Research Article

Statistics in Medicine

Received 2 March 2009, Accepted 9 November 2009 Published online 17 December 2009 in Wiley Interscience

(www.interscience.wiley.com) DOI: 10.1002/sim.3820

Specification of covariance structure in longitudinal data analysis for randomized clinical trials

Kaifeng Lu[a]*† and Devan V. Mehrotra[b]

Misspecification of the covariance structure for repeated measurements in longitudinal analysis may lead to biased estimates of the regression parameters and under or overestimation of the corresponding standard errors in the presence of missing data. The so-called sandwich estimator can 'correct' the variance, but it does not reduce the bias in point estimates. Removing all assumptions from the covariance structure (i.e. using an unstructured (UN) covariance) will remove such biases. However, an excessive amount of missing data may cause convergence problems for iterative algorithms, such as the default Newton–Raphson algorithm in the popular SAS PROC MIXED. This article examines, both through theory and simulations, the existence and the magnitude of these biases. We recommend the use of UN covariance as the default strategy for analyzing longitudinal data from randomized clinical trials with moderate to large number of subjects and small to moderate number of time points. We also present an algorithm to assist in the convergence when the UN covariance is used. Copyright © 2009 John Wiley & Sons, Ltd.

Keywords: covariance model; missing at random; missing completely at random; repeated measures; sandwich variance estimator

1. Introduction

Longitudinal data analysis has seen increasing use in randomized clinical trials [1–5]. The compound symmetry (CS) covariance structure (equal variances and equal pairwise correlations across fixed time points) is implicit in the classic linear model analyses of longitudinal data (e.g. split-plot analysis, linear mixed model with random subject effects, etc.). Classic analyses can provide incorrect inference if the CS assumption is violated [6, 7]. Approximate alternative tests based on adjusted degrees of freedom were proposed to 'protect' against departures from CS [8, 9]. The analyses were implemented in SAS PROC GLM [10]. More recently, it was suggested that prior experience or observed data can be used to select an appropriate covariance structure [11, 12]. SAS PROC MIXED was developed providing a variety of covariance structures including the commonly used CS, AR(1) (first-order autoregressive, equal variances and exponentially decreasing correlations), Toeplitz (equal variances and a separate correlation for each level of separation between the time points), and unstructured (UN). Repeated measurements in clinical trials are usually collected at a fixed and relatively small set of time points, which allows the variances and covariances to be modeled using the REPEATED statement of PROC MIXED.

Frison and Pocock [13] studied data from several trials covering a variety of diseases and outcome measures and reported no 'major' departure from the CS assumption. On the other hand, responses that are close together in time often have a higher correlation than those that are far apart; hence, the AR(1) model is a 'natural' choice from a time-series viewpoint when time periods are evenly spaced [14–16]. The Toeplitz structure includes both CS and AR(1) as special cases and may serve as a compromise between the assumption-free UN covariance model and the often overly simplistic CS or AR(1) model for longitudinal studies.

There are two approaches to making a model choice of the covariance structure. One is to compare models based on measures of fit that are adjusted for the number of covariance parameters, e.g. Akaike's information criterion (AIC) [17] or Schwarz's Bayesian information criterion (BIC) [18]. Another is to use likelihood ratio tests (LRT) to find whether additional parameters provide a statistically significant improvement in the model fit. The LRT approach can only be used when the covariance models are nested (e.g. CS is nested within Toeplitz, which, in turn, is nested within UN). Both approaches are theoretically appealing, but empirical

[a] *Clinical Biostatistics, Merck Research Laboratories, Rahway, NJ 07065, U.S.A.*
[b] *Clinical Biostatistics, Merck Research Laboratories, Upper Gwynedd, PA 19454, U.S.A.*
* *Correspondence to: Kaifeng Lu, Clinical Biostatistics, Merck Research Laboratories, Rahway, NJ 07065, U.S.A.*
† *E-mail: kaifeng_lu@merck.com*

evidence pointed to low success rates in selecting the correct covariance for short series lengths, small sample sizes, complex covariance structures, or rich sets of candidates [19, 20]. Type I error rates for the best AIC and BIC models were always higher than the target values [21]. In addition, 'optimal' modeling of the covariance structure is often unnecessary because the use of an UN covariance typically results in a negligible loss in efficiency when the sample sizes are reasonably large, as is the case with the late-stage clinical trials that motivated our research; we reinforce this important point later in the article.

For general linear models, the maximum likelihood (ML) estimator of the regression parameters $\boldsymbol{\beta}$ (given the variance–covariance matrix) is the same as the generalized least-squares (GLS) or generalized estimating equations (GEE) estimator of $\boldsymbol{\beta}$. An appealing property of the GEE estimator is that it yields a consistent estimate of $\boldsymbol{\beta}$ even if the assumed model for the covariances among the repeated measures is incorrect provided that there are no missing data or missing data are missing completely at random (MCAR). However, the model-based variance estimator of Cov($\hat{\boldsymbol{\beta}}$) is biased in this case; and valid inferences about $\boldsymbol{\beta}$ should be based on the so-called 'sandwich' (or 'robust') estimator of Cov($\hat{\boldsymbol{\beta}}$) [22–24]. The closer the 'working' covariance matrix approximates the true underlying covariance matrix, the greater the efficiency with which $\boldsymbol{\beta}$ can be estimated.

When missing data are missing at random (MAR), but not MCAR, standard GEE methods based on 'working' covariance matrix yield biased estimates of $\boldsymbol{\beta}$. To avoid such bias, an UN covariance is often used. However, an excessive amount of missing data can cause convergence issues for iterative numerical algorithms, such as the default Newton–Raphson algorithm used in SAS PROC MIXED. A more specific covariance structure may be adopted in such cases, in conjunction with the use of sandwich variance estimator. The sandwich can help on the variance side, but it does not affect the bias of the mean parameter estimates [25].

The following issues are investigated in this article, each in the presence of the missing data: (1) biases in the regression parameter estimates as the result of misspecification of the covariance structure; (2) comparison of the model-based and sandwich variance estimators; and (3) approaches to the convergence problem with UN covariance.

The remainder of this article is organized as follows. Notation and assumptions are described in Section 2. Sections 3 examines the first two issues for randomized, two-group longitudinal trials with incomplete bivariate and trivariate normal data. The results from the simulation studies are presented in Section 4. Real clinical trial data are analyzed in Section 5. The convergence problem is investigated in Section 6, and the article concludes with a brief discussion in Section 7.

2. Notation and assumptions

In randomized clinical trials, patients are often evaluated at the same set of time points and the number of time points is usually small. Let $\mathbf{A}$ denote the set of baseline variables including treatment assignment and suppose we have a total of T time points at which the measurement of a continuous endpoint is to be taken. Let $\mathbf{y}_i^{\mathrm{F}}=(y_{i1},\ldots,y_{iT})'$ denote the full data of repeated measurements from subject i. In the presence of missing data, say, as a result of patient drop-out, let R_{it} indicate whether the tth component, y_{it}, of $\mathbf{y}_i^{\mathrm{F}}$ is observed, i.e. $R_{it}=1$ if y_{it} is observed, and $R_{it}=0$ if y_{it} is missing. Therefore, $\mathbf{R}_i=(R_{i1},\ldots,R_{iT})'$ denotes the missing data pattern for subject i. We also denote the observed and the unobserved components of $\mathbf{y}_i^{\mathrm{F}}$ by $\mathbf{y}_i^{\mathrm{obs}}$ and $\mathbf{y}_i^{\mathrm{mis}}$, respectively. For instance, if $T=4$ and $\mathbf{R}_i=(1,1,1,0)'$, then $\mathbf{y}_i^{\mathrm{obs}}=(y_{i1},y_{i2},y_{i3})'$ and $\mathbf{y}_i^{\mathrm{mis}}=y_{i4}$. We observe n independent and identically distributed random vectors $\{(\mathbf{R}_i,\mathbf{y}_i^{\mathrm{obs}},\mathbf{A}_i):i=1,\ldots,n\}$. We assume that missing data are MAR; that is, the probability of $(\mathbf{R}_i=\mathbf{r})$ given $(\mathbf{y}_i^{\mathrm{F}},\mathbf{A}_i)$ does not depend on the unobserved components $\mathbf{y}_i^{\mathrm{mis}}$ for any given vector $\mathbf{r}$ of 0's and 1's. Furthermore, we assume that the parameters for the missing data mechanism are distinct from the parameters for the full data distribution, i.e. we assume an ignorable missing data mechanism so that the inference on the parameters for the full data distribution can be based on the observed data likelihood.

Assume $\mathbf{y}_i^{\mathrm{F}}|\mathbf{A}_i\sim\mathrm{N}(\mathbf{X}_i^{\mathrm{F}}\boldsymbol{\beta},\mathbf{V}_i^{\mathrm{F}})$, where $\mathbf{X}_i^{\mathrm{F}}$ denotes the design matrix constructed from the baseline covariates $\mathbf{A}_i$ and time points, and $\mathbf{V}_i^{\mathrm{F}}$ the covariance matrix for the full response vector $\mathbf{y}_i^{\mathrm{F}}$. Hereafter, for the ease of exposition, we write $\mathbf{y}_i^{\mathrm{obs}}$ as $\mathbf{y}_i$, and let $\mathbf{X}_i$ and $\mathbf{V}_i$ denote the design matrix and covariance matrix corresponding to $\mathbf{y}_i$. We have $\mathbf{y}_i|\mathbf{A}_i\sim\mathrm{N}(\mathbf{X}_i\boldsymbol{\beta},\mathbf{V}_i)$. The observed data ML estimator of $\boldsymbol{\beta}$ is the same as the GLS estimator, $\hat{\boldsymbol{\beta}}=(\sum_{i=1}^n\mathbf{X}_i'\hat{\mathbf{V}}_i^{-1}\mathbf{X}_i)^{-1}\sum_{i=1}^n\mathbf{X}'\hat{\mathbf{V}}_i^{-1}\mathbf{y}_i$, which is also the GEE solution to the observed data score equation, $\sum_{i=1}^n\mathbf{X}_i'\hat{\mathbf{V}}_i^{-1}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})=0$, where $\hat{\mathbf{V}}=\mathbf{V}(\hat{\boldsymbol{\omega}})$, and $\hat{\boldsymbol{\omega}}$ is the estimate of the variance–covariance parameters $\boldsymbol{\omega}$. For example, if $\mathbf{V}$ is UN, then $\boldsymbol{\omega}=\{\sigma_{st}:1\leqslant s\leqslant t\leqslant T\}$, i.e. no assumption is made on the structure of the covariance matrix; if $\mathbf{V}$ has an AR(1) structure, then $\boldsymbol{\omega}=(\sigma^2,\rho)$ and $\sigma_{st}=\sigma^2\rho^{|s-t|}$; if $\mathbf{V}$ has a CS structure, then $\boldsymbol{\omega}=\{\sigma^2,\rho\}$ and $\sigma_{st}=\sigma^2 I(s=t)+\sigma^2\rho I(s\neq t)$; if $\mathbf{V}$ has a Toeplitz structure, then $\boldsymbol{\omega}=\{\omega_1,\ldots,\omega_T\}$, where $\sigma_{st}=\omega_{|s-t|+1}$.

If the covariance structure of $\mathbf{V}$ is correctly specified, the consistency of $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ follows from the missing data likelihood theory [26]. If the covariance structure of $\mathbf{V}$ is misspecified, $\mathbf{V}=\mathbf{V}(\boldsymbol{\omega})$ is the 'working' covariance matrix and $\hat{\boldsymbol{\omega}}$ converges to $\boldsymbol{\omega}^*$ such that $\mathbf{V}^*=\mathbf{V}(\boldsymbol{\omega}^*)$ most closely approximates the true underlying covariance matrix. If there are no missing data or missing data are MCAR, the consistency of $\hat{\boldsymbol{\beta}}$ under misspecified covariance structures follows from the GEE theory.

In general, if the covariance structure of $\mathbf{V}$ is misspecified and the missing data are MAR, the parameter estimate $\hat{\boldsymbol{\beta}}$ is not consistent. To illustrate this point, we will first examine the problem of bivariate and trivariate normal data with monotone missingness. To include all randomized patients in the analysis, regardless of the availability of baseline or postrandomization data (as long as at least one measurement is available), we consider the longitudinal data analysis model proposed by Liang and Zeger [27], where the baseline measurement is included as a part of the response vector and is constrained to have the same distribution across treatment groups due to randomization used in clinical trials. This method will be referred to as the cLDA model henceforth.

3. Two or three repeated measurements with missing data

First consider the case with two treatment groups, $A=1$ for the control group and $A=2$ for the treatment group. We intend to collect two measurements on each subject, i.e. $\mathbf{y}_i^F=(y_{i1},y_{i2})'$. Assume that the baseline value, y_{i1}, is always observed, but the postbaseline value, y_{i2}, is only observed for a subset of patients and missing data are MAR.

Under the cLDA model, $\mathbf{y}_i^F|A_i=j\sim N(\boldsymbol{\mu}^{(j)},\Sigma)$ for $j=1,2$, where $\boldsymbol{\mu}^{(1)}=(\mu_1,\mu_2^{(1)})'$ and $\boldsymbol{\mu}^{(2)}=(\mu_1,\mu_2^{(2)})'$ are the mean vectors with baseline mean constrained to be the same across treatment groups, and $\Sigma=\{\sigma_{st}:s,t=1,2\}$ is the true covariance matrix for the two repeated measurements within each subject. Assume there were a total of n_j subjects randomized to group j, of which the first r_j subjects completed the study, and the remaining n_j-r_j subjects dropped out of the study resulting in a missing postbaseline value. Furthermore, suppose that we have misspecified the covariance matrix to be diagonal, i.e. the two repeated measurements within each subject are assumed to be independent. Let $\mathbf{V}=\text{diag}(\omega_1,\omega_2)$ denote the misspecified covariance matrix. The ML estimates under the misspecified model are

$$\hat{\mu}_1=\frac{1}{n_1+n_2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}y_{i1}^{(j)},\quad \hat{\omega}_1=\frac{1}{n_1+n_2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}(y_{i1}^{(j)}-\hat{\mu}_1)^2,$$

$$\hat{\mu}_2^{(j)}=\frac{1}{r_j}\sum_{i=1}^{r_j}y_{i2}^{(j)},\quad \hat{\omega}_2=\frac{1}{r_1+r_2}\sum_{j=1}^{2}\sum_{i=1}^{r_j}(y_{i2}^{(j)}-\hat{\mu}_2^{(j)})^2.$$

In other words, $\hat{\mu}_1$ and $\hat{\omega}_1$ are the mean and variance (uncorrected version) of baseline values among all randomized subjects, $\hat{\mu}_2^{(j)}$ the mean of postbaseline values among completers in group j, and $\hat{\omega}_2$ the variance of postbaseline values among all completers.

Let $n_j/(n_1+n_2)\to\pi_j$ for $j=1,2$ as $n_1\to\infty$, $n_2\to\infty$. By the weak law of large numbers, $\hat{\mu}_1\xrightarrow{p}\mu_1$, $\hat{\omega}_1\xrightarrow{p}\sigma_{11}$, and $r_j/n_j\xrightarrow{p}\phi_2^{(j)}$, where $\phi_2^{(j)}$ denotes the expected proportion of completers in group j. Let $\bar{y}_{t(2)}^{(j)}=(1/r_j)\sum_{i=1}^{r_j}y_{it}^{(j)}$ denote the observed mean value at time point t among completers (at time 2) in group j. By the weak law of large numbers, $\bar{y}_{t(2)}^{(j)}\xrightarrow{p}\mu_{t(2)}^{(j)}$, where $\mu_{t(2)}^{(j)}$ denotes the expected mean at time point t among completers in group j. Similarly, let $S_{st(2)}^{(j)}=\frac{1}{r_j}\sum_{i=1}^{r_j}(y_{is}^{(j)}-\bar{y}_{s(2)}^{(j)})(y_{it}^{(j)}-\bar{y}_{t(2)}^{(j)})$ denote the observed covariance between time points s and t among completers in group j. By the weak law of large numbers, $S_{st(2)}^{(j)}\xrightarrow{p}\sigma_{st(2)}^{(j)}$, where $\sigma_{st(2)}^{(j)}$ denotes the expected covariance between time points s and t among completers in group j. It follows that $\hat{\mu}_2^{(j)}\xrightarrow{p}\mu_{2(2)}^{(j)}$ and $\hat{\omega}_2\xrightarrow{p}\sum_{j=1}^{2}\pi_j\phi_2^{(j)}\sigma_{22(2)}^{(j)}/\sum_{j=1}^{2}\pi_j\phi_2^{(j)}$.

Obviously, misspecification of the covariance structure can result in bias of the complete-case estimator $\hat{\mu}_2^{(j)}$ under the MAR missing data mechanism. By the MAR assumption, $\mu_{2(2)}^{(j)}=\mu_2^{(j)}+\beta_{21\cdot1}(\mu_{1(2)}^{(j)}-\mu_1)$, where $\beta_{21\cdot1}=\sigma_{21}/\sigma_{11}$ denotes the population regression coefficient of postbaseline on baseline, and $\mu_{1(2)}^{(j)}-\mu_1$ the bias of baseline mean among completers in group j. If the probability of having y_{i2} missing only depends on y_{i1} but not on A_i, the argument of iterated expectations yields $\mu_{1(2)}^{(1)}=\mu_{1(2)}^{(2)}$. It follows that the bias of baseline mean among completers is the same across groups and $\mu_{2(2)}^{(2)}-\mu_{2(2)}^{(1)}=\mu_2^{(2)}-\mu_2^{(1)}$; hence, the estimated treatment difference based on completers, $\hat{\mu}_2^{(2)}-\hat{\mu}_2^{(1)}$, is (asymptotically) unbiased for the true treatment difference.

Let $\boldsymbol{\beta}=(\mu_1,\mu_2^{(1)},\mu_2^{(2)})'$ denote the mean parameters of the cLDA model, and $\boldsymbol{\beta}^*=(\mu_1,\mu_{2(2)}^{(1)},\mu_{2(2)}^{(2)})'$ the asymptotic limit of ML estimate $\hat{\boldsymbol{\beta}}=(\hat{\mu}_1,\hat{\mu}_2^{(1)},\hat{\mu}_2^{(2)})'$ under the misspecified model. Let $n=n_1+n_2$. Using the central limit theorem and the delta method, we can show that $n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)\xrightarrow{d}N(\mathbf{0},\mathbf{Q}_1)$, where

$$\mathbf{Q}_1=\begin{pmatrix}\sigma_{11} & \sigma_{12(2)}^{(1)} & \sigma_{12(2)}^{(2)}\\ \sigma_{12(2)}^{(1)} & \pi_1^{-1}\phi_2^{(1)-1}\sigma_{22(2)}^{(1)} & 0\\ \sigma_{12(2)}^{(2)} & 0 & \pi_2^{-1}\phi_2^{(2)-1}\sigma_{22(2)}^{(2)}\end{pmatrix}.$$

Let $\mathbf{X}$ and $\mathbf{V}$ denote the design matrix and the covariance matrix for the data from all subjects. It can be shown that the model-based variance estimator for $n^{1/2}\hat{\boldsymbol{\beta}}$ is given by

$$n(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\xrightarrow{p}\mathbf{Q}_2=\begin{pmatrix}\sigma_{11} & 0 & 0\\ 0 & \pi_1^{-1}\phi_2^{(1)-1}\omega_2 & 0\\ 0 & 0 & \pi_2^{-1}\phi_2^{(2)-1}\omega_2\end{pmatrix},$$

where ω_2 is the probability limit of $\hat{\omega}_2$.

Comparing $\mathbf{Q}_2$ with $\mathbf{Q}_1$, we see that the model-based variance estimate of $\hat{\mu}_2^{(j)}$ is (asymptotically) unbiased if and only if $\sigma_{22(2)}^{(1)}=\sigma_{22(2)}^{(2)}$, which holds if the probability of having y_{i2} missing does not depend on A_i. In fact, we can show that

$\sigma^{(j)}_{12(2)}=\beta_{21\cdot1}\sigma^{(j)}_{11(2)}$, $\sigma^{(j)}_{22(2)}=\sigma_{22\cdot1}+\beta^2_{21\cdot1}\sigma^{(j)}_{11(2)}$. If the probability of having y_{i2} missing does not depend on A_i, then $\sigma^{(1)}_{11(2)}=\sigma^{(2)}_{11(2)}$; hence, $\sigma^{(1)}_{12(2)}=\sigma^{(2)}_{12(2)}$, $\sigma^{(1)}_{22(2)}=\sigma^{(2)}_{22(2)}$, in which case, the model-based variance of the estimated treatment difference, $\hat{\mu}^{(2)}_2-\hat{\mu}^{(1)}_2$, is also asymptotically unbiased. However, the model-based variance of the estimated change from baseline within each group, $\hat{\mu}^{(j)}_2-\hat{\mu}_1$, is biased as the correlation between the repeated measurements is ignored in the misspecified model.

For the sandwich variance estimator, we have

$$\sum_{i=1}^{n}\mathbf{X}_i'\hat{\mathbf{V}}_i^{-1}\hat{\boldsymbol{\varepsilon}}_i\hat{\boldsymbol{\varepsilon}}_i'\hat{\mathbf{V}}_i^{-1}\mathbf{X}_i=\begin{pmatrix} n\hat{\omega}_1^{-1} & \hat{\omega}_1^{-1}\hat{\omega}_2^{-1}r_1S^{(1)}_{12(2)} & \hat{\omega}_1^{-1}\hat{\omega}_2^{-1}r_2S^{(2)}_{12(2)} \\ \hat{\omega}_1^{-1}\hat{\omega}_2^{-1}r_1S^{(1)}_{12(2)} & \hat{\omega}_2^{-2}r_1S^{(1)}_{22(2)} & 0 \\ \hat{\omega}_1^{-1}\hat{\omega}_2^{-1}r_2S^{(2)}_{12(2)} & 0 & \hat{\omega}_2^{-2}r_2S^{(2)}_{22(2)} \end{pmatrix},$$

where $\hat{\varepsilon}^{(j)}_{i1}=y^{(j)}_{i1}-\hat{\mu}_1$ and $\hat{\varepsilon}^{(j)}_{i2}=y^{(j)}_{i2}-\hat{\mu}^{(j)}_2$ denote the residuals. It follows that

$$n(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\left\{\sum_{i=1}^{n}\mathbf{X}_i'\hat{\mathbf{V}}_i^{-1}\hat{\boldsymbol{\varepsilon}}_i\hat{\boldsymbol{\varepsilon}}_i'\hat{\mathbf{V}}_i^{-1}\mathbf{X}_i\right\}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\xrightarrow{p}\mathbf{Q}_1.$$

This shows that the sandwich variance estimator is asymptotically unbiased.

Now consider the more complicated case where there are three repeated measurements to be collected on each subject within each of the two treatment groups. Assume that the distribution of the first measurement is identical across the two groups. The mean vector is $(\mu_1,\mu^{(1)}_2,\mu^{(1)}_3)'$ for the first group and $(\mu_1,\mu^{(2)}_2,\mu^{(2)}_3)'$ for the second group. Again consider the monotone missing data pattern. Under the MAR missing data mechanism, we can make valid inference based on the observed data likelihood.

Suppose we assume an AR(1) covariance structure while the true covariance structure is Toeplitz. We would like to assess the bias of the resulting mean parameter estimates and compare the model-based and the sandwich variance estimates. The theoretical results regarding the biases of mean parameter estimates and model-based variance estimates are summarized in Appendix A. The derivation of the results is very tedious and hence omitted. The consistency of the sandwich variance estimator is established for the general case in Appendix B. We can draw the following conclusions. The estimated change from baseline within each treatment group at the last time point is biased when the covariance structure is misspecified. The estimated treatment difference at the last time point is also biased even if the probability of missingness does not depend on treatment assignment. The model-based variance estimate is biased for both within-group changes from baseline and between-group treatment difference at the last time point. The empirical variance estimate is asymptotically unbiased.

For a concrete example, consider the following MAR missing data process. Let c_1 and c_2 be some thresholds for missingness. If $y_{i1}>c_1$, both y_{i2} and y_{i3} will be missing. If $y_{i1}\leqslant c_1$ but $y_{i2}>c_2$, y_{i3} will be missing. If $y_{i1}\leqslant c_1$ and $y_{i2}\leqslant c_2$, we observe all three repeated measurements. Let $\pi_1=\pi_2=0.5$ (i.e. equal allocation), $\mu_1=25$, $\mu^{(1)}_2=23$, $\mu^{(1)}_3=20$, $\mu^{(2)}_2=18$, $\mu^{(2)}_3=14$, $\sigma_{11}=\sigma_{22}=\sigma_{33}=40$, $c_1=30$, $c_2=25$, and we vary the values of $\rho_{12}=\rho_{23}=\rho_1$ and $\rho_{13}=\rho_2$ (i.e. Toeplitz structure) to assess the impact of misspecification of the covariance structure on the mean and the variance estimates. In order for the correlation matrix to be positive-definite, we need $\rho_2>2\rho_1^2-1$. Assuming the correlation between repeated measurements decreases as the time gap between repeated measurements increases, we also have $0\leqslant\rho_2\leqslant\rho_1<1$. To assess the relationship between the choice of ρ_1 and ρ_2 and the biases of the resulting estimates, the biases of mean parameter estimates and model-based standard error estimates were computed on the following grid of (ρ_1,ρ_2): $\rho_1=0.01$ to 0.99 by 0.01, $\rho_2=\max\{0,(2\rho_1^2-1)\}$ to ρ_1 by 0.01. The results are depicted in Figure 1.

It can be seen from Figure 1 that the biases change direction at the point of true AR(1) covariance structure, i.e. when $\rho_1^2=\rho_2$. If $\rho_1^2>\rho_2$, the biases of the mean parameter estimates are positive and the model-based variance estimates underestimate the true variability. If $\rho_1^2<\rho_2$, the biases of the mean parameter estimates are negative and the model-based variance estimates overestimate the true variability.

4. Simulation studies

An extensive simulation study was conducted to assess the impact of misspecification of the covariance structure on the mean parameter estimates and model-based and sandwich variance estimates under a variety of covariance structures for trials with five or nine repeated measurements with missing data. The parameters of interest include the between-group treatment difference at the last time point, the within-group changes from baseline at the last time point, and the Type I error rate and power for the hypothesis testing of treatment effect at the last time point.

Four underlying covariance structures were used to generate the simulation data: CS, AR(1), Toeplitz, and UN. The missing data were generated according to the following monotone missing data pattern: if $y_j>fy_{j-1}$, then $y_{j+1},y_{j+2},\ldots$ will be set to missing, where f is a tuning parameter that controls the amount of missing data. Obviously, the missing data are MAR. The cLDA model was used for the analysis. The ddfm=KR option was used in SAS PROC MIXED to account for the small sample bias and variability of $\hat{\mathbf{V}}$ in the variance estimate of $\hat{\boldsymbol{\beta}}$ [28]. The Kenward–Roger (KR) method adjusts the estimated covariance matrix of $\hat{\boldsymbol{\beta}}$ and computes the degrees of freedom of an approximating F distribution. Simulation studies have shown that this method performs well with fairly complicated covariance structures when sample sizes are moderate to small and the design is reasonably balanced [29]. Unfortunately, the KR method is not available for the sandwich variance estimator in PROC MIXED.

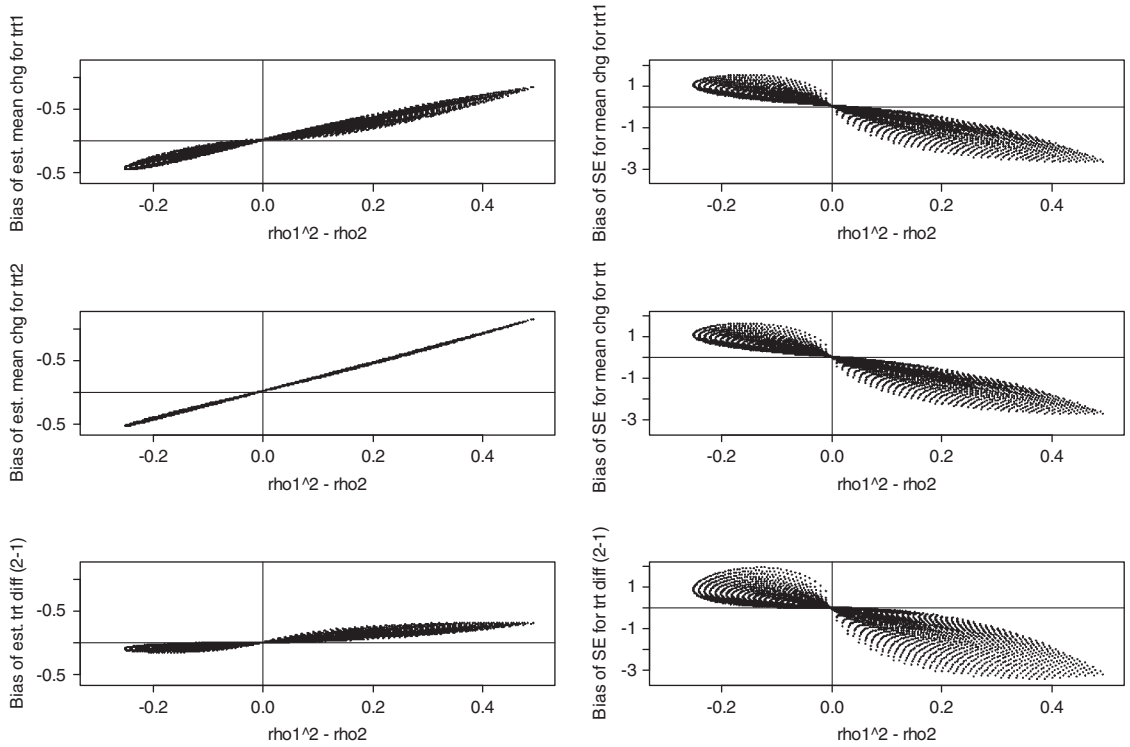

Figure 1. Asymptotic biases of mean parameter estimates and model-based standard error estimates due to misspecification of the covariance structure.

Five thousand repetitions were used in each simulation study. Each simulated data set was fit with each of the following five covariance structures: IND (independent, corresponding to TYPE=VC option of the REPEATED statement), CS, AR(1), TOEP, and UN; with and without the EMPIRICAL option for the first four structures and without the EMPIRICAL option for UN. The use of empirical sandwich variance estimator was indicated by superscript 'EMP'. Depending on the number of repeated measures per subject and the degree of correlation among repeated measures, two simulation setups were considered.

Setup I (*moderate sample size, moderate number of repeated measures, and moderate correlations*): We simulated data from two treatment groups (placebo and treatment) and five repeated measures per subject (including baseline). One hundred subjects per treatment group were considered. The response means over time were 25, 23, 20, 19, and 15 for the placebo group, and 25, 18, 14, 12, and 11 for the treatment group (under the alternative hypothesis H_1), at time points 1–5, respectively. The average correlation among repeated measures was $\sim$0.6 for each correlation structure. Specifically, $\sigma^2=85$, $\rho=0.6$ for the CS structure; $\sigma^2=60$, $\rho=0.72$ for the AR(1) structure; $\sigma^2=60$, $\rho_1=0.7$, $\rho_2=0.6$, $\rho_3=0.5$, $\rho_4=0.4$ for the Toeplitz structure; and, for the UN structure, the variances were 45, 50, 55, 65 and 70 at time points 1–5, and the correlation matrix was

$$\begin{pmatrix} 1 & & & & \\ 0.70 & 1 & & (\text{sym}) & \\ 0.55 & 0.60 & 1 & & \\ 0.45 & 0.50 & 0.60 & 1 & \\ 0.35 & 0.50 & 0.60 & 0.70 & 1 \end{pmatrix}.$$

The tuning parameters f for generating missing data were 1.45, 1.23, 1.19, and 1.35 for the CS, AR(1), Toeplitz, and UN structures, respectively. The amount of missing data at the last time point ranged from 27 to 37 per cent. The parameters were chosen to yield $\sim$90 per cent power for between-group comparison at the last time point.

Setup II (*small sample size, large number of repeated measures, and high correlations*): We simulated data from two treatment groups (placebo and treatment) and nine repeated measures per subject (including baseline). Fifty subjects per treatment group were considered. The model parameters were motivated by the hemoglobin A1c data from a phase III diabetes trial. The response means over time were 7.67, 7.60, 7.56, 7.52, 7.51, 7.50, 7.51, 7.53, and 7.55 for the placebo group, and 7.67, 7.14, 6.86, 6.82, 6.79, 6.79, 6.92, 7.05, and 7.11 for the treatment group (under H_1), at time points 1–9, respectively. The average correlation among repeated measures was $\sim$0.8 for each correlation structure. Specifically, $\sigma^2=0.79$, $\rho=0.8$ for the CS structure; $\sigma^2=0.79$, $\rho=0.94$ for the AR(1) structure; $\sigma^2=0.79$, $\rho_1=0.911$, $\rho_2=0.845$, $\rho_3=0.813$, $\rho_4=0.788$, $\rho_5=0.757$, $\rho_6=0.743$, $\rho_7=0.729$, $\rho_8=0.678$ for the Toeplitz structure; and, for the UN structure, the variances were 0.80, 0.72, 0.73, 0.73, 0.76, 0.76, 0.79, 0.90 and 0.93 at

Table I. Summary of convergence issues for fitting UN structure for simulation Setup II (small sample size, large number of repeated measures, and high correlations).

Hypothesis	True covariance structure	# simulations converged using Newton–Raphson algorithm	# simulations converged using Fisher-scoring algorithm (among those failed with Newton–Raphson algorithm)
Null	CS	4732	268
	AR(1)	5000	0
	TOEP	5000	0
	UN	5000	0
Alternative	CS	4695	305
	AR(1)	4990	10
	TOEP	4991	9
	UN	5000	0

Table II. Simulation results comparing different covariance structure specifications in terms of Type I error (per cent).

	Structure used for analysis								
True structure	IND	IND^{EMP}	CS	CS^{EMP}	AR(1)	$AR(1)^{EMP}$	TOEP	$TOEP^{EMP}$	UN
Setup I: moderate sample size, moderate number of repeated measures, and moderate correlations									
CS	4.5	5.2	4.7	5.0	3.6	5.2	4.8	5.2	4.6
AR(1)	4.6	4.9	(9.8)	5.2	4.4	5.0	4.5	5.2	4.5
TOEP	5.1	5.3	(8.6)	5.0	4.2	5.2	4.8	5.4	4.7
UN	(7.6)	5.0	(10.8)	4.9	(6.9)	5.1	(6.7)	5.4	4.7
Setup II: small sample size, large number of repeated measures, and high correlations									
CS	4.9	(6.5)	4.9	(6.0)	2.7	(6.3)	5.2	(6.8)	4.1
AR(1)	5.2	5.6	(17.6)	5.6	5.0	5.9	4.8	5.8	4.9
TOEP	5.3	(6.1)	(13.9)	5.8	2.9	(6.0)	5.0	(6.2)	4.9
UN	(7.1)	5.5	(18.8)	5.8	4.9	5.3	(7.0)	5.5	4.9

Note: Type I error in parentheses if >5.92 per cent (three standard errors of Monte Carlo variation above 5 per cent based on 5000 simulations, [19]).

time points 1–9, and the correlation matrix was

$$\begin{pmatrix} 1 & & & & & & & & \\ 0.855 & 1 & & & & & & (\text{sym}) & \\ 0.723 & 0.918 & 1 & & & & & & \\ 0.671 & 0.845 & 0.939 & 1 & & & & & \\ 0.661 & 0.822 & 0.892 & 0.942 & 1 & & & & \\ 0.627 & 0.786 & 0.858 & 0.891 & 0.941 & 1 & & & \\ 0.618 & 0.760 & 0.819 & 0.859 & 0.888 & 0.927 & 1 & & \\ 0.621 & 0.752 & 0.792 & 0.832 & 0.864 & 0.875 & 0.917 & 1 & \\ 0.610 & 0.740 & 0.780 & 0.830 & 0.850 & 0.860 & 0.910 & 0.944 & 1 \end{pmatrix}.$$

The tuning parameters f for generating missing data were 1.093, 1.070, 1.076, and 1.130 for the CS, AR(1), Toeplitz, and UN structures, respectively. The amount of missing data at the last time point ranged from 4 to 66 per cent. The parameters were chosen to yield ~80 per cent power for between-group comparison at the last time point.

The means for the treatment group were equal to those for the placebo group under the null hypothesis H_0. Two sets of data were generated for each case—one under H_0 and the other under H_1.

No convergence issues were encountered with the default Newton–Raphson algorithm for Setup I. Table I summarizes the convergence issues for Setup II. It can be seen that all convergence issues encountered by the default Newton–Raphson algorithm were resolved after using the Fisher scoring algorithm to start the first iteration.

Results for treatment comparison at the last time point in terms of Type I error under H_0 and power under H_1 are presented in Tables II and III, respectively.

Table III. Simulation results comparing different covariance structure specifications in terms of power (per cent).

	Structure used for analysis								
True structure	IND	IND^{EMP}	CS	CS^{EMP}	AR(1)	AR(1)^{EMP}	TOEP	TOEP^{EMP}	UN
Setup I: moderate sample size, moderate number of repeated measures, and moderate correlations									
CS	68.0	70.3	89.4	89.8	72.7	78.0	88.9	89.6	88.0
AR(1)	83.7	84.2	—	88.5	90.4	90.7	89.8	90.5	89.5
TOEP	78.1	78.7	—	86.7	87.4	88.6	90.0	90.5	89.3
UN	—	80.9	—	90.0	—	85.0	—	90.0	90.0
Setup II: small sample size, large number of repeated measures, and high correlations									
CS	29.4	—	79.9	—	30.8	—	77.5	—	67.5
AR(1)	50.8	51.8	—	86.8	79.6	80.4	78.6	80.0	76.9
TOEP	43.6	—	—	84.0	63.9	—	78.5	—	75.6
UN	—	60.0	—	79.5	78.9	79.1	—	79.8	79.4

Note: '—' power not reported due to generally inflated Type I error.

Table IV. Simulation results comparing different covariance structure specifications in terms of relative bias (per cent) for estimated between-group difference under H_1.

	Structure used for analysis				
True structure	IND	CS	AR(1)	TOEP	UN
Setup I: moderate sample size, moderate number of repeated measures, and moderate correlations					
CS	−2.6	−0.0	−3.5	−0.1	−0.0
AR(1)	−0.2	−1.5	−0.2	−0.2	−0.2
TOEP	−1.7	−4.9	−0.8	−0.2	−0.2
UN	−2.2	−0.6	−5.6	−0.8	0.0
Setup II: small sample size, large number of repeated measures, and high correlations					
CS	−4.2	−0.0	−11.3	−0.2	−0.4
AR(1)	−1.2	11.7	−0.8	−0.8	−0.8
TOEP	−1.8	6.6	−4.2	−0.6	−0.6
UN	−2.4	−1.3	−2.7	−2.9	−0.7

Table II shows that the misspecification of the covariance structure can inflate the Type I error. The inflation was substantial when the covariance structure was misspecified as CS. The UN covariance model maintained Type I error control in all cases. The sandwich estimator controlled the Type I error for large sample sizes. The inflation of Type I error observed for Setup II was due to the small sample size and the fact that the KR method is not available for sandwich variance estimator in PROC MIXED. This is also consistent with the finding in [30] for testing the significance of difference in mean trends across time. Additional simulations with a sample size of 100 subjects per group and otherwise the same parameters as Setup II confirmed the Type I error control of the sandwich variance estimator.

We did not consider the issue of relative power of tests based on models that provide nonconservative Type I error. Table III shows that the independent covariance structure was by far the least powerful among all the structures used for the analysis and the power loss was more dramatic for Setup II than for Setup I. Substantial power loss was also observed when the CS (Setups I & II) or TOEP (Setup II) structures were misspecified as AR(1), which can be attributed to the deflated Type I errors. The UN covariance model provided near-optimal power in most circumstances except for the case of CS structure in Setup II for which nontrivial power loss was observed. As pointed out in [31, page 228], the CS correlation model tends to be too simplistic for practical applications involving time-series data, as, in general, it is more realistic to assume a model in which the correlation between two observations decreases, in absolute value, with their distance. It is a useful model for applications involving short time series per group, or when all observations within a group are collected at the same time, as in split-plot experiments. Therefore, it is unlikely that a CS correlation model will actually hold for the long series in Setup II; hence, the observed power loss for the CS model in Setup II should be of little practical concern.

The percent bias of mean parameter estimates, percent bias of variance estimates (calculated as (average of estimated variances of $\hat{\beta}$/Monte Carlo variance of $\hat{\beta}-1$) $\times 100$), and empirical coverage of 95 per cent confidence interval (CI) for treatment difference at the last time point under H_1 are provided in Tables IV–VI, respectively.

The sandwich variance estimator does not affect the point estimate of mean parameters. Table IV shows that the misspecification of the covariance structure can result in biased estimate of mean parameters. Table V shows that the model-based variance estimator can dramatically underestimate or overestimate the true variability. In contrast, the sandwich variance estimator performed well for Setup I but tended to underestimate the true variability for Setup II. The underestimation can be attributed

Table V. Simulation results comparing different covariance structure specifications in terms of relative bias (per cent) of estimated variance for between-group difference under H_1.

	Structure used for analysis								
True structure	IND	IND^{EMP}	CS	CS^{EMP}	AR(1)	AR(1)^{EMP}	TOEP	TOEP^{EMP}	UN
Setup I: moderate sample size, moderate number of repeated measures, and moderate correlations									
CS	6.3	−0.3	2.4	0.4	17.3	−0.6	2.0	−1.0	2.4
AR(1)	0.8	−1.3	−29.5	−2.3	0.8	−1.3	−0.0	−3.2	−0.3
TOEP	1.3	−0.6	−24.6	−0.9	7.0	−0.7	0.3	−3.3	0.5
UN	−19.3	−0.8	−34.0	−0.5	−14.4	−0.4	−15.0	−1.5	1.1
Setup II: small sample size, large number of repeated measures, and high correlations									
CS	2.6	−6.9	0.6	−5.4	34.6	−4.8	0.5	−11.1	5.1
AR(1)	0.1	−3.2	−53.0	−4.1	−1.9	−5.7	−2.4	−7.2	−3.5
TOEP	0.2	−4.4	−42.5	−3.7	23.7	−4.6	−0.9	−6.9	−0.9
UN	−13.6	−0.8	−56.3	−4.1	−1.9	−3.5	−15.1	−4.3	−1.5

Table VI. Simulation results comparing different covariance structure specifications in terms of coverage (per cent) of 95 per cent CI for between-group difference under H_1.

	Structure used for analysis								
True structure	IND	IND^{EMP}	CS	CS^{EMP}	AR(1)	AR(1)^{EMP}	TOEP	TOEP^{EMP}	UN
Setup I: moderate sample size, moderate number of repeated measures, and moderate correlations									
CS	95.8	94.7	95.4	95.0	96.5	94.7	95.2	94.7	95.2
AR(1)	95.4	94.8	(90.0)	94.7	95.1	95.0	95.1	94.6	95.1
TOEP	94.9	94.6	(91.0)	94.5	95.5	94.5	94.9	94.5	95.0
UN	(92.0)	94.8	(89.1)	95.2	(92.5)	94.6	(92.9)	94.7	95.1
Setup II: small sample size, large number of repeated measures, and high correlations									
CS	95.2	(93.7)	95.0	(93.7)	97.2	(93.5)	95.0	(92.7)	95.8
AR(1)	94.6	(93.8)	(79.7)	(93.0)	94.7	(93.8)	94.8	(93.7)	94.6
TOEP	94.6	(93.4)	(85.6)	(93.4)	96.5	(93.7)	94.8	(93.8)	94.7
UN	(92.8)	94.6	(80.6)	94.3	94.9	94.4	(93.1)	94.2	94.8

Note: Coverage in parenthesis if <94.08 per cent (three standard errors of Monte Carlo variation below 95 per cent based on 5000 simulations).

to the small sample size and the fact that the KR method is not available for the sandwich variance estimator in PROC MIXED. Table VI shows that the biases in the point estimate or variance estimate can lead to undercoverage of 95 per cent CIs for the true parameter value. The UN structure provided unbiased estimates and adequate coverage in all cases.

The percent bias of mean parameter estimates, percent bias of associated variance estimates, and empirical coverage of 95 per cent CI for mean change from baseline at the last time point in the treatment group under H_1 are presented in Tables VII–IX, respectively. The results for the placebo group were similar and hence omitted.

Tables VII–IX show that the misspecification of the covariance structure was much more problematic for within-group effects than for between-group effects. For example, the misspecification of CS as IND introduced significant biases to the point estimates of within-group mean changes from baseline, but the within-group biases canceled for the two groups resulting in unbiased estimates of between-group treatment differences. The UN structure remained to be unbiased for the point estimates but seemed to underestimate the true variability resulting in slight undercoverage of 95 per cent CIs. Nevertheless, it had the best performance among all the misspecified covariance structures.

The results from additional simulation studies show that similar conclusions can be drawn when the missing data pattern is not monotone or the missingness probability depends on treatment assignment. The bias for between-group comparison is more pronounced when the missingness probability depends on treatment assignment.

5. An example

To illustrate the effect of fitting different covariance structures, we use data from a randomized clinical trial designed to compare the safety and activity of indinavir sulfate, 800 mg every 8 h ('t.i.d.') versus 1000 mg every 12 h ('b.i.d.') versus 1200 mg every 12 h ('b.i.d.'), in combination with zidovudine and lamivudine. Eligible HIV-positive patients were to be randomized in a 1:1:1 ratio to

Table VII. Simulation results comparing different covariance structure specifications in terms of relative bias (per cent) for estimated mean change from baseline for the treatment group under H_1.

	Structure used for analysis				
True structure	IND	CS	AR(1)	TOEP	UN
Setup I: moderate sample size, moderate number of repeated measures, and moderate correlations					
CS	−12.7	−0.2	−14.9	−0.2	−0.2
AR(1)	2.9	12.0	−0.2	−0.2	−0.2
TOEP	−0.3	8.1	−5.4	−0.2	−0.2
UN	−1.6	4.0	−6.0	−1.3	−0.1
Setup II: small sample size, large number of repeated measures, and high correlations					
CS	−36.1	−0.1	−63.7	−0.3	−0.3
AR(1)	−2.0	35.4	−0.1	−0.3	−0.3
TOEP	−10.7	28.1	−17.3	−0.3	−0.2
UN	−2.5	3.0	−2.3	−1.7	−0.1

Table VIII. Simulation results comparing different covariance structure specifications in terms of relative bias (per cent) of estimated variance for mean change from baseline for the treatment group under H_1.

	Structure used for analysis								
True structure	IND	IND^{EMP}	CS	CS^{EMP}	AR(1)	$AR(1)^{EMP}$	TOEP	$TOEP^{EMP}$	UN
Setup I: moderate sample size, moderate number of repeated measures, and moderate correlations									
CS	38.3	−1.6	0.1	−1.6	40.0	−1.0	−3.1	−5.6	−9.1
AR(1)	15.1	−3.6	−33.0	−3.2	−0.0	−1.6	−4.2	−6.3	−7.8
TOEP	20.5	−3.3	−28.5	−2.5	10.8	−1.6	−5.3	−7.8	−10.0
UN	1.9	−3.9	−34.0	−2.7	−2.1	−2.0	−8.4	−4.7	−6.1
Setup II: small sample size, large number of repeated measures, and high correlations									
CS	30.5	−7.2	−1.1	−7.3	62.0	−5.9	−2.5	−14.4	−2.7
AR(1)	34.3	−6.0	−53.7	−6.5	−1.2	−4.9	−3.6	−7.8	−7.0
TOEP	34.4	−5.1	−44.5	−7.4	30.7	−4.1	−4.5	−10.2	−8.2
UN	41.8	−2.4	−57.4	−5.2	3.1	−4.6	−14.0	−5.2	−2.3

Table IX. Simulation results comparing different covariance structure specifications in terms of coverage (per cent) of 95 per cent CI for mean change from baseline for the treatment group under H_1.

	Structure used for analysis								
True structure	IND	IND^{EMP}	CS	CS^{EMP}	AR(1)	$AR(1)^{EMP}$	TOEP	$TOEP^{EMP}$	UN
Setup I: moderate sample size, moderate number of repeated measures, and moderate correlations									
CS	(76.4)	(63.7)	95.1	94.6	(61.9)	(46.8)	94.5	(93.9)	(93.9)
AR(1)	95.0	(92.1)	(43.1)	(55.9)	95.0	94.4	94.7	94.1	94.2
TOEP	96.9	94.7	(65.0)	(74.7)	(89.7)	(86.7)	94.7	(93.9)	(94.0)
UN	94.9	(93.9)	(82.3)	(89.8)	(85.1)	(84.9)	(93.6)	(93.7)	94.5
Setup II: small sample size, large number of repeated measures, and high correlations									
CS	(89.8)	(80.2)	95.1	(93.0)	(58.4)	(34.7)	94.6	(92.0)	95.2
AR(1)	97.8	(93.5)	(37.1)	(58.5)	94.8	(93.9)	94.8	(93.2)	94.1
TOEP	97.1	(92.3)	(55.3)	(70.5)	(92.5)	(86.1)	94.7	(93.0)	94.3
UN	98.0	94.6	(78.9)	(93.8)	95.4	94.1	(93.3)	(93.9)	94.9

Note: Coverage in parenthesis if <94.08 per cent (three standard errors of Monte Carlo variation below 95 per cent based on 5000 simulations).

Table X. Treatment differences (in cells/mm^3) between 800 mg t.i.d. and 1200 mg b.i.d. doses using different covariance structures for the CD4 cell counts data from an indinavir study.

Structure	Estimated difference	Std. error	95 per cent CI	p-value (2-sided test)
IND	111	41	(29, 192)	0.008
IND^{EMP}	111	41	(29, 192)	0.008
CS	76	30	(17, 134)	0.011
CS^{EMP}	76	32	(12, 139)	0.020
AR(1)	91	39	(13, 168)	0.022
AR(1)^{EMP}	91	37	(18, 164)	0.015
TOEP	92	38	(18, 167)	0.016
TOEP^{EMP}	92	34	(26, 158)	0.006
UN	97	36	(23, 171)	0.011

one of the three treatment groups. The study results have been reported in [32]. For illustration, we focus on the 800 mg t.i.d. versus 1200 mg b.i.d comparison with respect to change from baseline at week 24 in CD4 cell counts.

There were 29 patients randomized to the 800 mg t.i.d. group and 30 patients randomized to the 1200 mg b.i.d. group. The percentages of missing data on CD4 cell counts at weeks 0, 2, 4, 8, 12, 16, 20 and 24 were 0, 14, 3, 17, 31, 31, 28 and 41 per cent for the 800 mg t.i.d. group, and 0, 7, 10, 13, 20, 30, 33 and 33 per cent for the 1200 mg b.i.d. group. We fitted the IND, CS, AR(1), Toeplitz, and UN structures in the cLDA model and used separate covariance matrices for the two groups (heteroscedasticity). The point estimates, standard errors, 95 per cent CIs, and p-values for treatment difference between 800 mg t.i.d. and 1200 mg b.i.d. with respect to change from baseline in CD4 cell counts at week 24 are shown in Table X. The estimated treatment difference and the standard errors based on the independent covariance structure were larger than those based on the other covariance structures. In contrast, the estimated treatment difference and the standard errors based on the CS covariance structure were smaller than those based on the other covariance structures.

Of note, the LRT assessing the adequacy of the Toeplitz structure as a possible simplification of the UN structure yielded a chi-square test statistic of 103.8 on 56 degrees of freedom ($p=0.0001$), which suggests that the UN structure should not be reduced to the Toeplitz, let alone, CS or AR(1) structures in this illustrative example. On the other hand, the AIC criterion favored the Toeplitz structure, whereas the BIC criterion favored the CS structure. In general, the BIC criterion imposes a heavier penalty on the number of parameters than the AIC criterion and tends to favor (potentially overly) parsimonious models.

6. Approaches to the convergence problem

As we have seen, the misspecification of the covariance structure could result in bias in the mean parameter estimates. On the other hand, with an excessive amount of missing data, high correlations among repeated measurements, or large number of time points, the default Newton–Raphson algorithm used by SAS PROC MIXED may not converge when using UN covariance matrix. One option is to use the Fisher scoring algorithm (via the SCORING option of the PROC MIXED statement) to obtain the initial values of covariance parameters [33]. Another option is to use the no-diagonal factor analytic structure (via the TYPE=FA0(T) option of the REPEATED statement, where T is the total number of time points), which effectively performs the Cholesky decomposition of the covariance matrix and is numerically more stable. When the above strategy fails, we propose a new method for obtaining appropriate initial values of covariance parameters.

The proposed method will be referred to as the successive univariate regression method and works as follows. Suppose there are a total of T time points where the measurement of a continuous response is to be collected, and the cLDA model will be used for the analysis. Let A denote the treatment group, and $\mathbf{X}$ the other baseline variables. The model for the full data collected on a subject is

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} \Bigg| (A=j, \mathbf{X}=\mathbf{x}) \sim \mathrm{N}\left(\begin{pmatrix} \alpha_1+\boldsymbol{\beta}_1'\mathbf{x} \\ \alpha_2^{(j)}+\boldsymbol{\beta}_2'\mathbf{x} \\ \vdots \\ \alpha_T^{(j)}+\boldsymbol{\beta}_T'\mathbf{x} \end{pmatrix}, \begin{pmatrix} \sigma_{11} & & & (\mathrm{sym}) \\ \sigma_{21} & \sigma_{22} & & \\ \vdots & \vdots & \ddots & \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_{TT} \end{pmatrix} \right). \quad (1)$$

The first step of the proposed method is to fit a series of univariate regression models. Specifically, we first fit $\{y_1|\mathbf{X}\}$, and then fit $\{y_t|y_1,\ldots,y_{t-1},A,\mathbf{X}\}$ successively for $t=2,\ldots,T$. Let $\hat{\sigma}_{11}$ denote the resulting variance estimate in regression model $\{y_1|\mathbf{X}\}$, $\hat{\sigma}_{tt\cdot 1\ldots t-1}$ the variance estimate in $\{y_t|y_1,\ldots,y_{t-1},A,\mathbf{X}\}$, and $\hat{\phi}_{ts}$ the estimated coefficient associated with y_s in $\{y_t|y_1,\ldots,y_{t-1},A,\mathbf{X}\}$ for $1\leqslant s<t\leqslant T$.

The second and final step is to construct the covariance parameters $\{\sigma_{st}: 1 \leqslant s,t \leqslant T\}$ from parameters in the univariate regression models. By (1),

$$(\phi_{t1},\ldots,\phi_{t,t-1}) = (\sigma_{t1},\ldots,\sigma_{t,t-1}) \begin{pmatrix} \sigma_{11} & & (\text{sym}) \\ \vdots & \ddots & \\ \sigma_{t-1,1} & \ldots & \sigma_{t-1,t-1} \end{pmatrix}^{-1},$$

$$\sigma_{tt\cdot 1\ldots t-1} = \sigma_{tt} - \sum_{s=1}^{t=1} \phi_{ts}\sigma_{ts}.$$

Consequently, we can successively apply the following matrix operations to obtain an initial estimate of the covariance parameters,

$$\begin{aligned} (\hat{\sigma}_{t1},\ldots,\hat{\sigma}_{t,t-1}) &= (\hat{\phi}_{t1},\ldots,\hat{\phi}_{t,t-1}) \begin{pmatrix} \hat{\sigma}_{11} & & (\text{sym}) \\ \vdots & \ddots & \\ \hat{\sigma}_{t-1,1} & \ldots & \hat{\sigma}_{t-1,t-1} \end{pmatrix}, \\ \hat{\sigma}_{tt} &= \hat{\sigma}_{tt\cdot 1\ldots t-1} + \sum_{s=1}^{t-1} \hat{\phi}_{ts}\hat{\sigma}_{ts}, \end{aligned} \tag{2}$$

for $t=2,\ldots,T$. Note that (2) can be rewritten as $\hat{\sigma}_{ts} = \sum_{k=1}^{t-1} \hat{\phi}_{tk}\hat{\sigma}_{sk}$ for $s=1,\ldots,t-1$, where $\hat{\sigma}_{sk} = \hat{\sigma}_{ks}$ for $k=s+1,\ldots,t-1$. Consequently, no matrix inversion is needed for constructing the covariance parameters. The positive-definiteness of the resulting covariance matrix is proven in Appendix C.

It is apparent that the method can be also used for the traditional longitudinal ANCOVA models with changes from baseline as the response and baseline as a covariate. In general, convergence is not an issue as the univariate regression models are noniterative; hence, an initial estimate of the covariance matrix can always be obtained using the proposed method. Once an initial estimate of the covariance matrix is obtained, it can be used in the PARMS statement of PROC MIXED to start the iterative algorithm. By construction, the initial estimate should be close to the optimal value; hence, the nonconvergence instances are expected to be reduced with the proposed method.

The proposed method is reminiscent of the likelihood method for monotone missing data [26] and a modified Cholesky decomposition of the covariance matrix Σ such that $\mathbf{T}\Sigma\mathbf{T}' = \mathbf{D}$, where $\mathbf{T}$ is a lower triangular matrix with 1's as diagonal entries and $\mathbf{D}$ is a diagonal matrix [34]. The below-diagonal entries of $\mathbf{T}$ are the negatives of the autoregressive coefficients, ϕ_{ts}, and the diagonal entries of $\mathbf{D}$ are the innovation variances, $\sigma_{tt\cdot 1\ldots t-1}$.

To evaluate the proposed method for the convergence problem, we simulated body weight data from a Phase III clinical trial in obese patients. Body weight was collected at 11 time points (including baseline). The mean body weight (in kg) was 99.7, 99.2, 98.7, 98.2, 97.7, 97.4, 97.2, 97.0, 96.8, 96.7, and 96.7 for the placebo group, and 99.7, 98.6, 97.6, 96.6, 95.7, 95.0, 94.5, 94.0, 93.6, 93.3, and 93.1 for the treatment group, at time points 1–11, respectively. The variance was (arbitrarily) set as 260 at each time point, and the correlation matrix was equal to

$$\begin{pmatrix} 1 & & & & & & & & & & \\ 0.997 & 1 & & & & & & & & & (\text{sym}) \\ 0.994 & 0.997 & 1 & & & & & & & & \\ 0.986 & 0.991 & 0.995 & 1 & & & & & & & \\ 0.977 & 0.984 & 0.989 & 0.996 & 1 & & & & & & \\ 0.967 & 0.975 & 0.982 & 0.991 & 0.996 & 1 & & & & & \\ 0.958 & 0.967 & 0.975 & 0.985 & 0.992 & 0.997 & 1 & & & & \\ 0.948 & 0.958 & 0.966 & 0.978 & 0.987 & 0.993 & 0.996 & 1 & & & \\ 0.939 & 0.949 & 0.958 & 0.971 & 0.981 & 0.987 & 0.992 & 0.996 & 1 & & \\ 0.931 & .942 & .951 & .964 & 0.974 & 0.982 & 0.988 & 0.993 & 0.997 & 1 & \\ 0.924 & 0.935 & 0.943 & 0.956 & 0.967 & 0.976 & 0.983 & 0.989 & 0.993 & 0.996 & 1 \end{pmatrix}.$$

It is common to have high correlations among repeated body weight measurements in obesity studies [1]. High correlations make the covariance matrix nearly singular and are one of the major causes for nonconvergence. The tuning parameter f for generating missing data was 1.028, which resulted in the cumulative drop-out rates of 14.1 and 9.5 per cent at the last time point for the placebo and treatment groups, respectively.

We used 150 subjects per group and 5000 simulations. For each simulated data set, we fitted a cLDA model with four methods to estimate the UN covariance, namely, the Newton–Raphson algorithm, the Fisher scoring algorithm, the factor analytic structure FA0(11), and the proposed successive univariate regression method. Of the 5000 simulations, only 3581 converged using the Newton–Raphson algorithm. Of the remaining 1419 cases, 1383 converged using the Fisher scoring algorithm. Of the remaining 36 cases, 29 converged using the factor analytic structure, and the remaining 9 converged using the proposed successive univariate regression method.

7. Discussion

In repeated measures analysis with missing data, if the covariance structure is misspecified, the estimate for the mean parameters will be biased even if the missing data are MAR. Both the model-based and sandwich variance estimator do not remove the bias of the mean parameter estimate. The sandwich variance estimate can correctly estimate the variance of the mean parameter estimate in large samples, whereas the model-based variance estimate can grossly overestimate or underestimate the true variability.

We recommend the use of UN covariance model (with ddfm=KR) as the default strategy for the analysis of continuous longitudinal data. Fitting an UN covariance will not introduce bias in the mean parameter estimates. It controls the Type I error, provides near-optimal power in most circumstances, and maintains adequate coverage for both between-group comparison and within-group treatment effects.

If the default Newton–Raphson algorithm used by SAS PROC MIXED fails to converge, the Fisher scoring algorithm, the factor analytic structure, and the proposed successive univariate regression method can be tried in turn. This usually resolves most convergence issues. In the rare case where the above strategy fails, it is prudent to examine the data configuration and the model specification. If no data or model issues are identified, progressively more specific covariance structures (e.g. Toeplitz $\rightarrow$ AR(1)+random effect [35, 36]) may be tried to minimize the impact of misspecification, and we recommend the sandwich variance estimator to be used with structured covariance models to control the Type I error for large samples.

For simplicity, SAS PROC MIXED has been used to implement the various variance–covariance models. Similar conclusions are expected to hold had other statistical software packages (e.g. the nlme library of the free software R) been used.

Appendix A: Results for three repeated measurements with missing data

Let $\Sigma=\{\sigma_{st}:s,t=1,2,3\}$ be the true covariance matrix for the repeated measurements. Assume it is misspecified to have an AR(1) structure with parameters (σ^2,ρ). It can be shown that the probability limit of the ML estimate $\hat{\rho}$ is determined by maximizing the profile log-likelihood function with respect to ρ:

$$l_p(\rho)=\sum_{j=1}^{2}\left[-\frac{1}{2}(1+\phi_2^{(j)}+\phi_3^{(j)})\{1+\log(2\pi)\}-\frac{1}{2}(1+\phi_2^{(j)}+\phi_3^{(j)})\log\hat{\sigma}(\rho)-\frac{1}{2}(\phi_2^{(j)}+\phi_3^{(j)})\log(1-\rho^2)\right],$$

where $\phi_t^{(j)}$ denotes the expected proportion of subjects with observed data at time t in group j, and

$$\hat{\sigma}^2(\rho)=\sum_{j=1}^{2}\pi_j\left[\sigma_{11}+\frac{\phi_2^{(j)}}{(1-\rho^2)}(\sigma_{22(2)}^{(j)}-2\rho\sigma_{12(2)}^{(j)}+\rho^2\sigma_{11(2)}^{(j)})+\frac{\phi_3^{(j)}}{(1-\rho^2)}(\sigma_{33(3)}^{(j)}-2\rho\sigma_{23(3)}^{(j)}+\rho^2\sigma_{22(3)}^{(j)})\right]\Bigg/\sum_{j=1}^{2}\pi_j(1+\phi_2^{(j)}+\phi_3^{(j)}),$$

where $\sigma_{uv(t)}^{(j)}$ denotes the conditional covariance between time u and time v among subjects with observed data at time t in group j.

Let $\boldsymbol{\beta}=(\mu_1,\mu_2^{(1)},\mu_3^{(1)},\mu_2^{(2)},\mu_3^{(2)})'$, and $\hat{\boldsymbol{\beta}}$ the estimates of $\boldsymbol{\beta}$ based on the misspecified model. It can shown that $\hat{\boldsymbol{\beta}}\xrightarrow{p}\boldsymbol{\beta}^*$, where $\mu_1^*=\mu_1$, $\mu_2^{(j)*}=\mu_{2(2)}^{(j)}-\rho(\mu_{1(2)}^{(j)}-\mu_1)$, $\mu_3^{(j)*}=\mu_{3(3)}^{(j)}-\rho(\mu_{2(3)}^{(j)}-\mu_2^{(j)*})$, and $\mu_{s(t)}^{(j)}$ denotes the conditional expectation at time s among subjects with observed data at time t in group j. Let $n=n_1+n_2$. It can be shown that $n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)\xrightarrow{d}\mathrm{N}(\mathbf{0},\mathbf{Q}_1)$, where the elements of $\mathbf{Q}_1$ are given by

$$q_{11}=\sigma_{11},$$

$$q_{12}=\rho\sigma_{11}+(\sigma_{12(2)}^{(1)}-\rho\sigma_{11(2)}^{(1)}),$$

$$q_{13}=\rho^2\sigma_{11}+\rho(\sigma_{12(2)}^{(1)}-\rho\sigma_{11(2)}^{(1)})+(\sigma_{13(3)}^{(1)}-\rho\sigma_{12(3)}^{(1)}),$$

$$q_{14}=\rho\sigma_{11}+(\sigma_{12(2)}^{(2)}-\rho\sigma_{11(2)}^{(2)}),$$

$$q_{15}=\rho^2\sigma_{11}+\rho(\sigma_{12(2)}^{(2)}-\rho\sigma_{11(2)}^{(2)})+(\sigma_{13(3)}^{(2)}-\rho\sigma_{12(3)}^{(2)}),$$

$$q_{22}=\rho^2\sigma_{11}+(2-\pi_1^{-1}\phi_2^{(1)-1})\rho(\sigma_{12(2)}^{(1)}-\rho\sigma_{11(2)}^{(1)})+\pi_1^{-1}\phi_2^{(1)-1}(\sigma_{22(2)}^{(1)}-\rho\sigma_{12(2)}^{(1)}),$$

$$q_{23}=\rho^3\sigma_{11}+(2-\pi_1^{-1}\phi_2^{(1)-1})\rho^2(\sigma_{12(2)}^{(1)}-\rho\sigma_{11(2)}^{(1)})+\pi_1^{-1}\phi_2^{(1)-1}\rho(\sigma_{22(2)}^{(1)}-\rho\sigma_{12(2)}^{(1)})+(1-\pi_1^{-1}\phi_2^{(1)-1})\rho(\sigma_{13(3)}^{(1)}-\rho\sigma_{12(3)}^{(1)})$$
$$+\pi_1^{-1}\phi_2^{(1)-1}(\sigma_{23(3)}^{(1)}-\rho\sigma_{22(3)}^{(1)}),$$

$$q_{24}=\rho^2\sigma_{11}+\rho(\sigma_{12(2)}^{(1)}-\rho\sigma_{11(2)}^{(1)})+\rho(\sigma_{12(2)}^{(2)}-\rho\sigma_{11(2)}^{(2)}),$$

$$q_{25}=\rho^3\sigma_{11}+\rho^2(\sigma_{12(2)}^{(1)}-\sigma_{11(2)}^{(1)})+\rho^2(\sigma_{12(2)}^{(2)}-\rho\sigma_{11(2)}^{(2)})+\rho(\sigma_{13(3)}^{(2)}-\rho\sigma_{12(3)}^{(2)}),$$

$$q_{33}=\rho^4\sigma_{11}+(2-\pi_1^{-1}\phi_2^{(1)-1})\rho^3(\sigma_{12(2)}^{(1)}-\rho\sigma_{11(2)}^{(1)})+\pi_1^{-1}\phi_2^{(1)-1}\rho^2(\sigma_{22(2)}^{(1)}-\rho\sigma_{12(2)}^{(1)})+(2-\pi_1^{-1}\phi_2^{(1)-1})\rho^2(\sigma_{13(3)}^{(1)}-\rho\sigma_{12(3)}^{(1)})$$
$$+2\pi_1^{-1}(\phi_2^{(1)-1}-\phi_3^{(1)-1})\rho(\sigma_{23(3)}^{(1)}-\rho\sigma_{22(3)}^{(1)})+\pi_1^{-1}\phi_3^{(1)-1}\sigma_{33(3)}^{(1)},$$

$$q_{34}=\rho^3\sigma_{11}+\rho^2(\sigma_{12(2)}^{(2)}-\rho\sigma_{11(2)}^{(2)})+\rho(\sigma_{13(3)}^{(1)}-\rho\sigma_{12(3)}^{(1)}),$$

$$q_{35}=\rho^4\sigma_{11}+\rho^3(\sigma_{12(2)}^{(1)}-\rho\sigma_{11(2)}^{(1)})+\rho^3(\sigma_{12(2)}^{(2)}-\rho\sigma_{11(2)}^{(2)})+\rho^2(\sigma_{13(3)}^{(1)}-\rho\sigma_{12(3)}^{(1)})+\rho^2(\sigma_{13(3)}^{(2)}-\rho\sigma_{12(3)}^{(2)}),$$

$$q_{44}=\rho^2\sigma_{11}+(2-\pi_2^{-1}\phi_2^{(2)-1})\rho(\sigma_{12(2)}^{(2)}-\rho\sigma_{11(2)}^{(2)})+\pi_2^{-1}\phi_2^{(2)-1}(\sigma_{22(2)}^{(2)}-\rho\sigma_{12(2)}^{(2)}),$$

$$q_{45}=\rho^3\sigma_{11}+(2-\pi_2^{-1}\phi_2^{(2)-1})\rho^2(\sigma_{12(2)}^{(2)}-\rho\sigma_{11(2)}^{(2)})+\pi_2^{-1}\phi_2^{(2)-1}\rho(\sigma_{22(2)}^{(2)}-\rho\sigma_{12(2)}^{(2)})+(1-\pi_2^{-1}\phi_2^{(2)-1})\rho(\sigma_{13(3)}^{(2)}-\rho\sigma_{12(3)}^{(2)})$$
$$+\pi_2^{-1}\phi_2^{(2)-1}(\sigma_{23(3)}^{(2)}-\rho\sigma_{22(3)}^{(2)}),$$

$$q_{55}=\rho^4\sigma_{11}+(2-\pi_2^{-1}\phi_2^{(2)-1})\rho^3(\sigma_{12(2)}^{(2)}-\rho\sigma_{11(2)}^{(2)})+\pi_2^{-1}\phi_2^{(2)-1}\rho^2(\sigma_{22(2)}^{(2)}-\rho\sigma_{12(2)}^{(2)})+(2-\pi_2^{-1}\phi_2^{(2)-1})\rho^2(\sigma_{13(3)}^{(2)}-\rho\sigma_{12(3)}^{(2)})$$
$$+2\pi_2^{-1}(\phi_2^{(2)-1}-\phi_3^{(2)-1})\rho(\sigma_{23(3)}^{(2)}-\rho\sigma_{22(3)}^{(2)})+\pi_2^{-1}\phi_3^{(2)-1}\sigma_{33(3)}^{(2)}.$$

The model-based variance estimate, $n(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\xrightarrow{d}\mathbf{Q}_2$, where $\mathbf{Q}_2$ is given by

$$\sigma^2\begin{pmatrix}
1 & \rho & \rho^2 & \rho & \rho^2\\
\rho & \rho^2+\dfrac{(1-\rho^2)}{\pi_1\phi_2^{(1)}} & \rho^3+\dfrac{\rho(1-\rho^2)}{\pi_1\phi_2^{(1)}} & \rho^2 & \rho^3\\
\rho^2 & \rho^3+\dfrac{\rho(1-\rho^2)}{\pi_1\phi_2^{(1)}} & \rho^4+\dfrac{(\phi_2^{(1)}+\phi_3^{(1)}\rho^2)(1-\rho^2)}{\pi_1\phi_2^{(1)}\phi_3^{(1)}} & \rho^3 & \rho^4\\
\rho & \rho^2 & \rho^3 & \rho^2+\dfrac{(1-\rho^2)}{\pi_2\phi_2^{(2)}} & \rho^3+\dfrac{\rho(1-\rho^2)}{\pi_2\phi_2^{(2)}}\\
\rho^2 & \rho^3 & \rho^4 & \rho^3+\dfrac{\rho(1-\rho^2)}{\pi_2\phi_2^{(2)}} & \rho^4+\dfrac{(\phi_2^{(2)}+\phi_3^{(2)}\rho^2)(1-\rho^2)}{\pi_2\phi_2^{(2)}\phi_3^{(2)}}
\end{pmatrix}.$$

The sandwich variance estimate, $n(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\{\sum_{i=1}^n\mathbf{X}_i'\hat{\mathbf{V}}_i^{-1}\hat{\boldsymbol{\varepsilon}}_i\hat{\boldsymbol{\varepsilon}}_i'\hat{\mathbf{V}}_i^{-1}\mathbf{X}_i\}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\xrightarrow{p}\mathbf{Q}_1$, where $\hat{\boldsymbol{\varepsilon}}_i=\mathbf{y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}}$ denotes the residual vector for subject i.

Appendix B: Consistency of sandwich variance estimator

To show the consistency of the sandwich variance estimator in more general settings, let $\mathbf{y}_i^{\mathrm{F}}=\mathbf{X}_i^{\mathrm{F}}\boldsymbol{\beta}+\boldsymbol{\varepsilon}_i^{\mathrm{F}}$ and suppose the covariance structure of $\boldsymbol{\varepsilon}_i^{\mathrm{F}}$ is misspecified as $\mathbf{V}(\boldsymbol{\omega})$, where $\boldsymbol{\omega}$ denotes the covariance parameters in the misspecified covariance model. In the presence of missing data on the repeated measurements, let $\mathbf{y}_i$ denote the observed responses, $\mathbf{X}_i$ and $\mathbf{V}_i$ the corresponding design matrix and covariance matrix. In addition, let $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\omega}}$ denote the estimates from the assumed likelihood function for the observed responses. We have $\hat{\boldsymbol{\beta}}=(\sum_{i=1}^n\mathbf{X}_i'\hat{\mathbf{V}}_i^{-1}\mathbf{X}_i)^{-1}\sum_{i=1}^n\mathbf{X}_i'\hat{\mathbf{V}}_i^{-1}\mathbf{y}_i$, where $\hat{\mathbf{V}}=\mathbf{V}(\hat{\boldsymbol{\omega}})$. Assume, without loss of generality, $\hat{\boldsymbol{\omega}}\xrightarrow{p}\boldsymbol{\omega}^*$ and $\hat{\boldsymbol{\beta}}\xrightarrow{p}\boldsymbol{\beta}^*$. By the weak law of large numbers,

$$\hat{\mathbf{J}}=n^{-1}\sum_{i=1}^n\mathbf{X}_i'\hat{\mathbf{V}}_i^{-1}\mathbf{X}_i\xrightarrow{p}\mathbf{J}^*=E(\mathbf{X}_i'\mathbf{V}_i^{*-1}\mathbf{X}_i),\quad n^{-1}\sum_{i=1}^n\mathbf{X}_i'\hat{\mathbf{V}}_i^{-1}\mathbf{y}_i\xrightarrow{p}E(\mathbf{X}_i'\mathbf{V}_i^{*-1}\mathbf{y}_i),$$

where $\mathbf{V}^* = \mathbf{V}(\boldsymbol{\omega}^*)$, and the expectations are taken with respect to the true underlying joint distribution of $(\mathbf{R}_i, \mathbf{y}_i, \mathbf{A}_i)$. It follows that

$$\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}^* = \mathbf{J}^{*-1} E(\mathbf{X}_i' \mathbf{V}_i^{*-1} \mathbf{y}_i),$$

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) = n^{-1/2} \sum_{i=1}^{n} \mathbf{J}^{*-1} \mathbf{X}_i' \mathbf{V}_i^{*-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}^*) + o_p(1).$$

By the central limit theorem, $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)$ is asymptotically normal with mean zero and asymptotic covariance matrix,

$$\mathbf{Q} = \mathbf{J}^{*-1} E[\mathbf{X}_i' \mathbf{V}_i^{*-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}^*)(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}^*)' \mathbf{V}_i^{*-1} \mathbf{X}_i] \mathbf{J}^{*-1},$$

which can be consistently estimated by the sandwich variance estimator. This completes the proof.

Appendix C: Positive-definiteness of covariance matrices constructed using the successive univariate regression method

We show that the covariance matrix constructed using the successive univariate regression method is positive-definite if $\hat{\sigma}_{11} > 0$ and $\hat{\sigma}_{tt \cdot 1 \ldots t-1} > 0$ for $t = 2, \ldots, T$. This can be proved by induction. Denote

$$\mathbf{M} = \begin{pmatrix} \hat{\sigma}_{11} & & (\text{sym}) \\ \vdots & \ddots & \\ \hat{\sigma}_{t-1,1} & \cdots & \hat{\sigma}_{t-1,t-1} \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} \hat{\sigma}_{t1} \\ \vdots \\ \hat{\sigma}_{t,t-1} \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} \hat{\phi}_{t1} \\ \vdots \\ \hat{\phi}_{t,t-1} \end{pmatrix}, \quad c = \hat{\sigma}_{tt \cdot 1 \ldots t-1},$$

where we have omitted the reference to time point for notational convenience. We have

$$\begin{pmatrix} \mathbf{M} & \mathbf{u} \\ \mathbf{u}' & \hat{\sigma}_{tt} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & c \end{pmatrix} + \begin{pmatrix} \mathbf{I} \\ \mathbf{v}' \end{pmatrix} \mathbf{M} (\mathbf{I} \ \ \mathbf{v})$$

For any nonzero t-dimensional vector $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1', \lambda_2)'$, where $\boldsymbol{\lambda}_1$ is a $(t-1)$-dimensional vector and λ_2 is a scalar, we have

$$(\boldsymbol{\lambda}_1', \lambda_2) \begin{pmatrix} \mathbf{M} & \mathbf{u} \\ \mathbf{u}' & \hat{\sigma}_{tt} \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda}_1 \\ \lambda_2 \end{pmatrix} = c\lambda_2^2 + (\boldsymbol{\lambda}_1 + \lambda_2 \mathbf{v})' \mathbf{M} (\boldsymbol{\lambda}_1 + \lambda_2 \mathbf{v}).$$

If $\lambda_2 \neq 0$, the first term is positive. If $\lambda_2 = 0$, then $\boldsymbol{\lambda}_1 \neq \mathbf{0}$; hence $\boldsymbol{\lambda}_1 + \lambda_2 \mathbf{v} \neq \mathbf{0}$; by assumption, $\mathbf{M}$ is positive-definite; hence the second term is positive. It follows that the quadratic form is positive for any $\boldsymbol{\lambda} \neq \mathbf{0}$. This completes the proof.

Acknowledgements

We are grateful to two anonymous referees for their helpful and constructive comments that led to significant improvement of the article. We would also like to thank James Kost and many other Merck colleagues for their review of an earlier version of the manuscript.

References

1. Gadbury GL, Coffey CS, Allison DB. Modern statistical methods for handling missing repeated measurements in obesity trials: beyond LOCF. *Obesity Reviews* 2003; **4**:175–184. DOI: 10.1046/j.1467-789X.2003.00109.x.
2. Mallinckrodt CH, Raskin J, Wohlreich MM, Watkin JG, Detke MJ. The efficacy of duloxetine: a comprehensive summary of results from MMRM and LOCF_ANOVA in eight clinical trials. *BMC Psychiatry* 2004; **4**:26. DOI: 10.1186/1471-244X-4-26.
3. Pi-Sunyer FX, Aronne LJ, Heshmati HM, Devin J, Rosenstock J. for the RIO-North America Study Group. Effect of rimonabant, a cannabinoid-1 receptor blocker, on weight and cardiometabolic risk factors in overweight or obese patients. RIO-North America: a randomized controlled trial. *The Journal of the American Medical Association* 2006; **295**:761–775.
4. Liu M, Wei L, Zhang J. Review of guidelines and literature for handling missing data in longitudinal clinical trials with a case study. *Pharmaceutical Statistics* 2006; **5**:7–18.
5. Lane P. Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. *Pharmaceutical Statistics* 2008; **7**(2):93–106. DOI: 10.1002/pst.267.
6. Box GEP. Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and of correlation between errors in the two-way classification. *Annals of Mathematical Statistics* 1954; **25**(3):484–498.
7. Huynh H, Feldt LS. Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. *Journal of the American Statistical Association* 1970; **65**(332):1582–1589.

8. Greenhouse SW, Geisser S. On methods in the analysis of profile data. *Psychometrika* 1975; **32**(2):95–112.
9. Huynh H, Feldt LS. Estimation of the Box correction for degrees of freedom from sample data in the randomized block and split plot designs. *Journal of Educational Statistics* 1976; **1**(1):69–82.
10. SAS Institute. *SAS/STAT 9.1 User's Guide*. SAS Institute Inc.: Cary, NC, 2004.
11. Laird NM, Ware JH. Random effects models for longitudinal data. *Biometrics* 1982; **38**(4):963–974.
12. Jennrich RI, Schluchter MD. Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* 1986; **42**:805–820.
13. Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Statistics in Medicine* 1992; **11**(13):1685–1704. DOI: 10.1002/sim.4780111304.
14. Littell RC, Pendergast J, Natarajan R. Modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine* 2000; **19**:1793–1819.
15. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. Wiley: Hoboken, NJ, 2004.
16. Brown H, Prescott R. *Applied Mixed Models in Medicine* (2nd edn). Wiley: Chichester, West Sussex, England, 2006.
17. Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974; **AC-19**:716–723.
18. Schwartz G. Estimating the dimension of a model. *Annals of Statistics* 1978; **6**:461–464.
19. Keselman HJ, Algina J, Kowalchuk RK, Wolfinger RD. A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements. *Communications in Statistics—Simulation and Computation* 1998; **27**(3):591–604.
20. Ferron J, Dailey R, Yi Q. Effects of misspecifying the first-level error structure in two-level models of change. *Multivariate Behavioral Research* 2002; **37**(3):379–403.
21. Gomez EV, Schaalje GB, Fellingham GW. Performance of the Kenward–Roger method when the covariance structure is selected using AIC and BIC. *Communications in Statistics—Simulation and Computation* 2005; **34**:377–392. DOI: 10.1081/SAC-200055719.
22. Huber PJ. The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Statistical Laboratory of the University of California, Berkeley, vol. 1, 21 June–18 July 1965, 27 December 1965–7 January 1966. University of California Press: Berkeley, CA, 1967; 221–233.
23. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**:13–22.
24. Diggle PJ, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. Clarendon Press: Oxford, U.K., 1994.
25. Freedman DA. On the so-called Huber sandwich estimator and robust standard errors. *The American Statistician* 2006; **60**(4):299–302. DOI: 10.1198/000313006X152207.
26. Little RJA, Rubin DB. *Statistical Analysis with Missing Data* (2nd edn). Wiley: Hoboken, NJ, 2002.
27. 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**:134–148.
28. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 1997; **53**:983–997.
29. Schaalje GB, McBride JB, Fellingham GW. Adequacy of approximations to distributions of test statistics in complex mixed linear models. *Journal of Agricultural, Biological, and Environmental Statistics* 2002; **7**(4):512–524.
30. Overall JE, Tonidandel S. Robustness of generalized estimating equations (GEE) tests of significance against misspecification of the error structure model. *Biometrical Journal* 2004; **46**(2):203–213. DOI: 10.1002/bimj.200210017.
31. Pinheiro JC, Bates DM. *Mixed-effects Models in S and S-PLUS*. Springer: New York, 2000.
32. Haas DW, Arathoon E, Thompson MA, de Jesus Pedro R, Gallant JE, Uip DE, Currier J, Noriega LM, Lewi DS, Uribe P, Benetucci L, Cahn P, Paar D, White Jr AC, Collier AC, Ramirez-Ronda CH, Harvey C, Chung MO, Mehrotra D, Chodakewitz J, Nguyen BY, Protocol 054/069 Study Teams. Comparative studies of two-times-daily versus three-times-daily indinavir in combination with zidovudine and lamivudine. *AIDS* 2000; **14**(13):1973–1978.
33. Mallinckrodt CH, Lane PW, Schnell D, Peng Y, Mancuso JP. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. *Drug Information Journal* 2008; **42**:303–319.
34. Pourahmadi M. Joint mean–covariance models with applications to longitudinal data: unconstrained parameterization. *Biometrika* 1999; **86**(3):677–690.
35. Diggle PJ. An approach to the analysis of repeated measures. *Biometrics* 1988; **44**:959–971.
36. Wolfinger R. Covariance structure selection in general mixed models. *Communications in Statistics—Simulation and Computation* 1993; **22**(4):1079–1106.