-analysis of several studies

Assume that IPD are available from r independent studies ($i = 1, \dots, r$) and let us extend Equation (1) assuming that the true underlying treatment effect at each time-point is fixed across studies. In this

situation the interaction terms study by treatment and study by treatment by time can be excluded from the model. For time as a factor, the model can be written

$$y_{hijk} = \alpha_{ik} + \beta_{hk} + \varepsilon_{hijk}, \quad (5)$$

where α_{ik} is the effect of treatment a at time t_k in study i , and β_{hk} represents the effect of treatment h minus the effect of treatment a at time t_k , which is common across all studies ($\beta_{ak} = 0$). If required, baseline assessments can be included as observations y_{hi0} , and model (5) defined for $k=0, 1, \dots, b$. Usually the parameters β_{h0} , $h=1, \dots, a$, would be set equal to zero as no differences due to treatment are expected at baseline.

For time as a continuous covariate, Equation (2) can be extended to become

$$y_{hijk} = \lambda_i + v_h + \alpha_i t_k + \beta_h t_k + \varepsilon_{hijk} \quad (6)$$

where λ_i is the intercept for treatment a in study i thus allowing for different prognostic groups entered across the trials, v_h is the intercept for treatment h minus the intercept for treatment a , which is common to all studies ($v_a = 0$). The parameter α_i is the time slope for treatment a in study i , and β_h is the difference in the time slopes between treatment h and treatment a , which is common to all studies ($\beta_a = 0$). If the baseline observations are to be included or a linear trend from time=0 is to be fitted, then v_h would be removed from the model.

In Equations (5) and (6), using the fully unstructured covariance matrix each study has a separate variance parameter for each time point within each treatment group, and a separate covariance parameter for each pair of time points within each treatment group (as in Equations (3) and (4)). Overall estimates of the mean treatment difference at each time point can be obtained: these would be $\hat{\beta}_{hk}$ if using Model (5), and $\hat{v}_h + \hat{\beta}_h t_k$ if using Model (6). T -tests and confidence intervals based on the t -distribution can be calculated. SAS PROC MIXED allows the option to inflate the estimated variance of the mean treatment difference to allow for the estimation of the variance and covariance components and to estimate the degrees of freedom for the t -test using Satterthwaite's procedure [12,13].

Two-step IPD meta-analysis of several studies

In the first step of the two-step IPD meta-analysis approach, each study is analyzed separately. When time is to be treated as a factor, Model (1) is fitted to

each study to produce estimates, variances, and covariances of the mean difference between treatment h and treatment a ($h = 1, \dots, a-1$) at each recorded time point. Let d_{hik} ($h = 1, \dots, a-1$) be the estimate of the mean difference between treatments h and a at time t_k in study i , that is $d_{hik} = \hat{\beta}_{hik}$. When time is to be treated as a continuous variable, Model (2) is fitted to each study to produce estimates, variances and covariances of the differences in the intercepts and time slopes between treatment h and treatment a ($h = 1, \dots, a-1$). Let f_{hi} ($h = 1, \dots, a-1$) be the estimate of the difference in intercepts, that is $f_{hi} = \hat{v}_{hi}$, and g_{hi} ($h = 1, \dots, a-1$) be the estimate of the difference in time slopes, that is $g_{hi} = \hat{\beta}_{hi}$.

The second step requires a multivariate meta-analysis framework [14], which allows the joint synthesis of the d_{hik} (or f_{hi} and g_{hi}) values across studies. The weighting for each d_{hik} (or f_{hi} and g_{hi}) is a function of its variance and covariance with other estimates from the same study. For time as a factor, the model is given by

$$d_{hik} = \beta_{hk} + \delta_{hik}, \quad (7)$$

where δ_{hik} is the residual error, and the δ_{hik} from the same study are correlated.

When time is a continuous variable, there are two model options. The first uses the study estimates of the differences in intercepts and the differences in slopes as the responses and fits the model given by

$$\begin{aligned} f_{hi} &= v_h + \gamma_{hi} \\ g_{hi} &= \beta_h + \xi_{hi} \end{aligned} \quad (8a)$$

where γ_{hi} and ξ_{hi} are residual errors, which are correlated within a study. A second option uses the estimates of the mean difference between treatments h and a at time t_k in each study as the responses and fits the model given by

$$d_{hik} = v_h + \beta_h t_k + \delta_{hik} \quad (8b)$$

where δ_{hik} is the residual error, and the δ_{hik} from the same study are correlated. In both cases, the effect of treatment h minus the effect of treatment a at time t_k is given by $v_h + \beta_h t_k$.

In common practice in the meta-analysis field, the variance and covariance terms for the d_{hik} (or f_{hi} and g_{hi}) calculated in the first step are used in the second step as the variance and covariance of the residual errors, and treated as if they were the true variances and covariances. The multivariate meta-analysis in the second step may be fitted using SAS PROC MIXED [11], as described elsewhere [15], in which the d_{hik} (or f_{hi} and g_{hi}) values are the observations in the dataset. When using the f_{hi} and g_{hi} values a study must have data

for at least two time points to be included in the analysis: if there is only one time point f_{hi} and g_{hi} cannot be calculated.

Overall estimates of the mean treatment difference at each time point can be obtained: these would be $\hat{\beta}_{hk}$ if using Model (7), and $\hat{v}_h + \hat{\beta}_h t_k$ if using Models (8a) or (8b). Assuming that the variances are known, hypothesis tests and confidence intervals may be based on the normal distribution.

One-step versus two-step approach

Although the two IPD approaches should provide similar results, they will not be identical for several reasons. First, the two-step approach allows for the possibility of study by treatment and study by treatment by time interactions when calculating the variance and covariance terms of the d_{hik} (or f_{hi} and g_{hi}), although the actual model fitted to the d_{hik} (or f_{hi} and g_{hi}) does not. The one-step method based on Models (5) or (6) does not include these interaction terms at all. The effect that this has on the results depends on the magnitude of these interaction terms.

A second factor which may produce a difference between the two IPD approaches is the amount of missing data at the patient-level within a study [16]. Under a missing at random assumption, Equations (1), (2), (5), and (6) can accommodate patients, who do not supply data at each time point in the study. Indeed, under this assumption the correlation between time points is used to improve the estimation of the parameters and this results in 'model-based' estimates (section 'Meta-analysis of aggregate data'). As parameter estimates and variances are dependent on the model terms as well as the structure of the covariance matrix, this may lead to differences in the estimates between the two IPD approaches. This factor is also important when we consider the aggregate data methods in section 'Meta-analysis of aggregate data'.

Meta-analysis of aggregate data

An aggregate data meta-analysis of longitudinal data clearly depends on the aggregate data available from trials [17]. If the estimates d_{hik} (or f_{hi} and g_{hi}) and their variances and covariances are available, then the second part of the two-step IPD analyses described in section 'Two-step IPD meta-analysis of several studies' can be undertaken directly. However, often trials only report means and standard deviations for each treatment group at each time point. In this section we thus describe

methods which are based only on these summary statistics, and discuss the implications that this has for the validity of the meta-analysis.

Let $\bar{y}_{hi.k}$ be the mean of the n_{hik} observations from treatment h at time t_k in study i , and $z_{hik} = \bar{y}_{hi.k} - \bar{y}_{ai.k}$. Then z_{hik} is an estimate of the effect of treatment h minus the effect of treatment a at time t_k , and the model for the z_{hik} based on Model (1) is given by

$$z_{hik} = \beta_{hik} + (\bar{\varepsilon}_{hi.k} - \bar{\varepsilon}_{ai.k}) \quad (9)$$

We refer to z_{hik} as a *raw* estimate of the mean treatment difference as it is based on raw means (this is in contrast to the model-based estimates of treatment difference discussed in section ‘One-step versus two-step approach’, which may be subtly different). Assuming that the treatment groups contain independent sets of patients, the variance of z_{hik} is given by

$$V(z_{hik}) = V(\bar{\varepsilon}_{hi.k} - \bar{\varepsilon}_{ai.k}) = \frac{\sigma_{hik}^2}{n_{hik}} + \frac{\sigma_{aik}^2}{n_{aik}} \quad (10)$$

Also, for $k \neq m$, the covariance between the observed mean difference at time point k and the observed mean difference at time point m is given by

$$\begin{aligned} \text{Cov}(z_{hik}, z_{him}) &= \text{Cov}(\bar{\varepsilon}_{hi.k}, \bar{\varepsilon}_{hi.m}) + \text{Cov}(\bar{\varepsilon}_{ai.k}, \bar{\varepsilon}_{ai.m}) \\ &= \frac{q_{hikm} \rho_{hihm} \sigma_{hik} \sigma_{him}}{n_{hik} n_{him}} \\ &\quad + \frac{q_{aikm} \rho_{aihm} \sigma_{aik} \sigma_{aim}}{n_{aik} n_{aim}} \end{aligned} \quad (11)$$

where q_{hikm} is the number of patients contributing to both time points k and m in treatment group h .

Meta-analysis of aggregate data – time as a factor

To perform a meta-analysis of the aggregate data which is close to the two-step IPD meta-analysis described in section ‘Two-step IPD meta-analysis of several studies’, the z_{hik} values and estimates of their variances and covariances based on Equations (10) and (11) need to be calculated for each study. These are used in the second step of the two-step IPD meta-analysis, so that model (7) is replaced by

$$z_{hik} = \beta_{hk} + (\bar{\varepsilon}_{hi.k} - \bar{\varepsilon}_{ai.k}). \quad (12)$$

If all patients provide data at all time points in trial i , the d_{hik} values from the IPD analysis of trial i , based on the fully unstructured covariance matrix, are equal to the raw z_{hik} values and they

have the covariance structure given by Equations (10) and (11). However, in the case when there are missing data at the patient-level, d_{hik} and z_{hik} may be subtly different; the d_{hik} are *model-based* estimates, which adjust for the missing patient data across time points, unlike the z_{hik} , and so the d_{hik} and their covariance structure will be a modification of Equations (10) and (11) [14]. Consequently, if there are any missing patient-level data the two-step IPD approach will use model-based estimates, of treatment difference, and their variances and covariances, whilst the aggregate data approach will only use *raw* estimates, and their variances and covariances. The magnitude of the differences between the model-based and raw mean treatment differences will depend on the pattern and amount of missing data. Large differences may subsequently cause a discrepancy in the results from the two-step IPD meta-analysis and the aggregate data meta-analysis of raw means.

Meta-analysis of aggregate data – time as a continuous variable

To perform a meta-analysis of the aggregate data with time treated as a continuous variable, the z_{hik} values and estimates of their variances and covariances based on Equations (10) and (11) could be used in the second step of the two-step IPD meta-analysis, with model (8b) replaced by

$$z_{hik} = v_h + \beta_h t_k + (\bar{\varepsilon}_{hi.k} - \bar{\varepsilon}_{ai.k}). \quad (13)$$

The effect of treatment h minus the effect of treatment a at time t_k is given by $v_h + \beta_h t_k$. Estimates of the mean treatment differences from the aggregate data approach may be similar, but not identical to those from the two-step IPD approach. When time is considered as a continuous variable, deviations from the fitted straight line with time are not incorporated in the variance and covariance terms for the aggregate data approach, whereas they are in the first step of the two-step IPD approach. Further, as discussed in section ‘Meta-analysis of aggregate data – time as a factor’, when there are missing patient data the model-based estimates from the IPD may differ subtly from the raw z_{hik} estimates.

Unavailable aggregate data

A major issue for the meta-analysis of raw means is that the covariance defined by Equation (11) cannot often be calculated, as σ_{hik} , q_{hikm} , and ρ_{hikm}

will rarely be available from publications, and so approximations are necessary. If s_{hik} denotes the standard deviation for treatment group h at time point k in study i , then this can be used as an estimate of σ_{hik} . The q_{hikm} values might be approximated by the minimum of n_{hik} and n_{him} . The ρ_{hikm} values could be approximated by one of two approaches. One could take a simplistic approach and assume one common correlation across all studies, treatments and time points. More appropriately, one could perform a sensitivity analysis to ascertain if and how the meta-analysis results change with the value of the correlation coefficient imputed [14]. In some situations it may also be more appropriate to assume different imputed correlations for each pair of time points, though this approach will become increasingly complicated as the number of time points increases. Note that setting the ρ_{hikm} equal to zero leads to a standard meta-analysis conducted at each time point separately. This is the common method in practice (section 'A review of current practice'), but it is a strong assumption to make as longitudinal data are often highly correlated [1]. It leads to an estimate of the mean difference between treatment h and treatment a at time point k given by

$$\sum_{i=1}^r \frac{w_{hik} z_{hik}}{w_{hik}} \text{ with variance } \frac{1}{\sum_{i=1}^r w_{hik}},$$

$$\text{where } w_{hij} = (s_{hik}^2/n_{hik} + s_{aik}^2/n_{aik})^{-1}.$$

Hypothesis tests and confidence intervals may be based on the normal distribution [18].

Illustration of the meta-analysis methods

To illustrate and compare the IPD and aggregate data meta-analysis methods, we now consider a systematic review of trials investigating the effects of selegiline versus placebo for the treatment of Alzheimer's disease [19]. Five trials provided IPD for the mini-mental state examination (MMSE), a measure of cognitive function, and in this article we focus on just these five studies as they allow us to empirically compare IPD and aggregate data approaches. In each of the trials the dosing schedule was the same (10 mg/day) and they all had a different length of treatment with no common time point. The MMSE can take values from between 0 and 30, with higher values being regarded as good, and is considered in our analyses to be approximately normally distributed.

To illustrate the methods presented in this article, the results have been grouped to create common time points. That is: month 1 = weeks 4 and 5; month 2 = weeks 8 and 9; month 4 = week 17; month 6 = weeks 24, 25, and 30; month 9 = week 35 and 43; month 12 = weeks 56 and 65. Some patients withdrew before the completion of all of the assessments and no single study gave results for each time-point; thus there are missing data in at the patient-level and also at the study-level. Table 1 taken from Table 9.6 of Whitehead [18] shows the raw means and standard deviations for each treatment at each time point in each study. From these values the raw mean differences and

Table 1 Raw summary statistics for MMSE data

Month	Study	Placebo			Selegiline			Treatment effect (Selegiline – placebo)	
		Number of subjects	Mean	Standard deviation	Number of subjects	Mean	Standard deviation	Mean, z_{hik}	Variance, $V(z_{hik})$
1	3	24	17.08	4.33	22	17.73	6.78	0.65	64.72
	4	166	12.33	5.61	165	13.07	5.40	0.74	60.63
	5	25	19.88	6.27	25	17.72	5.67	−2.16	71.46
2	1	20	18.30	4.40	18	18.78	6.28	0.48	58.80
	3	23	18.04	5.00	23	17.43	6.71	−0.61	70.02
	5	24	19.33	6.35	25	17.56	4.93	−1.77	64.63
4	3	20	17.20	4.49	23	18.00	6.28	0.80	59.60
	4	151	11.84	5.57	156	12.28	5.47	0.44	60.95
6	2	68	20.32	5.16	64	19.80	5.46	−0.52	56.44
	3	18	16.33	5.40	23	17.74	6.24	1.41	68.10
	4	139	11.14	5.95	144	11.23	5.68	0.09	67.66
9	1	18	16.17	6.22	17	17.12	6.38	0.95	79.39
	4	125	9.94	6.01	134	10.45	5.74	0.51	69.07
12	1	17	15.47	6.34	15	13.07	7.41	−2.40	95.10
	4	112	9.59	6.01	121	9.79	6.05	0.20	72.72

their variances (i.e., the z_{hik} and $V(z_{hik})$ values) can be calculated.

The meta-analysis models described in section ‘Methods for meta-analysis of longitudinal data’ were applied to the MMSE data to compare results from the one-step IPD analysis with the two-step IPD analysis of model-based estimates, and also the aggregate data meta-analysis of raw estimates. For illustrative purposes both time as a factor and time as a continuous variable were considered. However, it should be noted that the continuous model is perhaps not entirely appropriate for this data set. Figure 1 shows a plot of the differences between selegiline and placebo across time.

Time as a factor

IPD meta-analyses: one-step versus two-step using correct covariances

There are no statistically significant differences between the two treatments at any of the time points when time is treated as a factor (Table 2). As discussed in section ‘One-step versus two-step approach’ the results from the one-step (Equation (5)) and the two-step IPD meta-analysis (model (1) in each study followed by model (7) using the correct covariance estimates) can differ, but for this example the results are very similar.

IPD meta-analyses: one-step versus two-step imputing covariances

As part of our analyses, we considered how the two-step meta-analysis results would change if the

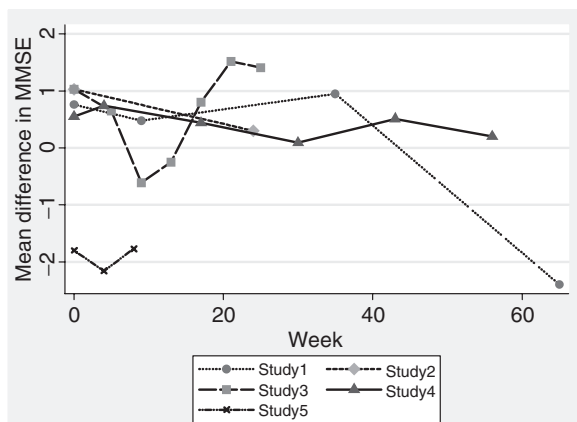


Figure 1 Plot of raw mean differences in MMSE (selegiline – placebo) against time

Table 2 Meta-analysis of MMSE data treating time as a factor: estimates (standard error) of difference between Selegiline and Placebo

Time point	One-step IPD Equation (5)	Two-step IPD Equation (1) in each study, followed by Equation (7)			Aggregate data Equation (12), covariance estimates from Equation (11) with $\sigma_{hik} = s_{hik}$		
		Correct covariance estimates for d_{hik}	Correlations of 0.8 between d_{hik}	Correlations of 0.4 between d_{hik}	Correlations of 0 between d_{hik}	$\rho_{hikm} = 0.8$	$\rho_{hikm} = 0.4$
1 month	0.31 (0.47)	0.30 (0.47)	0.30 (0.48)	0.34 (0.52)	0.43 (0.54)	0.40 (0.48)	0.39 (0.53)
2 months	-0.48 (0.62)	-0.47 (0.59)	-0.37 (0.65)	-0.58 (0.88)	-0.84 (0.97)	-0.22 (0.69)	-0.42 (0.91)
4 months	0.34 (0.48)	0.33 (0.47)	0.39 (0.49)	0.52 (0.55)	0.75 (0.57)	0.19 (0.51)	0.29 (0.57)
6 months	0.20 (0.49)	0.19 (0.48)	0.22 (0.48)	0.23 (0.50)	0.31 (0.50)	-0.10 (0.51)	-0.08 (0.52)
9 months	0.35 (0.53)	0.34 (0.52)	0.33 (0.53)	0.47 (0.60)	0.69 (0.63)	0.31 (0.60)	0.39 (0.67)
12 months	-0.02 (0.56)	-0.03 (0.55)	-0.10 (0.56)	0.05 (0.63)	0.29 (0.66)	-0.30 (0.66)	-0.21 (0.73)
							$\rho_{hikm} = 0$
							0.44 (0.54)
							-0.70 (0.99)
							0.49 (0.59)
							0.00 (0.53)
							0.56 (0.69)
							-0.04 (0.75)

model-based mean differences and their variances were available from each study, but without their covariances, as this situation may occur in practice. We thus fitted Equation (7) to the model-based estimates for each of correlations 0.8, 0.4, and 0 (Table 2). It can be seen that the pooled treatment difference estimates when the correlation is 0.8 are similar to those using the correct covariance estimates. This is not surprising as the actual correlations between time point estimates were of a similar magnitude. As the correlation coefficient becomes smaller, the treatment difference estimates move further away from those based on the correct covariance estimates. Indeed the assumption of zero correlation, as is usual in practice, produces estimates, and standard errors that are very different from those based on the correct covariance estimates. For example, assuming zero correlation gives a pooled treatment difference at 9 months of 0.69, with standard error of 0.63, compared to the true answer of 0.34, with standard error of 0.52. Importantly, though, the standard errors are conservative in this situation, a finding consistent with analytical assessment of longitudinal data in a non meta-analysis setting when time is taken as a factor [20].

Aggregate data meta-analyses

We now assume that trials only provide their raw means and standard deviations, and not their IPD. As discussed in section 'Unavailable aggregate data', to perform the aggregate data meta-analysis in this situation we needed to approximate σ_{hik} and q_{hikm} , and we then fitted meta-analysis model (12) for each of ρ_{hikm} equal to 0, 0.4, and 0.8 (Table 2). There is a considerable amount of missing patient data in the MMSE studies and thus the raw estimates of treatment difference differ considerably from the model-based estimates derived from IPD, as discussed in section 'Meta-analysis of aggregate data – time as a factor'. This leads to differences between the aggregate data meta-analysis results and those from the one-step and two-step IPD analyses (Table 2). For example, in the analyses assuming a correlation of 0.8, at 12 months the aggregate data meta-analysis gives a pooled treatment difference of -0.30 , with standard error of 0.66, compared to the true answer of -0.03 , with standard error of 0.55. However, importantly the standard error of the aggregate data meta-analysis results is again always conservative and, in this particular example, the conclusions from the aggregate data meta-analysis are the same as those from the IPD analyses,

Table 3 Meta-analysis of MMSE data treating time as a continuous variable: estimates (standard error) of difference between Selegiline and Placebo (Study 2 excluded*)

	One step IPD Equation (6)	Two-step IPD Equation (2) in each study, followed by Equation (8a)	Two-step IPD Equation (1) in each study, followed by Equation (8b)				Aggregate data Equation (13), covariance estimates from Equation (11) with $\sigma_{hik} = s_{hik}$		
			correct covariance estimates for d_{hik}	correlations of 0.8 between d_{hik}	correlations of 0.4 between d_{hik}	correlations of 0 between d_{hik}	$\rho_{hikm} = 0.8$	$\rho_{hikm} = 0.4$	$\rho_{hikm} = 0$
Intercept estimate	0.38 (0.53)	0.37 (0.52)	0.45 (0.51)	0.52 (0.49)	0.43 (0.48)	0.38 (0.46)	0.47 (0.51)	0.41 (0.49)	0.36 (0.47)
Slope estimate	-0.005 (0.036)	-0.005 (0.036)	-0.015 (0.035)	-0.020 (0.033)	-0.007 (0.055)	0.017 (0.067)	-0.018 (0.047)	-0.005 (0.063)	0.019 (0.073)
Time point									
1 month	0.37 (0.52)	0.37 (0.51)	0.43 (0.50)	0.50 (0.49)	0.42 (0.45)	0.40 (0.40)	0.45 (0.50)	0.40 (0.46)	0.38 (0.41)
2 months	0.37 (0.52)	0.36 (0.51)	0.42 (0.50)	0.48 (0.49)	0.41 (0.43)	0.41 (0.35)	0.43 (0.49)	0.40 (0.43)	0.40 (0.36)
4 months	0.36 (0.52)	0.35 (0.51)	0.39 (0.50)	0.44 (0.49)	0.40 (0.40)	0.45 (0.28)	0.40 (0.50)	0.39 (0.41)	0.44 (0.29)
6 months	0.35 (0.53)	0.34 (0.52)	0.36 (0.51)	0.40 (0.50)	0.39 (0.40)	0.48 (0.26)	0.36 (0.52)	0.38 (0.42)	0.48 (0.28)
9 months	0.33 (0.56)	0.32 (0.55)	0.31 (0.54)	0.34 (0.53)	0.37 (0.45)	0.53 (0.34)	0.31 (0.58)	0.36 (0.50)	0.53 (0.38)
12 months	0.32 (0.61)	0.31 (0.60)	0.26 (0.59)	0.28 (0.57)	0.35 (0.55)	0.58 (0.50)	0.25 (0.67)	0.35 (0.63)	0.59 (0.56)

*Study 2 had to be excluded from the analysis using Equation (8a) as it had only one time point result and thus did not provide an intercept and slope estimate. Study 2 could have been included in the other analyses presented, but to ensure comparability it was also removed from these.

namely that there is no evidence that selegiline is beneficial.

Time as a continuous variable

One-step IPD analysis versus two-step IPD analysis of intercept and slope

The results from the IPD meta-analyses when time is treated as a continuous variable and a linear regression is fitted are shown in Table 3. As a linear regression model could not be fitted to study 2, because it only provides data at one post-treatment time point, the two-step meta-analysis of intercepts, and slopes had to exclude this study. For comparability study 2 has also been excluded from all meta-analyses treating time as a continuous variable.

The one-step IPD meta-analysis gives results that are almost identical to the two-step analysis of model-based intercept and slope estimates (using the correct covariances). There is no evidence that the difference between selegiline and placebo follows a linear trend over time or that selegiline is better than placebo.

Two-step analysis of model-based or raw time point estimates

The two step method in section 'One-step IPD analysis versus two-step IPD analysis of intercept and slope' assumes that the slope and intercept terms are available from each study. These estimates are very rarely reported by trial authors and so, if the IPD are not available, the only option may be to fit a regression line to either the model-based or raw estimates of treatment difference at each time point (see Equations (8b) and (13), respectively). We thus fitted a regression line to both the model-based and raw time point estimates for the MMSE data (Table 3).

The two-step analysis using the model-based time point and correct covariance estimates gives a negative slope estimate of -0.015 , slightly larger than the negative slope estimate of -0.005 from than the two-step analysis based on model-based intercept, slope, and correct covariance estimates. If we also assume that the covariance estimates are unavailable, as likely in practice, the assumption of zero correlation between the model-based time point estimates causes the standard error of the pooled estimates of treatment difference to be too small (Table 3). For example, assuming zero correlation gives the standard error of the pooled treatment difference at 9 months to be 0.34 , whereas the true answer is 0.54 (compare columns

four and seven in '9 months' row in Table 3). This is important, as practitioners may place an overconfidence in their results in this situation. Furthermore, assuming zero correlation changes the slope of the fitted regression line from negative (-0.015) to positive (0.017), although there is still no evidence of a linear trend over time or that selegiline is beneficial.

The two-step meta-analysis using the raw treatment difference produces results that differ slightly to those from the meta-analysis of model-based estimates. The key observation again, though, is that assuming zero correlation seriously underestimates the standard error of pooled estimates. For example, at six months the true result (i.e., that from using the model-based estimates with correct covariances) is a pooled treatment difference of 0.36 , with standard error of 0.51 , but the analysis of raw means assuming zero correlation gives a pooled estimate of 0.48 , with standard error 0.28 . Thus, for time as continuous, it is clear that treating the correlation as zero does *not* lead to conservative standard errors, a finding consistent with analytical assessment of longitudinal data in a non meta-analysis setting when time is taken as continuous [20].

Discussion

When conducting a meta-analysis of longitudinal data, it is preferable to obtain the IPD from all of the studies that were identified from the systematic review. This facilitates meta-analysis models that correctly account for the correlation between repeated observations for the same patient in each trial. In this article we have presented such IPD meta-analysis models, which use either a one-step or a two-step approach. Application to the Alzheimer's data set confirms that these approaches are preferable to an analysis that assumes no correlation between repeated observations, for a number of reasons. First, the pooled estimates when assuming zero correlation can differ considerably from the correct answers, although in our particular example clinical conclusions were unaffected. Second, when time is treated as a factor, the analyses including correlation were more efficient, with the standard error of pooled estimates much smaller than in analyses assuming zero correlation. Third, when time is treated as continuous, the analyses assuming zero correlation severely underestimate the standard error of estimates, which in practice may lead to wrong conclusions. These findings are consistent with those regarding ignoring correlation in the analysis of single studies [20], and are especially important

given that our review of current practice (section 'A review of current practice') identified that the most common method for meta-analyzing longitudinal data is to synthesize each time-point independently. We thus strongly recommend that, wherever possible, practitioners should obtain IPD and then perform a meta-analysis that accordingly accounts for correlation.

We are aware, though, that it may not always be possible to obtain IPD from trialists, and so we have also presented possible methods for an aggregate data meta-analysis in this situation. The ideal aggregate data to obtain are the model-based study estimates, and their variance and covariance estimates, as these allow the IPD meta-analysis results to be closely replicated (Tables 2 and 3) even when there is a considerable amount of missing patient data. Model-based estimates may be available but without their relevant covariances (e.g., the covariance of treatment difference estimates between time points), so we have shown how to perform sensitivity analyses to investigate the robustness of meta-analysis conclusions to different values of imputed correlation. In doing so, we have assumed that the imputed correlation is the same for each pair of time points, which is simplistic. One could impute different correlations for each time-point, but this greatly increases the number of parameters to be imputed. As an aid, data from external studies in the same or similar indications could be used to inform the choice of values for the correlations and the extent of the variation across different pairs of time-points.

Though an aggregate data meta-analysis of model-based estimates is the ideal aggregate data approach, the model-based estimates may themselves be unavailable, and indeed often only treatment means and standard deviations at each time point are available in our experience. We have shown how these can be used to calculate raw mean differences and their variances, which can be considered as an approximation to the model-based estimates and their variances. Then essentially the same modeling framework can be used as we described for model-based estimates and their variances. Again, a range of correlation coefficients should be investigated in a sensitivity analysis. It is inadvisable to ignore the correlation between repeated observations and it should not be treated as zero without good reason. In our examples the assumption of zero correlation did not change the clinical conclusions and that in itself is an important finding, but it is not inconceivable that in other situations clinical conclusions may differ depending on the correlation.

Regardless of whether IPD are available, review authors need to consider carefully whether to fit

time as a factor or a continuous variable within the longitudinal models. We have considered both approaches in this model. However, the models presented for time as a continuous variable assume a linear effect of time. If this is an inappropriate assumption, then treating time as a factor will be preferable. Alternatively, other models could be considered, but as they become more complicated this will necessitate IPD.

When a meta-analysis of longitudinal data is carried out, it would be preferable if all the studies were of the same duration and each study measured and reported results at the same time points. In reality this is unlikely to happen, and so to use the models presented in this article the time points need to be grouped together, preferably based on clinical judgment and prespecified in the review protocol. If analyses are done *post-hoc*, then this should be reported as such with sensitivity analyses to the choice of groupings. Furthermore, when assessing multiple longitudinal data studies, it is perhaps inevitable that missing data will be an issue. The models presented assume that any missing data is missing at random, both at the patient-level and the study-level. It is clearly important for practitioners to clarify the type and amount of missing data in their studies, and consider the potential impact of missing data on their conclusions. For example, they should be aware that trials may only present time points showing a significant treatment difference, a problem known as within-study selective reporting [21].

All the models in this article assume fixed treatment effects across studies, and we found no evidence of between-study heterogeneity in treatment effect across the five trials in the Alzheimer's example. This assumption may not always be appropriate, but there are a number of important issues to consider before extension to random treatment effects. For example, should there be a between-study heterogeneity parameter at each time-point, and should there be a separate between-study correlation between each pair of time-points? Estimation of the between-study correlation may also be difficult in some situations [22]. Ishak *et al.* propose aggregate data models for meta-analysis of longitudinal data including random-effects, but note that correlation parameters were difficult to estimate [4]. This may in part be due to unavailable within-study correlations in their example. Future work should thus extend our IPD and aggregate data methods to include a random treatment effect across studies [23], and indeed develop meta-regression models to assess the impact of study-level covariates [24].

A Bayesian approach to our models is also possible [25] and would allow prior information

about the correlation coefficients to be included [26]. A further issue, which has not been considered in this article is how to perform a meta-analysis when some studies provide IPD and others only provide aggregate data [27]. Goldstein *et al* suggest using a multi-level approach to this problem [28], and this approach has recently been extended to multiple outcomes [29].

Acknowledgments

We thank the trialists of the selegiline studies who kindly provided the MMSE individual patient data. Whilst undertaking this work Richard Riley was funded by the UK Department of Health as a Research Scientist in Evidence Synthesis.

References

1. Diggle PJ. (ed). *Analysis of Longitudinal Data* (2nd edn). Oxford University Press, Oxford, 2002.
2. DuMouchel W. Repeated measures meta-analysis. *Bulletin of the International Statistical Institute Session 51, Tome LVII 1997*; Book 1: 285–88.
3. Pham B, Chin W, Miller B, Rocchi A. Repeated measures meta-analysis of clinical trials. Poster A10 6th International Cochrane Colloquium 1998.
4. Ishak KJ, Platt RW, Joseph L *et al*. Meta-analysis of Longitudinal Studies. *Clinical Trials* 2007; 4: 525–39.
5. Riley RD, Abrams KR, Lambert PC *et al*. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Statistics in Medicine* 2007; 26: 78–97.
6. Ishak KJ, Platt RW, Joseph L, Hanley JA. Impact of approximating or ignoring within-study covariances in multivariate meta-analyses. *Statistics in Medicine* 2008; 27: 670–86.
7. Simmonds MC, Higgins JPT, Stewart LA *et al*. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clinical Trials* 2005; 2: 209–17.
8. The Cochrane Library. Wiley, Chichester, Vol. 3, 2005.
9. Wolfinger R. Covariance structure selection in general mixed models. *Communications in Statistics, Simulation, and Computation* 1993; 22(4): 1079–106.
10. Littell RC, Prendergast J, Natarajan R. Modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine* 2000; 19: 1793–819.
11. SAS. *Version 9.1 for Windows*. SAS Institute Inc, Cary, NC, 2002–2003. p. 29.
12. Satterthwaite FF. Synthesis of variance. *Psychometrika* 1941; 6: 309–16.
13. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 1997; 53: 983–97.
14. Berkey CS, Anderson JJ, Hoaglin DC. Multiple-outcome meta-analysis of clinical trials. *Statistics in Medicine* 1996; 15: 537–57.
15. Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in metaanalysis: multivariate approach and meta-regression. *Statistics in Medicine* 2002; 21: 589–624.
16. Patel HI. Analysis of incomplete data from a clinical trial with repeated measurements. *Biometrika* 1991; 78: 609–19.
17. Dear KB. Iterative generalized least squares for meta-analysis of survival data at multiple times. *Biometrics* 1994; 50: 989–1002.
18. Whitehead A. *Meta-Analysis of Controlled Clinical Trials*. Wiley, Chichester, 2002.
19. Wilcock GK, Birks J, Whitehead A, Evans JG. The effect of selegiline in the treatment of people with Alzheimer's disease: a meta-analysis of published trials. *International Journal of Geriatric Psychiatry* 2002; 17: 175–83.
20. Dunlop DD. Regression for longitudinal data: A bridge from least squares. *The American Statistician* 1994; 48(4): 299–303.
21. Hutton JL, Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. *Applied Statistics* 2000; 49: 359–70.
22. Riley RD, Abrams KR, Sutton AJ *et al*. Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Methodology Research* 2007; 7: 3.
23. Lopes HF, Muller P, Rosner GL. Bayesian meta-analysis for longitudinal data models using multivariate mixture priors. *Biometrics* 2003; 59: 66–75.
24. Berkey CS, Hoaglin DC, Antczak-Bouckoms A *et al*. Metaanalysis of multiple outcomes by regression with random effects. *Statistics in Medicine* 1998; 17: 2537–50.
25. Riley RD, Abrams KR, Lambert PC *et al*. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Statistics in Medicine* 2007; 26: 78–97.
26. Nam IS, Mengersen K, Garthwaite P. Multivariate meta-analysis. *Statistics in Medicine* 22(14): 2309–33.
27. Riley RD, Look MP, Simmonds MC. Combining individual patient data and aggregate data in evidence synthesis: a systematic review identified current practice and possible methods. *Journal of Clinical Epidemiology* 2007; 60(5): 431–9.
28. Goldstein H, Yang M, Omar RZ *et al*. Meta-analysis using multilevel models with an application to the study of class size effects. *Applied Statistics* 2000; 49: 399–412.
29. Riley RD, Lambert PC, Staessen JA *et al*. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Statistics in Medicine* 2008; 27(11): 1870–93.