PharmaSUG 2012 - Paper SP07

Pattern-Mixture Models as Linear Combinations of Least Squares Means from MMRM with Delta Method Variance Estimation

Bohdana Ratitch, Quintiles, Montreal, Quebec, Canada
Michael O'Kelly, Quintiles, Dublin, Ireland

For studies with missing data, it is being increasingly recommended to provide sensitivity analyses that assume data are Missing Not at Random (MNAR). In certain therapeutic areas, MNAR assumptions about the missingness mechanism are even recommended for the primary analysis. MNAR assumptions can be modeled within several statistical frameworks, one of which is known as pattern-mixture models (PMMs). Certain PMM-based analyses for continuous outcomes can be formulated in such a way that the estimate of the difference between experimental and control treatments is expressed as a linear combination of Least Squares Means (LSMs) for different effects of a longitudinal model with correlated errors, weighted by the appropriate proportions of study drop-outs and completers. This approach requires special considerations for the estimation of the variance because the proportions of drop-outs and completers used in the linear combination of LSMs are themselves multinomial random variables and their variances need to be incorporated into the overall estimate. This can be done using a delta approximation method for variance estimation. In this paper, we present details of implementing such analyses (including delta variance estimation method) using exclusively SAS/STAT ® core functionality, such as PROC MIXED, data steps, and PROC FCMP. To illustrate this approach, we are using an example of MNAR assumptions that take into account the reasons for discontinuation from the study.

1. INTRODUCTION

Longitudinal models with correlated errors, otherwise widely known as mixed models with repeated measures (MMRMs), have been increasingly used for the analysis of clinical trials with missing data. A longitudinal model is often used even though the primary objective is to estimate a treatment effect and test a null hypothesis of no treatment effect at a single specific time-point (typically at the end of double-blind period). The advantage of using an MMRM analysis in this context (compared to ANCOVA at the primary time-point) is that longitudinal models include all randomized subjects regardless of whether they completed the study (provided data for the primary time-point) or not. Model estimation and inference is done without performing any imputation of the missing data for subjects who discontinued early, yet partial data available for these subjects is fully utilized and contributes to the estimation of effects at the end of the double blind period and to the variance-covariance structure of the longitudinal model.

A standard MMRM analysis relies on Missing at Random (MAR) assumptions regarding the mechanism behind missingness of the data. With MAR missingness depends only on the available data and is independent of the outcomes that are not observed. Under the MAR assumption, as long as a statistical model includes all observed variables that explain missingness, inference from that model will not be biased due to the missing data. In less technical terms, MAR assumptions imply the following: that the subjects with missing data (typically those that discontinue from the trial prematurely) follow the same trajectory (distribution) for the outcome variable after their discontinuation as the subjects with available data (study completers) with similar characteristics captured by observed data. If, on the other hand, missingness depends on the unobserved outcome and cannot be fully explained by available data, the mechanism is considered to be MNAR. In other words, with MNAR it is assumed that the post-discontinuation outcomes of subjects with missing data are systematically different from those with observed data even after accounting for all relevant observed factors in the statistical model.

While MAR assumptions may be reasonable for many studies, one can never rule out that the missingness mechanism for a study at hand is not MNAR (it is impossible to prove that missingness does not depend on unobserved data due to the fact that these data are missing). Because of this, regulatory and industry experts strongly recommend [1,2] that sensitivity analyses be performed for departures from MAR assumptions. In practice this requires analyzing the data under various MNAR scenarios that are clinically meaningful and plausible for a study at hand. In some cases, MAR assumptions can be considered as just not appropriate for the study, and MNAR assumptions have to be used for the primary analysis. We provide one example of MNAR assumptions in Section 2.

Importantly, we cannot be sure of the dependencies, if any, between observed data values and missing data values , so a typical solution for this problem is to make some explicit assumptions regarding these dependencies or about the distribution of the unobserved outcomes, and impose these assumptions on the statistical model.

Pattern-mixture models [3] provide one framework within which MNAR assumptions can be modeled and we provide a brief introduction to this framework in Section 2. There are different ways in which PMMs can be implemented [12]. One implementation strategy relies on Multiple Imputation (MI) (e.g., see reference [11] for a description of a general approach and reference [4] for an example of a particular PMM using MI functionality in SAS). Other approaches use maximum likelihood techniques as the basis for the estimation of an underlying statistical model, onto which additional (MNAR) assumptions are imposed [5, 7, 13]. In Section 2 of this paper, we describe one way of implementing PMMs in a maximum likelihood framework in SAS for the analysis of continuous outcomes. Unlike MI-based implementations, this approach does not involve any explicit imputation, but similar to standard MMRM analysis, partial data from subjects that discontinued early are used for estimation of the parameters in the longitudinal model. With this strategy, the estimate of the experimental-control treatment difference is formulated as a linear combination of LSMs obtained from a standard MMRM, where each LSM in the linear combination can be weighted by proportions of subjects in different completer/drop-out groups. This estimate can be obtained using SAS procedure PROC MIXED. However, PROC MIXED does not compute a standard error for this estimate in a way that would account for the fact that weights used in the linear combination of LSMs are not constants, but rather multinomial random variables (proportions of subjects in different subgroups). Since these proportions are estimated from data, standard errors associated with their estimation need to be accounted for in the overall standard error of the treatment difference estimate. This can be accomplished using a so called delta approximation method [6]. We provide a general algorithm for its calculation in Section 3 with a detailed example using SAS code in Section 5.

A maximum likelihood-based approach may be appealing to statisticians that prefer the “single answer” provided by an MMRM-type analysis to the random sampling (simulation) nature of an MI-based analysis, and who may have other concerns about MI, such as treatment of non-monotone missing data. It should be noted, however, that even when the underlying MNAR assumptions are the same (in terms of their clinical interpretation), MMRM-based implementation of PMMs and implementations based on explicit imputation (both single and multiple) can result in slightly different estimates of the mean difference between treatments as well as quite different standard errors of these estimates. We will further highlight this fact in Section 5.

The main objective of this paper is to demonstrate how PMM-type analyses based on linear combinations of LSMs from an MMRM can be implemented in SAS, and how standard errors for the corresponding estimates of the experimental-control treatment differences can be computed with the delta method using standard SAS functionality. The approach is illustrated using, as an example, a particular set of MNAR assumptions based on reasons for discontinuation form the study which results in a particular instance of a PMM and a corresponding linear combination of LSMs. However, general principles and implementation strategy described in Sections 2 and 3 can be applied to impose other MNAR assumptions as long as the final estimate of the difference between treatments can be formulated as a linear combination of LSMs obtained from a longitudinal model. This article does not suggest that the set of MNAR assumptions and the way of implementing them as illustrated in this paper have some general advantages over other methods, including those results of which are provided in Section 5. The choice of the assumptions and the most appropriate method to implement them depends on various characteristics of the real-world studies and would have to be considered on a case-by-case basis. We highlight some of the pros and cons of the MMRM-based and MI-based methods for PMM implementation in the conclusion section.

2. PATTERN-MIXTURE MODELS AS LINEAR COMBINATIONS OF LEAST SQUARES MEANS

PMMs are based on a joint modeling of (clinical) outcome and missingness. Assume that the entire outcome data matrix Y is divided into two components: $\mathrm{Y_{obs}}$, representing observed outcomes and $\mathrm{Y_{mis}}$, representing missing (unobserved) outcomes. Assume that X represents a set of (observed) covariates and R represents a matrix of indicators of missingness. PMMs use the following factorization of the joint probability of data and missingness:

$$\mathrm{p(Y_{obs}, Y_{mis}, R|X)} \;=\; \mathrm{p(R|X)}\;\; \mathrm{p(Y_{obs}, Y_{mis}|\,R, X)} \;=\; \mathrm{p(R|X)}\;\; \mathrm{p(Y_{obs}\,|\,R, X)}\;\; \mathrm{p(Y_{mis}\,|\,Y_{obs}, R, X)} \qquad (2.1)$$

The indicators of missingness R can also be regarded as indicators of belonging to a particular pattern of missingness or, in other words, a group of subjects that share certain characteristics related to missingness, for example, time of discontinuation from the clinical study, reason for discontinuation, or a treatment arm to which subjects were randomized. The probability distribution $\mathrm{p(R|X)}$ can be viewed as a probability distribution of various missingness patterns. The factor $\mathrm{p(Y_{obs}\,|\,R, X)}$ represents a model for available data within each pattern R, and $\mathrm{p(Y_{mis}\,|\,Y_{obs}, R, X)}$ represents a model for missing data conditioned on observed data within each pattern. The pattern-specific estimates are not typically of interest, so the average estimates across the missing data patterns are obtained to yield an overall result. Averaging is accomplished by the weighting of $\mathrm{p(Y_{obs}\,|\,R, X)}\;\; \mathrm{p(Y_{mis}\,|\,Y_{obs}, R, X)}$ with probabilities $\mathrm{p(R|X)}$.

Pattern-mixture models are, by definition, under-identified because the probability $\mathrm{p(Y_{mis}\,|\,Y_{obs}, R, X)}$ cannot be estimated without assumptions. In order to estimate PMMs, one needs to explicitly impose those assumptions, which are sometimes referred to as “identifying restrictions”, regarding the inestimable parameters. There are several methodologies for imposing these identifying restrictions.

In this paper, we consider an approach where an MMRM model is first estimated from observed data in a standard MAR-type modeling framework with treatment arm, time, treatment-by-time interaction, and possibly some covariates included as effects. From this model, one can estimate various LSMs for outcome in different treatment arms and at different time-points. Subsequently, in order to obtain LSMs under departures from the MAR assumptions used to estimate the above model, one can impose certain identifying restrictions in such a way that LSMs for drop-out groups (patterns) are assumed be equal to specific LSMs from the MAR-based model reflecting desired clinical assumptions. An overall estimate of the treatment effect (overall LSM over completers and drop-outs) is then defined as a linear combination of LSMs for different patterns weighted by the probability of subjects belonging to each respective pattern.

We will illustrate this approach using an analysis strategy the main features of which has been recommended in the past for trials with symptomatic treatment of chronic conditions (e.g., chronic pain). In such indications, regulatory agencies recommended the adoption of an approach where subjects who discontinue study treatment due to inability to tolerate it are assumed to revert to their baseline severity of symptoms and derive no further benefit from treatment after discontinuation. When single imputation methods were used in conjunction with ANCOVA, for this group of discontinued subjects, the outcome at the final time-point was imputed using the baseline observation carried forward (BOCF) method, while other subjects were imputed using last observation carried forward (LOCF). Using BOCF for those who discontinued due to intolerability resulted in a more conservative estimate of treatment effect, compared to using LOCF universally, because sometimes these subjects (often more numerous in the experimental treatment arm) did exhibit a considerable symptom relief prior to discontinuation, which would have been carried forward by the LOCF method to the final time-point, thus attributing a positive long-term treatment effect, contrary to regulatory recommendations and clinical objectives. However, a general concern with single imputation methods (both BOCF and LOCF) is that they can potentially underestimate the variance of the treatment effect estimate by not accounting for uncertainty about missing (imputed) data.

In this paper, we demonstrate how similar clinical assumptions can be modeled with MMRM-based PMM analysis as opposed to a hybrid single imputation with BOCF and LOCF. Assume that there are K+1 visits in a study, with 0 being a baseline visit and K being a final time-point used for the primary endpoint. For simplicity, we assume that there are two treatment arms – experimental and control. For the purposes of our analysis, we will consider two types of primary reasons leading to subjects discontinuing the study treatment: adverse events and lack of efficacy, which we will designate together as lack of therapeutic benefit (LTB), vs. other reasons. The primary reasons of discontinuation are assumed to be collected as part of the study data and are identified by the study investigator. A standard MMRM model will include time (with baseline included as a separate time-point), treatment arm and treatment-by-time interaction as fixed effects and baseline efficacy score as covariate. It can be implemented in SAS using PROC MIXED with REPEATED statement (time as repeated factor within subject).

Assume that $\mu_{t,k}$ represents an LSM estimated from this MMRM for a specific treatment arm and time-point, where subscript t denotes the treatment arm (t=0 for control and t=1 for experimental), and subscript k denotes the time-point (k=0,1,…,K). We will also define the following LSMs for each treatment arm t at the last time-point K, which is our primary analysis time-point.

- $\mu_{t,K}^{C} = \mu_{t,K}$: LSM that will be attributed to study completers, as denoted by the superscript C. For completers, the LSM is assumed to be equal to the LSM $\mu_{t,K}$ estimated from the MMRM model for the final time-point in each arm.
- $\mu_{t,K}^{LTB} = \mu_{t,0}$: LSM that will be attributed to drop-outs due to lack of therapeutic benefit, as denoted by the superscript LTB. For this group of drop-outs, the LSMs for the last post-baseline time-point K are assumed to be equal to the LSMs at baseline, $\mu_{t,0}$, thus not attributing any positive treatment effect as would, otherwise, be estimated by a post-baseline estimate of $\mu_{t,K}$ from the MMRM model.
- $\mu_{t,K}^{OTH} = \frac{1}{2}(\mu_{t,1} + \mu_{t,2}), k \geq 1$: LSM that will be attributed to drop-outs due to other reasons, as denoted by the superscript OTH. For this group of drop-outs, the LSMs for the last post-baseline time-point K are assumed to be equal to an average of LSMs at the first two post-baseline visits in the respective treatment arms. This way, the LSMs at the final time-point are effectively replaced by the LSMs from the earlier time-points to which these discontinued subjects contributed their data most. (Note, this assumption may be made more general and expressed as an average of all post-baseline visits except the final one).

Based on the LSMs defined above, an overall LSM, $\mu_{t,K}^{A}$, for a treatment arm t at the final time-point K (over completers and drop-outs as denoted by superscript A) is defined as follows.

$$\mu_{t,K}^{A} = \pi_{t,K}^{C} \times \mu_{t,K}^{C} + \pi_{t,K}^{LTB} \times \mu_{t,K}^{LTB} + \pi_{t,K}^{OTH} \times \mu_{t,K}^{OTH} =$$

$$= \pi_{t,K}^{C} \times \mu_{t,K} + \pi_{t,K}^{LTB} \times \mu_{t,0} + \pi_{t,K}^{OTH} \times \frac{1}{2}(\mu_{t,1} + \mu_{t,2}) \qquad (2.2)$$

where $\pi_{t,K}^{C}$, $\pi_{t,K}^{LTB}$,and $\pi_{t,K}^{OTH}$represent proportions of study completers and of drop-outs due to LTB and other reasons respectively, which can be estimated from observed frequencies of completers and drop-outs. As can be noted from

(2.2), this overall LSM for each treatment arm can be estimated using a linear combination of LSMs, $\mu_{t,k}$, obtained from a standard MAR-based MMRM; yet the overall LSMs $\mu_{t,K}^{A}$ reflect departures from the original MAR assumptions.

Once the overall LSMs are computed as defined in (2.2) for each treatment arm, a difference between LSMs of the two treatments can be estimated as

$$\mu_{d,K}^{A} = \mu_{1,K}^{A} - \mu_{0,K}^{A} \qquad (2.3)$$

Then, a test of the null hypothesis of no treatment effect can be performed using a Wald statistic

$$\mu_{d,K}^{A} / SE(\mu_{d,K}^{A}) \qquad (2.4)$$

where $SE(\mu_{d,K}^{A})$ is a standard error of the $\mu_{d,K}^{A}$ estimate. A special consideration should be given to the estimation of this standard error. The estimate should take into account the fact that the proportions $\pi_{t,K}^{C}$, $\pi_{t,K}^{LTB}$, and $\pi_{t,K}^{OTH}$ are estimates of the multinomial random variables, and thus their variances should be incorporated in the overall variance estimate. This can be done using a delta approximation method for variance estimation [6,7]. In the next section, we describe the delta method and how to implement it using information readily available from PROC MIXED.

3. DELTA APPROXIMATION METHOD FOR VARIANCE ESTIMATION

Assume an LSM of interest μ (e.g., $\mu_{d,K}^{A}$ as described in the previous section) is expressed as a linear combination of LSMs μ_1, μ_2,…, μ_P from a standard MMRM:

$$\mu = \pi_1 \times \mu_1 + \pi_2 \times \mu_2 + \cdots + \pi_P \times \mu_P \qquad (3.1)$$

where π_p are proportions of subjects in specific groups (patterns) defined, for example, based on a combination of treatment arm, study completion status, and reason for discontinuation. LSMs μ_p $(p = 1, \ldots, P)$ are estimated from an MMRM with a vector of parameter estimates $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_M)$. Each of the LSMs μ_p can be expressed in terms of model parameters and contrast coefficients c_m^p:

$$\mu_p = c_1^p \times \beta_1 + c_2^p \times \beta_2 + \cdots + c_M^p \times \beta_M \qquad (3.2)$$

Variance of the overall LSM, $Var(\mu)$, can be estimated using a delta approximation method [6,7] as follows.

$$Var(\mu) = \boldsymbol{\nabla}\mu \, [VarCov(\boldsymbol{\beta}) \oplus VarCov(\boldsymbol{\pi})] \, (\boldsymbol{\nabla}\mu)' \qquad (3.3)$$

where $\nabla\mu$ is a gradient vector of μ consisting of partial derivatives of μ with respect to model parameters β_m and proportions π_p. The elements of this gradient vector are as follows:

$$\boldsymbol{\nabla}\mu = \left(\frac{\partial\mu}{\partial\beta_1}, \frac{\partial\mu}{\partial\beta_2}, \ldots, \frac{\partial\mu}{\partial\beta_M}, \frac{\partial\mu}{\partial\pi_1}, \frac{\partial\mu}{\partial\pi_2}, \ldots, \frac{\partial\mu}{\partial\pi_P}\right) \qquad (3.4)$$

Variance-covariance matrix $VarCov(\boldsymbol{\beta})$ can be obtained directly from the estimates of the MMRM.

Variance-covariance matrix $VarCov(\boldsymbol{\pi})$ can be computed as variances and covariances of sampling proportions from a multinomial distribution:

$$\text{Diagonal elements (variances) } VarCov(\boldsymbol{\pi})[i,i] = \pi_i \frac{1-\pi_i}{N};\ i = 1, \ldots, P \qquad (3.5)$$

where N is the total number of subjects in the sample based on which π_i are estimated (e.g., number of subjects in a given treatment arm).

$$\text{Off-diagonal elements (covariances) } VarCov(\boldsymbol{\pi})[i,j] = -\frac{\pi_i \pi_j}{N};\ i,j = 1, \ldots, P, i \neq j \qquad (3.6)$$

Notation $[VarCov(\boldsymbol{\beta}) \oplus VarCov(\boldsymbol{\pi})]$ is used to indicate that the overall matrix is block-diagonal:

$$\begin{bmatrix} VarCov(\boldsymbol{\beta}) & 0 \\ 0 & VarCov(\boldsymbol{\pi}) \end{bmatrix} \qquad (3.7)$$

Proportions π_p are estimated from a sample of study subjects as $\frac{n_p}{N}$, i.e., the number of subjects in group *p* divided by a total number of subjects *N* over all categories (typically within each treatment arm).

To facilitate the computation of $Var(\mu)$ as defined by equation (3.3), we will use the following data structure that can be represented and manipulated as a regular SAS dataset.

Table 1: Data Structure with Elements for Delta Method Variance Calculation

Model Effect	Model Parameter Estimate	μ_1 LSM Coefficients	π_1 Estimate	…	μ_P LSM Coefficients	π_P Estimate	$\frac{\partial \mu}{\partial \beta_m}$
<Model Term, Level>	β_1	c_1^1	π_1	…	c_1^P	π_P	$\frac{\partial \mu}{\partial \beta_1}$
<Model Term, Level>	β_2	c_2^1	π_1	…	c_2^P	π_P	$\frac{\partial \mu}{\partial \beta_2}$
	…			…			
<Model Term, Level>	β_M	c_M^1	π_1	…	c_M^P	π_P	$\frac{\partial \mu}{\partial \beta_M}$
$\frac{\partial \mu}{\partial \pi_p}$			$\frac{\partial \mu}{\partial \pi_1}$	…		$\frac{\partial \mu}{\partial \pi_P}$	

The elements of the gradient vector $\nabla\mu$ are represented by the last column and last row of this data structure. They can be computed as follows using other values from this data structure.

$$\frac{\partial \mu}{\partial \beta_m} = \sum_{p=1}^{P} c_m^p \times \pi_p \qquad (3.8)$$

$$\frac{\partial \mu}{\partial \pi_p} = \sum_{m=1}^{M} c_m^p \times \beta_m \qquad (3.9)$$

PROC MIXED can be used to obtain parameter estimates $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_M)$ for the MMRM. We can use LSMESTIMATE statement in PROC MIXED to request the LSMs μ_p and the overall LSM μ, the result of which will be stored in the ODS output dataset LSMESTIMATES. Specification of LSMs is made easier with the new non-positional syntax of LSMESTIMATE statement [8]. When this simplified syntax is used (see examples in Section 5), option E allow us to obtain automatically all the contrast coefficients c_m^p (corresponding to a positional syntax with the full model parameterization) for the requested LSMs. These coefficients can be captured in the ODS dataset COEF. Parameter estimates β_m, and their variance-covariance matrix $VarCov(\boldsymbol{\beta})$ can also be readily obtained as ODS output datasets produced by PROC MIXED (SOLUTIONF and COVB respectively).

If all the necessary information is pulled from the output ODS datasets produced by PROC MIXED and stored in a dataset as described above, calculation of partial derivatives using equations (3.8) and (3.9) can be done in a SAS data step using arrays to compute $\frac{\partial \mu}{\partial \beta_m}$ on each row of the dataset, and using RETAIN statements to calculate $\frac{\partial \mu}{\partial \pi_p}$ in columns. Partial derivatives computed in this way can be transferred into a row vector (3.4) and stored as SAS dataset with one record and M+P variables.

Then matrix (3.7) needs to be constructed using ODS output dataset COVB for the upper left block and performing calculations in equations (3.5 and (3.6) for the lower right block. This matrix can be stored as a SAS dataset as well.

After that, the only thing that remains to be done is to perform vector and matrix multiplication operations according to equation (3.3). This can be done either by using PROC IML or implemented with core SAS functionality. In Appendix 1, we provide a macro that performs these operations using SAS/STAT procedure PROC FCMP.

This section presented a general definition and data structures needed to compute the variance of a linear combination of LSMs using the delta approximation method. In Section 5, we provide a detailed example and SAS code for the analysis based on the MNAR assumptions based on reasons for subject discontinuation described in the previous section.

4. EXAMPLE DATASET

Implementation of a PMM-based analysis will be illustrated using an example study where efficacy endpoint is stored in a SAS dataset, DATAIN, with the following variables:

- SUBJID – subject identification number;
- TRT – treatment arm (0=control and 1=experimental);

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：https://d.book118.com/238123033051006112