Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

RESEARCH ARTICLE

A note on the bias of standard errors when orthogonality of mean and variance parameters is not satisfied in the MMRM analysis

Kazushi Maruo*[1] | Ryota Ishii[2] | Yusuke Yamaguchi[3] | Masaaki Doi[4] | Masahiko Gosho[1]

[1]Department of Biostatistics, Faculty of Medicine, University of Tsukuba, Ibaraki, Japan
[2]Biostatistics Unit, Clinical and Translational Research Center, Keio University Hospital, Tokyo, Japan
[3]Data Science, Development, Astellas Pharma Inc., Tokyo, Japan
[4]Department of Biostatistics, Kyoto University School of Public Health, Kyoto, Japan

Correspondence
*Kazushi Maruo, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8575, Japan. Email: maruo@md.tsukuba.ac.jp

Summary

The mixed effect models for repeated measures (MMRM) analysis is sometimes used as a primary analysis in longitudinal randomized clinical trials. The standard error (SE) for the treatment effect in the MMRM analysis is usually estimated by assuming the orthogonality of the fixed effect and variance-covariance parameters, which is the orthogonality property of a multivariate normal distribution, because of default settings of most standard statistical software. However, this property might be lost when analysis models are misspecified and/or data include missing values with the mechanism of being missing at random. In this study, we investigated the effect of the assumption of the orthogonality property on the estimation of the SE for the MMRM analysis. From simulation and case studies, it was shown that the SE with the assumption of orthogonality property had non-negligible bias, especially when the analysis models assuming heteroscedasticity between treatment groups were applied. We also introduce the SAS code for the MMRM analysis without assuming the orthogonality property. Assuming the orthogonality property in the MMRM analysis would lead to invalid statistical inference, and it is necessary to be careful when applying the MMRM analysis with most standard software.

KEYWORDS:
Missing, Model Misspecification, Longitudinal Data, Randomized Clinical Trial, Statistical Software

1 | INTRODUCTION

In longitudinal randomized clinical trials, subjects are evaluated longitudinally at a number of time points and multivariate correlated outcome data are observed. Although a primary time point is usually specified, data must be treated as longitudinal multivariate data including the time points other than the primary one due to missingness. The mixed model for repeated measures (MMRM) analysis[1] provides valid inference results for the treatment effect, that is, the adjusted mean difference for a

continuous outcome between treatment groups at a specified time point, when the missing mechanism for the outcome variable is missing at random (MAR)[2]. Therefore, the MMRM analysis is sometimes used as the primary analysis method for incomplete longitudinal clinical trial data.

The MMRM analysis assumes multivariate normality for the error term in the analysis model. The orthogonality of mean and variance-covariance parameters is one of the important properties of the multivariate normal distribution. However, this property does not necessarily hold when an analysis model is misspecified. For example, the mean and variance are not orthogonal when a true error distribution is an exponential distribution. Furthermore, Kenward and Molenberghs[3] show that orthogonality of the mean and variance-covariance parameters is lost under the MAR mechanism even when an analysis model is correctly specified.

For the implementation of the MMRM analysis, packages for linear mixed effect models in most standard statistical software (e.g. `MIXED` procedure in `SAS` software or `lme4` package[4] in `R` software) are generally used. These packages estimate the naïve variance, that is, the variance estimator under the assumption that the analysis model is correctly specified, and the robust variance for the fixed effect parameters under the assumption of orthogonality property. Although this assumption is not valid if the missing mechanism is MAR and the true error distribution is not the multivariate normal distribution, the MMRM analysis is usually applied with the above software under the assumption of MAR missingness. Gosho and Maruo[5] compared the performances of several variance estimation methods for the MMRM analysis under the situation of heteroscedasticity between treatment groups. In their simulation studies, it was shown that the naïve and robust standard errors (SE) for the MMRM analysis with assumed heteroscedasticity between treatment groups had bias under the MAR missingness, even when the analysis model was correctly specified and the sample size was not small. Our subsequent investigation revealed that the bias was caused by estimating the SEs with standard software packages for linear mixed effect models, wherein the orthogonality property was assumed.

In this study, we evaluate the effect of the assumption of orthogonality for the MMRM analysis on the SE estimator of the treatment effect. The remainder of the paper proceeds as follows. In Section 2, we provide a brief explanation of the MMRM analysis and the variance estimator for the treatment effect with or without the assumption of orthogonality property. We describe the evaluation of the effect of the assumption of orthogonality through simulation and case studies in Sections 3 and 4, respectively. We then provide a method to apply the MMRM analysis without assuming the orthogonality property using `SAS` in Section 5 and conclude the paper in Section 6.

2 | MMRM ANALYSIS

2.1 | Analysis model

Now, we focus on the continuous outcome of a certain disease and consider a situation in which the efficacy of some treatments (group index: $g = 1, ..., G$) is compared based on a randomized, parallel group clinical trial. The outcomes are measured over time for each subject $i = 1, \dots, n$, and the number of planned measurement time points is T (time point index: $t = 1, \dots, T$). The outcome vector for the ith subject is denoted by $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{in_i})^{\mathrm{T}}$, where Y_{ij} is the jth observation of the ith subject measured at time t_{ij} ($t_{ij} \in \{1, \dots, T\}$), n_i is the number of measurements for the ith subject, and $^{\mathrm{T}}$ denotes the transpose. We have $n_i \leq T$ due to missingness. We then consider applying the following model:

$$\boldsymbol{Y}_i = X_i\boldsymbol{\beta} + W_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i, \tag{1}$$

where X_i is the $n_i \times p$ design matrix relating to the fixed effects that includes variables for treatment groups, time points, treatment by time interaction, and some covariates, and $\boldsymbol{\beta}$ is the p-dimensional parameter vector for the fixed effects. W_i is the $n_i \times q$ design matrix relating to random effects, and $\boldsymbol{b}_i$ is the q-dimensional vector of random effects distributed as $MVN_q(\mathbf{0}_q, D)$, where MVN_q denotes a q-dimensional multivariate normal distribution. $\boldsymbol{\epsilon}_i$ is the n_i-dimensional vector of random errors distributed as $MVN_{n_i}(\mathbf{0}_{n_i}, \Sigma_i)$. $\boldsymbol{b}_i$ and $\boldsymbol{\epsilon}_i$ are independent. From Formula (1), it is derived that $\boldsymbol{y}_i$ marginally follows $MVN_{n_i}(X_i\boldsymbol{\beta}, V_i)$, where $V_i = W_iDW_i^{\mathrm{T}} + \Sigma_i$. In the MMRM analysis, we usually have little interest in random effects and are interested in assessing fixed effects. In such cases, a simple formulation of the linear mixed model (1) can be implemented, wherein the random effects are not explicitly modeled, but rather are included as part of the covariance matrix V_i. In the following, we focus on such a "marginal" mean model. The MMRM analysis is usually defined as the inference on the difference of the model means (or least square means) between the treatment groups at the specified time point based on the marginal model.

The covariance parameter vector in $V = V_i$ for $n_i = T$ (i.e. subjects with no missing values) is denoted as $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m)^{\mathrm{T}}$. The dimension of $\boldsymbol{\alpha}$, m, depends on T and the specified covariance structure. For example, $m = 2$ when the specified structure is the compound symmetry (CS) or the first-order autoregression (AR(1)), and $m = T(T + 1)/2$ when the specified structure is unstructured (UN). The UN structure is usually preferred because no assumptions are made on the covariance structure (e.g. see O'Kelly and Ratitch[6]). Furthermore, we sometimes assume heteroscedasticity between treatment groups. In such cases, covariance matrices are fitted individually by each treatment group (e.g. see Gosho and Maruo[5]).

2.2 | Parameter inference

Inference on a misspecified model based on maximum likelihood estimation

First, we focus on complete data without missingness. Let $\boldsymbol{Y}_i$ $(i = 1, \ldots, n)$ be independent and identically distributed random variable vectors, and let the probability density function (pdf) of $\boldsymbol{Y}_i$ be $f(\boldsymbol{y}_i; \boldsymbol{\eta})$, where $\boldsymbol{\eta}$ is a k-dimensional parameter vector. Furthermore, $g(\boldsymbol{y}_i; \boldsymbol{\xi})$ is a pdf of a specified statistical model for $\boldsymbol{Y}_i$, where $\boldsymbol{\xi}$ is an l-dimensional parameter vector. Model g is generally misspecified unless f is known. Let $l_f(\boldsymbol{\eta})$ and $l_g(\boldsymbol{\xi})$ be log-likelihood functions of f and g for an arbitrary one unit (subject), respectively, and let $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\xi}}$ be maximum likelihood (ML) estimators for f and g, respectively. Under several regularity conditions, $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\xi}}$ converge in probability to $\boldsymbol{\eta}$ and $\boldsymbol{\xi}_{\boldsymbol{\eta}}$, respectively, where $\boldsymbol{\xi}_{\boldsymbol{\eta}}$ is a solution for the equation, $E_f\left[\partial l_g(\boldsymbol{\xi})/\partial \boldsymbol{\xi}\right] = \boldsymbol{0}_l$, and E_f denotes the expectation under the pdf f. Also, the asymptotic distribution of $\sqrt{n}\left((\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})^{\mathrm{T}}, \left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_{\boldsymbol{\eta}}\right)^{\mathrm{T}}\right)^{\mathrm{T}}$ is a $(k + l)$-variate normal distribution with mean $\boldsymbol{0}_{k+l}$ and the following variance-covariance matrix[7,8]:

$$\begin{pmatrix} \mathcal{I}_{\boldsymbol{\eta}}^{-} & \mathcal{I}_{\boldsymbol{\eta}}^{-}\boldsymbol{B} \\ \boldsymbol{B}^{\mathrm{T}}\mathcal{I}_{\boldsymbol{\eta}}^{-} & \mathcal{I}_{\boldsymbol{\xi}}^{-}\mathcal{J}_{\boldsymbol{\xi}}\mathcal{I}_{\boldsymbol{\xi}}^{-} \end{pmatrix},$$

where $^{-}$ denotes the generalized inverse and

$$\mathcal{I}_{\boldsymbol{\eta}} = E_f\left[-\frac{\partial^2}{\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}^{\mathrm{T}}} l_f(\boldsymbol{\eta})\right], \mathcal{I}_{\boldsymbol{\xi}} = E_f\left[-\frac{\partial^2}{\partial \boldsymbol{\xi}\partial \boldsymbol{\xi}^{\mathrm{T}}} l_g(\boldsymbol{\xi})\right]\Bigg|_{\boldsymbol{\xi}=\boldsymbol{\xi}_{\boldsymbol{\eta}}}, \mathcal{J}_{\boldsymbol{\xi}} = E_f\left[\left\{\frac{\partial}{\partial \boldsymbol{\xi}} l_g(\boldsymbol{\xi})\right\}\left\{\frac{\partial}{\partial \boldsymbol{\xi}} l_g(\boldsymbol{\xi})\right\}^{\mathrm{T}}\right]\Bigg|_{\boldsymbol{\xi}=\boldsymbol{\xi}_{\boldsymbol{\eta}}}, \boldsymbol{B} = \frac{\partial}{\partial \boldsymbol{\eta}}\boldsymbol{\xi}_{\boldsymbol{\eta}}^{\mathrm{T}}.$$

Now, we describe the variance-covariance estimator of $\hat{\boldsymbol{\xi}}$ for the specified model, which is the estimator of $(1/n)\mathcal{I}_{\boldsymbol{\xi}}^{-}\mathcal{J}_{\boldsymbol{\xi}}\mathcal{I}_{\boldsymbol{\xi}}^{-}$. The variance estimator of $\hat{\boldsymbol{\xi}}$, $V_{\boldsymbol{\xi}}^{(R)}$, is given by $V_{\boldsymbol{\xi}}^{(R)} = I_{\boldsymbol{\xi}}^{-}J_{\boldsymbol{\xi}}I_{\boldsymbol{\xi}}^{-}$, where

$$I_{\boldsymbol{\xi}} = -\frac{\partial^2}{\partial \boldsymbol{\xi}\partial \boldsymbol{\xi}^{\mathrm{T}}} l_g^{(n)}(\boldsymbol{\xi})\Big|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}}, J_{\boldsymbol{\xi}} = \sum_{i=1}^{n}\left\{\frac{\partial}{\partial \boldsymbol{\xi}} l_{g(i)}^{(n)}(\boldsymbol{\xi})\right\}\left\{\frac{\partial}{\partial \boldsymbol{\xi}} l_{g(i)}^{(n)}(\boldsymbol{\xi})\right\}^{\mathrm{T}}\Bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}}. \tag{2}$$

Moreover, $l_g^{(n)}(\boldsymbol{\xi})$ and $l_{g(i)}^{(n)}(\boldsymbol{\xi})$ are the log likelihood for all n subjects and ith subject, respectively. Because the expectation under the true model cannot be obtained unless the true model is known, arithmetic means are used, which are the observed information. If statistical model g is correctly specified, we obtain $\mathcal{I}_{\boldsymbol{\xi}} = \mathcal{J}_{\boldsymbol{\xi}}$; therefore, $V_{\boldsymbol{\xi}}^{(N)} = I_{\boldsymbol{\xi}}^{-}$ can be regarded as the variance estimator. $V_{\boldsymbol{\xi}}^{(N)}$ and $V_{\boldsymbol{\xi}}^{(R)}$ are called the naïve and robust variance estimators, respectively. The robust variance estimator has consistency for any model misspecification under certain regularity conditions, whereas the naïve variance estimator is not asymptotically valid unless the model is specified correctly. Note that the robust variance estimator based on the expected information under the specified model (i.e. based on E_g) is not asymptotically valid unless the model is correctly specified. These properties are true for missing data with the MAR mechanism (e.g. see Takai and Kano[9]).

Inference on MMRM

Let $\boldsymbol{\theta}$ denote the whole parameter vector for the MMRM, $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\alpha}^{\mathrm{T}})^{\mathrm{T}}$. Here, $\boldsymbol{\theta}$ is estimated with the ML or restricted ML (REML) method. We focus on the ML in this study. The naïve and robust variance-covariance estimators for the ML estimators

of θ, $\hat{\theta} = (\hat{\beta}^{\mathrm{T}}, \hat{\alpha}^{\mathrm{T}})^{\mathrm{T}}$, are given by $V_{\theta}^{(N)} = I_{\theta}^{-}$ and $V_{\theta}^{(R)} = I_{\theta}^{-} J_{\theta} I_{\theta}^{-}$, respectively, where I_{θ} and J_{θ} are obtained along with Equation (2). I_{θ} and J_{θ} are resolved to the components for β and α as follows:

$$I_{\theta} = \begin{pmatrix} I_{\beta\beta} & I_{\beta\alpha} \\ I_{\beta\alpha}^{\mathrm{T}} & I_{\alpha\alpha} \end{pmatrix}, \; J_{\theta} = \begin{pmatrix} J_{\beta\beta} & J_{\beta\alpha} \\ J_{\beta\alpha}^{\mathrm{T}} & J_{\alpha\alpha} \end{pmatrix}.$$

As for the detail of each component, refer to Lindstrom and Bates[10]. The naïve and robust variance estimators of $\hat{\beta}$, $V_{\beta}^{(N)}$ and $V_{\beta}^{(R)}$ are obtained as parts of $\beta\beta$ in $V_{\theta}^{(N)}$ and $V_{\theta}^{(R)}$ as follows:

$$\begin{aligned} V_{\beta}^{(N)} &= I_{\beta\beta}^{-} - I^{\beta\alpha} I_{\beta\alpha}^{\mathrm{T}} I_{\beta\beta}^{-}, \\ V_{\beta}^{(R)} &= V_{\beta}^{(N)} J_{\beta\beta} V_{\beta}^{(N)} + I^{\beta\alpha} J_{\beta\alpha}^{\mathrm{T}} V_{\beta}^{(N)} + V_{\beta}^{(N)} J_{\beta\alpha} I^{\beta\alpha\mathrm{T}} + I^{\beta\alpha} J_{\alpha\alpha} I^{\beta\alpha\mathrm{T}}, \end{aligned}$$

where

$$I^{\beta\alpha} = -I_{\beta\beta}^{-} I_{\beta\alpha} (I_{\alpha\alpha} - I_{\beta\alpha}^{\mathrm{T}} I_{\beta\beta}^{-} I_{\beta\alpha})^{-}.$$

If there are no missing data or the missing mechanism is missing completely at random (MCAR), and the analysis model is correctly specified, then we have $E[I_{\beta\alpha}] = E[J_{\beta\alpha}] = O$, that is the orthogonality property for the normal distribution. If we assume $I_{\beta\alpha} = J_{\beta\alpha} = O$ using the orthogonality property, the naïve and robust variance estimators of $\hat{\beta}$, $V_{\beta(O)}^{(N)}$ and $V_{\beta(O)}^{(R)}$ are given by

$$\begin{aligned} V_{\beta(O)}^{(N)} &= I_{\beta\beta}^{-} = \left\{ \sum_i X_i^{\mathrm{T}} \hat{V}_i^{-1} X_i \right\}^{-}, \\ V_{\beta(O)}^{(R)} &= I_{\beta\beta}^{-} J_{\beta\beta} I_{\beta\beta}^{-} = V_{\beta(O)}^{(N)} \left\{ \sum_i X_i^{\mathrm{T}} \hat{V}_i^{-1} \hat{\varepsilon}_i \hat{\varepsilon}_i^{\mathrm{T}} \hat{V}_i^{-1} X_i \right\} V_{\beta(O)}^{(N)}, \end{aligned}$$

respectively, where $\hat{\varepsilon}_i = \boldsymbol{Y}_i - X_i \hat{\beta}$. Although these variance estimators are widely used in ordinary statistical software, they are valid only when the above described strict conditions are true. When we use the MMRM analyses, however, the assumption of missingness is usually MAR, not MCAR.

Small sample adjustment

Let $l_R(\theta) = l_M(\theta) + \psi(\alpha)$ denote the log likelihood function for the REML method, where l_M is the log likelihood for the ML method, $\psi(\alpha) = -1/2 \log \left| \sum_i X_i^{\mathrm{T}} V_i^{-1} X_i \right| + c$, and c is the constant. Because $\psi(\alpha)$ does not depend on β, the naïve and robust variance estimators of $\hat{\beta}$ for the REML method are same as those for the ML method under the assumption of orthogonality property. However, J_{θ} in the robust variance estimator using the REML method cannot be calculated without the orthogonality property, because the contribution of each subject to $\psi(\alpha)$ cannot be evaluated. Thus, only the naïve variance estimator for the MMRM analysis with the REML method can be calculated when the orthogonality property cannot be assumed. Furthermore,

any adjustments of degree of freedom for statistical tests cannot be applied straightforwardly when the orthogonality property is not assumed.

There are several small sample bias adjusted estimators of the robust variance while assuming the orthogonality property [11,12], and Gosho et al. [13] compared the performance of these estimators in the context of the MMRM analysis. However, there is no small sample bias adjusted robust variance without assuming the orthogonality property.

Thus, we focus only on the ML method and asymptotic Wald type statistical tests in the following.

3 | SIMULATION STUDY

3.1 | Simulation design

We conducted a simulation study to evaluate the effect of the assumption of orthogonality property on the estimation of SE of the treatment effect for the MMRM analysis. The simulation design was as follows. The efficacy of two groups ($G = 2$), control ($g = 1$) and active ($g = 2$), were compared based on a parallel group, randomized clinical trial. The efficacy indexes were measured over time, where the number of measurement time points was 3 ($T = 3$). A higher value of the index meant a better state of the targeted disease, and the motivation of treatment was to increase the efficacy index. The primary endpoint was the measurement value of the efficacy index at the last time point ($t = 3$).

Random numbers of outcomes were generated with multivariate normal and exponential distributions. The heteroscedasticity settings were used by reference to the result of Gosho and Maruo[5]. As for the settings of the normal distribution, distributions for the control and active groups at each time point were set as $N(0, 1)$ and $N(d(1 + st/3), (1 + dst/3)^2)$, respectively, where N denotes a univariate normal distribution, $d = 0, 1$, and $s = 0, 1, 2$. The settings for the exponential distribution were as follows. The distributions for the control and active groups at each time point were set as $Ex(1)$ and $Ex(1 + kt/3)$, respectively, where Ex denotes a univariate exponential distribution and $k = 0, 1, 2$. The relationship between the time periods for the exponential distribution was specified with the Gaussian copula. For both types of distributions, the correlation structure for the three time points was the first order autoregression (AR(1)), where the correlation parameter was 0.7.

The missing structure was monotone and MAR. The missing probability was modeled by

$$\text{logit}\left\{Pr\left(R_{i(t)} = 1 | R_{i(t-1)} = 0\right)\right\} = \beta_I - y_{i(t-1)},$$

where β_I was the intercept term, $y_{i(t-1)}$ was the outcome for the ith subject at the $(t-1)$th time point, and $R_{i(t)}$ was the indicator random variable for the ith subject such that $R_{i(t)} = 1$ when $y_{i(t)}$ was missing, otherwise, $R_{i(t)} = 0$. We set $R_{i(1)} = 0$ (i.e. no missing for $t = 1$), and if $R_{i(t-1)} = 1$ then $R_{i(t)} = 1$ (i.e. monotone missing). β_I was calculated such that the missing proportion at the last time point ($t = 3$) became 0% (i.e. $\beta_I = -\infty$), 20%, and 40%. This missing structure meant dropout from the trial

due to lack of efficacy. We set the number of subjects for each group ($n/2$) as 200 so that the small sample bias for the ML methods and naïve and robust variance estimators would be negligible to evaluate only the effect of the model misspecification and assumption of orthogonality property. The number of all simulation settings was 21: 3 dropout settings (0%, 20%, and 40%) times 7 distribution settings ($d = 0$, $\{d = 1, s = 0\}$, $\{d = 1, s = 1\}$, $\{d = 1, s = 2\}$ for normal distribution, and $k = 1, 2, 3$ for exponential distribution).

Under each condition, we conducted 10,000 trial simulations. On each simulation, the MMRM methods using the CS and UN covariance structures with and without the assumption of heteroscedasticity between the groups were applied, where the mean structure was modeled by the group, time point, and group-by-time interaction effects, and marginal error distribution was specified as the normal distribution. The CS and UN structures assuming heteroscedasticity were denoted by CSG and UNG, respectively. Models with the CS and CSG structures were misspecified for all the settings. Models with the UN structure were misspecified when $\{d = 1, s = 1\}$ or $\{d = 1, s = 2\}$ for the normal distribution and for all settings of the exponential distribution. Models with the UNG structure were misspecified for the settings of the exponential distribution. The ML estimator of the model mean difference at the last time point ($t = 3$), $\hat{\delta}$, was calculated. The true value of δ is $\delta = d(s + 1)$ for the normal distribution and $\delta = k$ for the exponential distribution. The naïve and robust SEs of $\hat{\delta}$ with and without the assumption of orthogonality property were estimated. The naïve and robust SEs without the assumption are denoted by NV and RB, respectively, and those with the assumption are denoted by NVO and RBO, respectively. The number of SEs calculated for each simulation setting was 16: 4 analysis models (CS, UN, CSG, and UNG) times 4 SEs (NVO, NV, RBO, and RB).

We evaluated the simulated mean of $\hat{\delta}$ and four types of SEs, and simulated the standard deviation (SD) of $\hat{\delta}$ for each analysis model and simulation setting. The biases of SEs were evaluated with the percentage scale $100[\{\text{mean of SE}\}-\{\text{SD of } \hat{\delta}\}]/\{\text{SD of } \hat{\delta}\}$. We then evaluated the simulated coverage probabilities (CP) for the 95% Wald-type confidence intervals with each SE. We also calculated the correlations between the estimated fixed effects and variance-covariance parameters for each analysis model and simulation setting. In the analysis model with the CS structure, for example, 6 (number of the fixed effects) × 2 (number of the variance-covariance parameters) = 12 correlation coefficients were estimated. We then evaluated the range of the correlations.

3.2 | Simulation result

The simulation results for the normal and exponential distributions are shown in Tables 1 and 2, respectively.

The correlations between the fixed effect and variance-covariance parameters were almost zero (range: −0.02 to 0.02, which was considered as the simulation error) only when the error distribution was the normal distribution and there were no missing data. In the other settings, the correlations were not zero, and it was shown that the orthogonality property was lost when the analysis models were misspecified and/or under MAR missingness.

TABLE 1 Simulation results for normal distribution. Correlation: range of estimated correlations between fixed effect and variance-covariance parameters. δ: treatment effect; $d = 0$: no treatment effect, $d = 1$: non-zero treatment effect; $s = 0$: homoscedasticity, $s = 1, 2$: heteroscedasticity; true value of δ is $d(1 + s)$. NVO: naïve SE assuming orthogonality, NV: naïve SE without assuming orthogonality, RBO: robust SE assuming orthogonality, RB: robust SE without assuming orthogonality, CS: compound symmetry, UN: unstructured, CSG: CS structure assuming heteroscedasticity, UNG: UN structure assuming heteroscedasticity.

d	s	Dropout(%)	Covariance structure	$\hat{\delta}$	Bias of SE (%)				Coverage probability (%)				Correlation	
					NVO	NV	RBO	RB	NVO	NV	RBO	RB	Min.	Max.
0	0	0	CS	0.00	1.0	1.0	1.0	1.0	95.1	95.1	95.1	95.1	−0.02	0.02
		0	UN*	0.00	1.0	1.0	1.0	1.0	95.1	95.1	95.1	95.1	−0.02	0.02
		0	CSG	0.00	1.0	1.0	1.0	1.0	95.1	95.1	95.1	95.1	−0.01	0.02
		0	UNG*	0.00	1.0	1.0	1.0	1.0	95.1	95.1	95.1	95.1	−0.02	0.02
		20	CS	0.00	−2.5	−2.5	0.4	0.4	94.4	94.4	95.1	95.1	−0.10	0.06
		20	UN*	0.00	0.1	0.2	0.1	0.2	95.0	95.0	95.0	95.0	−0.15	0.02
		20	CSG	0.00	−4.1	−1.3	−1.3	2.3	94.0	94.7	94.7	95.5	−0.15	0.11
		20	UNG*	0.00	−3.8	−0.3	−3.8	−0.4	93.9	94.9	93.9	94.8	−0.21	0.15
		40	CS	0.00	−5.9	−5.8	−0.3	−0.2	93.3	93.4	94.8	94.8	−0.19	0.14
		40	UN*	0.00	−0.7	−0.4	−0.7	−0.4	94.7	94.7	94.6	94.7	−0.31	0.01
		40	CSG	0.00	−9.4	−1.8	−4.0	6.5	92.3	94.7	94.0	96.3	−0.26	0.19
		40	UNG*	0.00	−13.4	−0.8	−13.4	−1.2	90.8	94.7	90.8	94.6	−0.42	0.29
1	0	0	CS	1.00	−0.7	−0.7	−0.7	−0.7	94.6	94.6	94.6	94.6	−0.01	0.01
		0	UN*	1.00	−0.7	−0.7	−0.7	−0.7	94.6	94.6	94.6	94.6	−0.02	0.02
		0	CSG	1.00	−0.7	−0.7	−0.7	−0.7	94.6	94.6	94.6	94.6	−0.02	0.01
		0	UNG*	1.00	−0.7	−0.7	−0.7	−0.7	94.6	94.6	94.6	94.6	−0.02	0.02
		20	CS	1.01	−3.5	−3.2	−0.6	−0.3	93.9	94.0	94.8	94.8	−0.12	0.09
		20	UN*	1.00	−0.9	−0.6	−1.0	−0.6	94.6	94.6	94.6	94.7	−0.17	0.07
		20	CSG	1.01	−5.0	−1.8	−2.1	2.0	93.4	94.3	94.3	95.1	−0.19	0.15
		20	UNG*	1.00	−5.1	−1.0	−5.1	−1.1	93.8	94.7	93.8	94.7	−0.28	0.22
		40	CS	1.02	−6.1	−5.8	−0.5	−0.1	93.1	93.2	94.5	94.6	−0.19	0.14
		40	UN*	1.00	−1.7	−0.8	−1.8	−0.9	94.4	94.6	94.4	94.6	−0.32	0.11
		40	CSG	1.02	−10.1	−2.3	−4.5	6.2	91.9	94.2	93.3	95.8	−0.28	0.23
		40	UNG*	1.00	−16.7	−2.0	−16.7	−2.6	89.8	94.4	89.8	94.0	−0.50	0.42
1	1	0	CS	2.00	−13.1	−13.1	−1.0	−1.0	91.1	91.1	94.4	94.4	−0.01	0.01
		0	UN	2.00	−1.0	−1.0	−1.0	−1.0	94.4	94.4	94.4	94.4	−0.02	0.02
		0	CSG	2.00	−13.1	−13.1	−1.0	−1.0	91.1	91.1	94.4	94.4	−0.02	0.01
		0	UNG*	2.00	−1.0	−1.0	−1.0	−1.0	94.4	94.4	94.4	94.4	−0.02	0.02
		20	CS	2.09	−9.8	−9.8	−0.2	−0.2	88.4	88.5	91.8	91.9	−0.10	0.10
		20	UN	2.08	7.7	7.9	−0.5	−0.3	94.7	94.7	92.3	92.3	−0.27	0.05
		20	CSG	2.11	−14.7	−12.5	−1.0	1.0	83.9	85.0	89.9	90.8	−0.19	0.13
		20	UNG*	2.00	−4.2	−0.8	−4.2	−0.8	93.9	94.8	93.9	94.8	−0.29	0.18
		40	CS	2.18	−4.9	−4.9	−0.7	−0.7	84.3	84.3	86.1	86.0	−0.17	0.17
		40	UN	2.16	19.0	19.4	−1.3	−0.8	94.6	94.7	88.3	88.4	−0.46	0.08
		40	CSG	2.24	−18.3	−12.0	−3.3	3.1	68.7	72.9	78.0	81.7	−0.28	0.24
		40	UNG*	2.00	−16.2	−2.3	−16.2	−2.7	89.9	94.3	89.9	94.3	−0.53	0.35
1	2	0	CS	3.00	−18.7	−18.7	−1.0	−1.0	88.7	88.7	94.4	94.4	−0.01	0.01
		0	UN	3.00	−1.0	−1.0	−1.0	−1.0	94.4	94.4	94.4	94.4	−0.02	0.02
		0	CSG	3.00	−18.7	−18.7	−1.0	−1.0	88.7	88.7	94.4	94.4	−0.02	0.01
		0	UNG*	3.00	−1.0	−1.0	−1.0	−1.0	94.4	94.4	94.4	94.4	−0.02	0.02
		20	CS	3.20	−12.8	−12.8	−0.1	−0.2	80.2	80.2	86.3	86.3	−0.10	0.12
		20	UN	3.14	11.4	11.5	−0.4	−0.2	94.4	94.5	90.9	91.0	−0.34	0.07
		20	CSG	3.24	−19.6	−17.9	−0.4	0.6	71.3	72.6	82.9	83.4	−0.19	0.14
		20	UNG*	3.00	−4.0	−0.6	−4.0	−0.6	94.1	95.1	94.1	94.9	−0.29	0.13
		40	CS	3.38	−1.2	−1.0	−0.8	−0.9	71.5	71.6	71.2	71.1	−0.18	0.17
		40	UN	3.27	32.6	33.1	−1.4	−0.8	95.2	95.3	83.8	84.1	−0.55	0.09
		40	CSG	3.48	−22.6	−17.3	−2.4	1.6	42.5	46.7	58.4	61.2	−0.29	0.26
		40	UNG*	3.00	−16.0	−2.2	−16.0	−2.5	89.8	94.4	89.8	94.3	−0.56	0.30

*: Correctly specified models.

The ML estimators of treatment effects, $\hat{\delta}$, for the CS, UN, and CSG models had bias for the misspecified models when the treatment effect existed, and the missing probabilities were large. On the other hand, the biases of $\hat{\delta}$ for the UNG models were negligible for all the settings.

TABLE 2 Simulation results for exponential distribution. Correlation: range of estimated correlations between fixed effect and variance-covariance parameters. δ: treatment effect; $k = 0$: no treatment effect and homoscedasticity, $k = 1, 2$: non-zero treatment effect and heteroscedasticity; true value of δ is k. NVO: naïve SE assuming orthogonality, NV: naïve SE without assuming orthogonality, RBO: robust SE assuming orthogonality, RB: robust SE without assuming orthogonality, CS: compound symmetry, UN: unstructured, CSG: CS structure assuming heteroscedasticity, UNG: UN structure assuming heteroscedasticity.

k	Dropout(%)	Covariance structure	$\hat{\delta}$	Bias of SE (%)				Coverage probability (%)				Correlation	
				NVO	NV	RBO	RB	NVO	NV	RBO	RB	Min.	Max.
0	0	CS	0.00	0.3	0.3	0.2	0.2	95.1	95.1	95.1	95.1	−0.01	0.37
	0	UN	0.00	0.2	0.2	0.2	0.2	95.1	95.1	95.1	95.1	−0.29	0.50
	0	CSG	0.00	0.3	0.3	0.2	0.2	95.1	95.1	95.1	95.1	−0.36	0.52
	0	UNG	0.00	0.2	0.2	0.2	0.2	95.1	95.1	95.1	95.1	−0.49	0.70
	20	CS	0.00	−1.9	−1.9	0.3	0.3	94.3	94.3	95.0	95.0	−0.03	0.36
	20	UN	0.00	0.2	0.2	0.1	0.1	95.0	95.0	95.0	95.0	−0.26	0.50
	20	CSG	0.00	−1.3	−1.0	0.9	−1.2	94.6	94.7	95.2	94.7	−0.36	0.51
	20	UNG	0.00	4.2	4.6	4.2	0.3	95.9	95.9	95.9	95.1	−0.49	0.70
	40	CS	0.00	−5.3	−5.2	0.1	0.1	93.5	93.5	95.0	95.0	−0.07	0.37
	40	UN	0.00	−0.2	−0.1	−0.2	−0.2	95.0	95.0	94.9	94.9	−0.23	0.50
	40	CSG	0.00	−5.0	−3.8	0.3	−1.2	93.5	93.9	95.2	94.7	−0.37	0.52
	40	UNG	0.00	7.8	10.0	7.8	−0.1	96.5	97.0	96.5	95.0	−0.49	0.70
1	0	CS	1.00	−12.0	−12.0	0.1	0.1	91.2	91.2	94.7	94.7	0.00	0.35
	0	UN	1.00	0.1	0.1	0.1	0.1	94.7	94.7	94.7	94.7	−0.20	0.42
	0	CSG	1.00	−12.0	−12.0	0.1	0.1	91.2	91.2	94.7	94.7	−0.30	0.52
	0	UNG	1.00	0.1	0.1	0.1	0.1	94.7	94.7	94.7	94.7	−0.42	0.70
	20	CS	1.05	−14.3	−14.3	0.1	0.0	89.9	90.0	94.6	94.6	−0.03	0.35
	20	UN	1.04	1.9	2.0	0.8	0.1	95.1	95.2	94.9	94.8	−0.17	0.42
	20	CSG	1.05	−14.7	−14.5	0.3	−1.0	89.4	89.5	94.5	94.2	−0.31	0.51
	20	UNG	1.01	4.4	4.9	4.4	0.2	95.8	95.9	95.8	94.8	−0.42	0.70
	40	CS	1.11	−17.3	−17.3	−0.3	−0.3	85.7	85.7	93.0	93.0	−0.07	0.36
	40	UN	1.08	4.2	4.3	1.3	−0.2	95.2	95.2	94.3	93.8	−0.17	0.41
	40	CSG	1.11	−18.9	−18.0	−0.7	−1.2	84.3	84.8	92.3	92.2	−0.31	0.52
	40	UNG	1.02	7.8	10.4	7.8	−0.2	96.6	96.9	96.6	94.9	−0.42	0.70
2	0	CS	2.00	−17.8	−17.8	0.1	0.1	88.8	88.8	94.7	94.7	0.00	0.50
	0	UN	2.00	0.1	0.1	0.1	0.1	94.7	94.7	94.7	94.7	−0.14	0.55
	0	CSG	2.00	−17.8	−17.8	0.1	0.1	88.8	88.8	94.7	94.7	−0.26	0.52
	0	UNG	2.00	0.1	0.1	0.1	0.1	94.7	94.7	94.7	94.7	−0.42	0.70
	20	CS	2.09	−19.2	−19.2	0.1	−0.1	86.9	86.9	94.1	94.0	−0.03	0.49
	20	UN	2.06	4.2	4.2	1.5	0.1	95.7	95.7	95.0	94.8	−0.11	0.53
	20	CSG	2.10	−20.5	−20.3	0.1	−0.8	85.7	85.9	93.9	93.6	−0.26	0.51
	20	UNG	2.02	4.4	4.9	4.4	0.2	95.8	96.0	95.8	94.9	−0.38	0.70
	40	CS	2.22	−20.9	−20.9	−0.6	−0.6	78.6	78.6	89.8	89.8	−0.08	0.50
	40	UN	2.15	9.6	9.8	2.5	−0.3	95.7	95.7	93.5	92.8	−0.25	0.51
	40	CSG	2.24	−24.6	−23.8	−1.0	−1.2	74.5	75.2	88.9	88.8	−0.27	0.53
	40	UNG	2.05	7.3	10.0	7.3	−0.4	96.2	96.8	96.2	94.9	−0.36	0.70

No correctly specified models.

As for the NVO and NV methods, the SEs had bias when the analysis models were misspecified. The SEs using the RB method had little bias for almost all the settings. The discrepancies of SEs between the RBO and RB methods were very large for the UNG models when the missing probabilities were large. The discrepancies between the RBO and RB methods were small for the other models (CS, UN, and CSG). The model misspecification for the error distribution had little influence on the discrepancy in our settings.

Under the settings for correctly specified models, the deviations of the CP for the RBO method from the nominal level were up to about 5 points. On the contrary, the CP for the RB method was sufficiently close to the nominal level for the same settings. These results were caused by the bias of SEs due to the assumption of orthogonality. Under the model misspecification settings, the results depended on the bias of $\hat{\delta}$ and SEs.

TABLE 3 Summary statistics for AIDS clinical trial data. N: number of subjects, N_t: number of observations in each week.

Group	Week	N_t	Mean	Median	SD	Skewness
Alternating	Baseline	289	25.63	21.00	18.88	1.92
($N = 289$)	8	224	25.84	18.50	25.65	3.09
	16	235	24.66	18.00	24.12	3.52
	24	167	19.32	15.00	17.03	2.40
	32	196	18.33	14.00	16.06	2.15
Combination	Baseline	293	27.41	21.00	23.64	2.58
($N = 293$)	8	224	41.25	21.50	53.22	2.84
	16	251	35.49	21.00	40.57	2.19
	24	174	28.79	17.50	34.91	2.53
	32	187	30.39	18.00	37.67	2.98

4 | CASE STUDY

In this section, we evaluate the effect of non-orthogonality with real data[14,15]. The data are from a randomized, double-blind study of acquired immune deficiency syndrome (AIDS) patients with advanced immune suppression (cluster of differentiation 4 [CD4] counts of less than or equal to 50 cells/mm3). Patients in the AIDS Clinical Trial Group Study 193A were randomized to dual or triple combinations of human immunodeficiency virus-1 reverse transcriptase inhibitors. Specifically, patients were randomized to one of four daily regimens containing 600 mg of zidovudine: zidovudine alternating monthly with 400-mg didanosine; zidovudine plus 2.25 mg of zalcitabine; zidovudine plus 400 mg of didanosine; or zidovudine plus 400 mg of didanosine plus 400 mg of nevirapine (triple therapy). We focused on the measurements of CD4 counts, which were scheduled to be collected at the baseline and at eight-week intervals during follow-up. We also focused on the zidovudine alternating monthly with the didanosine group and the zidovudine plus didanosine group, wherein the motivation of our analysis was to compare the alternating and combination regimens for zidovudine and didanosine. We refer to these groups as alternating and combination groups, respectively. We focused on four time points: 8, 16, 24, and 32 weeks. As for the more detailed data handling process, see Maruo et al.[16]. The original data can be downloaded from https://content.sph.harvard.edu/fitzmaur/ala/ (Datasets->AIDS Clinical Trial Group (ACTG) Study 193A), and the reshaped data are available from the Supporting Information material of this article.

Table 3 shows the summary statistics for the data. Discrepancies between the means and medians were large, and the values of skewness were far from 0, which indicated that the normality assumption did not hold obviously. Heteroscedasticity between the groups was also apparent. The missing proportion for each group and week was not small (14% to 42%). Furthermore, the missing mechanism might not be MCAR. Thus, the ordinal MMRM analysis model was misspecified, and the orthogonality property would be lost in this case.

TABLE 4 Analysis results at week 32 based on the mixed effect models for repeated measures (MMRM) method for AIDS clinical trial data. $\hat{\delta}$: model mean differences between groups at week 32, NVO: naïve SE assuming orthogonality, NV: naïve SE without assuming orthogonality, RBO: robust SE assuming orthogonality, RB: robust SE without assuming orthogonality, Boot: SE based on the nonparametric bootstrap method, CS: compound symmetry, UN: unstructured, CSG: CS structure assuming heteroscedasticity, UNG: UN structure assuming heteroscedasticity.

Covariance structure	$\hat{\delta}$	SE				
		NVO	NV	RBO	RB	Boot
CS	8.65	3.07	3.08	2.46	2.37	2.41
CSG	8.15	3.07	3.07	2.45	2.34	2.35
UN	8.78	2.46	2.46	2.46	2.30	2.29
UNG	8.01	2.45	2.46	2.45	2.18	2.17

The MMRM models using the CS and UN covariance structures with and without the assumption of heteroscedasticity between groups were applied to the data, where the mean structure and marginal error distribution were the same as in the models in the simulation studies. Then, inferences on the model mean differences between groups at week 32 (treatment effect: δ) were obtained, where the SEs of δ were estimated based on the NVO, NV, RBO, and RB methods. The SEs were also estimated with the nonparametric bootstrap method, where the subject ID was resampled with replacement, and resampling was conducted with 10,000 iterations. Such nonparametric SE for mean estimation with the linear models can be regarded as the quasi-true value for a sufficiently large sample.

Table 4 shows the analysis results based on the MMRM methods. The treatment effects for the heteroscedasticity models were smaller than those for the homoscedasticity models. The SEs based on the NVO and NV methods were much different from those based on the RB method for the CS and CSG structures, which suggested that the naïve SEs had misspecification bias. The SEs based on the RB method were smaller than those based on the RBO method especially for the UNG setting and close to the those based on the nonparametric bootstrap method.

5 | APPLYING MMRM WITH SAS

In this section, we introduce the specifications for several procedures in the `SAS` software (`SAS/STAT` 14.3) that can be used to conduct the MMRM analysis. The MMRM analysis is probably most frequently conducted with the `MIXED` procedure, which is the procedure applying linear mixed effect models. However, the `MIXED` procedure only provides the naïve and robust SE assuming the orthogonality. The `GLIMMIX` procedure applies generalized linear mixed models (GLMM). Although the orthogonality property does not hold generally in GLMM, the `GLIMMIX` procedure also provides the SE assuming the orthogonality when applying the MMRM analysis. The `NLMIXED` procedure applies non-linear mixed effect models that include GLMM. The `NLMIXED` procedure does "not" assume the orthogonality property even when models based on normal distributions are applied.

Therefore, the NLMIXED procedure is one available option when we want to conduct the MMRM analysis without assuming the orthogonality property. A sample SAS code for the AIDS data analysis is provided as the Supporting Information, where the robust SEs for the UN and UNG models without assuming the orthogonality property are calculated with the NLMIXED procedure. Note that the codes used for the simulation and case studies in this article are not based on the NLMIXED procedure but our original programs. In the NLMIXED procedure, all specified parameters are estimated simultaneously, while the covariance parameters are estimated separately based on the profile likelihood in the MIXED and GLIMMIX procedures. Therefore, convergence problems for the parameter estimation process might occur, especially when the analysis model is complicated (e.g. the number of time points is large) unless the initial values for the parameters are adequately specified. In our sample code, the initial values used in the NLMIXED procedure are the solutions derived with the MIXED procedure. Furthermore, the marginal models are easily applied using the REPEATED statement in the MIXED procedure or the _RESIDUAL_ option in the GLIMMIX procedure. In the NLMIXED procedure, however, marginal models cannot be applied easily, and so code becomes rather complicated. See our sample code in the Supporting Information for further details.

6 | CONCLUSION

In this study, we investigated the effect of the assumption of orthogonality property on the variance estimation for the MMRM analysis. From the simulation and case studies, it was shown that the robust SEs had non-negligible bias due to the assumption of orthogonality property, especially for the analysis models assuming the UN structure and heteroscedasticity between treatment groups regardless of the actual degree of heteroscedasticity. The SEs for the analysis models assuming homoscedasticity did not have large bias in our simulation settings. However, the analysis models assuming homoscedasticity might lead to bias for the treatment effect estimation when heteroscedasticity between treatment groups is suspected (e.g. see Gosho and Maruo[5]). Therefore, robust SEs without assuming the orthogonality property for the analysis models with the UNG structure are recommended, especially when there is heteroscedasticity between treatment groups.

Although the model misspecification for the error distribution did not have a large influence on the bias of SE when assuming the orthogonality property in our simulation settings, this would not always be true. The simulation settings in this article were limited. However, they would be sufficient to reveal the risk of bias for assuming the orthogonality property in the MMRM analysis. The robust SE assuming the orthogonality property had underestimation bias for our simulation settings; on the other hand, the SE had overestimation bias for our case study. Note that the direction of bias of the robust SE assuming the orthogonality property depends on the situation. Although we investigated the effect of the assumption of the orthogonality property in the framework of MMRM analysis, this issue may arise in more general situations (i.e. likelihood-based linear mixed effect

modelling with multivariate normal distribution). In addition, the issue is not confined to missing data for longitudinal studies. The issue may also arise when the linear mixed model is used with unbalanced data. The effect of the assumption of the orthogonality property under these settings should be investigated in future work. The robust variance estimator with the REML method and the small sample bias adjustments for the robust variance estimator without assuming the orthogonality property in the linear mixed effect models have not yet been developed. Future research for these topics is therefore expected.

In conclusion, assuming the orthogonality property in the MMRM analysis—which is usually applied in the context of MAR, not MCAR—might lead to invalid statistical inferences. Therefore, it is necessary to be careful when applying the MMRM analysis with the most standard software package, especially for analysis models with heteroscedasticity between treatment groups. One available option to apply the MMRM analysis without assuming the orthogonality property is using the `NLMIXED` procedure of the SAS software.

ACKNOWLEDGEMENT

The authors wish to thank the reviewers for helpful comments on earlier version of this paper. This work was supported by JSPS KAKENHI Grant Number JP19K11849.

References

1. Mallinckrodt CH, Clark WS, David SR. Accounting for dropout bias using mixed-effects models. *Journal of Biopharmaceutical Statistics* 2001; 11: 9-21.

2. Siddiqui O, Hung HJ, O'Neill R. A comprehensive comparison based on simulation study and 25 NDA datasets. *Journal of Biopharmaceutical Statistics* 2009; 19: 227-246.

3. Kenward MG, Molenberghs G. Likelihood based frequentist inference when data are missing at random. *Statistical Science* 1998; 13: 236-247.

4. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 2015; 67: 1-48.

5. Gosho M, Maruo K. Effect of heteroscedasticity between treatment groups on mixed-effects models for repeated measures. *Pharmaceutical Statistics* 2018; 17: 587-592.

6. O'Kelly M, Ratitch B. *Clinical Trials with Missing Data: A Guide for Practitioners*. West Sussex: John Wiely & Sons . 2014.

7. Cox DR. Tests of separate families of hypotheses. In: Statistical Laboratory of the University of California. University of California Press; 1961; California.

8. White H. Maximum likelihood estimation of misspecified models. *Econometrica* 1982; 50: 1-25.

9. Takai K, Kano Y. Asymptotic inference with incomplete data. *Communications in Statistics-Theory and Methods* 2013; 42: 3174-3190.

10. Lindstrom MJ, Bates DM. Newton-Raphson and EM algorithms for linear mixed-effects models for repeated measures data. *Journal of the American Statistical Association* 1988; 83: 1014-1022.

11. Mancl LA, Derouen TA. A covariance estimator for GEE with improved small-sample properties. *Biometrics* 2001; 57: 126-134.

12. Kauermann G, Carroll RJ. A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association* 2001; 96: 1387-1396.

13. Gosho M, Hirakawa A, Noma H, Maruo K, Sato Y. Comparison of bias-corrected covariance estimators for MMRM analysis in longitudinal data with dropouts. *Statistical Methods in Medical Research* 2017; 26: 2389-2406.

14. Henry K, Erice A, Tierney C, et al. A randomized, controlled, double-blind study comparing the survival benefit of four different reverse transcriptase inhibitor therapies (three-drug, two-drug, and alternating drug) for the treatment of advanced AIDS. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 1998; 19: 339-349.

15. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis, 2nd Edition*. New York: Wiley . 2011.

16. Maruo K, Yamaguchi Y, Noma H, Gosho M. Interpretable inference on the mixed effect model with the Box–Cox transformation. *Statistics in Medicine* 2017; 36: 2420-2434.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article on the publisher's web site.

How to cite this article: Maruo K, Ishii R, Yamaguchi Y, Doi M, and Gosho M (2019), A note on the bias of standard errors when orthogonality of mean and variance parameters is not satisfied in the MMRM analysis, *Statistics in Medicine*, *2019;00:1–9*.