

Evaluation of overall treatment effect in MMRM

Peer-reviewed author version

Song, Tao; Dong, Qunming; Sankoh, Abdul J. & MOLENBERGHS, Geert (2013)

Evaluation of overall treatment effect in MMRM. In: JOURNAL OF BIOPHARMACEUTICAL STATISTICS, 23 (6), p. 1281-1293.

DOI: 10.1080/10543406.2013.834918

Handle: <http://hdl.handle.net/1942/16003>

Evaluation of overall treatment effect in MMRM

Tao Song, Qunming Dong, Abdul J. Sankoh, Geert Molenberghs

SUMMARY

In longitudinal clinical trials for drug development, the study objective is often to evaluate overall treatment effect across all visits. Despite careful planning and study conduct, the occurrence of missing data can not be completely eliminated. As a direct likelihood method, Mixed-effects Model for Repeated Measures (MMRM) has become one of the preferred approaches for handling missing data in such designs. MMRM is a full multivariate model in nature, which avoids potential bias as a predetermined model and operates in a more general missing at random (MAR) framework. However, if treatment effect is constant over time, overparameterization of treatment by time interaction in MMRM could result in loss of power. In this paper, we utilize MMRM estimates and propose an optimal weighting method for combining visit-specific estimates to maximize the power under MAR mechanism. For a special case when the underlying covariance is compound symmetry, we show that the optimal weighting method is asymptotically equal to MMRM. In other words, MMRM has optimal power under this special case. When the underlying covariance is of an unstructured pattern, the optimal weighting method has increased power under MAR and missing not at random (MNAR) mechanisms, and can lead to bias reduction under MNAR. These are especially true when the variance is greater at later time point, which could lead to a smaller weight. We present practical examples using the optimal weighting method to analyze two cystic fibrosis clinical trial datasets.

Keywords: Bias reduction; MMRM; Overall treatment effect; Optimal weight; Power.

1. INTRODUCTION

In longitudinal clinical trials designed to demonstrate effectiveness of a new therapeutic compound, the same outcome variable is often measured repeatedly at a number of scheduled visits. In practice, the treatment response outcome of interest is either change from baseline to last visit, treatment-specific slope (rate of change), or overall treatment effect across all visits. When treatment difference is anticipated to occur quickly and to remain relatively stable over time, treatment response outcome evaluated over time should be preferred. This is especially true for chronic diseases, such as cystic fibrosis, where the interest is in sustained treatment effect throughout the treatment period.

Despite careful planning and rigorous efforts to follow up each individual subject, the occurrence of missing data in clinical trials can not be completely eliminated. The presence of missing data could lead to reduced statistical power, and biased model estimates (Little and Rubin, 2002; Molenberghs and Kenward, 2007). Little and Rubin (2002) describe 3 classes of missing data mechanisms. Depending on the process that triggers the missing data, it can be classified as missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). For MCAR, missingness is independent of both the observed and unobserved outcomes. MAR, on the other hand, is a less restrictive assumption that allows the missingness to depend on the observed outcomes, but independent of the unobserved. MNAR is the least restrictive in that missingness can not be fully explained by observed outcomes, and depends on the unobserved outcomes.

For studies of missing data in controlled clinical trial setting, MAR is usually considered as a plausible underlying missing mechanism (Molenberghs and Kenward, 2007; Siddiqui et al., 2009), and various analysis approaches have been proposed to accommodate such an

assumption. Direct likelihood (Little and Rubin, 2002) or direct Bayesian approaches are commonly used because they are inherently valid, even ignoring the missing data mechanism. For non-likelihood based methods, such ignorability property can not be invoked. Generalized estimating equations (GEE) method (Liang and Zeger, 1986; Molenberghs and Verbeke, 2005) has also gained popularity because the estimator remains consistent even under model misspecification. The basic form of GEE is only valid under MCAR, but estimating equations can be adopted to accommodate MAR, and even MNAR. Robins et al. (1995) have developed weighted generalized estimating equations (weighted GEE), weighing observations by the inverse probability of being observed. Scharfstein et al. (1999) have further augmented the estimating equations to doubly robust version, by modeling inverse probability of being observed as a covariate. The doubly robust estimator has been proven to be consistent when either the missing data model or the response model is correctly specified. Next to GEE, pseudo-likelihood methods (Molenberghs and Verbeke, 2005) are good alternative to full likelihood methods due to their computational efficiency. Molenberghs et al. (2011) have developed weighted version and double-robust version in pseudo-likelihood inferential paradigms.

With the availability of flexible software, linear mixed-effects model (Laird and Ware, 1982) has become the most popular among direct likelihood methods. To avoid potential bias in the predetermined statistical model, a full multivariate model is usually considered as primary analysis in clinical study protocol. Mallinckrodt et al. (2001) referred to such a model as Mixed-effects Model for Repeated Measures (MMRM). Under various scenarios of MNAR, simulation studies have shown MMRM outperforms many of the traditional methods including the Last-Observation-Carried-Forward (LOCF) in terms of bias reduction and

control of the type I error rate for the statistical test (Mallinckrodt et al., 2001; Siddiqui et al., 2009).

The analytical solution of the overall treatment effect obtained from MMRM is equally weighted average of visit-specific estimates. A main-effect model is sometimes considered to evaluate overall treatment effect. In this case, the estimate is a weighted average of individual estimates adjusting for visit-specific information. The advantage of MMRM is that it can better satisfy the need for a predetermined model at study planning stage when *a priori* information is not sufficient. Under the context of missing data, MMRM operates in a more general MAR framework. When constancy assumption does not hold, MMRM remains valid under MAR while the main-effect model does not. However, if constancy condition were satisfied, MMRM could suffer power loss due to overparameterization of treatment by time interaction.

Utilizing MMRM estimates, this paper proposes an optimal weighting method for combining visit-specific estimates, with the ultimate objective of maximizing the power of testing for overall treatment effect under missing at random (MAR). The outline is as follows: The estimation approach of MMRM is discussed in Section 2.1. We propose an optimal weighting method in Section 2.2. In Section 3, we conduct simulation studies to compare the performance of the MMRM and the optimal weighting method under different missing data mechanisms. In Section 4, datasets from two cystic fibrosis clinical trials are used for illustration. Section 5 concludes with summary remarks and suggestion for application.

2. METHOD

2.1 Estimation in MMRM

In this section, we first describe how overall treatment effect is estimated. For each treatment group with n subjects, and each subject with K measurements $\bar{y}_i = (y_{i1}, \dots, y_{iK})'$, $i = 1, \dots, n$, MMRM is formulated as $\bar{y}_i \sim N(\bar{\mu}, V)$, where $\bar{\mu} = (\mu_1, \dots, \mu_K)'$ represent mean effect at each visit. V is a $K \times K$ completely unstructured covariance matrix. For two treatment groups, each group is assumed to have distinct mean parameters $(\mu_1^0, \dots, \mu_K^0)'$ and $(\mu_1^1, \dots, \mu_K^1)'$, but commonly allowed to have shared covariance matrix V . The main-effect model constrains visit-specific difference $\mu_j^E - \mu_j^P, j = 1, \dots, K$ to be a constant. For simplicity, we use one treatment group for future formulation. The restricted maximum likelihood (REML) estimate for complete data (Verbeke and Molenberghs, 2000) is given by

$$\hat{\bar{\mu}} = \left(\sum_{i=1}^n \hat{V}^{-1} \right)^{-1} \sum_{i=1}^n \hat{V}^{-1} \bar{y}_i = \frac{\sum_{i=1}^n \bar{y}_i}{n}.$$

The overall treatment effect Δ (obtained via SAS LSMEANS statement) is equally weighted average of individual estimates:

$$\hat{\Delta}_M = \left(\frac{1}{K}, \dots, \frac{1}{K} \right)_{1 \times K} \hat{\bar{\mu}} = \frac{\sum_{i=1}^n \sum_{j=1}^K y_{ij}}{nK}.$$

Now assume only r subjects have the K th measurements ($r < n$). The objective is to estimate μ_K in the presence of missing data. Little and Rubin (2001) have shown explicit expression of MMRM can be obtained via factored likelihood. Denote $(y_{i1}, \dots, y_{i,K-1})'$ as \bar{y}_{io} , and the conditional distribution $y_{iK} | \bar{y}_{io}$ as $N(\beta_0 + \beta_1' \bar{y}_{io}, \sigma_{K|o}^2)$. MMRM estimate is given by

$$\hat{\mu}_K = n^{-1} \left\{ \sum_{i=1}^r y_{iK} + \sum_{i=r+1}^n (\hat{\beta}_0 + \hat{\beta}_1' (y_{ik_1})_{(K-1) \times 1}) \right\} = n^{-1} \left\{ \sum_{i=1}^r y_{iK} + \sum_{i=r+1}^n \hat{y}_{iK} \right\}. \quad (1.1)$$

The estimator for μ_K is formulated in a way that looks like it “imputes” the unobserved $y_{iK}, i > r$, using the predictive distribution of the unobserved given the observed ones. Under MAR, the predictive distribution can be consistently estimated using r complete subjects, which leads to the consistent estimators for the marginal distribution even ignoring the missing data mechanism.

However, further information about the unobserved would be needed for MNAR mechanisms. For example, assume $y_{iK} | \bar{y}_{io}, i > r$ follows $N(\beta_0 + \bar{\beta}_1' \bar{y}_{io} + d, \sigma_{K|o}^2)$, where $d \neq 0$ represents the information to distinguish between completers and dropouts. In clinical trials, $d < 0$ could represent subjects drop out due to lack of efficacy, while $d > 0$ could represent subjects drop out due to good treatment response. When assuming a MAR, MMRM “imputes” the unobserved as $\hat{y}_{iK} = \hat{\beta}_0 + \hat{\beta}_1 y_{i1} + \dots + \hat{\beta}_{K-1} y_{i,K-1}, i > r$. The expected bias due to “imputing” the unobserved is equal to $-(n-r)d$.

2.2 Optimal weighting method

Overall treatment effect can be obtained by combining individual estimates $(\hat{\mu}_1, \dots, \hat{\mu}_K)'$ using the weighting scheme by Wei and Johnson (1985),

$$\hat{\Delta}_W = \hat{w}' \hat{\mu} = \sum_{i=1}^K \hat{w}_i \hat{\mu}_i, \text{ and test statistic is } (\hat{\Delta}_W - \Delta) / se[\hat{\Delta}_W] \quad (1.2)$$

$$\text{where } \hat{w} = \frac{\left(\text{Var}[\hat{\mu}] \right)^{-1} \mathbf{1}_{K \times 1}}{\mathbf{1}_{K \times 1}' \left(\text{Var}[\hat{\mu}] \right)^{-1} \mathbf{1}_{K \times 1}}, \text{ and } se[\hat{\Delta}_W] = \sqrt{\hat{w}' \text{Var}[\hat{\mu}] \hat{w}}. \quad (1.3)$$

In the Appendix, we show optimal weight $\bar{w} = \frac{(\text{Var}[\bar{\mu}])^{-1} \mathbf{1}_{K \times 1}}{\mathbf{1}_{K \times 1}' (\text{Var}[\bar{\mu}])^{-1} \mathbf{1}_{K \times 1}}$ maximizes the power

of the hypothesis test $H_0 : \Delta = 0$ versus $H_1 : \Delta > 0$ under MAR. In the above formulations, $\hat{\bar{\mu}}$ and $\text{Var}[\hat{\bar{\mu}}]$ are likelihood estimates from MMRM. Under MAR, they are consistent estimators for $\bar{\mu}$ and $\text{Var}[\bar{\mu}]$, and the statistic $(\hat{\Delta}_w - \Delta) / \text{se}[\hat{\Delta}_w]$ converges in distribution to a standard normal random variable, as $n \rightarrow \infty$. The weight \hat{w} is a function of likelihood estimates and is a consistent estimator of \bar{w} , thus establishing the consistency of the optimal weighting estimator $\hat{\Delta}_w$ for $\Delta_w = \bar{w}'\bar{\mu}$ under MAR.

Note: the distribution of $(\hat{\Delta}_w - \Delta) / \text{se}[\hat{\Delta}_w]$ under the null hypothesis is intractable but approximates to normality when sample size is sufficiently large. This sample size condition is usually met in a typical confirmatory clinical trial for drug development.

With complete data, $\text{Var}[\hat{\bar{\mu}}]$ is estimated as \hat{V} / n . The individual weight \hat{w}_i is then simply the sum of the i th column of the inverse covariance matrix \hat{V}^{-1} . The optimal weight is inversely proportional to visit-specific variance. We now discuss two special cases for the optimal weight. If V is assumed to be of independent structure, the weight simply reduces to $1 / \hat{\sigma}_i^2$, a well known procedure for combining information across K independent studies in meta analysis, where $\hat{\sigma}_i^2$ is the estimated variance for the i th element. When the covariance structure V is assumed to be of compound symmetry (CS), $\sigma^2 (1 \times I(k_1 = k_2) + \rho \times I(k_1 \neq k_2))_{K \times K}$; with complete data, $(\text{Var}[\hat{\bar{\mu}}])^{-1} / n$ approaches

$$\sigma^{-2} \left(\frac{1}{1-\rho} \left(1 - \frac{\rho}{1+(K-1)\rho} \right) \times I(k_1 = k_2) - \frac{1}{1-\rho} \left(\frac{\rho}{1+(K-1)\rho} \right) \times I(k_1 \neq k_2) \right)_{K \times K}, \text{ as } n \rightarrow \infty.$$

The optimal weight in (1.3) is then given by $\hat{w} = \frac{\left(\text{Var} \left[\hat{\mu} \right] \right)^{-1} \mathbf{1}_{K \times 1}}{\mathbf{1}_{K \times 1}' \left(\text{Var} \left[\hat{\mu} \right] \right)^{-1} \mathbf{1}_{K \times 1}} \rightarrow \left(\frac{1}{K}, \dots, \frac{1}{K} \right)_{1 \times K}'$, as

$n \rightarrow \infty$. In this case, the optimal weighting method is asymptotically equal to MMRM.

3. SIMULATION STUDIES

3.1 Simulation Set-up

We conduct simulation studies to evaluate the performance of the proposed optimal weighting method compared with MMRM in evaluating overall treatment effect under both MAR and MNAR mechanisms.

Assume the post-baseline outcome measurement for the experiment treatment group \bar{y}^E is generated from $N\left((\mu_1^E, \dots, \mu_K^E), V\right)$, $i = 1, \dots, n_E$. For placebo, outcome measurement \bar{y}^P is generated from $N\left((\mu_1^P, \dots, \mu_K^P), V\right)$, $i = 1, \dots, n_P$. We generate a sample of 200 subjects in each treatment group ($n_E = n_P = 200$), and $K = 4$. The differences between treatment groups are assumed to be a constant value Δ across all visits ($\mu_i^E - \mu_i^P = \Delta$, $i = 1, \dots, 4$). The marginal covariance matrix V is assumed to be of CS or UN, where UN covariance pattern is estimated from the second cystic fibrosis clinical trial dataset discussed in Section 4. To evaluate overall treatment effect, the hypothesis test is $H_0 : \Delta = 0$ versus $H_1 : \Delta > 0$.

The missingness mechanism is assumed to be related to the subject's measurement values at the previous and current clinical visits ($y_{i,j-1}$ and y_{ij}) via a logistic regression model

$$\log it \{P(d_{ij} = 1 | \bar{y}_i)\} = \beta_0 + \beta_1 y_{i,j-1} + \beta_2 y_{ij}, i = 1, \dots, n_E + n_P, j = 1, \dots, K,$$

where $d_{ij}=1$ if the i th subject drops out from the study at the j th visit, $d_{ij}=0$ otherwise. To generate a MAR mechanism, assume dropout is only related to $y_{i,j-1}$ ($\beta_1 \neq 0, \beta_2 = 0$), and a subject with a lower value of the measurement at the $(j-1)$ th visit has a higher probability to drop out at j th visit. To generate a MNAR mechanism, assume dropout is related to both $y_{i,j-1}$ and y_{ij} ($\beta_1 \neq 0, \beta_2 \neq 0$), and a subject with a lower increment between values of measurement at the current and previous visits have a higher probability to drop out (Verbeke and Molenberghs, 2000). The coefficients of the logistic regression ($\beta_0, \beta_1, \beta_2$) are chosen to achieve the desired amount of missing data.

A total of 16 scenarios are simulated by varying four input parameters described below.

In each scenario, 2,000 simulations are conducted. The four parameters are:

- i. V is assumed to be either of CS or UN;
- ii. True overall treatment difference is either $\Delta=1$ or $\Delta=1.5$;
- iii. Missingness is either MAR or MNAR mechanism;
- iv. $(\beta_0, \beta_1, \beta_2)$ are chosen to achieve either 25% or 40% of missing data.

When assessing the performance of the proposed method for evaluating overall treatment effect, we examine the type I error rate, statistical power and percent bias of the estimate. The type I error rate is estimated by the proportion of rejections of null hypothesis under H_0 . The statistical power is estimated by the proportion of rejections under H_1 . The percent bias is estimated by the average percent difference between the estimate and true value under H_1 .

3.2 Under MAR

Table 1 summarizes the simulation results under MAR. It can be observed that both optimal weighting method and MMRM achieved nominal type I error rate, with amount of missing data up to 40%. For statistical power, one can observe that optimal weighting method has similar power to MMRM for CS (average power: 0.73 versus 0.72), and optimal weighting method has increased power for each of the simulated scenarios for UN. Optimal weighting method has on average a 7% power gain under UN (average power: 0.75 versus 0.68). For percent bias, one can observe that both optimal weighting method and MMRM are unbiased under MAR as expected, although the optimal weighting method has slightly greater bias than MMRM.

3.3 Under MNAR

Table 2 summarizes the simulation results under MNAR. It can be observed that both optimal weighting method and MMRM have achieved nominal type I error rate under MNAR.

For statistical power, one can observe that optimal weighting method has similar power to MMRM for CS (average power: 0.72 versus 0.72). However, optimal weighting method is relatively more powerful for each of the simulated scenarios for UN (on average of 7% power gain). For percent bias, optimal weighting method has greater bias than MMRM for CS (average bias: -7.1% versus -6.5%), but has smaller bias for UN (average bias: -8.9% versus -11.1%). The greatest difference between the two methods occurs for UN when missingness is 40%, where optimal weighting method has 3.0% reduced bias as compared to MMRM. The underlying UN covariance results in a smaller weight at last time point (see Section 4 for the estimated weight in Case Study 2). Since missing data introduces most bias

at last time point, this probably benefits the optimal weighting method and it has reduced bias than MMRM.

3.4 Greater Variance for Later Time Point

As greater uncertainty is usually anticipated at later time points in clinical trials, we conduct simulations to investigate the impact. The variance at each visit is proportional to 1: 0.5: 1: 2. Note this will result in a smaller weight for the last time point. The simulation results are shown in Table 3. Under such scenarios, great bias reduction and increase of statistical power are observed. The greatest difference occurs when missingness is MNAR and missingness is 40%. The statistical power increases more than four-fold (0.838 versus 0.195), and the bias reduction of using optimal weighting method is more than one-eighth (-2.6% versus -22.1%).

4. THE CYSTIC FIBROSIS STUDY

Cystic fibrosis (CF) is a recessive genetic disease that affects approximately 70,000 children and adults in the world. Despite progress in the treatment of CF, there is currently no cure. The predicted median age of survival for a person with CF is in the mid-30s (Stern et al., 2008).

Although the disease affects multiple organs, the leading cause of mortality is the progressive loss of lung function. To illustrate the application of the proposed method, we utilize the data from two CF clinical trials.

4.1 Case Study 1

The first dataset is from a double-blind, placebo-controlled randomized study in which subjects received either placebo ($n_P=28$) or test drug E ($n_E=112$) for 16 weeks. Treatment response outcome is change from baseline in percent predicted FEV₁, which was measured at baseline (Day 1 pre-dose), and at 3 post-baseline visits: Week 2, Week 8, and Week 16. At the conclusion of the study, the percent of missing data was 7% for each treatment group. A total of 10 subjects dropped out of the study prematurely during treatment period and half of these subjects dropped out due to adverse events.

We first explore the structure of the marginal covariance matrix for this dataset. Table 4 summarizes model fitting and comparison for three models. All models assume a saturated treatment by time mean structure, with two baseline covariates: baseline FEV₁ and age. In model 1 (MMRM), the covariance matrix is completely unstructured (UN), yielding an additional 6 parameters. Models 2 and 3 assume the covariance structure to be of compound symmetry (CS) and first-order autoregressive (AR(1)) type, respectively. They can be considered as nested models, and likelihood ratio can be used to test the hypothesis of whether a parsimonious covariance structure is more appropriate. From the table, one can see that model 2 is favored by likelihood ratio test (p -values = 0.1586) and CS seems to be a plausible fit for this dataset. However, UN is still preferred in this setting. This is because as compared with a more parsimonious choice, loss of power under UN can be negligible for evaluating last visit treatment effect (Molenberghs and Kenward, 2007). In our simulations, we also find the power loss for evaluating overall treatment effect can be negligible for UN. MMRM is used to estimate the differences between E and placebo, and the visit-specific differences are displayed in the middle part of Table 4.

In this example, the main interest is evaluating overall treatment effect during the 16 weeks treatment period. The alternative hypothesis is that the overall treatment effect of change from baseline in percent predicted FEV₁ is different between E and placebo. We analyze the data using MMRM, the optimal weighting and ANOVA methods. ANOVA is conducted on the average of all available post-baseline measurements. The estimates of treatment effect and p -values are given in the lower part of Table 4. All three methods yield similar estimates, standard error and p -values. MMRM and ANOVA are similar probably due to the small percent of missing data. The similarity of MMRM and optimal weighting method probably lies in the presence of the underlying CS structure (note UN was still used in MMRM and optimal weighting method).

4.2 Case Study 2

For the second clinical trial dataset, which is also from a double-blind, placebo-controlled randomized study, where 26 subjects received placebo and 26 subjects received E for 24 weeks. Percent predicted FEV₁ was measured at 4 post-baseline visits: Week 2, 8, 16 and 24. All subjects completed treatment group E, while the percent of missing data was 11.5% for placebo. The covariance structure of this dataset is explored and results are shown in Table 5. One can see that models 2 and 3 (CS and AR(1)) are both rejected (p -values = 0.0453 and <0.0001, respectively) and model 1 (UN) provides the best fit to this dataset. The visit-specific differences between E and placebo are displayed in the middle part of Table 5.

Next, we evaluate overall treatment effect during the 24 weeks treatment period. The estimates of treatment effect and p -values are given in the lower half of Table 5. ANOVA and MMRM again give very similar analysis results. The optimal weighting method yields a

relatively smaller estimate. This is because the underlying UN covariance results in a smaller weight at Week 24, where the estimated treatment difference from MMRM is the greatest. The optimal weights are given as 0.41, -0.32, 0.64 and 0.27 at each visit.

5. CONCLUSIONS

With growing interests in using MMRM to analyze clinical trial data with missing data, we propose an optimal weighting method to address the question of comparing overall treatment effect. This method has optimal power to evaluate overall treatment effect under MAR. For a special case when the underlying covariance structure is compound symmetry (CS), we have shown optimal weighting method is asymptotically equal to MMRM with complete data.

In the simulation study we observe the proposed optimal weighting method performs generally comparable to MMRM, with regards to controlled type I error rate under MAR and MNAR missing mechanisms. When the underlying covariance is unstructured (UN) pattern, our simulation results show the optimal weighting method is relatively more powerful under MAR and MNAR, and has less bias under MNAR. This is especially so when the variance is greater at later time point, which could lead to a smaller weight.

The proposed optimal weighting method differs from weighted and double robust version non-likelihood methods in the following way. In non-likelihood methods, visit-specific estimates for population average achieve validity when inverse probability of being observed is utilized at subject level. In the proposed method, visit-specific estimates for population average are obtained based on direct likelihood, which are inherently valid under MAR. The weight is further a function of likelihood estimates, proven to achieve maximum power under MAR.

APPENDIX

In this appendix, we show the optimal weight in (1.3) (Section 3.2) maximizes the power of the hypothesis test $H_o : \Delta = 0$ versus $H_1 : \Delta > 0$ under MAR.

With the nominal significant level α , the power of the test (1.2) under H_1 is given by

$$\begin{aligned} P(\Delta_w / se[\Delta_w] \geq Z_\alpha \mid \Delta > 0) &= P\left[(\Delta_w - \Delta) / \sqrt{\bar{w}' Var[\bar{\mu}] \bar{w}} + \Delta / \sqrt{\bar{w}' Var[\bar{\mu}] \bar{w}} \mid \Delta > 0\right] \\ &= P\left[Z + \Delta / \sqrt{\bar{w}' Var[\bar{\mu}] \bar{w}}\right], \end{aligned}$$

where Z follows a standard normal distribution under MAR and Z_α satisfies $P(Z \geq Z_\alpha) = \alpha$.

From Cholesky decomposition, $Var[\bar{\mu}]$ is equal to LL' , where L is a lower triangular matrix with strictly positive diagonal entries. Further, from Cauchy-Schwarz inequality, we have

$$\begin{aligned} \bar{w}' Var[\bar{\mu}] \bar{w} \times \mathbf{1}_{K \times 1}' (Var[\bar{\mu}])^{-1} \mathbf{1}_{K \times 1} &= (\bar{w}' L) (\bar{w}' L)' \times (\mathbf{1}_{K \times 1}' L^{-1}) (\mathbf{1}_{K \times 1}' L^{-1})' \\ &\geq (\bar{w}' L) (\mathbf{1}_{K \times 1}' L^{-1})' = 1. \end{aligned}$$

Thus the quantity $\bar{w}' Var[\bar{\mu}] \bar{w}$ is minimized at $\bar{w} = \frac{(Var[\bar{\mu}])^{-1} \mathbf{1}_{K \times 1}}{\mathbf{1}_{K \times 1}' (Var[\bar{\mu}])^{-1} \mathbf{1}_{K \times 1}}$. Therefore, the

optimal weight in (1.2) maximizes the power of the test $H_o : \Delta = 0$ versus $H_1 : \Delta > 0$.

REFERENCES

- Laird, N. M., Ware J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics* 38(4):963-974.
- Liang K. Y., Zeger S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73:13–22.
- Little R. A., Rubin D. B. (2002). *Statistical Analysis with Missing Data (2nd edition)*. New York: Wiley.
- Mallinckrodt C. H., Clark W. S., David S. R. (2001). Accounting for dropout bias using mixed-effects models. *Journal of Biopharmaceuticals Statistics* 11:9–21.
- Molenberghs G., Verbeke G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.
- Molenberghs G., Kenward M. G. (2007). *Missing Data in Clinical Studies*. New York: Wiley.
- Molenberghs G., Kenward M. G., Verbeke G., Teshome B. (2011). Pseudo-likelihood estimation for incomplete data. *Statistica Sinica* 21:187-206.
- Robins J. M., Rotnitzky A., Zhao L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* 90:106–121.
- Scharfstein D. O., Rotnitzky A., Robins J. M. (1999). Adjusting for nonignorable drop-out using semi-parametric nonresponse models. *Journal of the American Statistical Association* 94:1096–1120 (with Rejoinder, 1135–146).
- Siddiqui O., Hung H. M., O'Neill R. (2009). MMRM vs. LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets. *Journal of Biopharmaceuticals Statistics* 19(2):227-246.
- Stern M., Wiedemann B., Wenzlaff P. (2008). From registry to quality management: The German Cystic Fibrosis Quality Assessment Project 1995 – 2006. *Eur Respir J.* 31(1):29-35.
- Verbeke G., Molenberghs G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.
- Wei L. J., Johnson W. E. (1985). Combining dependent tests with incomplete repeated measurements. *Biometrika* 72(2):359 –364.

Table 1. Simulation results under MAR to compare the optimal weighting method (1.2) and MMRM. Sample size $n = 200$ per treatment group and the number of replicates is 2,000.

Type I error		
Scenarios	Optimal weighting	MMRM
	CS	
25% missing	0.053	0.050
40% missing	0.054	0.052
	UN	
25% missing	0.054	0.048
40% missing	0.053	0.052

Power		
Scenarios	Optimal weighting	MMRM
	CS	
$\Delta=1$, and 25% missing	0.575	0.565
$\Delta=1.5$, and 25% missing	0.891	0.890
$\Delta=1$, and 40% missing	0.560	0.554
$\Delta=1.5$, and 40% missing	0.890	0.878
	UN	
$\Delta=1$, and 25% missing	0.594	0.530
$\Delta=1.5$, and 25% missing	0.916	0.858
$\Delta=1$, and 40% missing	0.575	0.498
$\Delta=1.5$, and 40% missing	0.901	0.836

Bias		
Scenarios	Optimal weighting	MMRM
	CS	
$\Delta=1$, and 25% missing	0.2%	0.3%
$\Delta=1.5$, and 25% missing	0.1%	0.2%
$\Delta=1$, and 40% missing	-0.2%	0.1%
$\Delta=1.5$, and 40% missing	-0.3%	0.1%
	UN	
$\Delta=1$, and 25% missing	-0.5%	0.1%
$\Delta=1.5$, and 25% missing	-0.3%	0.1%
$\Delta=1$, and 40% missing	-0.8%	-0.1%
$\Delta=1.5$, and 40% missing	-0.7%	-0.1%

Table 2. Simulation results under MNAR to compare the optimal weighting method (1.2) and MMRM. Sample size $n = 200$ per treatment group and the number of replicates is 2,000.

Type I error		
Scenarios	Optimal weighting	MMRM
	CS	
25% missing	0.052	0.054
40% missing	0.057	0.054
	UN	
25% missing	0.051	0.048
40% missing	0.050	0.048

Power		
Scenarios	Optimal weighting	MMRM
	CS	
$\Delta=1$, and 25% missing	0.569	0.563
$\Delta=1.5$, and 25% missing	0.889	0.887
$\Delta=1$, and 40% missing	0.549	0.539
$\Delta=1.5$, and 40% missing	0.873	0.874
	UN	
$\Delta=1$, and 25% missing	0.594	0.527
$\Delta=1.5$, and 25% missing	0.914	0.861
$\Delta=1$, and 40% missing	0.557	0.480
$\Delta=1.5$, and 40% missing	0.890	0.820

Bias		
Scenarios	Optimal weighting	MMRM
	CS	
$\Delta=1$, and 25% missing	-4.2%	-4.0%
$\Delta=1.5$, and 25% missing	-3.5%	-3.2%
$\Delta=1$, and 40% missing	-10.7%	-9.8%
$\Delta=1.5$, and 40% missing	-9.8%	-8.8%
	UN	
$\Delta=1$, and 25% missing	-6.1%	-7.3%
$\Delta=1.5$, and 25% missing	-5.0%	-6.0%
$\Delta=1$, and 40% missing	-13.0%	-16.3%
$\Delta=1.5$, and 40% missing	-11.6%	-14.6%

Table 3. Simulations to assess the impact of greater variance for later time point on the comparison of the optimal weighting method (1.2) and MMRM. Sample size n = 200 per treatment group and the number of replicates is 2,000.

Under MAR		
	Optimal weighting	MMRM
25% missing		
Alpha	0.052	0.051
Power	0.901	0.246
Bias (%)	1.0%	-0.4%
40% missing		
Alpha	0.054	0.051
Power	0.844	0.214
Bias (%)	0.8%	-1.4%

Under MNAR		
	Optimal weighting	MMRM
25% missing		
Alpha	0.050	0.051
Power	0.904	0.236
Bias (%)	1.7%	-12.9%
40% missing		
Alpha	0.054	0.054
Power	0.838	0.195
Bias (%)	-2.6%	-22.1%

Table 4. Analysis of CF clinical trial datasets: Case Study 1 (140 subjects, with $n_E=112$ and $n_P=28$).

Exploratory analysis for the covariance structures.

	Mean	Covar	par	-2l	Ref	G^2	df	p
1	unstr.	UN	14	2647.4				
2	unstr.	CS	10	2654.0	1	6.6	4	0.1586
3	unstr.	AR(1)	10	2666.1	1	18.7	4	0.0009

MMRM Estimation of visit-specific treatment difference between E and placebo.

	Week 2	Week 8	Week 16
Estimate	1.7	1.2	2.0
Standard Error	1.38	1.55	1.70
p value	0.21	0.45	0.25

Estimation of overall treatment difference between E and placebo.

	ANOVA	MMRM	Optimal weighting
Estimate	1.5	1.6	1.6
Standard Error	1.19	1.22	1.17
p value	0.20	0.18	0.18

Table 5. Analysis of CF clinical trial datasets: Case Study 2 (52 subjects, with $n_E=26$ and $n_P=26$).

Exploratory analysis for the covariance structures.

	Mean	Covar	par	-2l	Ref	G^2	df	p
1	unstr.	UN	20	1416.1				
2	unstr.	CS	12	1431.9	1	15.8	8	0.0453
3	unstr.	AR(1)	12	1457.8	1	41.7	8	<0.0001

MMRM Estimation of visit-specific treatment difference between E and placebo.

	Week 2	Week 8	Week 16	Week 24
Estimate	13.1	13.8	11.8	13.9
Standard Error	3.55	3.83	3.11	3.43
p value	<0.01	<0.01	<0.01	<0.01

Estimation of overall treatment difference between E and placebo.

	ANOVA	MMRM	Optimal weighting
Estimate	13.0	13.1	12.2
Standard Error	3.12	3.04	2.79
p value	<0.01	<0.01	<0.01