BMJ Open

To condition or not condition? Analysing 'change' in longitudinal randomised controlled trials

Cynthia J Coffman,[1,2] David Edelman,[1,3] Robert F Woolson[1,2]

ABSTRACT

Objective: The statistical analysis for a 2-arm randomised controlled trial (RCT) with a baseline outcome followed by a few assessments at fixed follow-up times typically invokes traditional analytic methods (eg, analysis of covariance (ANCOVA), longitudinal data analysis (LDA)). 'Constrained' longitudinal data analysis (cLDA) is a well-established unconditional technique that constrains means of baseline to be equal between arms. We use an analysis of fasting lipid profiles from the Group Medical Clinics (GMC) longitudinal RCT on patients with diabetes to illustrate applications of ANCOVA, LDA and cLDA to demonstrate theoretical concepts of these methods including the impact of missing data.

Methods: For the analysis of the illustrated example, all models were fit using linear mixed models to participants with only complete data and to participants using all available data.

Results: With complete data (n=195), 95% CI coverage are equivalent for ANCOVA and cLDA with an estimated 11.2 mg/dL (95% CI −19.2 to −3.3; p=0.006) lower mean low-density lipoprotein (LDL) cholesterol in GMC compared with usual care. With all available data (n=233), applying the cLDA model yielded an LDL improvement of 8.9 mg/dL (95% CI −16.7 to −1.0; p=0.03) for GMC compared with usual care. The less efficient, LDA analysis yielded an LDL improvement of 7.2 mg/dL (95% CI −17.2 to 2.8; p=0.15) for GMC compared with usual care.

Conclusions: Under reasonable missing data assumptions, cLDA will yield efficient treatment effect estimates and robust inferential statistics. It may be regarded as the method of choice over ANCOVA and LDA.

[1]Health Services Research, Durham Veterans Affairs Medical Center, Durham, North Carolina, USA
[2]Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, North Carolina, USA
[3]Department of Medicine, Duke University Medical Center, Durham, North Carolina, USA

Correspondence to
Dr Cynthia J Coffman; Cynthia.Coffman@duke.edu

Strengths and limitations of this study

- Clarification of the statistical methods available for longitudinal randomised controlled trials (RCTs) as well as analysis recommendations is warranted.
- In many longitudinal RCTs, participants are measured at baseline, then at the same follow-up occasions with a small number of follow-ups, for example, 2–4. In this design, how should baseline values be handled?
- In practical applications, constrained longitudinal data analysis, an appropriate generalisation of analysis of covariance, is the most straightforward to implement and under reasonable missing data assumptions will yield robust estimates of treatment effect differences and valid inferential statistics.

mean difference in LDL between treatment arms ranged from 6.9 to 11.2 mg/dL depending on the analysis method used. Reviewers questioned the primary analysis method which yielded an LDL improvement of 8.9 mg/dL for the GMC arm compared with usual care (p=0.03), and requested an alternative, and seemingly plausible but less powerful analysis that yielded an LDL improvement of 7.2 mg/dL for the GMC arm compared with usual care (p=0.15). Thus, the interpretation of the intervention effect varied significantly depending on the analytic technique used. Based on this experience and others of a similar nature, clarification of methods available as well as analysis recommendations is warranted and may be useful, not only to statistical analysts, but clinical researchers evaluating longitudinal RCTs.

Longitudinal study designs

In many longitudinal RCTs, research participants are measured at the same follow-up measurement occasions and the number of follow-up occasions is small, for example, 2–4. Typically, baseline outcomes are measured

INTRODUCTION

In a recent longitudinal randomised controlled trial (RCT) designed to examine the effect of Group Medical Clinics (GMC) on cardiovascular outcomes in patients with diabetes, the statistical inference on the effect of the GMC intervention on low-density lipoprotein (LDL) levels was dependent on the method of analysis applied.[1] Estimates of

prior to randomisation. In analyses, baseline outcomes can be ignored, used to calculate change scores, used conditionally as covariates or can be part of the outcome response vector. In this design, how should baseline values be handled? Is the question of interest conditional, given one's baseline what is the difference in outcome between treatment arms, or is the question unconditional, what is the treatment difference in change? Or does it matter? These fundamental questions and their implications on the method of analysis were introduced over 50 years ago in a sentinel paper by Lord[2] subsequently leading to the term 'Lord's paradox. In Lord's hypothetical example, the difference between the conditional and unconditional questions and how analyses are applied yielded different answers and statistical inference when applied to the same data set. In spite of extensive literature[3–8] for this type of data, there remains a lack of clarity in biomedical research applications regarding longitudinal statistical analysis.

As the issue is addressed here, a major complication requiring consideration in the analysis strategy of a longitudinal RCT is missing data. Attrition (dropout) or intermittent missingness occurs even in studies with few measurement occasions as we are addressing. Reducing the quantity of missing data in an RCT enhances the reliability of results;[9] however, rarely is it possible to eliminate missing data completely? The assumptions and methods applied for handling missing data when analysing 'change' in a longitudinal RCT can greatly affect the robustness, validity and power of results and need to be clearly understood.

The general consensus in the statistical literature for a longitudinal RCT is that the conditional approach, that is, analysis of covariance (ANCOVA), is the most powerful and robust method[4][8] to address the fundamental questions of interest. We take this view, and elaborate on its implementation, taking into account the additional complication of missing data, and enumerate the benefits of unconditional constrained longitudinal data analysis (cLDA) models, illustrating that they generalise the ANCOVA approach in the longitudinal trials' setting.

Analysis strategies

In a longitudinal RCT of the type addressed here, the outcome of interest is measured at each of the defined assessment periods of the trial (each participant in each arm of the trial). In this setting define the outcome as Y_{it}, the assessed value at each time point, t for each individual i (where i = 1..n, t = 0..k). In its simplest form, a longitudinal RCT is a pre/post-type of study where there are assessments at two time points, one pre-treatment baseline, Y_{i0}, and one post-treatment, Y_{i1}.

In the more general longitudinal RCT study design with multiple but few follow-up time points, response profile modelling[6] also known as mixed model repeated measures[10] is generally used as it allows for arbitrary patterns in the mean responses over time (ie, time is

dummy coded) as well as for the covariance of responses and can be used for fitting ANCOVA, cLDA or longitudinal data analysis (LDA) models. From a modelling perspective, the pre-treatment baseline assessment can be viewed as either a fixed predictor of post-treatment outcome or as an outcome assessment. For the former, the conditional ANCOVA approach yields estimates of treatment differences over time given the observed baseline values where post-treatment outcome measurements are the response variables. For the latter, an unconditional model (LDA or cLDA) yields estimates of treatment differences unconditionally, and in the context of a longitudinal RCT, both the pre-treatment and post-treatment outcome measurements are response variables.

Conditional versus unconditional models

A fundamental difference between a conditional or unconditional analysis is in the modelling of the pre-treatment assessment. In unconditional analysis, baseline is part of the response vector requiring additional assumptions for modelling baseline. In an LDA, there are no modelling constraints on the baseline; separate baseline means are assumed and fit for each randomised group. The general test used for treatment difference over time in LDA is equivalent to a change score type of analysis; comparing change from baseline to follow-up between randomised groups. In contrast for cLDA baseline means are constrained to be equal between the randomised groups; a common baseline mean is assumed and fit across randomised groups. In an RCT, baseline precedes any treatment deliverance and under expectation the baseline means are equal. The test for treatment difference over time in the cLDA is essentially equivalent to a test for treatment difference in an ANCOVA when no outcome data are missing.[5][11–13]

Conditional model

An ANCOVA model in a two-arm (j=1, 2) RCT with a pre/post design will have study outcome measures at baseline and one follow-up time (t=0,1). However, there will only be one response variable per participant (i=1, …,n_j) as the baseline (Y_0) is a covariate in the model. The model is written as:

$$Y_{ij1} = \mu + \alpha Y_{ij0} + \beta\, I(j = 2) + e_{ij1} \qquad (1)$$

The marginal mean at follow-up time (t=1) is conditional on the baseline Y_{ij0}. The parameter α is the slope for the baseline, and β is the effect of treatment (j=2) at time (t=1) compared with treatment (j=1) adjusted for the baseline effect.

Unconditional models
LDA model
An LDA model in a two-arm (j=1, 2) RCT with a pre/post design will have study outcome measures at both

the baseline and follow-up time (t=0,1). The model is written as:

$$Y_{ijt} = \mu + \gamma\, I(j=2) + \tau\, I(t=1) \\ + \delta\, I(j=2, t=1) + e_{ijt} \qquad (2)$$

The parameter γ is the difference in baseline means at time (t=0) between arms, τ is the mean difference in change from baseline to follow-up for arm (j=1), and δ is the difference in change from baseline to follow-up between arms.

cLDA model

A cLDA model in a two-arm (j=1, 2) RCT with a pre/post design will have study outcome measures at both the baseline and follow-up time (t=0,1). The model is written as:

$$Y_{ijt} = \mu' + \tau'\, I(t=1) + \delta'\, I(j=2, t=1) + e_{ijt} \qquad (3)$$

The parameter τ' is the mean change in outcome from baseline to follow-up time in arm (j=1), and δ' is the mean difference in change from baseline to follow-up between arms, however, since baseline means are assumed equivalent this is the mean difference between arms at the follow-up time.

Conditional and unconditional model comparison

No missing baseline or follow-up data: When there is no missing baseline or follow-up data, Liang and Zeger[5] and Liu et al[11] have shown that ANCOVA and cLDA models produce identical point estimates for treatment differences. Using our model notation we can state under the case of no missing baseline or follow-up data:

$$\beta_{\text{ANCOVA}} = \delta'_{\text{cLDA}} \qquad (4)$$

Similarly, for reasonably sized trials the variance of the treatment effect differences will be approximately equivalent between ANCOVA and cLDA models (see online supplementary appendix A).

Frison and Pocock[3] have shown that estimated treatment differences from ANCOVA or of POST-treatment means or CHANGE from baseline from an LDA model have the same expected value. However, they demonstrate the general superiority of inferential power for ANCOVA over both POST and CHANGE from baseline analysis. We can see from plotting equations of estimates of variance in figure 1 that ANCOVA and cLDA are superior to LDA-POST and LDA-CHANGE over the range of correlations between pre and post measurements (see online supplementary appendix A for variance estimates). For high correlations between pre-measurements and post-measurements the variances of LDA-CHANGE and ANCOVA/cLDA are similar.

The conditional and unconditional models described above are easily extendible from the pre/post type
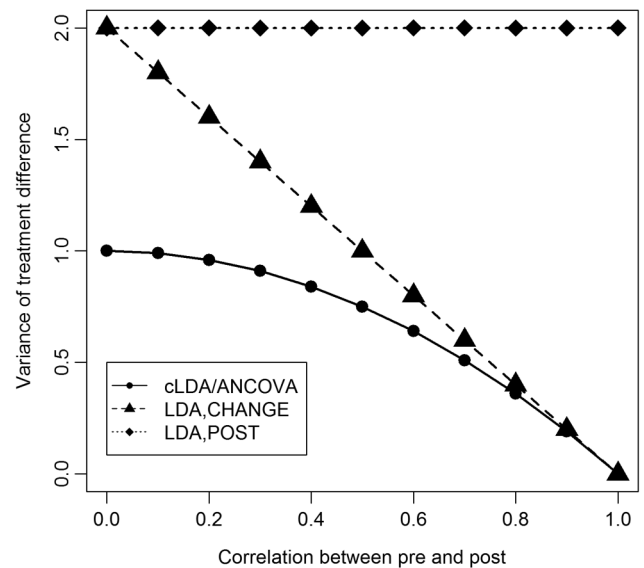

Figure 1 Comparison of variance of treatment difference estimates over the range of correlations between pre and post measurements for LDA, cLDA, ANCOVA and SPO methods; plot generated from variance estimate formulas given in online supplementary appendix A. ANCOVA, analysis of covariance; cLDA, constrained longitudinal data analysis; LDA, longitudinal data analysis; SPO, simple post only.

design with only one follow-up time point to multiple follow-up time points by additional dummy variables to represent follow-up times. Similarly, comparison of models and estimates from the pre/post design apply to the follow-up time points. In the longitudinal RCTs with few measurement occasions, the comparison of interest is generally the treatment difference at the last follow-up time point (T)—the theory as described above for the post-time measurement in a pre/post design applies to analysing the differences between treatment arms at any specific follow-up time.

Motivating example: group medical clinics study

We present an analysis of fasting lipid profiles from the GMC longitudinal RCT on patients with diabetes to illustrate applications of ANCOVA, LDA and cLDA models and demonstrate theoretical concepts described above including the impact of missing data.[1] This study randomised 239 participants at a 3:2 ratio to the GMC arm (n=133) and usual care arm (n=106). Fasting lipid profiles are secondary outcomes from the GMC study measured at baseline, midpoint (median follow-up 6.8 months) and end of study (median follow-up 12.8 months).[14]

For illustration, we focus analysis on the baseline and 12-month LDL cholesterol (LDL-C) measurements. For notation, the two arms will be denoted with j=G for GMC or j=U for usual care and time with t=0 for baseline and t=12 for 12 months. All analyses were conducted using SAS V.9.2 (SAS Institute, Cary, North Carolina, USA).

METHODS

For completeness of analyses approaches, in addition to the ANCOVA, LDA and cLDA models, we also conducted simple post only (SPO) and simple change score analysis (SACS). For SPO, we compared mean 12-month LDL-C ($\widehat{Y}_{G,12}$) for GMC (j=G) to mean usual care (j=U) 12-month LDL-C ($\widehat{Y}_{U,12}$) and for SACS, we compared mean LDL-C change from baseline ($\widehat{C}_{G,12}$) for GMC (j=G) to the mean change from baseline LDL-C ($\widehat{C}_{U,12}$) for usual care (j=U) using two-sample t-tests. We then applied the ANCOVA, LDA-CHANGE and cLDA models to, $Y_{j,0}$ and $Y_{j,12}$, the assessed value of LDL-C. All three types of models were fit using linear mixed models[15] and the response profile modelling approach. Models were fit using PROC MIXED with restricted maximum likelihood estimation and unstructured covariance.

Completers analysis

The first set of analyses was applied to participants that had both baseline and 12-month measurements (completers) to demonstrate theoretical comparisons of models as described without the added complication of missing data.

RESULTS

Among the 239 patients randomised in the study, 195 participants had LDL-C measurements at both baseline and 12 months. All methods yield statistically significant differences in LDL-C between arms (table 1) with change score analysis methods (SACS and LDA) right at p=0.05. SPO had the largest estimated difference in mean LDL-C between GMC and usual care. As expected, estimated mean treatment differences at 12 months and 95% CI coverage are equivalent for the ANCOVA and cLDA with an estimated 11.2 mg/dL lower mean LDL-C in GMC compared with usual care. The SACS and LDA estimated differences and 95% CI coverage are equivalent with an estimated 10.1 mg/dL lowering between baseline and 12 months of mean LDL-C for GMC compared with usual care. LDA and SACS analyses yield a wider 95% CI than ANCOVA and cLDA methods. For completers, the ANCOVA and cLDA yield the most robust results and illustrate their superiority (as shown in figure 1) over SPO and change score methods (both SACS and LDA), as the correlation between baseline and 12-month LDL-C values was estimated in the 0.50 range.

All participants analysis

The second set of analyses was applied to all participants (ie, including those with either missing baseline or 12-month measurements) to compare methods and illustrate the impact of missing data. SPO participants with missing 12-month LDL-C are deleted and SACS participants with either missing baseline or end of study measurements are deleted. Similarly, for ANCOVA with only two time points, participants with missing baseline or 12-month measurements are deleted. For LDA and cLDA, all available data were used; no participants were deleted due to missing data. The estimation procedure used in the mixed model framework for longitudinal analysis yields unbiased estimates of parameters when missing outcomes are assumed to be ignorable, that is, when missing values are related to either observed covariates or response variables but not to unobserved variables.[16 17]

RESULTS

Among the 239 patients randomised, 6 participants had no baseline or 12-month LDL-C, so they are excluded yielding 233 patients for the all participants analysis. The estimated treatment differences diverge across analysis methods as well as statistical significance using a α=0.05 (table 2). ANCOVA and SACS are equivalent to completers analyses, as the 38 cases with either a missing baseline or 12-month LDL-C are deleted. ANCOVA and cLDA methods no longer have equivalent estimates of treatment differences. All 233 participants contribute at least one measurement to cLDA analysis so that participants missing either a baseline or 12-month LDL-C contribute to estimated treatment difference of an 8.9 mg/dL lower mean LDL-C at 12 months in GMC compared with usual care (see table 2). The discrepancy in the estimated treatment difference between ANCOVA and cLDA is due to missing data and assumptions that are made about missing data. As discussed previously, mixed-effect models yield unbiased estimates of treatment effects under the assumption that the missing data

Table 1 Completers only (n=195 participants)—pre/post analyses

Model	Outcome (Y_t and/or C_t)	GMC (n=117)	Usual care (n=78)	GMC vs usual care (95% CI)	p Value
Post-only	12 months (Y_{12})	81.9	94.1	−12.1 (−21.5 to −2.7)	0.01
SACS	12 months−baseline (C_{12})	−12.9	−2.8	−10.1 (−20.2 to 0.0)	0.05
ANCOVA	12 months (Y_{12})	82.3	93.5	−11.2 (−19.2 to −3.3)	0.006
LDA	Baseline (Y_0)	94.8	96.9		
	12 months (Y_{12})	81.9	94.1	−10.1 (−20.2 to 0.0)	0.05
cLDA	Baseline (Y_0)	95.7	95.7		
	12 months (Y_{12})	82.3	93.5	−11.2 (−19.2 to −3.3)	0.006

ANCOVA, analysis of covariance; cLDA, constrained longitudinal data analysis; GMC, Group Medical Clinics; LDA, longitudinal data analysis; SACS, simple change score analysis.

Table 2 All available data (n=233 participants)—pre/post analyses

Model	Outcome (Y_t and/or C_t)	N	GMC	Usual care	GMC vs usual care (95% CI)	p Value
Post-only	12 months (Y_{12})	204	89.7	96.7	−6.9 (−14.2 to 0.4)	0.07
SACS	12 months−baseline (C_{12})	195	−12.9	−2.8	−10.1 (−22.0 to −0.8)	0.05
ANCOVA	12 months (Y_{12})	195	83.4	94.6	−11.2 (−19.2 to −3.3)	0.006
LDA*	Baseline (Y_0)	233	96.7	99.6		
	12 months (Y_{12})		83.5	93.6	−7.2 (−17.2 to 2.8)	0.15
cLDA*	Baseline (Y_0)	233	98.0	98.0		
	12 months (Y_{12})		84.0	92.9	−8.9 (−16.7 to −1.0)	0.03

*Baseline LDL-C is missing for 9 participants and 12-month LDL-C is missing for 29 participants.
ANCOVA, analysis of covariance; cLDA, constrained longitudinal data analysis; GMC, Group Medical Clinics; LDA, longitudinal data analysis; LDL-C, low-density lipoprotein cholesterol; SACS, simple change score analysis.

are conditional on observed quantities. Although this assumption cannot be specifically tested, the discrepancy in estimates between ANCOVA and cLDA model yields some evidence that data are missing conditional on observed quantities. If data were missing completely at random (not conditional on observed quantities) then estimated treatment differences between ANCOVA with case deletion and cLDA should be similar. Similarly, differences in estimates from the POST-only as well as LDA models between completers only and participants using all available data provide evidence that participants with missing observations may not be a completely random sample from the study population. Therefore, any analyses based on completers only (including the ANCOVA model) may be biased.

As shown for ANCOVA and cLDA, LDA and SACS methods estimated differences are discrepant due to missing data and assumptions that are made about missing data. LDA is using all available data and fit using mixed-effects model, whereas in SACS participants with missing data are deleted. Similar to completers analysis, LDA methods yield wider 95% CI than both cLDA and ANCOVA.

DISCUSSION
Missing data implications: the case for cLDA
The illustrative example presented above clearly demonstrates the statistical theory described for the analysis methods and the impact of missing data[3 4 16] on the inference and interpretations of results. In general, the implications of how missing data are handled in a trial range from sample size and statistical power loss to potentially biased estimates of treatment effects. Olsen et al[18] discuss principled methods for handling missing data in longitudinal sleep disorder trials that are relevant to the types of longitudinal RCTs we are discussing. The default method for handling missing data in many software packages is case deletion; if any variable or outcome is missing for a participant, then the participant is deleted from analysis. In order for estimates of treatment effect differences to be unbiased when case deletion occurs, we have to assume that the subset of participants with complete data is a random sample of the entire study sample. Therefore, for

ANCOVA, SACS and SPO analysis above, we are making this assumption. For many longitudinal RCTs it is likely that the probability of dropout is related to some observable characteristic or quantity (eg, treatment arm assignment) and therefore the subset of participants with complete data is not a random sample of the entire study. Under this assumption unbiased estimates of treatment effects can be achieved with mixed-effect models or by performing multiple imputation.[18] Methods for handling missing data when we cannot assume that missing data can be characterised by observable quantities are beyond the scope of this paper.

The choice of whether to use multiple imputation is up to the analyst and based on the untestable assumption that all predictors of missing data can be included in the mixed-effect model. In many longitudinal RCTs it is a reasonable assumption that treatment group assignment, time of assessment and available outcome assessments along with potential baseline stratification variables would be the predictors of missing outcome data. If the number of potential predictor variables of missing data is expansive and unreasonable to include in the mixed-effects model then it may be necessary to perform a multiple imputation analysis.

Multiple imputation is a more complicated process for handling missing data that is also based on the untestable assumption that all predictors of missing data are included in the imputation model.[17] In the multiple imputation framework, comparisons of conditional and unconditional models would be equivalent to what was discussed above for no missing baseline or follow-up data. Once imputed data sets are created, there are no missing outcome data; the uncertainty from estimating imputed values is accounted for in the SE estimates.

For ease of implementation and robustness of results[15 16] mixed-effect models with maximum likelihood estimates are a good first-line choice for handling missing data. Even when applying mixed-effects models, the treatment of baseline values can impact results because of differences in case deletion. In the pre/post design if the baseline is a covariate (ANCOVA), participants with the missing covariate will be deleted. Similarly, with only one follow-up time, if the follow-up time is missing then there is no outcome data for the participant

to be included. However, with unconditional models since both baseline and follow-up are part of the outcome vector, if a participant is missing either measurement, their available data will be included and participant is not deleted. When missing data occur and mixed-effect models are used, cLDA models are the optimal choice for providing the most precise estimate of treatment differences under a reasonable assumption that missing data are related to observable characteristics. Finally, it should also be noted that these methods will not diverge for every data set to which they may be applied to—however as certain missing data assumptions cannot specifically be tested, cLDA models are the optimal choice across a range of conditions and data sets.

SUMMARY

It does not appear widely appreciated that cLDA can often be regarded as the method of choice for the analysis of a longitudinal RCT with few measurement occasions. Except for small samples it is equivalent to ANCOVA when there are no missing data, and cLDA is an appropriate generalisation of ANCOVA under reasonable missing data assumptions. Without question, potential baseline imbalance between treatment arms has implications for the analysis of longitudinal RCTs and is a source of confusion. In an RCT, baseline differences can be attributed to random chance assuming there were no problems or issues with the randomisation process and no significant measurement error issues. LDA or CHANGE score analysis is sometimes viewed as a more intuitive analysis; however, for an RCT most often the best method is an ANCOVA as far as bias, precision and power.[19] In a longitudinal RCT, when within-participant correlations are not high, change score or LDA (an unconditional analysis) is not the most powerful analyses. Frison and Pocock[3] found that ANCOVA treatment difference estimates have small bias under the assumption that measurement error is a small proportion of the between-participant variance; however, bias in a change score analysis may be somewhat larger especially if correlations between pre- and post-measurements are small.

cLDA is sometimes erroneously viewed as more problematic when there is baseline imbalance in outcomes between treatment arms. However, cLDA and ANCOVA are equivalent when analysing complete data. cLDA generalises the ANCOVA approach and both are superior to an LDA in many cases. Therefore, the primary analytic issue is not necessarily whether or not to perform conditional analysis. ANCOVA is a conditional analysis and cLDA is an unconditional analysis, yet both are powerful methods that can be applied to examine treatment differences over time in a longitudinal RCT. In practical applications, cLDA is the most straightforward to implement and under reasonable missing data assumptions will yield robust estimates of treatment effect differences and inferential statistics. In most cases, it is the method of choice.

REFERENCES

1. Edelman D, Fredrickson SK, Melnyk SD, et al. Medical clinics versus usual care for patients with both diabetes and hypertension: a randomized trial. *Ann Intern Med* 2010;152:689–96.
2. Lord FM. A paradox in the interpretation of group comparisons. *Psychol Bull* 1967;68:304–5.
3. Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Stat Med* 1992;11:1685–704.
4. Senn S. Change from baseline and analysis of covariance revisited. *Stat Med* 2006;25:4334–44.
5. Liang KY, Zeger SL. Longitudinal data analysis of continuous and discrete responses for pre-post designs. *Sankhyā* The Indian Journal of Statistics, Series B 2000;62:134–48.
6. Fitzmaurice GM, Laird NM, Ware JH. *Applied longitudinal analysis*. Hoboken, NJ: Wiley, 2011.
7. Crager MR. Analysis of covariance in parallel-group clinical trials with pretreatment baselines. *Biometrics* 1987;43:895–901.
8. Vickers AJ, Altman DG. Statistics notes: analysing controlled trials with baseline and follow up measurements. *BMJ* 2001;323:1123–4.
9. Fleming TR. Addressing missing data in clinical trials. *Ann Intern Med* 2011;154:113–17.
10. Mallinckrod CH, Lane PW, Schnell D, et al. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. *Drug Info J* 2008;42:303–19.
11. Liu GF, Lu K, Mogg R, et al. Should baseline be a covariate or dependent variable in analyses of change from baseline in clinical trials? *Stat Med* 2009;28:2509–30.
12. Lu K, Mehrotra DV, Liu G. Sample size determination for constrained longitudinal data analysis. *Stat Med* 2009;28:679–99.
13. Kenward MG, White IR, Carpenter JR. Should baseline be a covariate or dependent variable in analyses of change from baseline in clinical trials? by G. F. Liu, K. Lu, R. Mogg, M. Mallick and D. V. Mehrotra, Statistics in Medicine 2009;28:2509–2530. *Stat Med* 2010;29:1455–6.
14. Crowley MJ, Melnyk SD, Ostroff JL, et al. Can group medical clinics improve lipid management in diabetes? *Am J Med* 2014;127:145–51.
15. Verbeke G, Molenberghs G. *Linear mixed models for longitudinal data*. New York: Springer, 2000.
16. Little RJA, Rubin DB. *Statistical analysis with missing data*. Wiley, 2002.
17. Schafer JL. *Analysis of incomplete multivariate data*. London: Chapman & Hall, 1997.
18. Olsen MK, Stechuchak KM, Edinger JD, et al. Move over LOCF: principled methods for handling missing data in sleep disorder trials. *Sleep Med* 2012;13:123–32.
19. Egbewale BE, Lewis M, Sim J. Bias, precision and statistical power of analysis of covariance in the analysis of randomized trials with baseline imbalance: a simulation study. *BMC Med Res Methodol* 2014;14:49.