Mixed Model with Repeated Measures: How Many Observations to Include in the Model?

Ziliang Li * Qing Li †

Abstract

For clinical trials where repeated measurements from a subject are obtained over the course of study, the Mixed Model with Repeated Measure (MMRM) approach is frequently used to analyze the aggregated data and explore the longitudinal profile of the investigational medication. A common practice when applying the MMRM is to model all non-missing measurements as response. However, due to the complexity of the estimation algorithm and the lack of closed form solution of the estimates, the effect of intermediate measurements on the efficacy estimates is not transparent. The understanding of such effect is of practical importance for study planning and results interpretation. This work focuses on some hypothetical experiments where the MMRM is employed to analyze the change from baseline treatment effect based on repeated observations. Theoretical and numerical properties of the treatment effect estimates, with or without the intermediate observation(s) as response will be discussed. Technical considerations when pre-specifying the analysis method will be highlighted.

Key Words: MMRM, intermediate measurements

1 Introduction

For clinical trials where repeated measurements from a subject are obtained over the course of study, the Mixed Model with Repeated Measure (MMRM) approach is frequently used to analyze the aggregated data and explore the longitudinal profile of the investigational medication. A common practice when applying the MMRM is to model all non-missing measurements as response. However, due to the complexity of the estimation algorithm and the lack of closed form solution of the estimates, the effect of intermediate measurements on the efficacy estimates is not transparent. The understanding of such effect is of practical importance for study planning and results interpretation. Maxwell (1998) and Venter et al. (2002) explored this topic, focusing on the MMRM approach and the Analysis of Covariance (ANCOVA) method under a linear growth model assumption. Their study suggest when testing for treatment effects, decent power gain can be achieved by including at least three or more intermediate measurements, compared to a simple pre-post design.

In this work, we study the analytic and numeric properties of the treatment effect estimates derived under three commonly used linear mixed models. The primary focus is to investigate how the point estimate and the associated standard error depend on the intermediate measurements. Further, the efficiency gain and/or bias reduction (if any) will be evaluated under the context of numeric stability of each model through simulations.

This paper is structured as follows. Section 2 introduces the notations and three commonly used models in practice. Section 3 gives detailed derivation of placebo-adjusted change from baseline treatment effect estimates and shows that conditioning on the the variance-covariance parameters, the estimates do not depend on intermediate observations in all three scenarios. Section 4 discusses briefly the impact of missing data.

*Merck & Co., Inc., 2000 Galloping Hill Rd, Kenilworth, NJ 07033
†Merck & Co., Inc., 126 E Lincoln Ave, Rahway, NJ 07065

2 Model Specifications

Following the notations in Verbeke and Molenberghs (2009), assume there are N subjects, each having n_i observations y_{ik} for $i = 1, \ldots, N$ and $k = 1, \ldots, n_i$. Denote $\mathbf{Y}_i = (y_{i1}, \ldots, y_{in_i})'$, the linear mixed model for the repeated measurements y_{ik} is given by

$$\begin{cases} {}_{n_i}\mathbf{Y}_{i\,1} &= {}_{n_i}\mathbf{X}_{i\,\mathrm{p}}\,\boldsymbol{\beta}_{\,1} + {}_{n_i}\mathbf{Z}_{\mathbf{i}\,\mathrm{q}}\,\mathbf{b}_{i\,1} + {}_{n_i}\boldsymbol{\varepsilon}_{i\,1} \\ {}_{\mathrm{q}}\mathbf{b}_{i\,1} &\sim N(\mathbf{0}, \mathbf{D}) \\ {}_{n_i}\boldsymbol{\varepsilon}_{i\,1} &\sim N(\mathbf{0}, \boldsymbol{\Sigma}_i) \end{cases} \tag{1}$$

where $\mathbf{X}_i$ is a $n_i \times p$ design matrix for the fixed effects $\boldsymbol{\beta}$, $\mathbf{Z}_i$ is a $n_i \times q$ design matrix for the the random effects b_i. The random vectors $\mathbf{b}_1, \ldots, \mathbf{b}_N$ and $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_N$ are independently distributed as normal with mean $\mathbf{0}$ and covariance matrices $\mathbf{D}$ and $\boldsymbol{\Sigma}_i$, respectively.

This gives the marginal model for the data,

$$\mathbf{Y}_i \sim N\Big(\mathbf{X}_i\boldsymbol{\beta},\ \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \boldsymbol{\Sigma}_i\Big) \tag{2}$$

Let $\mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \boldsymbol{\Sigma}_i$ and $\mathbf{W}_i = \mathbf{V}_i^{-1}$. Assume that there are at most ℓ observations per subject (i.e., the largest $\boldsymbol{\Sigma}_i$ is of dimension $\ell \times \ell$), let $\boldsymbol{\alpha}$ denote all $q(q+1)/2$ different parameters in $\mathbf{D}$ and $\ell(\ell+1)/2$ parameters in $\boldsymbol{\Sigma}$. The maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$, conditional on $\boldsymbol{\alpha}$, is given by

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{W}_i\mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i'\mathbf{W}_i\mathbf{Y}_i \tag{3}$$

where $\boldsymbol{\alpha}$ can be estimated (thus substituted) by MLE or restricted MLE (REML).

In this paper, three model specifications will be discussed to cover different scenarios of clinical trials.

(1) The longitudinal data analysis (LDA) model for a parallel design with total 7 time points (baseline, Visit 1, Visit 2, ... , Visit 6),

$$\begin{aligned} y_{ik} = \ & \text{subject}_i + \beta_0 + \beta_1 \text{trt}_i + \beta_2 \text{visit}_{i1} + \beta_3 \text{visit}_{i2} + \beta_4 \text{visit}_{i3} \\ & + \beta_5 \text{visit}_{i4} + \beta_6 \text{visit}_{i5} + \beta_7 \text{visit}_{i6} \\ & + \beta_8 \text{trt}_i\, \text{visit}_{i1} + \beta_9 \text{trt}_i\, \text{visit}_{i2} + \beta_{10} \text{trt}_i\, \text{visit}_{i3} \\ & + \beta_{11} \text{trt}_i\, \text{visit}_{i4} + \beta_{12} \text{trt}_i\, \text{visit}_{i5} + \beta_{13} \text{trt}_i\, \text{visit}_{i6} \\ & + \varepsilon_{ik}, \end{aligned} \tag{4}$$

where

$$\text{trt}_i = \begin{cases} 1 & \text{for treatment} \\ 0 & \text{for placebo} \end{cases} \quad \text{and} \quad \text{visit}_{ik} = \begin{cases} 1 & \text{for visit } k \\ 0 & \text{otherwise} \end{cases}.$$

This is the same model used in Dinh and Yang (2011) for their simulation, except that in this paper Visit 0 (for the baseline observation) is used as the reference level for estimating other fixed effects.

The change from baseline treatment effect can be derived as follows,

$$\begin{aligned} & \mathrm{E}[y_{16} - y_{10} | \text{trt}_1 = 1] - \mathrm{E}[y_{26} - y_{20} | \text{trt}_2 = 0] \\ & = \big[(\beta_0 + \beta_1 + \beta_7 + \beta_{13}) - (\beta_0 + \beta_1)\big] - \big[(\beta_0 + \beta_7) - (\beta_0)\big] \\ & = \beta_{13} \end{aligned}$$

(2) Constrained longitudinal data analysis (cLDA) model for a parallel design with total 7 time points,

$$\begin{aligned} y_{ik} = \ & \text{subject}_i + \beta_0 \qquad\qquad + \beta_1 \text{visit}_{i1} + \beta_2 \text{visit}_{i2} + \beta_3 \text{visit}_{i3} \\ & + \beta_4 \text{visit}_{i4} + \beta_5 \text{visit}_{i5} + \beta_6 \text{visit}_{i6} \\ & + \beta_7 \text{trt}_i \ \text{visit}_{i1} + \beta_8 \text{trt}_i \ \text{visit}_{i2} + \beta_9 \text{trt}_i \ \text{visit}_{i3} \\ & + \beta_{10} \text{trt}_i \ \text{visit}_{i4} + \beta_{11} \text{trt}_i \ \text{visit}_{i5} + \beta_{12} \text{trt}_i \ \text{visit}_{i6} \\ & + \varepsilon_{ik}. \end{aligned} \tag{5}$$

It is clear that the only difference of cLDA vs. LDA is that β_1 in the LDA is "fixed" to be 0 and therefore there is one less fixed effect (parameter) in the cLDA model. The change from baseline treatment effect can be similarly derived, which is given by β_{12}.

(3) The LDA model for a simple 2-treatment with 2-period (2×2) crossover design with a total of 3 time points (baseline, Visit 1 and Visit 2) in each period.

$$\begin{aligned} y_{ijk} = \ & \text{subject}_i + \beta_0 + \beta_1 \text{period}_j + \beta_1 \text{trt}_i + \beta_3 \text{visit}_{i1} + \beta_4 \text{visit}_{i2} \\ & + \beta_5 \text{trt}_i \ \text{visit}_{i1} + \beta_6 \text{trt}_i \ \text{visit}_{i2} + \varepsilon_{ik}, \end{aligned} \tag{6}$$

where

$$\text{period}_j = \begin{cases} 1 & \text{for period } j \\ 0 & \text{otherwise} \end{cases}.$$

The change from baseline treatment effect under this model is given by β_6. Notice in this model, *subject* is included as a random effect for all observations from a single subject, thus creating within-subject correlations among observations across periods. See more discussions in Section 3.3.

3 Explicit Expression of $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$

The derivation in this section assumes N is even and there are equal number of subjects in each treatment group. Further assume there is no missing data at any time point for each subject.

3.1 $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ for the LDA Model

With the model specification of (4), the corresponding linear mixed model in the form of (1) is given as follows,

$$_7\mathbf{Y}_{i\ 1} \ = \ _7\mathbf{X}_{i\ 14}\boldsymbol{\beta}_{\ 1} \ + \ _7\mathbf{Z}_{i\ 1} \cdot b_i \ + \ _7\boldsymbol{\varepsilon}_{i\ 1}, \tag{7}$$

where for $i = 1, \ldots, N$,

- Random variables $b_i \sim N(0, \tau^2)$ and random vectors $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, where b_i's and $\boldsymbol{\varepsilon}_i$'s are independent.
- The design matrix for the random effect is simply

$$\mathbf{Z}_i = (1, 1, 1, 1, 1, 1, 1)' \triangleq \mathbf{1}.$$

- The design matrix $\mathbf{X}$ for subjects in the treatment group is given by

$$\mathbf{X}_T \triangleq \left[\begin{array}{cc:cccccc:cccccc} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & & & & & & 1 & & & & & \\ 1 & 1 & & 1 & & & & & & 1 & & & & \\ 1 & 1 & & & 1 & & & & & & 1 & & & \\ 1 & 1 & & & & 1 & & & & & & 1 & & \\ 1 & 1 & & & & & 1 & & & & & & 1 & \\ 1 & 1 & & & & & & 1 & & & & & & 1 \end{array}\right],$$

and for subjects in the placebo group is given by

$$\mathbf{X}_P \triangleq \left[\begin{array}{cc:cccccc:cccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & & & & & & 0 & & & & & \\ 1 & 0 & & 1 & & & & & & 0 & & & & \\ 1 & 0 & & & 1 & & & & & & 0 & & & \\ 1 & 0 & & & & 1 & & & & & & 0 & & \\ 1 & 0 & & & & & 1 & & & & & & 0 & \\ 1 & 0 & & & & & & 1 & & & & & & 0 \end{array}\right].$$

Since there is no missing data, all subjects in either treatment group share the same design matrix ($\mathbf{X}_T$ and $\mathbf{X}_P$) and their covariance matrices are the same, too. Therefore, $\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{W}_i\mathbf{X}_i$ in (3) can be written as

$$\begin{aligned}\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{W}_i\mathbf{X}_i &= \sum_{i=1}^{N/2} \mathbf{X}_T'\mathbf{W}\mathbf{X}_T + \sum_{i=1}^{N/2} \mathbf{X}_P'\mathbf{W}\mathbf{X}_P \\ &= \frac{N}{2}\Big(\mathbf{X}_T'\mathbf{W}\mathbf{X}_T + \mathbf{X}_P'\mathbf{W}\mathbf{X}_P\Big).\end{aligned}$$

In the rest of this note, $\sum_{i=1}^{N} \mathbf{X}'\mathbf{W}\mathbf{X}$ and $\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{W}_i\mathbf{X}_i$ will be used interchangeably unless stated otherwise.

Assume $\text{Cov}[\boldsymbol{\varepsilon}] = \boldsymbol{\Sigma}$ is an unstructured variance-covariance matrix where

$$\underset{7}{\Sigma} = \underset{7}{\begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{02} & & \cdots & & \sigma_{06} \\ \sigma_{10} & \sigma_1^2 & & & & & \\ \sigma_{20} & & \sigma_2^2 & & & & \\ & & & \sigma_3^2 & & & \vdots \\ \vdots & & & & \sigma_4^2 & & \\ & & & & & \sigma_5^2 & \\ \sigma_{60} & & \cdots & & & & \sigma_6^2 \end{bmatrix}}, \tag{8}$$

where $\sigma_{ij} = \sigma_{ji} = \rho_{ij}\sigma_i\sigma_j$. Then

$$\begin{aligned}\mathbf{V}_i &= \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \boldsymbol{\Sigma}_i \\ &= \mathbf{1}\cdot\tau^2\cdot\mathbf{1}' + \boldsymbol{\Sigma} \\ &= \underset{7}{\begin{bmatrix} \sigma_0^2+\tau^2 & \sigma_{01}+\tau^2 & \sigma_{02}+\tau^2 & & \cdots & & \sigma_{06}+\tau^2 \\ \sigma_{10}+\tau^2 & \sigma_1^2+\tau^2 & & & & & \\ \sigma_{20}+\tau^2 & & \sigma_2^2+\tau^2 & & & & \\ & & & \sigma_3^2+\tau^2 & & & \vdots \\ \vdots & & & & \sigma_4^2+\tau^2 & & \\ & & & & & \sigma_5^2+\tau^2 & \\ \sigma_{60}+\tau^2 & & & \cdots & & & \sigma_6^2+\tau^2 \end{bmatrix}}.\end{aligned} \tag{9}$$

To derive the explicit expression of $\hat{\beta}_{13}$ using the result (3), we first need to derive the explicit form of $(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{W}_i \mathbf{X}_i)^{-1}$. It is clear that inverting a matrix is typically done numerically as the algebraic expression is usually extremely complicated. However, due to the special structure of $\mathbf{X}$ (sparse matrix with entries being 1 or 0), it turns out the inverse of $\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{W}_i \mathbf{X}_i$ can be derived explicitly. The key approach is to use row operations (and row operation only) to further simplify them and then apply block matrix calculations. The calculations are divided into three steps,

① Derive a matrix $\mathbf{T}_L$ to perform row operations to simplify $\mathbf{X}_T'$ and $\mathbf{X}_P'$.

② Calculate $(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{W}_i \mathbf{X}_i)^{-1}$ using block matrix inverse.

③ Derive $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ and then extract $\hat{\beta}_{13}$.

For step ①, it consists of performing three sets of row operations sequentially.

(1.1) Subtract ROW 1 by ROW 2, then ROW 3 by ROW 9, ROW 4 by ROW 10, ..., ROW 8 by ROW 14. This can be achieved by left multiplying $\mathbf{X}_T'$ ($\mathbf{X}_P'$) with

$$\mathbf{T}_1 = \left[\begin{array}{cc|cccccc|cccccc} 1 & -1 & & & & & & & & & & & & \\ 0 & 1 & & & & & & & & & & & & \\ \hline & & 1 & & & & & & -1 & & & & & \\ & & & 1 & & & & & & -1 & & & & \\ & & & & 1 & & & & & & -1 & & & \\ & & & & & 1 & & & & & & -1 & & \\ & & & & & & 1 & & & & & & -1 & \\ & & & & & & & 1 & & & & & & -1 \\ \hline & & & & & & & & 1 & & & & & \\ & & & & & & & & & 1 & & & & \\ & & & & & & & & & & 1 & & & \\ & & & & & & & & & & & 1 & & \\ & & & & & & & & & & & & 1 & \\ & & & & & & & & & & & & & 1 \end{array}\right]. \quad (10)$$

(1.2) Subtract ROW 2 by ROW 9 to ROW 14. Subtract ROW 1 by ROW 3 to ROW 8. The corresponding matrix is $\mathbf{T}_2$,

$$\mathbf{T}_2 = \left[\begin{array}{cc|cccccc|cccccc} 1 & 0 & -1 & -1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 & -1 & -1 \\ \hline & & 1 & & & & & & & & & & & \\ & & & 1 & & & & & & & & & & \\ & & & & 1 & & & & & & & & & \\ & & & & & 1 & & & & & & & & \\ & & & & & & 1 & & & & & & & \\ & & & & & & & 1 & & & & & & \\ \hline & & & & & & & & 1 & & & & & \\ & & & & & & & & & 1 & & & & \\ & & & & & & & & & & 1 & & & \\ & & & & & & & & & & & 1 & & \\ & & & & & & & & & & & & 1 & \\ & & & & & & & & & & & & & 1 \end{array}\right] \quad (11)$$

(1.3) Next, switch ROW 2 with ROW 3, then ROW 3 with ROW 4, continue till ROW 7

with ROW 8. This can be achieved by the following matrix $\mathbf{T}_3$,

$$\mathbf{T}_3 = \left[\begin{array}{cccccccc|cccccc} 1 & & & & & & & & & & & & & \\ 0 & 0 & 1 & & & & & & & & & & & \\ & & 0 & 1 & & & & & & & & & & \\ & & & 0 & 1 & & & & & & & & & \\ & & & & 0 & 1 & & & & & & & & \\ & & & & & 0 & 1 & & & & & & & \\ & & & & & & 0 & 1 & & & & & & \\ & 1 & & & & & & 0 & & & & & & \\ \hline & & & & & & & & 1 & & & & & \\ & & & & & & & & & 1 & & & & \\ & & & & & & & & & & 1 & & & \\ & & & & & & & & & & & 1 & & \\ & & & & & & & & & & & & 1 & \\ & & & & & & & & & & & & & 1 \end{array}\right]. \quad (12)$$

Denote

$$\mathbf{T}_L = \mathbf{T}_3\mathbf{T}_2\mathbf{T}_1 = \left[\begin{array}{cccccccc|cccccc} 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ & 0 & 1 & & & & & & -1 & & & & & \\ & & 0 & 1 & & & & & & -1 & & & & \\ & & & 0 & 1 & & & & & & -1 & & & \\ & & & & 0 & 1 & & & & & & -1 & & \\ & & & & & 0 & 1 & & & & & & -1 & \\ & & & & & & 0 & 1 & & & & & & -1 \\ & 1 & & & & & & 0 & -1 & -1 & -1 & -1 & -1 & -1 \\ \hline & & & & & & & & 1 & & & & & \\ & & & & & & & & & 1 & & & & \\ & & & & & & & & & & 1 & & & \\ & & & & & & & & & & & 1 & & \\ & & & & & & & & & & & & 1 & \\ & & & & & & & & & & & & & 1 \end{array}\right]. \quad (13)$$

Then we have

$$\mathbf{T}_L \cdot \mathbf{X}_T' = \begin{bmatrix} \mathbf{0}_{7\times 7} \\ \\ \mathbf{I}_{7\times 7} \end{bmatrix} \quad \text{and} \quad \mathbf{T}_L \cdot \mathbf{X}_P' = \begin{bmatrix} \mathbf{I}_{7\times 7} \\ \\ \mathbf{0}_{7\times 7} \end{bmatrix} \quad (14)$$

Now proceed to step ② and calculate $(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{W}_i\mathbf{X}_i)^{-1}$. Notice that

$$\begin{aligned} \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{W}_i\mathbf{X}_i\right)^{-1} &= \frac{2}{N}\left(\mathbf{X}_T'\mathbf{W}\mathbf{X}_T + \mathbf{X}_P'\mathbf{W}\mathbf{X}_P\right)^{-1} \\ &= \frac{N}{2}\,\mathbf{T}_L'\left(\mathbf{T}_L\mathbf{X}_T'\mathbf{W}\mathbf{X}_T\mathbf{T}_L' + \mathbf{T}_L\mathbf{X}_P'\mathbf{W}\mathbf{X}_P\mathbf{T}_L'\right)^{-1}\mathbf{T}_L \\ &= \frac{2}{N}\mathbf{T}_L'\left(\begin{bmatrix} \mathbf{0}_{7\times 7} \\ \\ \mathbf{I}_{7\times 7} \end{bmatrix}\mathbf{W}\begin{bmatrix}\mathbf{0}_{7\times 7} & \mathbf{I}_{7\times 7}\end{bmatrix} + \begin{bmatrix} \mathbf{I}_{7\times 7} \\ \\ \mathbf{0}_{7\times 7} \end{bmatrix}\mathbf{W}\begin{bmatrix}\mathbf{I}_{7\times 7} & \mathbf{0}_{7\times 7}\end{bmatrix}\right)^{-1}\mathbf{T}_L \\ &= \frac{2}{N}\mathbf{T}_L'\begin{bmatrix} \mathbf{W}^{-1} & \mathbf{0}_{7\times 7} \\ \\ \mathbf{0}_{7\times 7} & \mathbf{W}^{-1} \end{bmatrix}\mathbf{T}_L \end{aligned}$$

Next, proceed to step ③ and derive the explicit form of $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ and subsequently $\hat{\beta}_{13}$. Let $\mathbf{Y}_{i|1}$ denote the vector observations from subject i who are on the treatment group

($\text{trt}_i = 1$) and $\mathbf{Y}_{i|0}$ for subject i on placebo group ($\text{trt}_i = 0$). Recall (3),

$$\begin{aligned}
\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) &= \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{W}_i \mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{W}_i \mathbf{Y}_i \\
&= \left(\sum_{i=1}^{N/2} \mathbf{X}_T' \mathbf{W} \mathbf{X}_T + \sum_{i=1}^{N/2} \mathbf{X}_P' \mathbf{W} \mathbf{X}_P\right)^{-1} \left(\sum_{i=1}^{N/2} \mathbf{X}_T' \mathbf{W}_i \mathbf{Y}_{i|1} + \sum_{i=1}^{N/2} \mathbf{X}_P' \mathbf{W}_i \mathbf{Y}_{i|0}\right) \\
&= \frac{2}{N} \mathbf{T}_L' \begin{bmatrix} \mathbf{W}^{-1} & \mathbf{0}_{7\times 7} \\ \mathbf{0}_{7\times 7} & \mathbf{W}^{-1} \end{bmatrix} \mathbf{T}_L \left(\mathbf{T}_L\right)^{-1} \left(\sum_{i=1}^{N/2} \begin{bmatrix} \mathbf{0}_{7\times 7} \\ \mathbf{W} \end{bmatrix} \mathbf{Y}_{i|1} + \sum_{i=1}^{N/2} \begin{bmatrix} \mathbf{W} \\ \mathbf{0}_{7\times 7} \end{bmatrix} \mathbf{Y}_{i|0}\right) \\
&= \frac{2}{N} \mathbf{T}_L' \begin{bmatrix} \sum_{i=1}^{N/2} \mathbf{Y}_{i|0} \\ \sum_{i=1}^{N/2} \mathbf{Y}_{i|1} \end{bmatrix}. \qquad (15)
\end{aligned}$$

To obtain the form of $\hat{\beta}_{13}$, left multiply $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ by $\mathbf{c} = {}_1(0,0,0,\ldots,1)_{14}$,

$$\hat{\beta}_{13} = \mathbf{c} \cdot \hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \mathbf{c} \cdot \frac{2}{N} \mathbf{T}_L' \begin{bmatrix} \sum_{i=1}^{N/2} \mathbf{Y}_{i|0} \\ \sum_{i=1}^{N/2} \mathbf{Y}_{i|1} \end{bmatrix}.$$

Therefore

$$\begin{aligned}
\hat{\beta}_{13} &= \frac{2}{N} (1,0,0,0,0,0,0,-1,-1,0,0,0,0,0,0,1) \cdot \begin{bmatrix} \sum_{i=1}^{N/2} \mathbf{Y}_{i|0} \\ \sum_{i=1}^{N/2} \mathbf{Y}_{i|1} \end{bmatrix} \\
&= (\bar{y}_{6|1} - \bar{y}_{0|1}) - (\bar{y}_{6|0} - \bar{y}_{0|0}), \qquad (16)
\end{aligned}$$

where $\bar{y}_{6|1}$ ($\bar{y}_{0|1}$) is the average of all observations at Visit 6 (baseline) from subjects who are on treatment, while $\bar{y}_{6|0}$ ($\bar{y}_{0|0}$) is the average of all observations at Visit 6 (baseline) from subjects who are on placebo. It is straight forward to use the matrix algebra to derive $\text{Var}(\hat{\beta}_{13}) = \frac{4}{N}\left(\sigma_0^2 - 2\rho_{06}\sigma_0\sigma_6 + \sigma_6^2\right)$.

3.2 $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ for the cLDA Model

The corresponding linear mixed model is almost identical to (7) in the LDA model,

$${}_7\mathbf{Y}_{i\,1} = {}_7\mathbf{X}_{i\,13}\boldsymbol{\beta}_{1} + {}_7\mathbf{Z}_{i\,1} \cdot b_i + {}_7\boldsymbol{\varepsilon}_{i\,1}, \qquad (17)$$

where the random variables b_i, $\boldsymbol{\varepsilon}_i$ and the design matrix $\mathbf{Z}_i$ are defined in the same way as their counterparts in (7). The key difference is that marginal treatment effect is *not* included in the cLDA model to reflect the constraint that measurements at baseline have the same mean for all subjects.

Thus, the design matrix $\mathbf{X}$ in cLDA for subjects in the treatment group is given by

$$\mathbf{X}_T \triangleq \left[\begin{array}{c:cccccc:cccccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & & & & & & 1 & & & & & \\
1 & & 1 & & & & & & 1 & & & & \\
1 & & & 1 & & & & & & 1 & & & \\
1 & & & & 1 & & & & & & 1 & & \\
1 & & & & & 1 & & & & & & 1 & \\
1 & & & & & & 1 & & & & & & 1
\end{array}\right],$$

and for subjects in the placebo group is given by

$$\mathbf{X}_P = \left[\begin{array}{c:cccccc:cccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & & & & & & 0 & & & & & \\ 1 & & 1 & & & & & & 0 & & & & \\ 1 & & & 1 & & & & & & 0 & & & \\ 1 & & & & 1 & & & & & & 0 & & \\ 1 & & & & & 1 & & & & & & 0 & \\ 1 & & & & & & 1 & & & & & & 0 \end{array}\right].$$

Similar to the LDA model, the change from baseline treatment effect estimator is given by $\hat{\beta}_{12}$, with the MLE of $\beta(\alpha)$ given by

$$\begin{aligned} \hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) &= \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{W}_i \mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{W}_i \mathbf{Y}_i \\ &= \frac{2}{N}\Big(\mathbf{X}_T' \mathbf{W} \mathbf{X}_T + \mathbf{X}_P' \mathbf{W} \mathbf{X}_P\Big)^{-1} \left(\sum_{i=1}^{N/2} \mathbf{X}_T' \mathbf{W}_i \mathbf{Y}_{i|1} + \sum_{i=1}^{N/2} \mathbf{X}_P' \mathbf{W}_i \mathbf{Y}_{i|0}\right). \end{aligned}$$

Following the similar steps outlined in Section 3.1, step (1) is to derive the transform matrix $\mathbf{T}_L$ to simplify $\mathbf{X}_T'$ and $\mathbf{X}_P'$. It turns out one operation is sufficient: subtract ROW 1 by ROW 2 to ROW 7. This can be achieved by the following matrix $\mathbf{T}_L$,

$$\mathbf{T}_L = \left[\begin{array}{ccccccc:ccccccc} 1 & -1 & -1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 1 & & & & & & & & & & & & \\ & & 1 & & & & & & & & & & & \\ & & & 1 & & & & & & & & & & \\ & & & & 1 & & & & & & & & & \\ & & & & & 1 & & & & & & & & \\ & & & & & & 1 & & & & & & & \\ \hdashline & & & & & & & 1 & & & & & & \\ & & & & & & & & 1 & & & & & \\ & & & & & & & & & 1 & & & & \\ & & & & & & & & & & 1 & & & \\ & & & & & & & & & & & 1 & & \\ & & & & & & & & & & & & 1 & \end{array}\right]. \tag{18}$$

Applying this operation gives

$$\mathbf{T}_L \cdot \mathbf{X}_T' = \left[\begin{array}{ccccccc} & & & \mathbf{I}_{7\times 7} & & & \\ & & & & & & \\ 0 & 1 & & & & & \\ 0 & & 1 & & & & \\ 0 & & & 1 & & & \\ 0 & & & & 1 & & \\ 0 & & & & & 1 & \\ 0 & & & & & & 1 \end{array}\right] \quad \text{and} \quad \mathbf{T}_L \cdot \mathbf{X}_P' = \left[\begin{array}{c} \mathbf{I}_{6\times 7} \\ \\ \mathbf{0}_{7\times 7} \\ \\ \end{array}\right],$$

Let $\mathbf{M}'$ and $\mathbf{0}'$ denote the submatrices,

$$\mathbf{M}' = \left[\begin{array}{ccccccc} 0 & 1 & & & & & \\ 0 & & 1 & & & & \\ 0 & & & 1 & & & \\ 0 & & & & 1 & & \\ 0 & & & & & 1 & \\ 0 & & & & & & 1 \end{array}\right] \quad \text{and} \quad \mathbf{0}' = \left[\begin{array}{ccccccc} 0 & 0 & & & & & \\ 0 & & 0 & & & & \\ 0 & & & 0 & & & \\ 0 & & & & 0 & & \\ 0 & & & & & 0 & \\ 0 & & & & & & 0 \end{array}\right],$$

the MLE $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ can be re-written as

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \mathbf{T}'_L \left(\frac{N}{2} \begin{bmatrix} \mathbf{I} \\ \mathbf{M}' \end{bmatrix} \mathbf{W} \begin{bmatrix} \mathbf{I} & \mathbf{M} \end{bmatrix} + \frac{N}{2} \begin{bmatrix} \mathbf{I} \\ \mathbf{0}' \end{bmatrix} \mathbf{W} \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \right)^{-1} \cdot \left(\sum_{i=1}^{N/2} \begin{bmatrix} \mathbf{I} \\ \mathbf{M}' \end{bmatrix} \mathbf{W}\, \mathbf{Y}_{i|1} + \sum_{i=1}^{N/2} \begin{bmatrix} \mathbf{I} \\ \mathbf{0}' \end{bmatrix} \mathbf{W}\, \mathbf{Y}_{i|0} \right). \tag{19}$$

With some simplifications,

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) \;=\; \mathbf{T}'_L \frac{2}{N} \left(\sum_{i=1}^{N/2} \begin{bmatrix} \frac{1}{2} & \mathbf{0} \\ \frac{1}{2}\frac{\mathbf{V}_{06}}{v_{00}} & -\frac{1}{2}\mathbf{I} \\ -\frac{\mathbf{V}_{06}}{v_{00}} & \mathbf{I} \end{bmatrix} Y_{i|1} + \sum_{i=1}^{N/2} \begin{bmatrix} \frac{1}{2} & \mathbf{0} \\ -\frac{1}{2}\frac{\mathbf{V}_{06}}{v_{00}} & \frac{1}{2}\mathbf{I} \\ \frac{\mathbf{V}_{06}}{v_{00}} & -\mathbf{I} \end{bmatrix} Y_{i|0} \right). \tag{20}$$

To get $\hat{\beta}_{12}$, let $\mathbf{c} = {}_1(0, 0, \ldots, 0, 1)_{13}$ and it is clear that $\mathbf{c} \cdot \mathbf{T}'_L \;=\; (0, 0, \ldots, 0, 1)$, thus

$$\begin{aligned} \hat{\beta}_{12} &= \frac{2}{N} (0, 0, \ldots, 0, 1) \cdot \hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) \\ &= \left(\bar{y}_{6|1} - \bar{y}_{6|0}\right) - \frac{\rho_{06}\sigma_0\sigma_6 + \tau^2}{\sigma_0^2 + \tau^2} \left(\bar{y}_{0|1} - \bar{y}_{0|0}\right). \end{aligned} \tag{21}$$

Let $\mathbf{L}' = \; (-\frac{v_{06}}{v_{00}}, 0, 0, 0, 0, 0, 1)$. Then we have

$$\begin{aligned} \text{Var}(\hat{\beta}_{12}) &= \frac{4}{N^2} \left[\text{Var}\left(\sum_{i=1}^{N/2} \mathbf{L}' \cdot \mathbf{Y}_{i|1} \right) \;+\; \text{Var}\left(\sum_{i=1}^{N/2} -\mathbf{L}' \cdot \mathbf{Y}_{i|0} \right) \right] \\ &= \frac{4}{N} \left(\sigma_6^2 + \tau^2 - \frac{(\rho_{06}\sigma_0\sigma_6 + \tau^2)^2}{\sigma_0^2 + \tau^2} \right). \end{aligned}$$

In the absence of random effect where $\tau = 0$, the variance simplifies to $\sigma_6^2(1 - \rho_{06}^2)$.

3.3 $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ for the Crossover Model

Recall the LDA model (6) for a simple 2-treatment with 2-period (2×2) crossover design with total 3 observation time points (baseline, Visit 1 and Visit 2),

$$\begin{aligned} y_{ijk} = \; & \text{subject}_i + \beta_0 + \beta_1 \text{period}_j + \beta_2 \text{trt}_i + \beta_3 \text{visit}_{i1} + \beta_4 \text{visit}_{i2} \\ & + \beta_5 \text{trt}_i\, \text{visit}_{i1} + \beta_6 \text{trt}_i\, \text{visit}_{i2} + \varepsilon_{ik}, \end{aligned}$$

where

$$\text{period}_j = \begin{cases} 1 & \text{for period } j \\ 0 & \text{otherwise} \end{cases},$$

with the change from baseline treatment effect given by β_6. Under this specification, assume $\text{subject}_i \;\sim\; N(0, \tau^2)$, the covariance matrix for the 6 observations from a single subject is given by

$$V = \left[\begin{array}{ccc:ccc} \sigma_0^2 + \tau^2 & \sigma_{01} + \tau^2 & \sigma_{02} + \tau^2 & \tau^2 & \tau^2 & \tau^2 \\ \sigma_{10} + \tau^2 & \sigma_1^2 + \tau^2 & \sigma_{12} + \tau^2 & \tau^2 & \tau^2 & \tau^2 \\ \sigma_{20} + \tau^2 & \sigma_{21} + \tau^2 & \sigma_2^2 + \tau^2 & \tau^2 & \tau^2 & \tau^2 \\ \hdashline \tau^2 & \tau^2 & \tau^2 & \sigma_0^2 + \tau^2 & \sigma_{01} + \tau^2 & \sigma_{02} + \tau^2 \\ \tau^2 & \tau^2 & \tau^2 & \sigma_{10} + \tau^2 & \sigma_1^2 + \tau^2 & \sigma_{12} + \tau^2 \\ \tau^2 & \tau^2 & \tau^2 & \sigma_{20} + \tau^2 & \sigma_{21} + \tau^2 & \sigma_2^2 + \tau^2 \end{array}\right]. \tag{22}$$

As a special case, if the random effect *subject* is nested within *period*, then the off-diagonal block of V given in (22) will be all 0. In this section, the derivation shall proceed with the general non-zero off-diagonal elements followed by the discussion of nested random effect. Similar to the previous sections, denote the $\mathbf{V}^{-1}$ by $\mathbf{W}$. Due to the symmetric nature of $\mathbf{V}$, $\mathbf{W}$ has the following block structure,

$$\mathbf{W} = \left[\begin{array}{ccc:ccc} w_{11} & w_{12} & w_{13} & w_{14} & w_{15} & w_{16} \\ w_{12} & w_{22} & w_{23} & w_{24} & w_{25} & w_{26} \\ w_{13} & w_{23} & w_{33} & w_{34} & w_{35} & w_{36} \\ \hdashline w_{14} & w_{15} & w_{16} & w_{11} & w_{12} & w_{13} \\ w_{24} & w_{25} & w_{26} & w_{12} & w_{22} & w_{23} \\ w_{34} & w_{35} & w_{36} & w_{13} & w_{23} & w_{33} \end{array}\right] = \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{12} & \mathbf{W}_{11}. \end{bmatrix} \tag{23}$$

Consider that N subjects are evenly divided into two treatment sequences, Sequence 1: treatment followed by placebo and Sequence 2: placebo followed by treatment. Each subject has 6 observations, denoted as $\mathbf{Y}_i = (y_{i10}, y_{i11}, y_{i12}, y_{i20}, y_{i21}, y_{i22})$. Assume there is no missing data, then the MLE for $\beta(\alpha)$ under model (6) is given by

$$\begin{aligned} \hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) &= \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{W}_i \mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{W}_i \mathbf{Y}_i \\ &= \left(\sum_{i=1}^{N/2} \mathbf{X}_{TP}' \mathbf{W} \mathbf{X}_{TP} + \sum_{i=1}^{N/2} \mathbf{X}_{PT}' \mathbf{W} \mathbf{X}_{PT}\right)^{-1} \left(\sum_{i=1}^{N/2} \mathbf{X}_{TP}' \mathbf{W} \mathbf{Y}_{i|1} + \sum_{i=1}^{N/2} \mathbf{X}_{PT}' \mathbf{W} \mathbf{Y}_{i|2}\right) \\ &= \frac{2}{N}\Big(\mathbf{X}_{TP}' \mathbf{W} \mathbf{X}_{TP} + \mathbf{X}_{PT}' \mathbf{W} \mathbf{X}_{PT}\Big)^{-1} \left(\sum_{i=1}^{N/2} \mathbf{X}_{TP}' \mathbf{W} \mathbf{Y}_{i|1} + \sum_{i=1}^{N/2} \mathbf{X}_{PT}' \mathbf{W} \mathbf{Y}_{i|2}\right), \end{aligned} \tag{24}$$

where $\mathbf{Y}_{i|1}$ and $\mathbf{Y}_{i|2}$ denote the 6 observations for subject i in the sequence 1 and 2, respectively. Here $\mathbf{W}$ is the inverse of $\mathbf{V}$ given by (22), $\mathbf{X}_{TP}$ is the design matrix for subjects in sequence 1,

$$\mathbf{X}_{TP} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \tag{25}$$

and $\mathbf{X}_{PT}$ is the design matrix for subjects in sequence 2,

$$\mathbf{X}_{PT} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}. \tag{26}$$

Similar steps as in previous sections will be taken to derive $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ with some added tricks. Step (1) is to perform row operations to simplify both design matrices.

(1.1) Subtract ROW 1 by ROW 3, then ROW 4 by ROW 6 and ROW 5 by ROW 7. This can be achieved by left multiplying $\mathbf{X}_{TP}'$ (and $\mathbf{X}_{PT}'$) by

$$\mathbf{T}_1 = \begin{bmatrix} 1 & & -1 & & & & \\ & 1 & & & & & \\ & & 1 & & & & \\ & & & 1 & & -1 & \\ & & & & 1 & & -1 \\ & & & & & 1 & \\ & & & & & & 1 \end{bmatrix}. \tag{27}$$

(1.2) Switch ROW 2 with ROW 4 and ROW 3 with ROW 5. The corresponding row operation matrix is $\mathbf{T}_2$,

$$\mathbf{T}_2 = \begin{bmatrix} 1 & & & & & & & \\ & 0 & & 1 & & & & \\ & & 0 & & 1 & & & \\ & 1 & & 0 & & & & \\ & & 1 & & 0 & & & \\ & & & & & 1 & & \\ & & & & & & 1 & \end{bmatrix}. \tag{28}$$

(1.3) Subtract ROW 1 by ROW 2 and 3, then subtract ROW 5 by ROW 6 and 7, with the corresponding matrix $\mathbf{T}_3$,

$$\mathbf{T}_3 = \begin{bmatrix} 1 & -1 & -1 & & & & & \\ & 1 & & & & & & \\ & & 1 & & & & & \\ & & & 1 & & & & \\ & & & & 1 & -1 & -1 & \\ & & & & & 1 & & \\ & & & & & & 1 & \end{bmatrix}. \tag{29}$$

(1.4) Finally, roll ROW 4 up to ROW 1. This can be achieved by $\mathbf{T}_4$,

$$\mathbf{T}_4 = \begin{bmatrix} & & & 1 & & & & \\ 1 & & & & & & & \\ & 1 & & & & & & \\ & & 1 & & & & & \\ & & & & 1 & -1 & -1 & \\ & & & & & 1 & & \\ & & & & & & 1 & \end{bmatrix}. \tag{30}$$

Applying these four operations gives

$$\mathbf{T}_4\mathbf{T}_3\mathbf{T}_2\mathbf{T}_1 \cdot \mathbf{X}'_{TP} = \begin{bmatrix} \mathbf{1}'_{1\times 3} & \mathbf{0}'_{1\times 3} \\ \mathbf{0}_{3\times 3} & \mathbf{I}_{3\times 3} \\ \mathbf{I}_{3\times 3} & \mathbf{0}_{3\times 3} \end{bmatrix} \quad \text{and} \quad \mathbf{T}_4\mathbf{T}_3\mathbf{T}_2\mathbf{T}_1 \cdot \mathbf{X}'_{PT} = \begin{bmatrix} \mathbf{1}'_{1\times 3} & \mathbf{0}'_{1\times 3} \\ \mathbf{I}_{3\times 3} & \mathbf{0}_{3\times 3} \\ \mathbf{0}_{3\times 3} & \mathbf{I}_{3\times 3} \end{bmatrix}.$$

Multiplying all four row operation matrices gives,

$$\mathbf{T}_L = \begin{bmatrix} 0 & 1 & & & & & & \\ 1 & 0 & -1 & -1 & -1 & 1 & 1 & \\ & & 0 & 1 & & -1 & & \\ & & & 0 & 1 & & -1 & \\ & & 1 & & 0 & -1 & -1 & \\ & & & & & 1 & & \\ & & & & & & 1 & \end{bmatrix}. \tag{31}$$

Step ② is to compute the inverse of $\mathbf{X}'_{TP}\mathbf{W}\mathbf{X}_{TP} + \mathbf{X}'_{PT}\mathbf{W}\mathbf{X}_{PT}$. From step ①,

$$\begin{aligned} &\left(\mathbf{X}'_{TP}\mathbf{W}\mathbf{X}_{TP} + \mathbf{X}'_{PT}\mathbf{W}\mathbf{X}_{PT}\right)^{-1} \\ &= \mathbf{T}'_L \cdot \begin{bmatrix} 2\mathbf{1}'\mathbf{W}_{11}\mathbf{1} & \mathbf{1}'\mathbf{W}_{11} + \mathbf{1}'\mathbf{W}_{12} & \mathbf{1}'\mathbf{W}_{11} + \mathbf{1}'\mathbf{W}_{12} \\ \mathbf{W}_{11}\mathbf{1} + \mathbf{W}_{12}\mathbf{1} & 2\mathbf{W}_{11} & 2\mathbf{W}_{12} \\ \mathbf{W}_{11}\mathbf{1} + \mathbf{W}_{12}\mathbf{1} & 2\mathbf{W}_{12} & 2\mathbf{W}_{11} \end{bmatrix}^{-1} \mathbf{T}_L \\ &= \mathbf{T}'_L\, \mathbf{M}_1^{-1}\, \mathbf{T}_L \end{aligned}$$

It can be shown that

$$\mathbf{M}_1^{-1} = \mathbf{T}\begin{bmatrix} \delta^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}\mathbf{V}_{11} & \frac{1}{2}\mathbf{V}_{12} \\ \mathbf{0} & \frac{1}{2}\mathbf{V}_{12} & \frac{1}{2}\mathbf{V}_{11} \end{bmatrix}\mathbf{T}' = \begin{bmatrix} \frac{1}{\delta} & -\frac{1}{2\delta}\mathbf{1}' & -\frac{1}{2\delta}\mathbf{1}' \\ -\frac{1}{2\delta}\mathbf{1} & \frac{1}{4\delta}+\frac{1}{2}\mathbf{V}_{11} & \frac{1}{4\delta}+\frac{1}{2}\mathbf{V}_{12} \\ -\frac{1}{2\delta}\mathbf{1} & \frac{1}{4\delta}+\frac{1}{2}\mathbf{V}_{12} & \frac{1}{4\delta}+\frac{1}{2}\mathbf{V}_{11} \end{bmatrix},$$

where $\delta = \mathbf{1}'\mathbf{W}_{11}\mathbf{1} - \mathbf{1}'\mathbf{W}_{12}\mathbf{1}$. Plug these results into (24),

$$\begin{aligned}\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) &= \frac{2}{N}\Big(\mathbf{X}'_{TP}\mathbf{W}\mathbf{X}_{TP} + \mathbf{X}'_{PT}\mathbf{W}\mathbf{X}_{PT}\Big)^{-1}\left(\sum_{i=1}^{N/2}\mathbf{X}'_{TP}\mathbf{W}\mathbf{Y}_{i|1} + \sum_{i=1}^{N/2}\mathbf{X}'_{PT}\mathbf{W}\mathbf{Y}_{i|2}\right) \\ &= \frac{2}{N}\mathbf{T}_L\left(\sum_{i=1}^{N/2}\begin{bmatrix} \frac{1}{2\delta}\big(\mathbf{1}'\mathbf{W}_{11}-\mathbf{1}'\mathbf{W}_{12}\big) & \frac{1}{2\delta}\big(\mathbf{1}'\mathbf{W}_{12}-\mathbf{1}'\mathbf{W}_{11}\big) \\ -\frac{1}{2\delta}\mathbf{1}\mathbf{1}'\mathbf{W}_{11}+\frac{1}{4\delta}\mathbf{W}_{112} & -\frac{1}{2\delta}\mathbf{1}\mathbf{1}'\mathbf{W}_{12}+\frac{1}{4\delta}\mathbf{W}_{112}+\frac{1}{2}\mathbf{I} \\ -\frac{1}{2\delta}\mathbf{1}\mathbf{1}'\mathbf{W}_{11}+\frac{1}{4\delta}\mathbf{W}_{112}+\frac{1}{2}\mathbf{I} & -\frac{1}{2\delta}\mathbf{1}\mathbf{1}'\mathbf{W}_{12}+\frac{1}{4\delta}\mathbf{W}_{112} \end{bmatrix}\cdot\mathbf{Y}_{i|1}\right. \\ &\quad \left. + \sum_{i=1}^{N/2}\begin{bmatrix} \frac{1}{2\delta}\big(\mathbf{1}'\mathbf{W}_{11}-\mathbf{1}'\mathbf{W}_{12}\big) & \frac{1}{2\delta}\big(\mathbf{1}'\mathbf{W}_{12}-\mathbf{1}'\mathbf{W}_{11}\big) \\ -\frac{1}{2\delta}\mathbf{1}\mathbf{1}'\mathbf{W}_{11}+\frac{1}{4\delta}\mathbf{W}_{112}+\frac{1}{2}\mathbf{I} & -\frac{1}{2\delta}\mathbf{1}\mathbf{1}'\mathbf{W}_{12}+\frac{1}{4\delta}\mathbf{W}_{112} \\ -\frac{1}{2\delta}\mathbf{1}\mathbf{1}'\mathbf{W}_{11}+\frac{1}{4\delta}\mathbf{W}_{112} & -\frac{1}{2\delta}\mathbf{1}\mathbf{1}'\mathbf{W}_{12}+\frac{1}{4\delta}\mathbf{W}_{112}+\frac{1}{2}\mathbf{I} \end{bmatrix}\cdot\mathbf{Y}_{i|2}\right),\end{aligned}$$

where $\mathbf{W}_{112} = \mathbf{W}_{11} + \mathbf{W}_{12}$. Let $\mathbf{c} = (0,0,0,0,0,1)$ and $\mathbf{c}\cdot\mathbf{T}'_L = (0,1,0,-1,-1,0,1)$. Then,

$$\begin{aligned}\hat{\beta}_6 &= \mathbf{c}\cdot\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) \\ &= \frac{\big(\big(\bar{y}_{1,1,2}-\bar{y}_{1,1,0}\big)-\big(\bar{y}_{2,0,2}-\bar{y}_{2,0,0}\big)\big)}{2} + \frac{\big(\big(\bar{y}_{2,1,2}-\bar{y}_{2,1,0}\big)-\big(\bar{y}_{1,0,2}-\bar{y}_{1,0,0}\big)\big)}{2}.\end{aligned}$$

The subscript-triplet represents (period, treatment, visit), e.g., $\bar{y}_{1,1,2}$ is the average of observations at Visit 2 from subjects who are in period 1 and on treatment (trt = 1), and $\bar{y}_{2,0,0}$ is the average of observations at Visit 0 from subjects who are in period 2 and on placebo (trt = 0). From this result we can see that, the change-from-baseline treatment effect estimator under this LDA model is simply the average of two change-from-baseline estimators based on subjects in each sequence.

The variance of the estimate can be derived similarly as in the previous sections. Let $\mathbf{L}' = \Big(-\frac{1}{2},0,\frac{1}{2},\frac{1}{2},0,-\frac{1}{2}\Big)$, and we have

$$\begin{aligned}\text{Var}(\hat{\beta}_6) &= \frac{4}{N^2}\frac{N}{2}\Big(\mathbf{L}'\mathbf{V}\mathbf{L} + (-\mathbf{L}')\mathbf{V}(-\mathbf{L})\Big) \\ &= \frac{4}{N}\frac{\sigma_0^2 - 2\rho_{02}\sigma_0\sigma_2 + \sigma_2^2}{2}\end{aligned} \tag{32}$$

4 Impact of Missing Data

As shown in Section 3, in all three scenarios considered, the placebo-adjusted treatment effect does not depend on intermediate observations, assuming that the variance-covariance parameters are given. In other words, the MMRM model works exactly the same as the pre-post design with ANCOVA model, in calculating the point estimate with known variance parameters. In this sense, for a MMRM model, if the missing data only occurs at intermediate observations, the best method to obtain the point estimate is to ignore all intermediate data and only use the baseline and last visit data. This would give the same estimate if the "true" values of the missing observations are observed.

However, estimates of the variance-covariance parameters $\boldsymbol{\alpha}$ will be affected by the intermediate measurements, regardless. Depending on the choice of $\boldsymbol{\Sigma}$ and the dimension

of $\boldsymbol{\alpha}$, including more intermediate observations may gain efficiency for simple $\boldsymbol{\Sigma}$ (e.g. AR(1)) since fewer parameters need to be estimated (σ^2 and ρ in case of AR(1)). But if conservative structure of $\boldsymbol{\Sigma}$ is specified, e.g., heterogenous Toeplitz or unstructured, similar efficiency gain may not be likely since the additional observations are used to compensate the increased number of parameters.

Of note, if the missing data happens at baseline or at the last visit, then the handling would be completely different and will be studied further.

5 Numerical Studies

As discussed in the previous sections, the focus of this paper is to study the impact of intermediate observations on the point estimates of the change from baseline treatment effect. Two simulation studies are conducted with the following specifications based on the model (17) with 5 time points and $N = 40$ at each group with no missing data:

- $\boldsymbol{\beta} = (0.3, 0.8, 0.2, 0.4, 0.4, 0.4, 0.3, 0.7, 1.1, 1.6)$
- $b_i \overset{\text{iid}}{\sim} N(0, 1.2^2)$
- Two covariance matrices are specified. The simple AR(1):

$$\begin{bmatrix} 2.0000 & 0.6000 & 0.1800 & 0.0540 & 0.0162 \\ 0.6000 & 2.0000 & 0.6000 & 0.1800 & 0.0540 \\ 0.1800 & 0.6000 & 2.0000 & 0.6000 & 0.1800 \\ 0.0540 & 0.1800 & 0.6000 & 2.0000 & 0.6000 \\ 0.0162 & 0.0540 & 0.1800 & 0.6000 & 2.0000 \end{bmatrix}. \tag{33}$$

  Thus $\text{Var}(Y_0) = \text{Var}(Y_4) = 2 + 1.2^2 = 3.44$ and $\text{Cov}(Y_0, Y_4) = 0.0162 + 1.2^2 = 1.4562$. And the unstructured covariance matrix:

$$\begin{bmatrix} 1.0000 & 0.8764 & 0.7100 & 0.5367 & 0.3098 \\ 0.8764 & 1.2000 & 1.0369 & 0.8818 & 0.6788 \\ 0.7099 & 1.0369 & 1.4000 & 1.2700 & 1.0998 \\ 0.5367 & 0.8818 & 1.2700 & 1.8000 & 1.6628 \\ 0.3098 & 0.6788 & 1.0998 & 1.6628 & 2.4000 \end{bmatrix}. \tag{34}$$

  Thus $\text{Var}(Y_0) = 1 + 1.2^2 = 2.44$, $\text{Var}(Y_4) = 2.4 + 1.2^2 = 3.84$ and $\text{Cov}(Y_0, Y_4) = 0.3098 + 1.2^2 = 1.7498$.

- The numerical study is based on 1000 simulations.

In each setup we compare the mean and standard deviation of the point estimates for treatment effect and the associated variance parameters. In addition, number of times the model converges properly will be tallied.

In the first setup where AR(1) is used as covariance matrix, the points estimates for β_{10}, σ_0^2, σ_5^2 and σ_{05} are summarized as follows:

Visits Included	CR	$\hat{\beta}_{10}$ (SD)	$\hat{\sigma}_0^2$ (SD)	$\hat{\sigma}_5^2$ (SD)	$\hat{\sigma}_{05}$ (SD)
Baseline, Visit 1 to 4	100	1.608 (0.439)	3.438 (0.361)	3.438 (0.361)	1.448 (0.358)
Baseline, Visit 2 to 4	100	1.608 (0.439)	3.436 (0.373)	3.436 (0.373)	1.449 (0.383)
Baseline, Visit 3 to 4	100	1.608 (0.439)	3.437 (0.402)	3.437 (0.402)	1.457 (0.418)
Baseline, Visit 4	100	1.680 (0.439)	3.437 (0.444)	3.437 (0.444)	1.454 (0.431)

Here "CR" stands for the "Convergence Rate", or the percentage of times that a model converges. As expected, the point estimates are close to the true values (seen in the simulation setup) and are also close to each other among these four models. However, due to the

simple covariance structure, more intermediate observations leads to reduced variability of the variance estimates. For the second setup where unstructured covariance matrix is used, the results are as follows:

Visits Included	CR	$\hat{\beta}_{10}$ (SD)	$\hat{\sigma}_0^2$ (SD)	$\hat{\sigma}_5^2$ (SD)	$\hat{\sigma}_{05}$ (SD)
Baseline, Visit 1 to 4	17.1	1.534 (0.450)	2.470 (0.394)	3.407 (0.548)	1.582 (0.392)
Baseline, Visit 2 to 4	97.3	1.577 (0.438)	2.440 (0.453)	3.806 (0.681)	1.758 (0.473)
Baseline, Visit 3 to 4	100	1.577 (0.435)	2.441 (0.450)	3.838 (0.704)	1.776 (0.482)
Baseline, Visit 4	100	1.577 (0.435)	2.441 (0.450)	3.838 (0.704)	1.776 (0.482)

Again, the point estimates are very close among these four models, however the simplified models with fewest time points (e.g., Baseline Visit 4) is stable and still produces close results to models with one or more intermediate observations. Here we observe including more time points increases the complexity of the model and thus offsets the benefit of the additional data.

6 Conclusion and Discussion

This work considers three commonly used models for clinical trials and shows that, the placebo-adjusted change from baseline treatment effect does not depend on the intermediate observations, given the variance-covariance parameters. For the derivation in Section 3, N is assumed to be even and equal sample size is assumed for each treatment group. These assumptions are not essential. It can be shown that, for an odd N and unequal treatment groups sizes, the conclusion still holds. Also, the derivation can be generalized to any number of measurements on a subject.

The assumption of no missing data is more essential which the derivation in Section 3 heavily depends on. Section 4 discusses the case of missing intermediate observations. Other cases of missing data will be considered in the future research.

REFERENCES

P. Dinh and P. Yang. Handling baselines in repeated measures analyses with missing data at random. *Journal of Biopharmaceutical Statistics*, 21(2):326–341, 2011.

J. Jiang. REML estimation: Asymptotic behavior and related topics. *The Annals of Statistics*, 24(1):255–286, 1996.

S.E. Maxwell. Longitudinal designs in randomized group comparisons: When will intermediate observations increase statistical power? *Psychological Methods*, 3(3):275, 1998.

S. Portnoy. Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *The Annals of Statistics*, 16(1):356–366, 1988.

A. Venter, S.E. Maxwell, and E. Bolig. Power in randomized group comparisons: The value of adding a single intermediate time point to a traditional pretest-posttest design. *Psychological methods*, 7(2):194, 2002.

G. Verbeke and G. Molenberghs. *Linear mixed models for longitudinal data*. Springer Verlag, 2009.