Journal of Evaluation in Clinical Practice ISSN 1356-1294

Meta-analysis of repeated measures study designs

Jaime L. Peters PhD[1] and Kerrie L. Mengersen PhD[2]

[1]Research Associate, School of Mathematical Sciences, Queensland University of Technology, Australia
[2]Research Professor, School of Mathematical Sciences, Queensland University of Technology, Australia

Keywords
aggregate data, follow-up, longitudinal, meta-analysis, repeated measures

Correspondence
Kerrie Mengersen
School of Mathematical Sciences
Queensland University of Technology
GPO Box 2434
Brisbane Qld 4001
Australia
E-mail: k.mengersen@qut.edu.au

Accepted for publication: 10 March 2008

doi:10.1111/j.1365-2753.2008.01010.x

Abstract

Rationale, aims and objectives Repeated measures studies are found in many areas of research, particularly in areas of healthcare. There is currently little information available to inform the method of meta-analysis of repeated measures studies so that the structural dependence of the data is appropriately accommodated and the findings are meaningful.
Method Using a published meta-analysis on the impact of diet advice on weight reduction of obese or overweight individuals, we demonstrate possible approaches for repeated measures meta-analysis. These approaches differ in terms of the type of result obtained (e.g. effect at a particular time-point, trend over time, change between time-points) and the data needed for the analysis (e.g. means, regression slope estimates). Some approaches involve violating assumptions of independence in the data structure and so to investigate the impact of this violation a simulation study is carried out.
Results The different approaches described for the meta-analyses of repeated measures studies can all provide useful effect estimates depending on the question to be addressed by the meta-analysis. However, violation of the independence assumption in some approaches can lead to biased estimates.
Conclusions In practice, the methods available to carry out meta-analyses of repeated measures studies will not only depend upon the question of interest, but also on the data available from the primary studies.

Introduction

Repeated measures studies are designed to record measurements or observations of a unit, say an individual or site, at a number of time-points in order to assess follow-up, trend or change over time. This study design is used in diverse disciplines, for instance to model and analyse forest fuels at different sites [1], the health status of animals in toxicology experiments [2,3], depression in children [4], the short or long-term effects of smoking cessation interventions [5] and bone mineral density measurements with ongoing increased milk intake [6]. However, the analysis of this type of design is not straightforward as the unit of analysis is not the observation, *per se*, but the unit on which the observations are made, for instance the individual (human or animal) in health studies or the study site in ecology. Thus the temporal non-independence between measurements must be considered as the same individuals, or sites, are being measured at each time-point.

There are a number of approaches to the primary analysis of this type of data. These range from (i) assessing the effect at one particular time-point only; (ii) calculating and assessing summary statistics of the individual measurements at each time-point, ignoring the temporal dependence; and (iii) using more complex analyses such as multivariate analysis of variance or multilevel models including growth and growth-mixture models [7]. Unlike the first two approaches, these more sophisticated models can account for the dependencies induced by the study design [8].

In the past 20 or so years, there has been an explosion of primary analyses in all disciplines. To help inform policy and decision making on a specific question, this research has to be identified, reviewed, combined and critically assessed. Secondary analyses in the form of systematic reviews and meta-analyses have been commonly used in psychology and education [9], and in medicine [10,11] to allow formal evaluation of the relevant evidence for a particular question of interest, but are increasingly being adopted in other areas of research [12]. Meta-analysis allows for a quantitative summary of the evidence from multiple studies about a measurable parameter (e.g. mean, correlation) of interest [10]. Advantages include greater statistical power than in a single study, the potential to be more generalizable, as well as the potential for more precise estimates [13,14]. Meta-analyses also allow a framework for investigation of publication and other biases, and for possible sources of heterogeneity between studies [13,14].

Although standard techniques are available for the meta-analysis of most types of studies, to date there has been little guidance available to researchers for the meta-analysis of repeated measures. Although examples of repeated measures meta-analysis

exist, they often assume that the raw data from each study, the individual participant data (IPD), are available [11,15,16], or only compare two time-points, for example, pre- versus post-treatment [4,17]. Meta-analysis of IPD is considered the 'gold standard' [18]. When the IPD are available a one- or two-stage approach could be taken to the meta-analysis of repeated measures studies [19]. The two-stage approach involves analysing each individual study and then combining the summary estimates as in a usual meta-analysis. The one-stage approach includes all the IPD in the synthesis model with study as an identifier [19]. However, it is often very difficult to obtain all of the IPD for a meta-analysis and properly account for study-specific issues such as design, missing data, covariates, biases and confounders. Hence, a meta-analysis of IPD is rarely feasible [20]. The case where only two time-points are considered in a repeated measures meta-analysis requires a more simple approach. In this situation, consideration of the difference between an individual's measurements at the two time-points will lead to a single summary effect (e.g. mean difference), which can be combined using standard meta-analysis methods. Often, however, more than two time-points are measured in a study, in which case this pairwise difference approach requires adjustment for multiple testing and does not completely address the desired inferential questions.

The Cochrane Handbook [21] makes a number of recommendations for the meta-analysis of repeated measures data, including the use of IPD and assessment of one particular time-point. Other suggestions are to calculate and combine a summary effect for each individual across time (e.g. the mean effect or some measure of trend over time), perform separate analyses at each time-point or select and meta-analyse results for just the final time-point in each study. It may be unlikely that all primary studies analyse and present results in a standard way, so there needs to be some consideration of how to deal with results that are not reported in the same format. The choice of method will depend on the question of interest (i.e. an outcome at a particular time-point or trend over time) and the data available from the primary studies. In the absence of guidance on this, the aim of this paper is to outline and illustrate a number of possible approaches. We discuss the considerations to be made and limitations of the approaches in addition to areas for further work. To reflect the most likely scenarios for meta-analysis of repeated measures studies, in the rest of this paper, we assume that the IPD is not available and that there are more than two time-points to be considered. We now introduce the illustrative meta-analysis used in this paper.

Methods

Illustrative meta-analysis

Pirozzo *et al*. [22] report a Cochrane systematic review and meta-analysis of the impact of advice about low-fat diets on the weight reduction of obese or overweight individuals compared with other weight-loss interventions. Six randomized controlled trials were included in the meta-analysis, each of them having at least 6-month follow-up from the intervention. The main outcome of interest was the difference in weight loss between subjects given low-fat diet advice and control subjects at 6, 12 and 18 months follow-up: a negative value for the mean difference indicates more weight loss in the treatment group, whereas a positive mean difference indicates more weight loss in the control group.

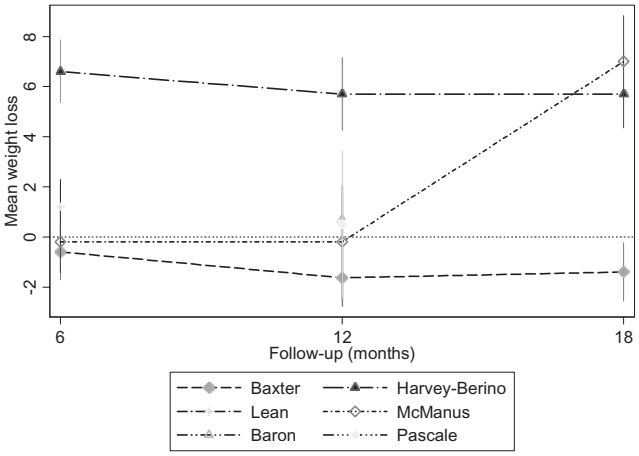

Figure 1 Mean difference in weight loss between treatment and control group from the Pirozzo *et al.* meta-analysis (vertical lines indicate ± standard errors of each estimate).

Only three of the six studies measured and reported weight loss at all three time-points [23–27]; one study only reported weight loss at 6 months [28] and the remaining two studies only reported weight loss at 12 months [29,30]. The reported mean differences at each time-point for the six studies are shown in Fig. 1. Pirozzo *et al*. [22] report observing between-study heterogeneity and use random effects inverse-variance meta-analysis models [31] to combine the mean difference in weight loss between treated and control subjects at each time-point. The reported pooled mean differences [and 95% confidence intervals (CIs)] at the different follow-up times are as follows where n is the number of studies contributing to the pooled effect: at 6 months ($n = 4$): 1.72 (−1.39, 4.83); at 12 months ($n = 5$): 1.06 (−1.62, 3.75); at 18 months ($n = 3$): 3.66 (−1.84, 9.15).

Meta-analysis methods and models

In this section, a number of approaches for the meta-analysis of repeated measures studies are described. We also consider the type of question the different methods attempt to answer and the format of the data needed to apply each approach. We present the results of frequentist meta-analysis and, where more appropriate, Bayesian meta-analysis models to combine the evidence. Given the heterogeneity between the studies identified by Pirozzo *et al*., random effect models are described here. The frequentist meta-analyses are implemented in Stata [32], while the Bayesian models are estimated in WinBUGS [33].

Meta-analysis models

Recall that a random effects model, where y_i is the effect estimate in study i and σ_i^2 is its associated variance, may be written by

$$y_i = \theta_i + \varepsilon_i; \quad \varepsilon_i \sim N(0, \sigma_i^2); \quad \theta_i \sim N(\mu, \tau^2)$$
$$\mathrm{var}(y_i) = \sigma_i^2 + \tau^2 \qquad (1)$$

where θ_i is the true effect in study i, τ^2 represents the estimate of between-study heterogeneity and μ is the pooled effect [10]. The

amount of between-study heterogeneity can be investigated using I^2, which assesses the percentage of total variation across studies that is due to between-study heterogeneity rather than chance [34].

In a Bayesian context, prior distributions are placed on the overall effect, μ, and the estimate of between-study variance, τ^2 [35]. We first consider vague prior distributions, so that a Bayesian random effects meta-analysis model can be given by

$$\begin{array}{ll} y_i \sim N(\theta_i, \sigma_i^2) & \mu \sim N(0,100000) \\ \theta_i \sim N(\mu, \tau^2) & \tau \sim N(0,100), \tau > 0 \end{array}. \quad (2)$$

The sensitivity of these prior distributions is assessed in a later section. The WinBUGs code used in this paper is available in the Appendix.

Methods of repeated measures meta-analysis

Five methods for the meta-analysis of repeated measures studies are described in order of the type of outcome required: effect at a particular time-point, trend over time or change between time-points. The methods depend on the type of data available from the primary studies and the aim of the meta-analysis. A summary of the aims, models, data requirements and assumptions of each of these five methods is given in Table 1. We assume that the available data are all presented in the same format from each primary study, however, that may not necessarily be the case and we return to this point in the discussion.

The first method we describe involves the calculation and combination of summary estimates from primary studies at a particular time-point and is referred to as the *relevant time-point meta-analysis* (RTM). This time-point may be chosen for its clinical importance (and will have been defined before any analyses were carried out), or as the time-point where most data are available from the primary studies (although the potential for introducing bias by defining the relevant time-point after the data are assessed should be kept in mind [21]). We assume that each study reports the same type of effect outcome at this time-point, e.g. a mean effect and some measure of its precision.

The second approach calculates and combines the effect in each study at either the first or final time-point, regardless of whether this time-point is different across studies. Thus, any studies that were excluded from the RTM because they did not report at that particular time-point will be included in this *first/final time-point meta-analysis* (FTM). However, the interpretation and clinical importance of the resultant pooled estimate may be limited as the primary studies may have different first/final time-points. Examples of this approach can be seen in Spittle *et al*. [36] and Cahill *et al*. [37] who both combine study estimates at the final time-point.

These above two methods only consider one time-point. The third approach described applies the RTM to all time-points and is referred to as the *all time-points meta-analysis* (ATM). There are many examples of this approach in the literature (see Franz *et al*. [38], Tan *et al*. [39] and Dansinger *et al*. [40]). The pooled estimate at a time-point can be qualitatively or quantitatively compared with estimates at other time-points (as demonstrated later). However, assumptions of independence between the data points may not be valid as some subjects contribute data to pooled estimates at more than one time-point, while others only contribute to one time-point. Furthermore, this approach may lead to a trend being observed at the aggregate level, which may not reflect such a trend at the individual level. This is known as the ecological fallacy [41].

The fourth method we describe allows some analysis of trend over time at the study level, as well as at the population level. In this approach, regression is used to model the outcome over time and provide a summary of the trend, so we refer to this as the *trend meta-analysis* (TM). Various approaches can be taken depending on the form of the available data. If the primary studies have used multilevel modelling and report a slope estimate for trend, these can be combined using the random effects meta-analysis models defined above where y_i is the slope estimate from study i (and σ_i^2 is the associated variance of this estimate). This analysis would maintain the time dependencies within each study. If, as in the Pirozzo *et al*. example, only the means and variances are available at each time-point a study-specific regression analysis can be undertaken on these means and the resultant slope estimates combined across studies. To carry out this analysis, the Bayesian model given in (2) could be extended as follows:

$$\begin{array}{ll} y_{it} \sim N(\theta_{it}, \sigma_{it}^2) & \alpha_i \sim N(0,100000) \\ \theta_{it} = \alpha_i + \beta_i time_{it} & \mu \sim N(0,100000) \\ \beta_i \sim N(\mu, \tau^2) & \tau \sim N(0,100), \tau > 0 \end{array} \quad (3)$$

where y_{it} is the observed effect at time-point t in study i, σ_{it}^2 is its variance, θ_{it} is the true effect at time-point t in study i, α_i and β_i are, respectively, the intercept and slope for the study-specific linear regression slope, $time_{it}$ is the time-point in study i at time t (= 0, 6, 12, 18 months in the Pirozzo *et al*. example), μ is the overall trend and τ^2 is the variance in trend between studies.

This model allows inclusion of studies that do not contribute to the trend analysis, but do provide estimates at certain time-points. This 'borrowing of strength' is one advantage of the Bayesian approach to evidence synthesis [42].

In the final approach, we apply a method that is relevant when the change between time-points is of interest, and so refer to this method as the *change in time meta-analysis* (CTM). Once the difference between estimates at different time-points in a primary study have been obtained, the approach follows that given earlier for the ATM; the difference in means in the primary studies are combined at each time-point. This could be done in two ways: where the difference between each successive time-point is calculated and combined (e.g. Burke *et al*. [43]), or where the difference from baseline to each time-point is calculated (e.g. Earl and Albarracín [44]).

Simulation study

As pointed out in the introduction the correct analysis of repeated measures studies is complicated by the unit of analysis being the individual, say, rather than the observation taken on that individual. Some primary analyses ignore this structure, so assumptions of independence are violated. We simulate repeated measures IPD for a meta-analysis to demonstrate the impact of assuming independence across time-points as the dependencies increase over time at the individual level. Three simple meta-analyses are simulated, each containing five studies with each study reporting some continuous measurements on 20 individuals at four time-points. For simplicity all time-points are the same across the five studies and there is no missing data. A linear trend was assumed

Table 1 Summary of repeated measures meta-analysis models, their aims, data requirements and assumptions

Method	Aim	Data required from primary studies	Model*	Assumptions/considerations
Relevant time-point meta-analysis (RTM)	To assess evidence at one particular time-point	Summary effect (e.g. mean, slope estimate) and some measure of variance at time-point of interest	For every study i, y_i is summary estimate with variance σ_i^2 $y_i \sim N(\theta_i, \sigma_i^2)\ \mu \sim [-,-]$ $\theta_i \sim N(\mu, \tau^2)\quad \tau \sim [-,-]$	Only interested in that one particular time-point, not trend
First/final time-point meta-analysis (FTM)	To assess evidence at first or final time-point for each study	Summary effect (e.g. mean, slope estimate) and some measure of variance at first or final time-point	For every study i, y_i is summary estimate with variance σ_i^2 $y_i \sim N(\theta_i, \sigma_i^2)\ \mu \sim [-,-]$ $\theta_i \sim N(\mu, \tau^2)\quad \tau \sim [-,-]$	Only interested in first or final time-point in each study. Limits interpretation as first or final time-points may be different in each study.
All time-points meta-analysis (ATM)	To assess evidence at every time-point reported by the primary studies	Summary effect (e.g. mean, slope estimate) and some measure of variance at each time-point	For every study i and time-point t, y_{it} is summary estimate with variance σ_{it}^2 $y_{it} \sim N(\theta_{it}, \sigma_{it}^2)\ \mu \sim [-,-]$ $\theta_{it} \sim N(\mu_t, \tau_t^2)\ \tau \sim [-,-]$	Independence of time-points assumed. Overlap of time-points in primary studies.
Trend meta-analysis (TM)	To investigate any trend over time	Slope estimate and some measure of variance calculated within primary study	For every study i, y_i is slope estimate with variance σ_i^2 $y_i \sim N(\theta_i, \sigma_i^2)\ \mu \sim [-,-]$ $\theta_i \sim N(\mu, \tau^2)\quad \tau \sim [-,-]$	Assuming the primary analyses are carried out correctly, this method maintains the temporal dependencies
		Summary effect (e.g. mean, slope estimate) and some measure of variance at each time-point	For every study i and time-point t, y_{it} is summary estimate with variance σ_{it}^2 $y_{it} \sim N(\theta_{it}, \sigma_{it}^2)\ \alpha_i \sim [-,-]$ $\theta_{it} = \alpha_i + \beta_i time_{it}\quad \mu \sim [-,-]$ $\beta_i \sim N(\mu, \tau^2)\quad \tau \sim [-,-]$	Violates the assumption of independence between time-points
Change in time meta-analysis (CTM)	To investigate the change between consecutive time-points	Summary effect (e.g. mean, slope estimate) and some measure of variance at each time-point	For every study i, y_i is difference in summary estimate with variance σ_i^2 $y_i \sim N(\theta_i, \sigma_i^2)\ \mu \sim [-,-]$ $\theta_i \sim N(\mu, \tau^2)\quad \tau \sim [-,-]$	Independence of time-points assumed. Overlap of time-points in primary studies.

*[–,–] indicates a prior distribution is to be placed on the parameter, when a Bayesian meta-analysis model is implemented.

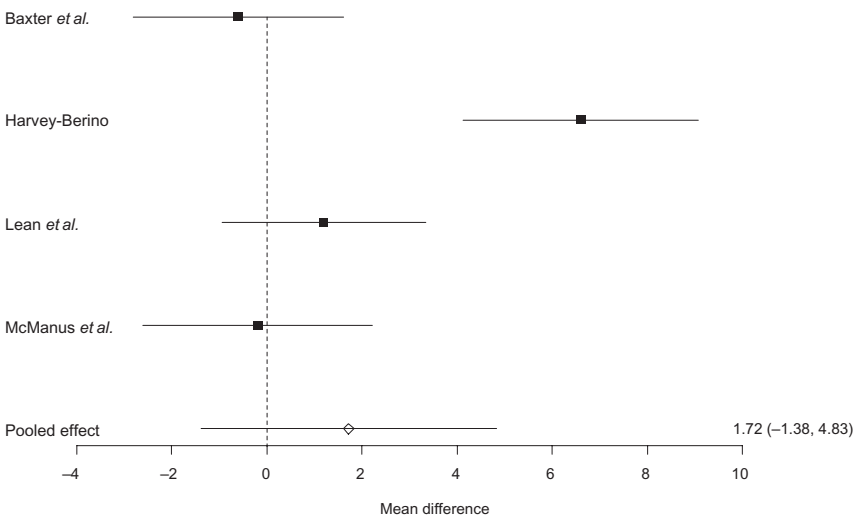

Figure 2 Observed and pooled estimates in *relevant time-point meta-analysis* (RTM).

for each individual and data were simulated using a random effects model. The three meta-analyses differ in how well the data fit the linear regression model, i.e. the variability around the regression line for each of the meta-analyses. These simulated datasets are meta-analysed in three ways: by calculating the slope estimate from the pooled means at each time-point (i.e. using the ATM approach), by calculating and combining the study-specific slope estimates based on the means at each time-point in each study (i.e. using the TM approach), and by carrying out an IPD analysis using a 3-level hierarchical model.

Results

RTM

The aim of this approach is to combine the available evidence at one particular time-point. Pirozzo *et al*. [22] did not define a clinically important time-point but for illustration we define 6 months as the time-point of interest. This choice has the immediate effect of excluding two of the six studies from the meta-analysis as they did not report at 6 months. The mean differences in weight loss between treated and control groups reported in each study at 6 months were combined using equation (1). The mean differences from each study as reported in Pirozzo *et al*. and the overall mean and associated 95% CIs are shown in Fig. 2. The pooled estimate at the 6 month time-point is 1.72 (95% CI: −1.38, 4.83), the wide 95% CI reflects the variability between study effects and the fact that only four estimates are being combined. In this meta-analysis, I^2 is estimated to be 86% (95%CI: 66, 94), indicating large between-study heterogeneity [34].

FTM

We now consider a meta-analysis of the final time-point with the Pirozzo *et al*. example. Six months is the final time-point for one study, 12 months for two studies and 18 months for the remaining three studies. The observed study effects and pooled effect, μ, are shown in Fig. 3. The pooled effect from this meta-analysis is slightly larger and less variable than that of the previous meta-analysis (Fig. 2).

As mentioned, the studies have different final time-points and although the statistical heterogeneity may be somewhat smaller (even though it is still high) in this meta-analysis ($I^2 = 80\%$; 95% CI: 56, 91) compared with the RTM, the clinical heterogeneity may also be quite large making the pooled value difficult to interpret.

ATM

This method in which each of the time-points reported in the primary studies is considered separately is the approach reported by Pirozzo *et al*. in their meta-analysis. Our re-analysis of the data at each time-point using the frequentist model in (1) is shown in Fig 4.

The pooled estimates have a great deal of uncertainty associated with them owing to the small number of studies in these meta-analyses and the variability between the studies; the I^2 estimate ranges from 75% to 91% for the three pooled effects. The results of the Bayesian meta-analysis model [equation (2)] are very similar, except that, as expected, there is more variability associated with the Bayesian pooled estimates: 1.72 [95% credibility interval (CrI): −4.07, 7.59] at 6 months; 1.04 (−3.18, 5.24) at 12 months; 3.63 (−5.79, 13.26) at 18 months.

A quantitative comparison of these estimates was undertaken by fitting a regression model to the pooled estimates at each time-point. In Fig. 5, the pooled estimates from the ATM are plotted alongside the slope estimate from an unweighted linear regression model which has been forced through the origin (as there is no difference in weight loss at 0 months). The slope estimate based on these three pooled estimates from the frequentist analysis is 0.18 (0.07, 0.28) suggesting an increase over time in the difference in weight loss between controls and treated subjects, i.e. with controls losing more weight than treated subjects as time goes on. The slope estimate from the Bayesian analysis is very similar: 0.18 (0.09, 0.26).

If the intercept is allowed to be non-zero the slope estimate (95% CI) from a frequentist analysis suggest that there is little

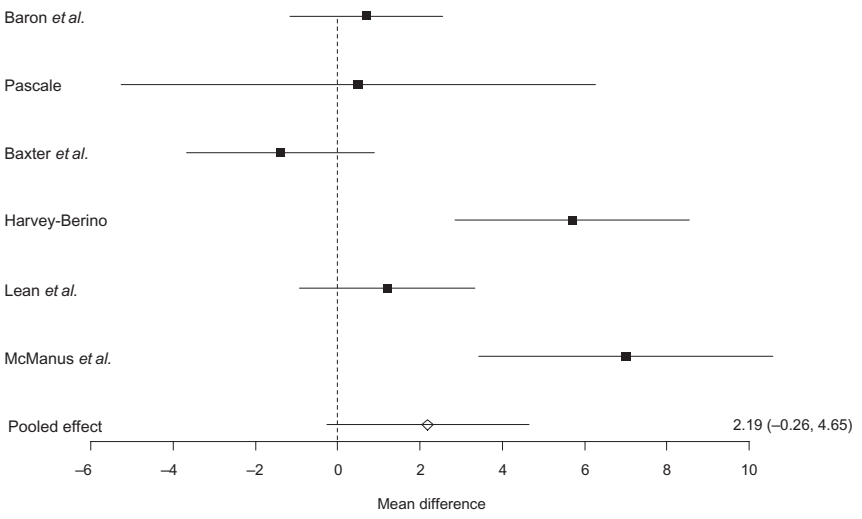

Figure 3 Observed, shrunken and pooled estimates in *final time-point meta-analysis* (FTM).

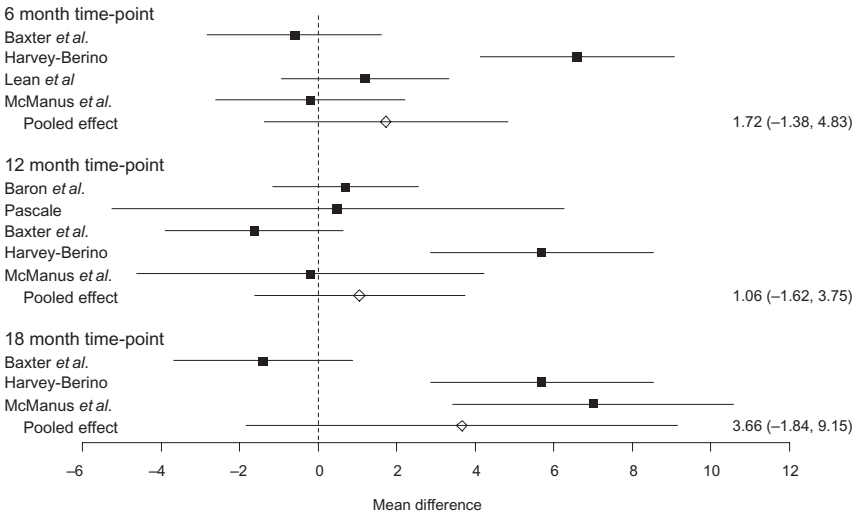

Figure 4 Observed and pooled estimates at (1) 6 months, (2) 12 months, and (3) 18 months follow-up for the all-time points meta-analysis (ATM).

evidence of a time trend: 0.17 (–0.13, 0.47). (The results of a Bayesian analysis are similar to this.)

TM

The TM allows assessment of trend at the study level, as well as at the population level. For the Pirozzo *et al*. meta-analysis, only the means and variances at each time-point are available, so equation (3) is used to calculate and combine study-specific slope estimates to assess trend over time where the slopes are forced through the origin. The study-specific estimates are shown in Fig. 6, with a pooled slope estimate (95% CrI), μ, of 0.15 (–0.12, 0.42) overlaid (the shaded area). From this analysis of the data, the probability that the slope is greater than zero, $P(\mu > 0)$, is 0.90.

CTM

The CTM focuses on the change between estimates at successive time-points. Here we again define the mean difference in weight

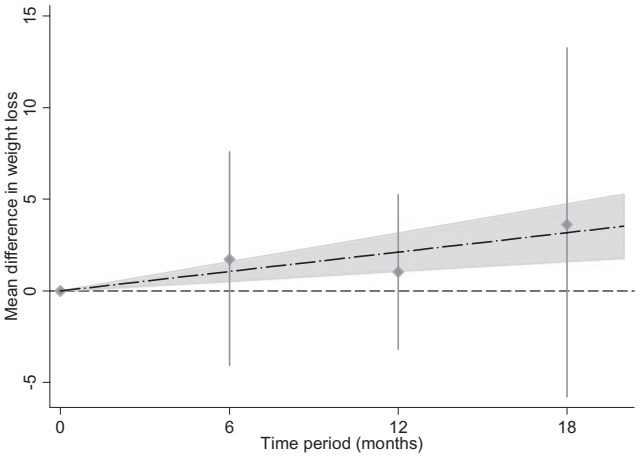

Figure 5 Pooled estimates at each time-point and associated regression slope from a Bayesian analysis of the data.

loss between control and treated subjects at 0 months to be zero and consider four different values for the standard deviation at time zero (SD_0) in each primary study: $SD_0 = 0.001$; $SD_0 = 1$; $SD_0 = \max(SD_t;\ t = 1, \ldots, 3)$; $SD_0 = 1.2\max(SD_t;\ t = 1, \ldots, 3)$. The different values had little impact on the results of the meta-analyses, so we report here the results where $SD_0 = 0.001$.

Only three studies contribute to this analysis. The results of a frequentist analysis suggest that there is greater difference between the mean weight loss of the control and treated groups between 12 and 18 months with controls losing more weight (see Table 2). As can be seen, however, there is a great deal of between-study heterogeneity.

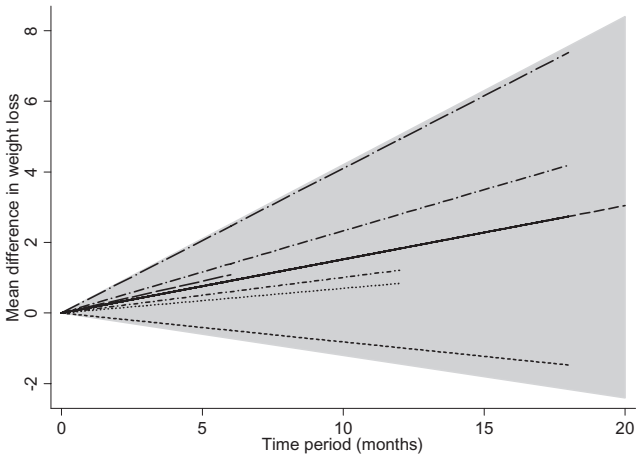

Figure 6 Study-specific (dashed) and pooled slope (bold) estimates with 95% CrI displayed. CrI, credibility interval.

Table 2 Change in effect at different time-points for the Pirozzo *et al.* meta-analysis

Change-point	Median estimate (95% CrI)	I^2
1 ($t_1 - t_0$)	1.93 (–2.66, 6.53)	100%
2 ($t_2 - t_1$)	–0.68 (–1.25, –0.10)	81%
3 ($t_3 - t_2$)	2.41 (–1.56, 6.38)	100%

CrI, credibility interval.

Violating the independence assumption

In the analysis of the Pirozzo *et al*. meta-analysis, the Bayesian slope estimate based on the pooled means at each time-point (from the ATM) is slightly larger and much more precise (0.18; 95% CrI: 0.09, 0.26) than the Bayesian slope estimate based on the pooled slopes across studies (calculated in the TM) (0.15; 95% CrI –0.12, 0.42). This trend is also seen in the results of the simulation study. Table 3 shows that the slopes based on the pooled estimates at each time-point (ATM approach) are the most precise estimates. The precision of the slopes based on the study-specific slope estimates (TM approach) are of a similar magnitude to the precision of the IPD calculated pooled slopes, although there is greater variability in the estimate from the IPD meta-analysis. These results suggest that violating the independence assumption when meta-analysing repeated measures data can lead to estimates of effect that appear to be more precise than the whole evidence suggests. With this in mind, it would be more appropriate to calculate a slope (if that is the outcome of interest) by pooling the study-specific slope estimates based on the mean measurements at each time-point within the individual studies, rather than pooling data across studies at each time-point and then calculating the slope estimate. In other words, estimating a trend based on a TM is preferred to calculating a trend based on an ATM.

Sensitivity of prior distributions in Bayesian analysis

For the TM approach described in this paper, the sensitivity of the prior distribution placed on τ^2 in equation (2) was assessed. The following three alternative prior distributions were investigated:

$$\frac{1}{\tau^2} \sim gamma(0.01, 0.01)$$

Table 3 Application of different repeated measures Bayesian meta-analysis methods to the simulated datasets

Repeated measures meta-analysis method	Mean (standard deviation) slope estimate	Median (95% CrI) slope estimate	Estimate of between-study heterogeneity
Dataset 1			
Slope estimate based on results of ATM	1.36 (0.29)	1.36 (0.79, 1.93)	NA*
Slope estimate from TM	1.38 (0.44)	1.37 (0.58, 2.20)	0.15
Individual participant analysis	1.40 (0.44)	1.40 (0.55, 2.26)	0.42
Dataset 2			
Slope estimate based on results of ATM	1.33 (0.26)	1.33 (0.82, 1.84)	NA*
Slope estimate from TM	1.36 (0.41)	1.35 (0.61, 2.15)	0.15
Individual participant analysis	1.39 (0.43)	1.39 (0.57, 2.23)	0.31
Dataset 3			
Slope estimate based on results of ATM	1.31 (0.25)	1.31 (0.83, 1.80)	NA*
Slope estimate from TM	1.34 (0.39)	1.33 (0.63, 2.11)	0.15
Individual participant analysis	1.38 (0.42)	1.38 (0.58, 2.20)	0.30

ATM, all time-points meta-analysis; CrI, credibility interval; TM, trend meta-analysis.
*An estimate of between-study heterogeneity cannot be obtained with this analysis as the slope is calculated from the pooled means at each time-points (between-study heterogeneity at each time-point can, however, be obtained).

$$\frac{1}{\tau^2} \sim gamma(0.1, 0.1)$$

$$\tau \sim U(0, 100)$$

The posterior means of the pooled effect were unchanged for these different prior distributions (results not shown). However, the half normal prior (as used in the original analysis) and the uniform prior gave estimates of the between-study heterogeneity parameter, τ^2, that were greater than those obtained from either of the gamma prior distributions. Not surprisingly, these larger estimates of τ^2 lead to wider posterior intervals around the pooled estimates for the half normal and uniform prior distributions. The sensitivity of the between-study heterogeneity estimate is expected given the relatively few data in this meta-analysis and the consequent influence of the prior. Nevertheless, the general conclusions are robust to changes in the prior distribution placed on τ^2.

Discussion

In this paper, we have examined a number of approaches for the meta-analysis of repeated measures studies. These approaches differ in terms of the type of result obtained (e.g. effect at a particular time-point, trend over time, change between time-points) and the data needed for the analysis (e.g. IPD, means, slope estimates). These differences and the specific research question of the meta-analysis will ultimately lead to the choice of method to use. In an ideal world, one would want to use the IPD from each study. Difficulties in obtaining the full IPD for a meta-analysis limit the applicability and generality of analyses; if IPD and the more commonly seen 'aggregate data' from studies are available, these can also be combined [45].

The methods described in this paper can be generalized to other modelling scenarios, for example, fixed effects rather than random effects meta-analysis models and more complex hierarchical meta-analysis models.

A number of the meta-analysis approaches described in this paper use the observations as the unit of analysis, not the individuals. This is also a criticism often levelled at analyses of primary repeated measures studies [46]. Often the data regarding the individuals is not available for a meta-analysis, so it is difficult to carry out the meta-analysis with the individual as the unit of analysis. Hence, this assumption may often be violated, which may lead to deflated variance estimates and consequently biased pooled mean estimates.

In the illustrative meta-analysis of Pirozzo *et al*. [22], only six studies were included, and for some of the meta-analysis approaches only three studies were applicable. It is difficult to say how many studies are too few for a meta-analysis, although a number of authors have suggested a minimum of 10, particularly for assessment of between-study heterogeneity or publication bias [21,47]. In a Bayesian framework, the lack of data will mean that priors that are intended to be vague may actually be more informative than planned. With this in mind, the sensitivity of 'vague' priors should be assessed, especially when multilevel models are being considered [48]. Our sensitivity analysis indicated that the prior distributions placed on the between-study heterogeneity parameter in the TM approach, τ^2, could affect the estimate of between-study heterogeneity, but in this example, they had little impact on the pooled estimate obtained.

As the aim of this paper was to compare methods, it was assumed that the data from studies for a meta-analysis are all reported in the same style, for example, in Pirozzo *et al*. [22], means and standard deviations at the time-points were reported. In general, different studies might report different types of summary estimates; for instance, some studies may model the time trend using one particular model, while another study may use a different model. The question of comparability of study estimates obtained from different models is problem specific. Further work is required to examine how the different types of summary estimates from each repeated measures study could be meaningfully combined in a meta-analysis.

We have not considered the important issue of publication bias. If one summary estimate from each study is used for a meta-analysis, say a RTM, FTM or TM, then current methods for detecting and assessing publication bias [49] and between-study heterogeneity [34,41] may be used. A problem arises when more than one summary estimate from a study is included in the meta-analysis. In the Pirozzo *et al*. paper, only six studies were involved, so publication bias and between-study heterogeneity would be difficult to assess anyway. However, as techniques develop for the synthesis of repeated measures studies, these and other important aspects of meta-analysis should not be overlooked.

References

1. Graham, J. B. & McCarthy, B. C. (2006) Forest floor fuel dynamics in mixed-oak forests of south-eastern Ohio. *International Journal of Wildland Fire*, 15, 479–488.
2. Anning, P. B., Finney, S. J., Singh, S., Winlove, C. P. & Evans, T. W. (2004) Fluids reverse the early lipopolysaccharide-induced albumin leakage in rodent mesenteric venules. *Intensive Care Medicine*, 30, 1944–1949.
3. Tamura, R. N. & Buelke-Sam, J. (1992) The use of repeated measures analyses in developmental toxicology studies. *Neurotoxicology and Teratology*, 14, 205–210.
4. Twenge, J. M. & Nolen-Hoeksema, S. (2002) Age, gender, race, socioeconimc status, and birth cohort differences on the children's depression inventory: a meta-analysis. *Journal of Abnormal Psychology*, 111 (4), 578–588.
5. Colby, S. M., Monti, P. M., O'Leary Tevyaw, T., Barnett, N. P., Spirito, A., Rohsenow, D. J., Riggs, S. & Lewander, W. (2005) Brief motivational intervention for adolescent smokers in medical settings. *Addictive Behaviors*, 5, 865–874.
6. Cadogan, J., Eastell, R., Jones, N. & Barker, M. E. (1997) Milk intake and bone mineral acquisition in adolescent girls: randomised, controlled intervention trial. *British Medical Journal*, 315, 1255–1260.
7. Hoeksma, J. B. & Kelderman, H. (2006) On growth curves and mixture models. *Infant and Child Development*, 15, 627–634.
8. Goldstein, H., Browne, W. & Rasbash, J. (2002) Tutorial in biostatistics: multilevel modelling of medical data. *Statistics in Medicine*, 21, 3291–3315.
9. Hedges, L. V. & Olkin, I. (1985) Statistical Methods for Meta-Analysis. Orlando, FL: Academic Press, Inc.
10. Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A. & Song, F. (2000) Methods for Meta-Analysis in Medical Research. Chichester: Wiley.
11. Whitehead, A. (2002) Meta-analysis of Controlled Clinical Trials. Chichester: Wiley.
12. Petticrew, M. (2001) Systematic reviews from astronomy to zoology: myths and misconceptions. *British Medical Journal*, 322, 98–101.

13. Blettner, M., Sauerbrei, W., Shlehofer, B., Scheuchenpflug, T. & Friedenreich, C. (1999) Traditional reviews, meta-analyses and pooled analyses in epidemiology. *International Journal of Epidemiology*, 28, 1–9.
14. Fleiss, J. L. & Gross, A. J. (1991) Meta-analysis in epidemiology, with special reference to studies of the association between exposure to environmental tobacco smoke and lung cancer: a critique. *Journal of Clinical Epidemiology*, 44, 127–139.
15. Inoue, L. Y. T., Etzioni, R., Slate, E. H., Morrell, C. & Penson, D. F. (2004) Combining longitudinal studies of PSA. *Biostatistics*, 5 (3), 483–500.
16. Lopes, H. F., Müller, P. & Rosner, G. L. (2003) Bayesian meta-analysis for longitudinal data models using multivariate mixture priors. *Biometrics*, 59, 66–75.
17. Horowitz, J. L. & Garber, J. (2006) The prevention of depressive symptoms in children and adolescents: a meta-analytic review. *Journal of Consulting and Clinical Psychology*, 74 (3), 401–415.
18. Stewart, L. A. & Parmar, M. K. B. (1993) Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet*, 341, 418–422.
19. Simmonds, M. C., Higgins, J. P. T., Stewart, L. A., Tierney, J. F., Clarke, M. J. & Thompson, S. G. (2005) Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clinical Trials*, 2, 209–217.
20. Stewart, L. A. & Tierney, J. F. (2002) To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Evaluation and the Health Professions*, 25, 76–97.
21. Higgins, J. P. T. & Green, S. (2005) Cochrane Handbook for Systematic Reviews of Interventions. Version 4.2.5. [updated May 2005]. *The Cochrane Collaboration 3*. Available at: http://www.cochranehandbook.org (last accessed 5 June 2008).
22. Pirozzo, S., Summerbell, C., Cameron, C. & Glasziou, P. (2006) Advice on low-fat diets for obesity. *Cochrane Database of Systematic Reviews*, 2, CD003640.
23. Harvey-Berino, J. (1998) The efficacy of dietary fat vs total energy restriction for weight loss. *European Journal of Clinical Nutrition*, 51, 243–248.
24. Jeffery, R. W., Hellerstedt, W., French, S. & Baxter, J. (1995) A randomized trial of counseling for fat restriction versus calorie restriction in the treatment of obesity. *International Journal of Obesity*, 19, 132–137.
25. McManus, K., Antinoro, L. & Sacks, F. (2001) A randomized controlled trial of moderate-fat, low-energy diet compared with a low fat, low-energy diet for weight loss in overweight adults. *International Journal of Obesity*, 25, 1503–1511.
26. Shah, M., Baxter, J., McGovern, P. & Garg, A. (1996) Nutrient and food intake in obese women on a low-fat or low-calorie diet. *American Journal of Health Promotion*, 10, 179–182.
27. Shah, M., McGovern, P., French, S. & Baxter, J. (1994) Comparison of a low-fat, ad libitum complex-carbohydrate diet with a low-energy diet in moderately obese women. *American Journal of Clinical Nutrition*, 59, 980–984.
28. Lean, M. E. J., Han, T. S., Prvan, T., Richmond, P. R. & Avenell, A. (1997) Weight loss with high and low carbohydrate 1200 kcal diets in free living women. *European Journal of Clinical Nutrition*, 51, 243–248.
29. Baron, J. A., Schori, A., Crow, B., Carter, R. & Mann, J. I. (1986) A randomized controlled trial of low carbohydrate and low fat/high fibre diets for weight loss. *American Journal of Public Health*, 76 (11), 1293–1296.
30. Pascale, R. W., Wing, R. R., Butler, B. A., Mullen, M. & Bononi, P. (1995) Effects of behavioral weight loss program stressing calorie restriction versus calorie plus fat restriction in obese individuals with NIDDM or a family history of diabetes. *Diabetes Care*, 18, 1241–1248.
31. DerSimonian, R. & Laird, N. (1986) Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7, 177–188.
32. StataCorp (2004) Stata Statistical Software. Release 8.2. College Station, TX: Stata Corporation.
33. Spiegelhalter, D. J., Thomas, A., Best, N. G. & Lunn, D. (2003) WinBugs User Manual: Version 1.4. Cambridge: MRC Biostatistics Unit.
34. Higgins, J. P. T., Thompson, S. G., Deeks, J. J. & Altman, D. G. (2003) Measuring inconsistency in meta-analyses. *British Medical Journal*, 327, 557–560.
35. Spiegelhalter, D. J., Abrams, K. R. & Myles, J. P. (2004) Bayesian Approaches to Clinical Trials and Health-care Evaluation. Chichester: Wiley.
36. Spittle, A. J., Orton, J., Doyle, L. W. & Boyd, R. (2007) Early developmental intervention programs post hospital discharge to prevent motor and cognitive impairments in preterm infants. *Cochrane Database of Systematic Reviews*, 2, CD005495.
37. Cahill, K. & Ussher, M. (2007) Cannabinoid type I receptor antagonisis (rimonabant) for smoking cessation. *Cochrane Database of Systematic Reviews*, 4, CD005353.
38. Franz, M. J., Vanwormer, J. J., Crain, A. L., Boucher, J. L., Histon, T., Caplan, W., Bowman, J. D. & Pronk, N. P. (2007) Weight-loss outcomes: a systematic review and meta-analysis of weight-loss clinical trials with a minimum of 1-year follow-up. *Journal of the American Dietetic Association*, 107 (10), 1755–1767.
39. Tan, A., Liao, C., Mo, Z. & Cao, Y. (2007) Meta-analysis of holmium laser enucleation versus transurethral resection of the prostate for symptomatic prostatic obstruction. *British Journal of Surgery*, 94 (10), 1201–1208.
40. Dansinger, M. L., Tatsioni, A., Wong, J. B., Chung, M. & Balk, E. M. (2007) Meta-analysis the effect of dietary counselling for weight loss. *Annals of International Medicine*, 147 (1), 41–50.
41. Thompson, S. G. & Higgins, J. P. T. (2002) How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, 21, 1559–1573.
42. Higgins, J. P. T. & Whitehead, A. (1996) Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine*, 15, 2733–2749.
43. Burke, S. P., Silveira, A. M., Goldsmith, L. J., Yancey, J. M., Van Stewart, A. & Scarfe, W. C. (1998) A meta-analysis of mandibular intercanine width in treatment and postretention. *Angle Orthodontist*, 68 (1), 53–60.
44. Earl, A. & Albarracin, D. (2007) Nature, decay, and spiraling of the effects of fear-inducing arguments and HIV counseling and testing: A meta-analysis of the short- and long-term outcomes of HIV-prevention interventions. *Health Psychology*, 26 (4), 496–506.
45. Riley, R. D., Simmonds, M. C. & Look, M. P. (2007) Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. *Journal of Clinical Epidemiology*, 60, 431–439.
46. Altman, D. G. & Bland, J. M. (1997) Statistics notes: units of analysis. *British Medical Journal*, 314, 1874.
47. Sutton, A. J., Duval, S. J., Tweedie, R. L., Abrams, K. R. & Jones, D. R. (2000) Empirical assessment of effect of publication bias on meta-analyses. *British Medical Journal*, 320, 1574–1577.
48. Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R. & Jones, D. R. (2005) How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine*, 24 (15), 2401–2428.
49. Rothstein, H. R., Sutton, A. J. & Borenstein, M. (2005) Publication Bias in Meta-Analysis. Chichester: Wiley.

Appendix: WinBUGS code

Equation (2)
Used for the all time-point meta-analysis (ATM).

```
model{
        for(i in 1:N){
                  prec[i] <-1/(sd_diff[i]*sd_diff[i])
                  mean_diff[i] ~ dnorm(theta[i],prec[i])
                  theta[i] ~ dnorm(psi, tau)
                  }
        psi ~ dnorm(0, 0.00001)
        tau <-1/var
        var <-pow(std, 2)
        std ~ dnorm(0, 0.01)I(0,)
        pr.gr.zero <-step(psi) – equals(0, psi)
        }
     }
```

Where mean_diff, sd_diff and N are the data.

Equation (3)
Used for the trend meta-analysis (TM) when study-specific slope estimates are not reported in each study.

```
model{
       for(i in 1:N){
                 for(j in 1:T[i]){
                           mean.diff[i,j] ~dnorm(theta[i,j],prec[i,j])
                           prec[i,j]<-1/(sd_diff[i,j]* sd_diff[i,j])
                           theta[i,j]<-beta[i]*time_period[i,j]
                           }
                 beta[i] ~ dnorm(beta.p, tau)
                 }
        beta.p ~ dnorm(0, 0.00001)
        tau <-1/var
        var <-pow(sd,2)
        sd ~ dnorm(0, 0.01)I(0,)
        pr.gr.zero <-step(beta.p) – equals(0, beta.p)
        }
```

Where mean_diff, sd_diff, time_period, N and T are the data.