STATISTICS IN MEDICINE
Statist. Med. 2009; **28**:2509–2530
Published online 17 July 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3639

Should baseline be a covariate or dependent variable in analyses of change from baseline in clinical trials?

Guanghan F. Liu*,†, Kaifeng Lu, Robin Mogg, Madhuja Mallick and Devan V. Mehrotra

Clinical Biostatistics, UG1CD-44, Merck Research Laboratories, 351 N. Sumneytown Pike, North Wales, PA 19454, U.S.A.

SUMMARY

In randomized clinical trials, a pre-treatment measurement is often taken at baseline, and post-treatment effects are measured at several time points post-baseline, say $t=1,\ldots,T$. At the end of the trial, it is of interest to assess the treatment effect based on the mean change from baseline at the last time point T. We consider statistical methods for (i) a point estimate and 95 per cent confidence interval for the mean change from baseline at time T for each treatment group, and (ii) a p-value and 95 per cent confidence interval for the between-group difference in the mean change from baseline. The manner in which the baseline responses are used in the analysis influences both the accuracy and the efficiency of items (i) and (ii). In this paper, we will consider the ANCOVA approach with change from baseline as a dependent variable and compare that with a constrained longitudinal data analysis (cLDA) model proposed by Liang and Zeger (*Sankhya*: *Indian J. Stat.* (*Ser B*) 2000; **62**:134–148), in which the baseline is modeled as a dependent variable in conjunction with the constraint of a common baseline mean across the treatment groups. Some drawbacks of the ANCOVA model and potential advantages of the cLDA approach are discussed and illustrated using numerical simulations. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: baseline; repeated measures; ANCOVA; constrained LDA

1. INTRODUCTION

Baseline values are commonly measured in clinical trials to help the assess drug effects after randomization. The baseline measurement can potentially be used for several purposes, including subject selection in studies targeting a study population with a certain disease condition, and serving as a basis to measure the treatment effect in terms of the change from the baseline. When there is only one post-randomization measurement, treatment effects on mean change from baseline

*Correspondence to: Guanghan F. Liu, Clinical Biostatistics, UG1CD-44, Merck Research Laboratories, 351 N. Sumneytown Pike, North Wales, PA 19454, U.S.A.
†E-mail: guanghan_frank_liu@merck.com

Received 20 March 2008
Accepted 20 May 2009

are typically assessed using an analysis of covariance (ANCOVA) model with either the post-randomization value or the calculated change from baseline value as the dependent variable and baseline as a covariate [1–5]. Note that in this ANCOVA model the treatment comparison would be the same with either the post-baseline value or the change from baseline value as the dependent variable; for simplicity, we use the latter in the discussions of this paper and recognize that there are some philosophical differences among these two endpoints discussed in the literature [5]. This ANCOVA model is one of the most commonly used statistical methods for the analysis of change from baseline data in clinical trials. Under the assumption of bivariate normality for baseline and post-baseline measurements, estimates and statistical tests from the ANCOVA model conditional on baseline values are unbiased and valid even when the baseline is a random variable [1]. When baseline measurements are correlated to post-baseline measurements, adjusting for baseline using ANCOVA has been shown to remove conditional bias in treatment group comparisons due to chance imbalances [2] and improve efficiency over unadjusted comparisons [1–3]. Many discussions can be found in the literature on this topic and on the adjustment for baseline with measurement error (see e.g. Chambless and Roeback [6], Yanez *et al.* [7], Chan *et al.* [8], and Senn [9, 10]).

When there are repeated measurements post-randomization, a longitudinal data analysis (LDA) model may be used for the treatment comparison. Conventionally, the change from baseline values at each time point after baseline are calculated and the baseline is included in the LDA model as a covariate. This model will hereafter be referred as the longitudinal ANCOVA model or simply called ANCOVA model.

Alternatively, Liang and Zeger [11] proposed a constrained full likelihood approach in which the baseline value as well as the post-randomization values are modeled as dependent variables; the 'constraint' is that the baseline mean responses for the treatment groups are assumed equal, which is reasonable due to randomization. Because of this constraint, the baseline mean in this model is independent of treatment (i.e. baseline is not an 'outcome' of treatment in the analysis model). This model will be referred as the constrained longitudinal data analysis (cLDA) model in this paper. When there are no missing data in a pre–post design with one post-randomization measurement, Liang and Zeger showed that both models produce identical point estimates for the treatment difference.

In this paper, we compare the ANCOVA model with the cLDA model with respect to (i) the point estimate and 95 per cent confidence interval for the mean change from baseline at a give time point (e.g. last time point in the study) for each treatment group, and (ii) the p-value and 95 per cent confidence interval for the between-group difference at a given time point in terms of the mean change from baseline. To reflect the real clinical trial situation in which every new trial will enroll different patients and therefore will have different baseline values, we assume throughout this paper that the baseline values are random. The statistical properties will be evaluated under this assumption across trials for the ANCOVA and cLDA models. We acknowledge the broad discussions on the role of baseline measurements, and whether the randomness of baseline measurements should be taken into account in clinical trials, found in the literature (e.g. [6–9, 12]). However, the philosophical perspective on the role of baseline measurements in clinical trials is beyond the scope of this paper.

The paper is organized as follows. The two competing statistical models are briefly described in Section 2. The characteristics with respect to the estimate and test of treatment group differences and model adjusted group mean changes, as well as distributional assumptions, are discussed in Section 3. Simulation results to illustrate the findings are provided in Section 4. The methods

DOI: 10.1002/sim

are applied to a real clinical trial data set in Section 5, followed by a discussion and concluding remarks in Section 6.

2. METHODS FOR ANALYSIS OF LONGITUDINAL DATA IN TERMS OF CHANGE FROM BASELINE

2.1. Longitudinal ANCOVA model

Suppose responses for a study endpoint are measured at baseline ($t=0$) and at T post-baseline time points in a clinical trial. Let Y_{ijt} be the response for subject i, with treatment assignment j, at time t. The marginal mean at time t conditional on baseline Y_{ij0} can be modeled as:

$$E(Y_{ijt}|Y_{ij0})=\alpha_t Y_{ij0}+\beta_{jt} I(\text{treatment}=j) I(\text{time}=t), \quad t=1, 2, \ldots, T \tag{1}$$

The slope, α_t, can be different for each time t but is the same across treatment groups, and β_{jt} is the effect for treatment j at time t after adjusting for the baseline effect. Because the baseline is a covariate, the model can also be written in terms of the change from baseline, $E(Y_{ijt}-Y_{ij0}|Y_{ij0})=(\alpha_t-1)Y_{ij0}+\beta_{jt} I(\text{treatment}=j) I(\text{time}=t)$. The standard analysis for this model assumes that conditional on baseline, the post-baseline values or the change from baseline values are multivariate normally distributed. The baseline is treated as fixed in this analysis model. Both subjects who have missing baseline values and subjects who have only baseline values are excluded from the model fit. When $T=1$, this model corresponds to the commonly used ANCOVA model for a pre–post study design.

With repeated measures (i.e. $T>1$), we will call this model a longitudinal ANCOVA model. An unstructured covariance matrix can be used to account for within subject correlation at times $t>0$. A separate covariance matrix can be specified for each treatment group; however because the baseline value is not a part of the response vector, the covariance between baseline and post-baseline responses in the ANCOVA model are specified by the coefficients of $\{\alpha_t, t=1, \ldots, T\}$ and are the same across the treatment groups. For convenience, we focus on the last time point ($t=T$) and assume that the study has two treatment arms, a test drug ($j=A$) and a control ($j=B$). The comparison of interest is the treatment effect on the mean change from baseline at the last time point:

$$\theta_T=\beta_{AT}-\beta_{BT}$$

For the ANCOVA model, this parameter can also be interpreted as the mean treatment difference at last time point.

Suppose $\hat{\alpha}_T, \hat{\beta}_{AT}$, and $\hat{\beta}_{BT}$ are the maximum likelihood estimates for the parameters at time T from model (1). Then the treatment effect is estimated as

$$\hat{\theta}_T=\hat{\beta}_{AT}-\hat{\beta}_{BT}$$

The model adjusted group mean changes from baseline at the last time point are defined as the estimated group mean changes at the overall mean baseline level

$$\hat{\theta}_{AT}=(\hat{\alpha}_T-1)\tilde{Y}_{\bullet\bullet 0}+\hat{\beta}_{AT}$$

$$\hat{\theta}_{BT}=(\hat{\alpha}_T-1)\tilde{Y}_{\bullet\bullet 0}+\hat{\beta}_{BT}$$

DOI: 10.1002/sim

for test drug and control, respectively, where $\tilde{Y}_{\bullet\bullet 0}$ is the overall baseline mean for subjects included in the model fit from both treatment groups. These estimates are the best-linear unbiased estimates of the marginal mean changes for each treatment group at $\tilde{Y}_{\bullet\bullet 0}$ from model (1). Note that the treatment difference can also be estimated as the difference of the model adjusted group mean changes, $\hat{\theta}_T = \hat{\theta}_{AT} - \hat{\theta}_{BT}$.

2.2. cLDA model

A cLDA model was proposed by Liang and Zeger [11]. Utilizing the same notation as above, the cLDA model includes the baseline value as part of the response vector. The marginal mean response can be modeled as:

$$E(Y_{ijt}) = \gamma_0 + \gamma_{jt} I(\text{treatment} = j) I(\text{time} = t \text{ and } t > 0), \quad t = 0, 1, 2, \ldots, T \tag{2}$$

where γ_0 is the mean response at $t = 0$, which is constrained to be the same for both treatment groups due to randomization, γ_{jt} is the effect for treatment j at time t after adjusting for the baseline value and can be interpreted as the mean change from baseline for a given treatment group. The cLDA model assumes that the baseline and post-baseline values are jointly multivariate normally distributed. An unstructured covariance matrix can be used in this model to account for within subject correlation at times $t \geqslant 0$ (including baseline). This allows more flexibility than the ANCOVA model because a separate covariance matrix can be specified for each treatment group, therefore, the covariance between baseline and post-baseline measurements are not necessary the same among the treatment groups.

Let $\eta_{AT} = \gamma_{AT} + \gamma_0$ and $\eta_{BT} = \gamma_{BT} + \gamma_0$ be the marginal group means at time T for treatment A and B, respectively. Suppose $\hat{\gamma}_0$, $\hat{\gamma}_{AT}$, and $\hat{\gamma}_{BT}$ are maximum likelihood estimates for the parameters at baseline and time T from model (2). Then the treatment effect on the mean change from baseline at time T is estimated as:

$$\hat{\gamma}_T = \hat{\gamma}_{AT} - \hat{\gamma}_{BT} = (\hat{\gamma}_{AT} + \hat{\gamma}_0) - (\hat{\gamma}_{BT} + \hat{\gamma}_0) = \hat{\eta}_{AT} - \hat{\eta}_{BT}$$

Here, the $\hat{\gamma}_{AT}$ and $\hat{\gamma}_{BT}$ are the model adjusted group mean change estimates for test drug and control at time T, respectively. Under the constraint, the baseline term cancels out because the estimated baseline mean is the same for both treatment groups.

3. COMPARISONS BETWEEN ANCOVA AND CONSTRAINED LDA MODELS

3.1. Treatment difference between groups

We first show that the treatment difference parameters are the same between the ANCOVA and cLDA models defined above. Under the ANCOVA model, the treatment difference is the conditional mean difference on responses between treatment groups. Let Y_{jT} be the random variable of outcome at time T for treatment j, and Y_0 be the baseline variable. The treatment difference from the ANCOVA model is

$$E(Y_{iAT} | Y_{iA0} = x) - E(Y_{iBT} | Y_{iB0} = x) = (\alpha_T x + \beta_{AT}) - (\alpha_T x + \beta_{BT}) = \beta_{AT} - \beta_{BT} \tag{3}$$

which is the conditional mean difference of responses between treatment groups at time T and is independent of any specific value, x, for the baseline.

DOI: 10.1002/sim

Let

$$\Sigma = \begin{pmatrix} \sigma_{00} & \cdots & \sigma_{0T} \\ \vdots & \ddots & \vdots \\ \sigma_{0T} & \cdots & \sigma_{TT} \end{pmatrix}$$

denote the unconditional covariance matrix for the repeated measurements including baseline. Using the conditional expectation formula, we can associate the parameters in Model (2) with that in Model (1)

$$\alpha_T = \frac{\sigma_{0T}}{\sigma_{00}}, \quad \beta_{jT} = (\gamma_0 + \gamma_{jT}) - \frac{\sigma_{0T}}{\sigma_{00}}\gamma_0 \quad \text{for } j = A, B$$

It follows that $\theta_T = \beta_{AT} - \beta_{BT} = \gamma_{AT} - \gamma_{BT} = \gamma_T$; that is, the conditional mean difference between treatment groups in the ANCOVA model is the same as the unconditional mean difference between treatment groups in the cLDA model.

The maximum likelihood estimates for the parameters are consistent from both models when there are no missing data. In fact, the point estimates are shown to be the same analytically (see Appendix A):

$$\hat{\gamma}_T = \hat{\theta}_T = (\bar{y}_{\cdot AT} - \bar{y}_{\cdot BT}) - \hat{\alpha}_T(\bar{y}_{\cdot A0} - \bar{y}_{\cdot B0})$$

where $\hat{\alpha}_T = \hat{\sigma}_{0T}/\hat{\sigma}_{00}$.

For the ANCOVA model, the variance estimate for $\hat{\theta}_T$ is based on the conditional distribution given baseline values. Specifically, let $\underline{Y}_0$ denote the collection of baseline values, we have

$$\hat{\text{var}}_{\text{ANCOVA}}(\hat{\theta}_T | \underline{Y}_0) = \hat{\text{var}}(\bar{y}_{\cdot AT} - \bar{y}_{\cdot BT} | \underline{Y}_0) + \hat{\text{var}}(\hat{\alpha}_T | \underline{Y}_0)(\bar{y}_{\cdot A0} - \bar{y}_{\cdot B0})^2$$

where we used the fact that $\hat{\text{cov}}(\bar{y}_{\cdot AT} - \bar{y}_{\cdot BT}, \hat{\alpha}_T | \underline{Y}_0) = 0$. The first term becomes

$$\left(\frac{1}{n_A} + \frac{1}{n_B}\right)\hat{\sigma}_{TT\cdot 0}$$

which does not depend on $\underline{Y}_0$. Note

$$\hat{\text{var}}(\hat{\alpha}_T | \underline{Y}_0) = \frac{1}{(n_A + n_B)S_{00}}\hat{\sigma}_{TT\cdot 0}$$

where

$$S_{00} = \frac{1}{n_A + n_B}\sum_{j=A}^{B}\sum_{i=1}^{n_J}(y_{ij0} - \bar{y}_{\cdot j0})^2$$

Consequently,

$$\hat{\text{var}}_{\text{ANCOVA}}(\hat{\theta}_T | \underline{Y}_0) = \left\{\frac{1}{n_A} + \frac{1}{n_B} + \frac{(\bar{y}_{\cdot A0} - \bar{y}_{\cdot B0})^2}{(n_A + n_B)S_{00}}\right\}\hat{\sigma}_{TT\cdot 0}$$

The last term inside the bracket measures the difference of the observed baseline mean between the two treatment groups.

DOI: 10.1002/sim

For the cLDA model, the variance estimate for $\hat{\gamma}_T$ is based on model (2) in which baseline values are part of the responses. As shown in Appendix A (cf. (A10)), the variance estimate is

$$\hat{\text{var}}_{\text{cLDA}}(\hat{\gamma}_T) = \left(\frac{1}{n_A} + \frac{1}{n_B}\right)\hat{\sigma}_{TT\cdot 0}$$

Therefore, the estimated variance from the ANCOVA model is always greater than or equal to that from the cLDA model for any given data set:

$$\hat{\text{var}}(\hat{\gamma}_T) \leqslant \hat{\text{var}}(\hat{\theta}_T | \underline{Y}_0) \tag{4}$$

An analytic proof of this inequality for the general case with missing post-baseline data is given in Appendix A (cf. (A1) and (A6)). This result is also confirmed in the simulation study in Section 4.

Therefore, when there are no missing data the point estimates of the treatment differences are identical for both models. The variance estimates, however, are different because the variance estimate from the ANCOVA model is a conditional variance. Because of (4), the cLDA model will be at least as powerful as the ANCOVA model. The equality of (4) is true only if the observed baseline means are the same for both treatment groups. In practice, the difference of the variance estimates from the two models is fairly small because randomization will typically result in very similar observed baseline means. Therefore, the power gain from the cLDA over the ANCOVA model is generally small. This is also confirmed in the simulations in Section 4.

When there are missing data, it is known that the methods based on maximum likelihood are consistent under the missing at random (MAR) assumption. Therefore, the parameter estimates and statistical inference under cLDA, which is based on a full likelihood function, are asymptotically unbiased when data are from multivariate normal distribution. However, the analysis based on the ANCOVA model may be biased if there are subjects with either missing data at baseline or missing data at all post-baseline measurements. When missing data occur at baseline, it is generally unrelated to treatment or any observed outcome, and therefore likely to be missing completely at random. As such, excluding these subjects from the analysis will not generally cause any bias. However, the ANCOVA model can produce biased estimates when subjects with missing data at all post-baseline time points are excluded from the analysis. This is because the missing data at post-baseline may depend on observed baseline value. The baseline mean after excluding those subjects may be different from the baseline mean of all subjects in the study. For the treatment difference, the impact is small because the estimated difference may be asymptotically independent of baseline values as shown in (3). In fact, when $T = 1$, the estimated treatment difference is still the same from ANCOVA and cLDA models as shown in Appendix A. However, the impact on the model adjusted group mean estimates can be significant. We explore the details in the following subsections.

3.2. Model adjusted group mean estimates

Under the ANCOVA model, the model adjusted group mean estimate is obtained as

$$\hat{\theta}_{jT} = (\hat{\alpha}_T - 1)\tilde{Y}_{\bullet\bullet 0} + \hat{\beta}_{jT}$$

where $\hat{\alpha}_T, \hat{\beta}_{AT}$, and $\hat{\beta}_{BT}$ are maximum likelihood estimates for the parameters at time T from model (1). In general, the parameter estimates for α_T, β_{AT}, and β_{BT} are asymptotically unbiased conditional on baseline values. For the model adjusted group mean change at time T, we have

$$E(\hat{\theta}_{jT}) = E[E((\hat{\alpha}_T - 1)\tilde{Y}_{\bullet\bullet 0} + \hat{\beta}_{jT} | \underline{Y}_0)] = (\alpha_T - 1)E(\tilde{Y}_{\bullet\bullet 0}) + \beta_{jT}$$

DOI: 10.1002/sim

When there are no missing data, we have $\tilde{Y}_{\bullet\bullet 0}=\bar{Y}_{\bullet\bullet 0}$ and $E(\bar{Y}_{\bullet\bullet 0})=E(Y_{ij0})$. Therefore,

$$E(\hat{\theta}_{jT})=(\alpha_T-1)E(Y_{ij0})+\beta_{jT}=E(Y_{ijT}-Y_{ij0})=\theta_{jT}$$

That is, the model adjusted group mean change estimate is unbiased. However, this may not be true when there are missing data. Specifically, when the missing data depend on the observed baseline values under MAR, subjects included in the analysis model may have a different baseline distribution from that of all the subjects randomized, therefore, $E(\tilde{Y}_{\bullet\bullet 0})\neq E(\bar{Y}_{\bullet\bullet 0})=E(Y_{ij0})$. Consequently, the estimate $\hat{\theta}_{jT}$ can be biased. This is confirmed in the simulations in Section 4.

When there are no missing data, the conditional variance of $\hat{\theta}_{jT}$, $\text{var}(\hat{\theta}_{jT}|\underline{Y}_0)$, obtained from the ANCOVA model also underestimates the overall variance of $\hat{\theta}_{jT}$ if $\underline{Y}_0$ is a random vector rather than a constant vector. In fact, when there are no missing data, the model adjusted group mean change from the ANCOVA model is $\hat{\theta}_{jT}=(\bar{y}_{\cdot jT}-\hat{\alpha}_T\bar{y}_{\cdot j0})+(\hat{\alpha}_T-1)\bar{y}_{\cdot\cdot 0}$. As shown in Appendix A (cf. (A14)), the estimated variance for $\hat{\theta}_{jT}$ is

$$\hat{\text{var}}_{\text{ANCOVA}}(\hat{\theta}_{jT}|\underline{Y}_0)=\left\{\frac{1}{n_j}+\frac{(\bar{y}_{\cdot j0}-\bar{y}_{\cdot\cdot 0})^2}{(n_A+n_B)S_{00}}\right\}\hat{\sigma}_{TT\cdot 0}$$

The unconditional variance estimate for $\hat{\theta}_{jT}$ can be obtained from the cLDA model because $\hat{\theta}_{jT}=\hat{\gamma}_{jT}$ when there are no missing data (cf., (A12)). The model adjusted group mean change estimate $\hat{\gamma}_{jT}$ is unbiased under MAR missing mechanism. Its variance estimate (as shown in Appendix A, cf. (A13)) is

$$\hat{\text{var}}_{\text{cLDA}}(\hat{\theta}_{jT})=\frac{1}{n_j}\hat{\sigma}_{TT\cdot 0}+\frac{1}{n_A+n_B}(\hat{\alpha}_T-1)^2\hat{\sigma}_{00}$$

Because the observed baseline means will be very similar between treatment groups due to randomization,

$$\hat{\text{var}}_{\text{ANCOVA}}(\hat{\theta}_{jT}|\underline{Y}_0)\cong\frac{1}{n_j}\hat{\sigma}_{TT\cdot 0}$$

Consequently, we see that the ANCOVA model underestimates the variance of model adjusted group mean changes. The relative underestimation can be defined as

$$\frac{\hat{\text{var}}_{\text{cLDA}}(\hat{\theta}_{jT})-\hat{\text{var}}_{\text{ANCOVA}}(\hat{\theta}_{jT}|\underline{Y}_0)}{\hat{\text{var}}_{\text{cLDA}}(\hat{\theta}_{jT})}\cong\frac{\dfrac{1}{n_A+n_B}(\hat{\alpha}_T-1)^2\hat{\sigma}_{00}}{\dfrac{1}{n_j}\hat{\sigma}_{TT\cdot 0}+\dfrac{1}{n_A+n_B}(\hat{\alpha}_T-1)^2\hat{\sigma}_{00}}$$

which is equal to $(1-\rho_{0T})/(3+\rho_{0T})$ when $\sigma_T=\sigma_0$ and $n_A=n_B$, where ρ_{0T} is the correlation coefficient between baseline and response at time T. The underestimation is more severe when the correlation coefficient between baseline and post-baseline values becomes smaller.

Note that the above discussion is also true asymptotically when there are missing data. The proof for the general case in the presence of missing post-baseline data is given in Appendix A (cf. (A2) and (A8)).

DOI: 10.1002/sim

3.3. Distributional assumption

There are some differences with respect to the distributional assumption between the ANCOVA and cLDA models. The ANCOVA model assumes that conditional on baseline the post-baseline measurements are multivariate normally distributed, while the cLDA model assumes that both the baseline and post-baseline measurements are jointly multivariate normally distributed since the baseline value is treated as part of the response vector. Therefore, intuitively the cLDA model has a stronger assumption with respect to the baseline measurement. However, the ANCOVA model assumes that the baseline values are fixed for each trial, which may cause restriction for generalizing the results to replicated trials.

When the baseline value is used as part of the inclusion/exclusion criteria for clinical trials, the distribution of baseline values can be skewed. The normality assumption may be violated. We investigate the impact of this non-normality problem for both models via simulations in Section 4.

Table I summarizes the differences between the ANCOVA model and cLDA model with respect to the assumption and properties of estimates for treatment difference and model adjusted group mean changes.

4. SIMULATIONS

Simulation studies were undertaken to assess the performance of the ANCOVA and the cLDA models under a variety of scenarios. Treatment difference and individual treatment group mean estimates were compared with respect to bias, confidence interval coverage, Type I error rate and power. To reflect the real clinical trial situation in which every new trial will enroll different patients with different baseline values, the baseline values in the simulations were generated at random in each replication.

Without loss of generality, we simulated data from two treatment groups (test drug and control) and four repeated measures per subject (including baseline). We first generate repeated measures from multivariate normal with given means and a covariance matrix as follows:

$$\Sigma=\sigma^2\begin{bmatrix}1 & 0.7 & 0.4 & 0.2\\ 0.7 & 1 & 0.7 & 0.4\\ 0.4 & 0.7 & 1 & 0.7\\ 0.2 & 0.4 & 0.7 & 1\end{bmatrix}$$

To study the impact of the normality assumption at baseline, the study data were generated under three different scenarios for the baseline variable:

(i) from a normal distribution;
(ii) from a truncated normal distribution; and
(iii) from a truncated t-distribution with 3 degrees of freedom.

For all cases, we set the control group mean vector $\mu=(3.0\ 2.5\ 2.3\ 2.0)$. For scenario (ii), the baseline value was left-truncated at 2, that is, the vector of repeated measures was taken if the baseline value was greater than 2. This corresponds to a situation where only subjects with baseline values >2 are enrolled in a study. For scenario (iii), the baseline value was calculated as that from (ii) divided by an independently generated chi-square random variable with 3 degrees of freedom.

DOI: 10.1002/sim

Table I. Properties of estimates for treatment difference and model adjusted group mean changes between the longitudinal ANCOVA and cLDA models.

	ANCOVA	cLDA
Assumption	Conditional on baseline, post-baseline measures are multivariate normal variables	Baseline and post-baseline measures are jointly multivariate normal variables
Baseline	Is treated as fixed and modeled as a covariate	Is random and modeled as a dependent variable with same means across treatment groups
Missing data handling	Excludes subjects with missing baseline or missing all post-baseline measures	Includes subjects with either baseline or any of of the post-baseline measures
Treatment difference at time T	$\theta_T = E(Y_{AT} - Y_{BT} \mid Y_0)$ $= E(Y_{AT} - Y_{BT})$ Both parameters are the mean difference of measurement at time T between groups; Point estimates from both models are the same when there are no missing data or for $T = 1$ with monotone missing data, but can be different with missing data for $T > 1$; In general, $\hat{\text{var}}(\hat{\eta}_T) \leqslant \hat{\text{var}}(\hat{\theta}_T \mid \underline{Y}_0)$ This inequality is also true for estimated variances obtained from maximum likelihood approaches	$\gamma_T = E(Y_{AT} - Y_{BT})$
Model adjusted group mean changes	$\hat{\theta}_{jT} = (\hat{\alpha}_T - 1)\tilde{Y}_{\bullet\bullet 0} + \hat{\beta}_{jT}$ is an estimate for the conditional mean change, $E(Y_{jT} - Y_{j0} \mid \underline{Y}_0)$, at $\tilde{Y}_{\bullet\bullet 0}$; $\hat{\theta}_{jT}$ is asymptotically unbiased when there are no missing data; but can be biased after excluding subjects with missing data; Estimated variance for $\hat{\theta}_{jT}$ always under estimates the unconditional variance	$\hat{\gamma}_{jT}$ is an estimate for mean change, $E(Y_{jT} - Y_{j0})$; $\hat{\gamma}_{jT}$ is asymptotically unbiased; $\hat{\gamma}_{jT} = \hat{\theta}_{jT}$ when there are no missing data; Estimated variance for $\hat{\gamma}_{jT}$ is asymptotically unbiased

DOI: 10.1002/sim

To study the impact of variability, two values of variances, $\sigma^2 = 1$ and 4, were considered. With a larger variance the resulting distribution from the left-truncation on baseline is more severely deviated from a normal distribution. The mean vectors under the alternative were chosen as $\mu = (3.0\ 2.2\ 1.9\ 1.3)$ and $\mu = (3.0\ 2.0\ 1.5\ 0.7)$ for $\sigma^2 = 1$ and 4, respectively. For scenarios (ii) and (iii), the truncation would change the mean vectors as well as the correlation between baseline and post-baseline measures. The post-baseline means under the truncation would be different from the means without truncation. In general, the impact from the truncation on post-baseline means would be less severe than that on the baseline means because of a regression-to-the mean effect. To ensure that the appropriate parameters were used in calculating bias and confidence interval coverage in the simulations, the true parameter values for treatment difference and group means at last time point were obtained numerically from simulated data from all replications for each scenario. Data for a sample size of 50 subjects per treatment group were generated, so there was approximately 80–90 per cent power under each scenario when no data were missing. Therefore, with a replication of 5000 simulations the true parameters were obtained from 250 000 observations and should have very good precision. For all the analyses, an unstructured covariance matrix and the Satterthwaite approximation for the degrees of freedom were used in both the cLDA and longitudinal ANCOVA models.

Three missing data scenarios were considered (a) no missing data; (b) low to moderate amount of missing data; and (c) moderate to high amount of missing data. The percentages of data missing by time are shown in Table II under different hypotheses. The baseline measurements (time 0) were missing completely at random. To reflect real clinical trials, the percentage of missing baseline data was set to be low; about 2–4 per cent. For post-baseline data, a monotone MAR missing data mechanism was considered. Subjects dropped out of the study when the previous measurement (observed) was greater than a given cut off value. For each scenario of the simulation studies, different cut off values were chosen at each time point and treatment group to have approximate missing data percentages as given in Table II.

In the simulations, all subjects with any available data from baseline to post-baseline were used in the cLDA model. For the ANCOVA model, subjects who had either missing baseline data or missing all post-baseline data were excluded. Results under the null hypothesis are provided in Table III for the treatment difference. Under the alternative hypothesis, the results for treatment difference as well as for the individual treatment group mean estimates are provided in Tables IV and V. The conclusions from the simulation results are as follows:

Table II. Summary of data missing percentage in simulation studies.

	Baseline (per cent)	**Time 1** (per cent)	**Time 2** (per cent)	**Time 3** (per cent)
(1) No missing data	0	0	0	0
(2) Lower/moderate missing proportion				
Placebo/Treatment (under H_0)	~2	~3	~8	~15
Placebo (under H_1)	~2	~3	~8	~15
Treatment (under H_1)	~2	~5	~15	~25
(3) Moderate/high missing proportion				
Placebo/Treatment (under H_0)	~4	~6	~15	~25
Placebo (under H_1)	~4	~6	~15	~25
Treatment (under H_1)	~4	~9	~22	~35

DOI: 10.1002/sim

Table III. Bias, coverage and type I error for treatment difference from the cLDA and longitudinal ANCOVA models under H_0 (5000 replications).

		Bias		Coverage (per cent)		Type I Error* (per cent)	
Var	Distribution	cLDA	ANCOVA	LDA	ANCOVA	cLDA	ANCOVA
No missing data							
$\sigma^2=1$	Normal	0.000	0.000	94.8	95.1	2.4	2.4
	Trunc N	0.000	0.000	94.9	95.1	2.8	2.6
	Trunc T	0.000	0.000	94.9	95.1	2.7	2.7
$\sigma^2=4$	Normal	−0.001	−0.001	94.8	95.1	2.4	2.4
	Trunc N	0.001	0.001	94.5	94.6	2.7	2.7
	Trunc T	0.000	0.000	94.7	94.8	2.6	2.6
With low/moderate missing data							
$\sigma^2=1$	Normal	−0.000	−0.000	94.8	95.0	2.5	2.5
	Trunc N	−0.000	0.000	95.0	94.9	2.9	2.7
	Trunc T	−0.000	−0.000	94.8	94.9	2.8	2.8
$\sigma^2=4$	Normal	−0.000	−0.000	94.8	95.0	2.5	2.5
	Trunc N	0.001	−0.000	94.6	94.8	2.6	2.4
	Trunc T	0.002	0.002	94.4	94.7	2.6	2.5
With moderate/high missing data							
$\sigma^2=1$	Normal	−0.001	−0.002	94.7	94.8	2.6	2.5
	Trunc N	0.001	0.001	94.8	94.8	2.8	2.7
	Trunc T	−0.001	−0.000	94.2	94.5	3.0	2.9
$\sigma^2=4$	Normal	−0.002	−0.003	94.5	95.0	2.6	2.4
	Trunc N	0.002	−0.000	94.3	95.0	2.7	2.4
	Trunc T	0.005	0.003	94.3	94.5	2.8	2.7

*The type I error was based on a one-sided $\alpha=2.5$ per cent. With 5000 replications, a value of 3.1 per cent would be within the 2 times the standard error of the simulation.

For between-treatment difference (Tables III and IV):

- Both the ANCOVA and the cLDA models provide unbiased estimate for the treatment difference in all scenarios.
- The 95 per cent confidence intervals for the treatment difference are adequately covered in all cases.
- Under the null hypothesis, both the ANCOVA and the cLDA models control type-I error rate.
- In general, the estimates and statistical tests are robust against deviation from the normality assumption of the baseline values.
- Under the alternative hypothesis, the cLDA model is consistently more powerful than the ANCOVA model. In the simulated cases, the power gain is less than 1 per cent without missing data, about 1–2 per cent for low/moderate amount of missing data, and about 2–4 per cent for moderate/high amount of missing data.

In the simulated cases shown, only about 2–4 per cent of subjects had missing data at baseline and 3–6 per cent dropped out at time 1. Additional simulations showed that when we increased the missing data at baseline or time 1, the power gain for the cLDA model increased as expected (results not shown).

DOI: 10.1002/sim

Table IV. Bias, coverage and power for treatment difference from the cLDA and longitudinal ANCOVA models under H_1 (5000 replications).

		Bias		Coverage (per cent)		Power (per cent)	
Var	Distribution	cLDA	ANCOVA	cLDA	ANCOVA	cLDA	ANCOVA
No missing data							
$\sigma^2=1$	Normal	−0.001	−0.001	94.8	95.1	88.4	88.0
	Trunc N	0.000	0.000	94.9	95.1	89.7	89.2
	Trunc T	0.001	0.001	95.0	95.3	89.3	89.1
$\sigma^2=4$	Normal	−0.001	−0.001	94.8	95.1	83.9	83.4
	Trunc N	0.001	0.001	94.5	94.6	84.3	83.8
	Trunc T	−0.000	−0.000	94.4	94.6	84.2	83.8
With low/moderate missing data							
$\sigma^2=1$	Normal	−0.001	−0.001	94.3	94.8	81.6	79.6
	Trunc N	0.001	0.001	94.4	94.6	82.4	81.0
	Trunc T	0.001	0.001	94.6	94.7	82.2	80.5
$\sigma^2=4$	Normal	−0.003	−0.003	94.3	94.7	75.4	73.6
	Trunc N	0.001	0.001	94.6	94.8	76.5	74.6
	Trunc T	0.006	0.006	94.6	94.8	75.7	73.8
With moderate/high missing data							
$\sigma^2=1$	Normal	−0.001	−0.002	94.5	949	77.7	74.1
	Trunc N	−0.000	0.000	94.6	94.9	78.3	74.9
	Trunc T	0.001	0.002	94.5	94.7	78.0	74.5
$\sigma^2=4$	Normal	−0.003	−0.004	94.4	94.9	70.4	67.5
	Trunc N	0.005	0.003	94.1	94.5	71.2	67.9
	Trunc T	0.011	0.010	94.3	95.2	70.1	66.7

For model adjusted group mean estimates (Table V)

- In both missing data and no missing data cases, confidence intervals for the model adjusted group mean changes are not covered at the appropriate $100(1-\alpha)$ per cent level in the ANCOVA model, while the confidence intervals from cLDA models provide appropriate coverage.
- When there are no missing data, both models are robust against departures from normality in terms of producing unbiased point estimates. The severity of variance under-estimation in the ANCOVA model depends on the variability of the baseline values and the correlation between baseline and post-baseline values. The latter has more impact on the relative underestimation of variance (in the simulations for the truncated normal or truncated t-distributions, the correlation coefficient between baseline and post-baseline values for $\sigma^2=1$ is smaller than that for $\sigma^2=4$, and with the same σ^2 the correlation coefficient for the truncated t-distribution is smaller than that of the truncated normal distribution, resulting in a relatively more severe variance underestimation).
- In the presence of missing data, the cLDA model provides fairly robust results against mild deviation from normality of the baseline value. The resulting estimates are unbiased and confidence intervals cover at the appropriate $100(1-\alpha)$ per cent level.
- With missing data, the estimates from the ANCOVA model are biased. This is because the baseline mean from the observed data can be biased from the true baseline mean. As such,

DOI: 10.1002/sim

Table V. Bias and coverage for individual treatment group mean estimates from the cLDA and longitudinal ANCOVA models under H_1 (5000 replications).

		Treatment means at last time point			
		Bias		Coverage (per cent)	
Var	Distribution	cLDA	ANCOVA	cLDA	ANCOVA
No missing data					
$\sigma^2=1$	Normal	0.000	0.000	94.9	92.0
	Trunc N	−0.000	−0.000	95.1	93.4
	Trunc T	0.000	0.000	93.7	60.5
$\sigma^2=4$	Normal	0.001	0.001	95.0	92.3
	Trunc N	−0.001	−0.001	94.7	93.5
	Trunc T	0.000	0.000	94.4	76.5
With low/moderate missing data					
$\sigma^2=1$	Normal	0.001	0.072	94.5	90.5
	Trunc N	−0.001	0.066	94.9	92.3
	Trunc T	0.007	0.473	93.1	28.4
$\sigma^2=4$	Normal	0.001	0.147	94.5	90.6
	Trunc N	−0.000	0.118	94.5	92.4
	Trunc T	0.014	0.614	94.2	54.9
With moderate/high missing data					
$\sigma^2=1$	Normal	−0.000	0.125	94.8	88.1
	Trunc N	−0.000	0.110	94.4	89.6
	Trunc T	0.011	0.685	93.5	9.3
$\sigma^2=4$	Normal	0.001	0.249	94.7	88.1
	Trunc N	−0.008	0.206	94.3	90.4
	Trunc T	0.020	0.889	93.3	34.2

the confidence interval can be misleading. The coverage can be severely impacted as a result of the biased parameter estimates. In the simulations, we saw more bias for the case with truncated-t baseline values. This is because the missing data causes more severe baseline differences from excluding those subjects with missing all post-baseline data.

We also conducted simulations under a moderate/high missing data scenario with 'large' sample sizes to see whether the bias and under-coverage problem of the ANCOVA model can be improved. Table VI provides results for $N=250$ subjects per treatment group and $\sigma^2=1$ under the alternative hypothesis. Clearly, the estimates for the treatment difference are unbiased and the confidence interval coverage is adequate for both models. For individual treatment means, the cLDA model shows slight improvement on the bias and confidence interval coverage compared with the results from $N=50$. However, the ANCOVA model still provides biased parameter estimates. The coverage of the confidence interval becomes worse because of the smaller variability from a larger sample size coupled with a biased parameter estimate. In addition, simulations were also done using different correlation values between baseline and post-baseline measurements. The conclusions are similar as above (results not shown). In general, the lower the correlation between baseline and post-baseline values, the worse the under-coverage of the confidence interval for the ANCOVA model. This is consistent with the analytic results given in Appendix.

DOI: 10.1002/sim

Table VI. Bias and coverage for individual treatment group mean estimates with $n=250$, $\sigma^2=1$, under H_1 and moderate/high missing data (2000 replications).

	Bias		Coverage (per cent)	
Distribution	cLDA	ANCOVA	cLDA	ANCOVA
For treatment difference				
Normal	−0.000	0.000	95.8	95.8
Trunc N	0.000	0.001	94.3	94.5
Trunc T	0.001	0.002	94.9	95.3
For treatment mean				
Normal	−0.001	0.124	95.0	68.2
Trunc N	−0.001	0.110	93.6	73.5
Trunc T	0.013	0.701	93.2	0.0

5. AN APPLICATION

To illustrate the difference between the ANCOVA and cLDA models, we apply these methods in a clinical trial for concomitant use of two vaccines. Subjects in the study were randomized into two groups: one received two vaccines concomitantly at months 2, 4, and 6 of age (concomitant use group), another received vaccines in a staggered schedule (staggered use group), that is, one vaccine was given about two to four weeks after another. One of the primary endpoints was the antibody titers measured at about 42 days after the last dose of the test vaccine. Antibody titers were also measured at baseline prior to randomization.

A subset of subjects from one country in the study was taken for this example. In this subset, a total of 366 subjects were randomized, 184 in the concomitant use group, and 182 in the staggered use group. Table VII summarizes the missing data information for these subjects. No subject had both baseline and post-randomization data missing. In general, the drop-out rates were low. The missing data at baseline are mostly due to assay error and are likely missing completely at random. The missing data post-randomization are mostly due to early discontinuation (the causes include adverse experience, lost to follow-up, and withdrawal of consent, etc.) and are likely to be either missing completely at random or missing at random.

The statistical analyses were based on natural log transformed antibody titers measured at baseline and post-baseline. In the ANCOVA model, only the 'completers' were included in the analysis. In contrast, all subjects with data were included in the cLDA model. The estimated geometric mean titer (GMT) and 95 per cent confidence interval for each group, and for the geometric mean ratio between treatment groups are given in Table VIII.

In this example:

- The ANCOVA and cLDA models have similar results for the treatment comparisons in terms of GMT ratio and confidence interval.
- For the individual group GMT estimates, however, the confidence interval from the ANCOVA model is clearly shorter than that from the cLDA model. The two treatment group confidence intervals from the ANCOVA model do not overlap. In contrast, the confidence intervals of the GMT from the cLDA model do overlap. This result is consistent with the theoretical and

DOI: 10.1002/sim

Table VII. Number (per cent) of subjects with missing data in the analysis data set.

Treatment group	Completers	Only baseline missing	Only post-randomization missing	Total N
Concomitant use	172 (93.5 per cent)	9 (4.9 per cent)	3 (1.6 per cent)	184
Staggered use	166 (91.2 per cent)	15 (8.2 per cent)	1 (0.6 per cent)	182

Table VIII. Estimated geometric mean titer (GMT) and 95 per cent confidence interval from ANCOVA and cLDA models.

Method	Concomitant use (95 per cent CI)	Staggered use (95 per cent CI)	GMT ratio* (95 per cent CI)
ANCOVA	1019.5 (823.6, 1261.9)	1611.9 (1297.2, 2002.8)	0.63 (0.47, 0.86)
cLDA	1001.0 (731.7, 1369.5)	1559.6 (1140.0, 2133.6)	0.64 (0.48, 0.86)

*GMT ratio of the concomitant use group over the staggered use group.

simulation results showing that the ANCOVA model produces an under-estimated variance for the individual group mean estimate.

6. DISCUSSION

We discuss two different models to fit longitudinal continuous data for clinical trials where the primary objective is to assess treatments with respect to mean change from baseline. One is the commonly used longitudinal ANCOVA model in which the baseline measurement is included as a covariate in the analysis model. Another is a constrained longitudinal data analysis (cLDA) model in which both baseline and post-baseline measures are considered as dependent variables. In the latter model, the baseline means are constrained to be the same across the treatment groups due to randomization. Both analytic assessment and numerical simulations confirm that the cLDA model is more appropriate for the analysis of this type of data in clinical trials.

One advantage of using cLDA model is that the method provides appropriate variance estimates for the model adjusted group mean estimates, which is often presented in scientific publications with the estimated treatment difference. The model adjusted group mean estimates give a 'best' prediction of the mean outcome for each treatment group. Although the individual group mean estimates are not free from systematic selection bias because the patients enrolled in a clinical trial are, in general, not a random sample from the target population, it provides an estimate of the mean outcome for the population similar to the patients in the study under the trial condition. Compared with the ANCOVA model, the cLDA model provides more appropriate confidence interval estimates for those who choose to present those quantities in scientific research papers.

It is worth noting that the LDA model without the constraint of equal baseline means among treatment groups can also be used. It has been shown (Liang and Zeger [11]) that the constrained model can be more efficient in the context of randomized clinical trials where proper randomization ensures baseline balance (therefore the assumption of equal baseline means is true by design). A similar property is assumed in the commonly used ANCOVA model with one post-baseline measurement [4]. Recently, Overall and Tonidandel [13] investigated the methods on comparing the rate of change with repeated measures when dropouts depend on baseline values.

DOI: 10.1002/sim

In general, under similar modeling conditions, the cLDA model is more efficient than the longitudinal ANCOVA model. The efficiency loss of the ANCOVA model is partially from treating the baseline values as fixed. Without accounting for the variability in the baseline values, the Gauss–Markov theorem does not apply (e.g. Popper [14], Senn *et al.* [15]) and the longitudinal ANCOVA model underestimates the variance of the model adjusted group mean estimates by conditioning on the baseline variables. The severity of the variance under-estimation depends on the amount of variability in the baseline measurement and the amount of correlation between the baseline and post-baseline measures. The latter has more of an impact on the relative underestimation. When the missing data depend on the baseline value, the model adjusted group mean estimates from the ANCOVA model may also be biased. Simulation results show that when baseline values deviate from a normal distribution or have lower correlation with post-baseline values, the variability of the model adjusted group mean estimates is more severely underestimated by the longitudinal ANCOVA model. The cLDA model can overcome this drawback as it is based on a full likelihood function that takes the variability of the baseline measures into consideration. Therefore, the cLDA model provides appropriate variance and confidence interval estimates. The cLDA model also provides more flexibility in handling missing data by including all observed data, which, in general, results in more power when testing treatment differences compared with the longitudinal ANCOVA model. In the cLDA model, it is also possible to model the correlation between baseline and post-baseline measures differently for different treatment groups. This helps address the differential baseline-dependent dropout issues as discussed in Overall and Tonidandel [13].

We have assumed that the baseline values are random throughout this paper. It should be noted that if the baseline values are considered as fixed, then the longitudinal ANCOVA model is appropriate. This ANCOVA model answers the question of what is the treatment effect for patients with the baseline values fixed as in the current clinical trial.

In this paper, we have considered the parameter of interest to be the mean change from baseline effect at a given time point such as the last visit time point T. In clinical trials, this parameter is often estimated using a full analysis set that includes all subjects randomized and having at least one measurement during the study. This parameter measures the treatment effect for a subject if the subject takes the treatment up to time T. This is the scientific question typically of interest to a sponsor. In reality, some subjects that entered the study may never reach the time point T due to premature discontinuation or death. There are philosophical and public health debates on whether this parameter is of interest in a public health setting. These questions are out of the scope for this paper. Some alternative approaches have been proposed in the literature (e.g. Shih and Quan [16]).

APPENDIX A: PROPERTIES OF TREATMENT DIFFERENCE AND GROUP MEANS FOR ANCOVA AND cLDA MODELS

Because missing data at baseline are usually missing completely at random, we will consider the case where the baseline is always observable while some post-baseline values might be missing. Let g index the observed data pattern for a subject, i.e. g is the set of post-baseline time points at which the measurement was made on the subject. Let $\dim(g)$ denote the number of elements in g. To define the incidence matrix, I_g, write $g=\{t_1,\ldots,t_{\dim(g)}\}$, then I_g is a $\dim(g)\times T$ matrix of zeros and ones such that $I_g[i,t_i]=1$ for $i=1,\ldots,\dim(g)$. Let n_{jg} denote the number of subjects in treatment group j with observed data pattern g, P_{jg} the set of subjects in treatment group

DOI: 10.1002/sim

j with observed data pattern g, and G_j the collection of observed data patterns for treatment group $j\,(j=A, B)$.

For the cLDA model, write the mean parameters as $\psi=(\gamma_0, \eta_{A1}, \ldots, \eta_{AT}, \eta_{B1}, \ldots, \eta_{BT})'$, then the design matrix for a subject with observed data pattern g can be written as

$$X_{Ag}=\begin{pmatrix}1 & 0 & 0\\ 0 & I_g & 0\end{pmatrix}, \quad X_{Bg}=\begin{pmatrix}1 & 0 & 0\\ 0 & 0 & I_g\end{pmatrix}$$

for treatment groups A and B, respectively, where zero entries have proper dimensions so that X_{jg} is a $(\dim(g)+1)\times(2T+1)$ matrix. The observed response vector for a subject with observed data pattern g is given by

$$y_{ijg}=\begin{pmatrix} y_{ij0}\\ I_g y_{ij1}\end{pmatrix}$$

where y_{ij1} denotes the $T\times 1$ vector of post-baseline values. We first assume that the covariance matrix of y_{ijg} is known. Partition the covariance matrix according to the baseline and post-baseline time points,

$$\Sigma=\begin{pmatrix}\sigma_{00} & \Sigma_{01}\\ \Sigma_{10} & \Sigma_{11}\end{pmatrix}$$

then the covariance matrix for a subject with observed data pattern g can be written as

$$V_g=\begin{pmatrix}\sigma_{00} & \Sigma_{01}I_g'\\ I_g\Sigma_{10} & I_g\Sigma_{11}I_g'\end{pmatrix}$$

It follows that

$$V_g^{-1}=\begin{pmatrix}\sigma_{00}^{-1}+\sigma_{00}^{-1}\Sigma_{01}I_g'(I_g\Sigma_{11\cdot 0}I_g')^{-1}I_g\Sigma_{10}\sigma_{00}^{-1} & -\sigma_{00}^{-1}\Sigma_{01}I_g'(I_g\Sigma_{11\cdot 0}I_g')^{-1}\\ -(I_g\Sigma_{11\cdot 0}I_g')^{-1}I_g\Sigma_{10}\sigma_{00}^{-1} & (I_g\Sigma_{11\cdot 0}I_g')^{-1}\end{pmatrix}$$

where $\Sigma_{11\cdot 0}=\Sigma_{11}-\Sigma_{10}\sigma_{00}^{-1}\Sigma_{01}$ is the conditional covariance matrix of post-baseline measurements given the baseline. With some algebra, it can be shown that the covariance matrix for the mean parameters estimate is given by

$$\mathrm{var}(\hat{\psi})=\left(\sum_{j=A}^{B}\sum_{g\in G_j} n_{jg}X_{jg}'V_g^{-1}X_{jg}\right)^{-1}=\begin{pmatrix}z & C_A' & C_B'\\ C_A & D_A & 0\\ C_B & 0 & D_B\end{pmatrix}^{-1}$$

DOI: 10.1002/sim

where

$$z=\sum_{j=A}^{B}\sum_{g\in G_j} n_{jg}(\sigma_{00}^{-1}+\sigma_{00}^{-1}\Sigma_{01}W_g\Sigma_{10}\sigma_{00}^{-1})$$

$$C_j=-\sum_{g\in G_j} n_{jg}W_g\Sigma_{10}\sigma_{00}^{-1}$$

$$D_j=\sum_{g\in G_j} n_{jg}W_g$$

and $W_g=I_g'(I_g\Sigma_{11\cdot 0}I_g')^{-1}I_g$. By the formula for inverting block matrices, we have

$$\mathrm{var}(\hat{\psi})=\begin{pmatrix}(n_A+n_B)^{-1}\sigma_{00} & (n_A+n_B)^{-1}\Sigma_{01} & (n_A+n_B)^{-1}\Sigma_{01}\\ (n_A+n_B)^{-1}\Sigma_{10} & D_A^{-1}+(n_A+n_B)^{-1}\Sigma_{10}\sigma_{00}^{-1}\Sigma_{01} & (n_A+n_B)^{-1}\Sigma_{10}\sigma_{00}^{-1}\Sigma_{01}\\ (n_A+n_B)^{-1}\Sigma_{10} & (n_A+n_B)^{-1}\Sigma_{10}\sigma_{00}^{-1}\Sigma_{01} & D_B^{-1}+(n_A+n_B)^{-1}\Sigma_{10}\sigma_{00}^{-1}\Sigma_{01}\end{pmatrix}$$

Let $\eta_j=(\eta_{j1},\ldots,\eta_{jT})'$, then the variances for between-group differences and within-group mean changes are given by

$$\mathrm{var}(\hat{\eta}_A-\hat{\eta}_B)=D_A^{-1}+D_B^{-1} \tag{A1}$$

$$\mathrm{var}(\hat{\eta}_j-1_T\hat{\gamma}_0)=D_j^{-1}+(n_A+n_B)^{-1}(1_T-\Sigma_{10}\sigma_{00}^{-1})\sigma_{00}(1_T-\Sigma_{10}\sigma_{00}^{-1})' \tag{A2}$$

For the point estimate, we have

$$\sum_{j=A}^{B}\sum_{g\in G_j}\sum_{i\in P_{jg}} X_{jg}'V_g^{-1}y_{ijg}=\begin{pmatrix}\sum_{j=A}^{B}\{\sigma_{00}^{-1}n_j\bar{y}_{\cdot j0}-\sigma_{00}^{-1}\Sigma_{01}(F_{j1}-F_{j0}\Sigma_{10}\sigma_{00}^{-1})\}\\ F_{A1}-F_{A0}\Sigma_{10}\sigma_{00}^{-1}\\ F_{B1}-F_{B0}\Sigma_{10}\sigma_{00}^{-1}\end{pmatrix}$$

where

$$F_{jt}=\sum_{g\in G_j}\sum_{i\in P_{jg}} W_g y_{ijt}$$

for $j=A, B$ and $t=0, 1$. It follows that

$$\hat{\gamma}_0=\bar{y}_{\cdot\cdot 0},\quad \hat{\eta}_j=(M_{j1}-M_{j0}\Sigma_{10}\sigma_{00}^{-1})+\Sigma_{10}\sigma_{00}^{-1}\bar{y}_{\cdot\cdot 0}$$

where $M_{jt}=D_j^{-1}F_{jt}$, M_{j0} and M_{j1} can be viewed as matrix-weighted averages of baseline and post-baseline values, respectively. Hence, the estimates for between-group differences and within-group mean changes are given by

$$\hat{\eta}_A-\hat{\eta}_B=(M_{A1}-M_{B1})-(M_{A0}-M_{B0})\Sigma_{10}\sigma_{00}^{-1} \tag{A3}$$

$$\hat{\eta}_j-1_T\hat{\gamma}_0=(M_{j1}-M_{j0}\Sigma_{10}\sigma_{00}^{-1})-(1_T-\Sigma_{10}\sigma_{00}^{-1})\bar{y}_{\cdot\cdot 0} \tag{A4}$$

DOI: 10.1002/sim

Now consider the ANCOVA model with parameters $\phi=(\alpha_1,\ldots,\alpha_T,\beta_{A1},\ldots,\beta_{AT},\beta_{B1},\ldots,\beta_{BT})'$. The design matrix for a subject with observed data pattern g can be written as

$$X_{iAg}=(y_{iA0}I_g\ \ I_g\ \ 0),\quad X_{iBg}=(y_{iB0}I_g\ \ 0\ \ I_g)$$

for treatment groups A and B, respectively. The observed response vector and the corresponding covariance matrix for a subject with observed data pattern g are given by

$$y_{ijg}=I_g\ \ y_{ij1},\quad V_g=I_g\Sigma_{11\cdot 0}I_g'$$

Let $\beta_j=(\beta_{j1},\ldots,\beta_{jT})'$, then following similar steps as outlined for the cLDA model, we obtain

$$\mathrm{var}(\hat{\phi})=\begin{pmatrix} Q_0^{-1} & -Q_0^{-1}M_{A0}' & -Q_0^{-1}M_{B0}' \\ -M_{A0}Q_0^{-1} & D_A^{-1}+M_{A0}Q_0^{-1}M_{A0}' & M_{A0}Q_0^{-1}M_{B0}' \\ -M_{B0}Q_0^{-1} & M_{B0}Q_0^{-1}M_{A0}' & D_B^{-1}+M_{B0}Q_0^{-1}M_{B0}' \end{pmatrix}$$

and

$$\hat{\alpha}=Q_0^{-1}Q_{01},\quad \hat{\beta}_j=M_{j1}-M_{j0}Q_0^{-1}Q_{01}$$

where

$$Q_0=\sum_{j=A}^{B}\sum_{g\in G_j}\sum_{i\in P_{jg}}(y_{ij0}I_T-M_{j0})'W_g(y_{ij0}I_T-M_{j0})$$

$$Q_{01}=\sum_{j=A}^{B}\sum_{g\in G_j}\sum_{i\in P_{jg}}(y_{ij0}I_T-M_{j0})'W_g(y_{ij1}-M_{j1})$$

Note that $(n_A+n_B)^{-1}Q_0$ can be viewed as matrix-weighted variance for the baseline values, and $(n_A+n_B)^{-1}Q_{01}$ can be viewed as matrix-weighted covariance between baseline and post-baseline values. Consequently, for the between-group differences, we have

$$\hat{\beta}_A-\hat{\beta}_B=(M_{A1}-M_{B1})-(M_{A0}-M_{B0})Q_0^{-1}Q_{01} \tag{A5}$$

$$\mathrm{var}(\hat{\beta}_A-\hat{\beta}_B)=D_A^{-1}+D_B^{-1}+(M_{A0}-M_{B0})'Q_0^{-1}(M_{A0}-M_{B0}) \tag{A6}$$

For the within-group mean changes, we have

$$\hat{\beta}_j+(\hat{\alpha}-1_T)\tilde{y}_{\cdot\cdot 0}=(M_{j1}-M_{j0}Q_0^{-1}Q_{01})-(1_T-Q_0^{-1}Q_{01})\tilde{y}_{\cdot\cdot 0} \tag{A7}$$

$$\mathrm{var}\{\hat{\beta}_j+(\hat{\alpha}-1_T)\tilde{y}_{\cdot\cdot 0}\}=D_j^{-1}+(\tilde{y}_{\cdot\cdot 0}I_T-M_{j0})Q_0^{-1}(\tilde{y}_{\cdot\cdot 0}I_T-M_{j0})' \tag{A8}$$

where $\tilde{y}_{\cdot\cdot 0}$ is the mean baseline value of all subjects included in the ANCOVA model, and the variances are calculated conditional on the baseline values.

Comparing (A1) and (A6), we see that the variance for treatment difference from the cLDA model is always less than (or equal to if $M_{A0}=M_{B0}$, which is unlikely due to the randomness of y_{ij0} and potentially different observed data patterns between the two treatment groups) that from the ANCOVA model provided that the baseline value is not missing. In the above derivations,

DOI: 10.1002/sim

we assumed that the covariance matrix is known. The results are also true when the covariance matrix is estimated using maximum likelihood approach. The estimates of $\Sigma_{11\cdot 0}$ from the two models are the same because both models use the conditional likelihood of post-baseline values given the baseline for the estimation of $\Sigma_{11\cdot 0}$ (there may be some slight numeric differences in the estimates of $\Sigma_{11\cdot 0}$ if the residual maximum likelihood (REML) method is used for estimating the covariance parameters, but the extra term in (A6) generally more than makes up for such numeric differences).

Consider the special case when there are no missing data. Let

$$S_{uv}=(n_A+n_B)^{-1}\sum_{j=A}^{B}\sum_{i=1}^{n_j}(y_{iju}-\bar{y}_{\cdot ju})(y_{ijv}-\bar{y}_{\cdot jv})',\quad u,v=0,1$$

then, for the between-group differences, we have

$$\hat{\eta}_A-\hat{\eta}_B=\hat{\beta}_A-\hat{\beta}_B=(\bar{y}_{\cdot A1}-\bar{y}_{\cdot B1})-\hat{\alpha}(\bar{y}_{\cdot A0}-\bar{y}_{\cdot B0}) \tag{A9}$$

$$\hat{\mathrm{var}}_{\mathrm{cLDA}}(\hat{\eta}_A-\hat{\eta}_B)=\left(\frac{1}{n_A}+\frac{1}{n_B}\right)\hat{\Sigma}_{11\cdot 0} \tag{A10}$$

$$\hat{\mathrm{var}}_{\mathrm{ANCOVA}}(\hat{\beta}_A-\hat{\beta}_B)=\left\{\frac{1}{n_A}+\frac{1}{n_B}+\frac{(\bar{y}_{\cdot A0}-\bar{y}_{\cdot B0})^2}{(n_A+n_B)S_{00}}\right\}\hat{\Sigma}_{11\cdot 0} \tag{A11}$$

and for the within-group mean changes, we have

$$\hat{\eta}_j-1_T\hat{\gamma}_0=\hat{\beta}_j+(\hat{\alpha}-1_T)\bar{y}_{\cdot\cdot 0}=(\bar{y}_{\cdot j1}-\hat{\alpha}\bar{y}_{\cdot j0})-(1_T-\hat{\alpha})\bar{y}_{\cdot\cdot 0} \tag{A12}$$

$$\hat{\mathrm{var}}_{\mathrm{cLDA}}(\hat{\eta}_j-1_T\hat{\gamma}_0)=\frac{1}{n_j}\hat{\Sigma}_{11\cdot 0}+\frac{1}{n_A+n_B}(1_T-\hat{\alpha})\hat{\sigma}_{00}(1_T-\hat{\alpha})' \tag{A13}$$

$$\hat{\mathrm{var}}_{\mathrm{ANCOVA}}\{\hat{\beta}_j+(\hat{\alpha}-1_T)\bar{y}_{\cdot\cdot 0}\}=\left\{\frac{1}{n_j}+\frac{(\bar{y}_{\cdot j0}-\bar{y}_{\cdot\cdot 0})^2}{(n_A+n_B)S_{00}}\right\}\hat{\Sigma}_{11\cdot 0} \tag{A14}$$

where

$$\hat{\alpha}=S_{10}/S_{00},\quad \hat{\Sigma}_{11\cdot 0}=S_{11}-S_{10}S_{01}/S_{00}$$

$$\hat{\sigma_{00}}=(n_A+n_B)^{-1}\sum_{j=A}^{B}\sum_{i=1}^{n_j}(y_{ij0}-\bar{y}_{\cdot 00})^2$$

.

Therefore, the point estimates for between-group differences from the two models are identical, and the variances are asymptotically equivalent because $(\bar{y}_{\cdot A0}-\bar{y}_{\cdot B0})^2/S_{00}\to 0$ as $n_A, n_B\to\infty$. However, although the point estimates for within-group mean changes are identical for the two models, the ANCOVA model underestimates the true variability because $(\bar{y}_{\cdot j0}-\bar{y}_{\cdot\cdot 0})^2/S_{00}\to 0$, $(1_T-\hat{\alpha})\hat{\sigma}_{00}(1_T-\hat{\alpha})'\to(1_T-\Sigma_{10}\sigma_{00}^{-1})\sigma_{00}(1_T-\Sigma_{10}\sigma_{00}^{-1})'$ as $n_A, n_B\to\infty$.

DOI: 10.1002/sim

Now consider the special case when $T=1$. Let k_j denote the number of subjects with post-baseline value in treatment group $j\,(j=A,B)$, and

$$\tilde{y}_{\cdot jt}=k_j^{-1}\sum_{i=1}^{k_j}y_{ijt},\quad j=A,B,\ t=0,1$$

$$S_{uv}=(k_A+k_B)^{-1}\sum_{j=A}^{B}\sum_{i=1}^{k_j}(y_{iju}-\tilde{y}_{\cdot ju})(y_{ijv}-\tilde{y}_{\cdot jv})',\quad u,v=0,1$$

then, for the between-group difference, we have

$$\hat{\eta}_A-\hat{\eta}_B=\hat{\beta}_A-\hat{\beta}_B=(\tilde{y}_{\cdot A1}-\tilde{y}_{\cdot B1})-\hat{\alpha}(\tilde{y}_{\cdot A0}-\tilde{y}_{\cdot B0}) \tag{A15}$$

$$\hat{\mathrm{var}}_{\mathrm{cLDA}}(\hat{\eta}_A-\hat{\eta}_B)=\left(\frac{1}{k_A}+\frac{1}{k_B}\right)\hat{\sigma}_{11\cdot 0} \tag{A16}$$

$$\hat{\mathrm{var}}_{\mathrm{ANCOVA}}(\hat{\beta}_A-\hat{\beta}_B)=\left\{\frac{1}{k_A}+\frac{1}{k_B}+\frac{(\tilde{y}_{\cdot A0}-\tilde{y}_{\cdot B0})^2}{(k_A+k_B)S_{00}}\right\}\hat{\sigma}_{11\cdot 0} \tag{A17}$$

and for the within-group mean changes, we have

$$\hat{\eta}_j-\hat{\gamma}_0=(\tilde{y}_{\cdot j1}-\hat{\alpha}\tilde{y}_{\cdot j0})-(1-\hat{\alpha})\bar{y}_{\cdot\cdot 0} \tag{A18}$$

$$\hat{\mathrm{var}}_{\mathrm{cLDA}}(\hat{\eta}_j-\hat{\gamma}_0)=\frac{1}{k_j}\hat{\sigma}_{11\cdot 0}+\frac{1}{n_A+n_B}(1-\hat{\alpha})^2\hat{\sigma}_{00} \tag{A19}$$

$$\hat{\beta}_j+(\hat{\alpha}-1)\tilde{y}_{\cdot\cdot 0}=(\tilde{y}_{\cdot j1}-\hat{\alpha}\tilde{y}_{\cdot j0})-(1-\hat{\alpha})\tilde{y}_{\cdot\cdot 0} \tag{A20}$$

$$\hat{\mathrm{var}}_{\mathrm{ANCOVA}}\{\hat{\beta}_j+(\hat{\alpha}-1)\tilde{y}_{\cdot\cdot 0}\}=\left\{\frac{1}{k_j}+\frac{(\tilde{y}_{\cdot j0}-\tilde{y}_{\cdot\cdot 0})^2}{(k_A+k_B)S_{00}}\right\}\hat{\sigma}_{11\cdot 0} \tag{A21}$$

where

$$\hat{\alpha}=S_{10}/S_{00},\quad \hat{\sigma}_{11\cdot 0}=S_{11}-S_{10}S_{01}/S_{00}$$

$$\hat{\sigma}_{00}=(n_A+n_B)^{-1}\sum_{j=A}^{B}\sum_{i=1}^{n_j}(y_{ij0}-\bar{y}_{\cdot 00})^2$$

.

Therefore, the point estimates for between-group differences from the two models are identical, and the variances are asymptotically equivalent if the probability of missing post-baseline value does not depend on the treatment group so that $\tilde{y}_{\cdot A0},\tilde{y}_{\cdot B0}\to\tilde{\gamma}_0$, where $\tilde{\gamma}_0$ denotes the asymptotic mean among completers. However, if the probability of missing post-baseline value depends on the baseline, we have $\tilde{\gamma}_0\neq\gamma_0$, hence the point estimates for within-group mean changes from the ANCOVA are biased.

Assume $n_j/(n_A+n_B)\to\pi_j$, $k_j/n_j\to\phi_j$ as $n_A\to\infty$ and $n_B\to\infty$, then the asymptotic variance of the within-group mean from ANCOVA model is $\sigma_{11\cdot 0}/(\pi_j\phi_j)$, which is always smaller than $\sigma_{11\cdot 0}/(\pi_j\phi_j)+(1-\sigma_{01}/\sigma_{00})^2\sigma_{00}$, the asymptotic variance of the within-group mean from the

DOI: 10.1002/sim

cLDA model. When $\sigma_{11}=\sigma_{00}, \sigma_{11\cdot 0}=\sigma_{11}-\sigma_{01}^2/\sigma_{00}=\sigma_{11}(1-\rho^2)$, therefore the under estimation of the variance for ANCOVA model is $(1-\sigma_{01}/\sigma_{00})^2\sigma_{00}=(1-\rho)^2\sigma_{00}$, which is proportional to the baseline variance.

ACKNOWLEDGEMENTS

We thank the associate editor and three reviewers for their constructive comments, suggestions and providing additional references that led to significant improvement of the paper.

REFERENCES

1. Crager MR. Analysis of covariance in parallel-group clinical trials with pretreatment baselines. *Biometrics* 1987; **43**:895–901.
2. Senn SJ. Covariate imbalance and random allocation in clinical trials. *Statistics in Medicine* 1989; **8**:467–475.
3. Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Statistics in Medicine* 1992; **11**:1685–1704.
4. Wei L, Zhang J. Analysis of data with imbalance in the baseline outcome variable for randomized clinical trials. *Drug Information Journal* 2001; **35**:1201–1214.
5. Vickers AJ, Altman DG. Analysing controlled trials with baseline and follow up measurements. *British Medical Journal* 2001; **323**:1123–1124.
6. Chambless LE, Roeback JR. Methods for assessing difference between groups in change when initial measurement is subject to intra-individual variation. *Statistics in Medicine* 1993; **12**:1213–1237.
7. Yanez ND, Kronmal RA, Shemanski LR. The effects of measurement error in response variables and tests of association of explanatory variables in change models. *Statistics in Medicine* 1998; **17**:2597–2606.
8. Chan SF, Macaskill P, Irwig L, Walter SD. Adjustment for baseline measurement error in randomized controlled trials induces bias. *Controlled Clinical Trials* 2004; **25**:408–416.
9. Senn S. Letter and comments to editor Re: methods for assessing difference between groups in change when initial measurement is subject to intra-individual variation. *Statistics in Medicine* 1994; **13**:2280–2285.
10. Senn S. Change from baseline and analysis of covariance revisited. *Statistics in Medicine* 2006; **25**:4334–4344.
11. Liang KY, Zeger S. Longitudinal data analysis of continuous and discrete responses for pre–post designs. *Sankhyā*: *The Indian Journal of Statistics* (*Series B*) 2000; **62**:134–148.
12. Senn S, Stevens L, Chaturvedi N. Repeated measures in clinical trials: simple strategies for analysis using summary measures. *Statistics in Medicine* 2000; **19**:861–877.
13. Overall JE, Tonidandel S. Analysis of data from a controlled repeated measurements design with baseline-dependent dropouts. *Methodology*: *European Journal of Research Methods for the Behavioral and Social Sciences* 2007; **3**:58–66.
14. Popper SJ. The Gauss–Markov theorem and random regressors. *The American Statistician* 1991; **45**:269–272.
15. Senn SE, Graf E, Caputo A. Stratification for the propensity score compared with linear regression techniques to assess the effect of treatment or exposure. *Statistics in Medicine* 2007; **26**:5529–5544.
16. Shih WJ, Quan H. Testing for treatment differences with dropouts present in clinical trials—a composite approach. *Statistics in Medicine* 1997; **16**:1225–1239.

DOI: 10.1002/sim