A Comparison of Frequentist and Bayesian Model Based Approaches for Missing Data Analysis: Case Study with a Schizophrenia Clinical Trial

G. Frank LIU, Baoguang HAN, Xin ZHAO, and Qun LIN

Missing data are common in clinical trials and could lead to biased estimation of treatment effects. The National Research Council (NRC) report suggests that sensitivity analysis on missing data mechanism should be a mandatory component of the primary reporting of findings from clinical trials, and regulatory agencies are requesting more thorough sensitivity analyses from sponsors. However, recent literature research showed that missing data were almost always inadequately handled. This is partially due to the lack of standard software packages and straightforward implementation platform. With recent availability of flexible Bayesian software packages such as WinBUGS, SAS Proc MCMC, and Stan, it is relatively simple to develop Bayesian methods to address complex missing data problems while incorporating the uncertainty. In this article, we present a case study from the DIA Bayesian Scientific Working Group (BSWG) on Bayesian approaches for missing data analysis. We illustrate how to use Bayesian approaches to fit a few commonly used frequentist missing data models. The properties, advantage, and flexibility of the Bayesian analysis methods will be discussed using a case study based on a schizophrenia clinical trial. Supplementary materials for this article are available online.

Key Words: Bayesian; Longitudinal clinical trial; Missing data; Sensitivity analysis.

1. Introduction

Missing data are inevitable in most longitudinal clinical trials. Extensive research has been devoted to statistical methods on how to handle missing data in clinical trials. Traditionally, single imputation methods such as last observations carried forward and baseline observations carried forward have been used to handle missing data. The single imputation method may provide a conservative estimate for the mean response, but it may not be conservative for estimating and testing the treatment difference. Maximum likelihood (ML) or multiple imputation (MI) based methods are recommended in more recent research papers (e.g., Carpenter and Kenward 2007; EMA guideline 2010; NRC Report 2010). One assumption for ML or MI-based method is that the missing data are missing at random (MAR). This assumption cannot be verified from the observed data. Therefore, sensitivity analysis is suggested for evaluating the robustness of the analysis results against the missing data and MAR assumption. This is recommended in both the European Medicine Agency guideline on missing data in confirmatory trial
A Comparison of Frequentist and Bayesian Model Based Approaches

(EMA 2010) and U.S. Food and Drug Administration-commissioned panel report, the prevention and treatment of missing data in clinical trials, issued by National Research Council of the National Academies (NRC 2010).

A great amount of research has been devoted in statistical methods for handling missing data in the last two to three decades. Extensive literature and details may be found in comprehensive textbooks (Little and Rubin, 2002; Molenberghs and Kenward, 2007; Daniels and Hogan, 2008). Although there is no universally best method for handling missing data, many methods have been proposed and applied in clinical trials, including ML, MI, and fully Bayesian (FB). However, recent literature research showed that missing data were almost always inadequately handled (Sterne et al. 2009). This is partially due to the lack of standard software packages and straightforward implementation platform.

The Bayesian method provides a natural approach to modeling the uncertainty of missing data in longitudinal trials and have been considered in statistical literature (Best et al. 1996; Carpenter, Pocock, and Lamm 2002; Daniels and Hogan 2008; Hogan, Daniels, and Hu 2014; Ibrahim et al. 2005). In fact, there is a close connection between Bayesian methods and other popular missing data methods such as ML and MI (Chen and Ibrahim 2013). With a large sample size, the implementation of the Bayesian method with noninformative priors on all parameters will lead to ML estimates. The imputation step in MI is based on sampling from a posterior predictive distribution. Therefore, Bayesian methods are generally considered as more powerful for dealing with various missing data problem including missing data in both covariates and response (Ibrahim et al. 2005). However, the application of Bayesian methods in real clinical trials is still not common due to lack of computational software and regulatory considerations. With significant advancement in computation and statistical software, the Bayesian method provides a feasible alternative approach for analysis of longitudinal clinical trials. To investigate the potential application of Bayesian methods with missing data, a subteam was formed within the Drug Information Association Bayesian Scientific Working Group (DIA BSWG) to tackle the problem. Considering the wide variety of methods were readily available for missing data analysis (e.g., Little and Rubin 2002; Daniels and Hogan 2008), the subteam decided to work on several case studies with real datasets from clinical trials to investigate and demonstrate the application of Bayesian approaches. In this article, we illustrate Bayesian approaches along with commonly used frequentist methods for handling missing data for a continuous endpoint measured in a real schizophrenia clinical trial. The properties of the analysis methods will be discussed using this case study dataset.

Some descriptions about the case study data and prespecified analysis methods are given in Section 2. Section 3 provides common frequentist sensitivity analysis models including selection model, shared parameter model, and pattern mixture model. In Section 4, we demonstrate how to use Bayesian methods to fit the similar sensitivity analysis models with some additional flexibility to modify the model specification. Some discussions are provided in Section 5.

2. Summary of Study Design and Primary Analysis

2.1 Study Design and Missing Data

The case study data are from a multicenter, randomized, double-blind clinical trial on patients with schizophrenia. The study had three treatment groups: test, active, and placebo with a 2:1:2 randomization ratio. A total of about 200 patients were enrolled from four eastern European countries. The primary endpoint for efficacy was the mean change from baseline in the Positive and Negative Syndrome Scale (PANSS) total score. The PANSS was measured at baseline (end of placebo lead-in period), Day 4, and Weeks 1, 2, 3, and 4. The primary time point for treatment comparison was at Week 4.

In the study, 19%–33% of the patients dropped out from the study depending on the treatment group prior to Week 4. The majority of the patients dropped out due to lack of efficacy, especially in the test drug and placebo groups. A Kaplan–Meier curve for time to discontinuation is given in Figure 1. A summary of patient dropout by reason is provided in Table 1.
Table 1. Summary of patient dropout by reason

| Treatment Group | Active N = 45 | Placebo N = 78 | Test N = 81 |
|-----------------|--------------|---------------|-------------|
| Completer       | n 36         | n 57          | n 54        |
|                 | % 80.0       | % 73.1        | % 66.7      |
| Dropout by reason |             |               |             |
| Adverse Event   | 0            | 0             | 3           |
| Lack Of Efficacy| 4            | 15            | 16          |
| Physician Decision | 0       | 0             | 1           |
| Pregnancy       | 1            | 0             | 0           |
| Progressive Disease | 1        | 0             | 1           |
| Protocol Violation | 1         | 0             | 0           |
| Withdrawal by Subject | 2       | 6             | 6           |

are given in Figure 2; and for test drug and placebo groups in Figure 3. The numbers of dropout patients at each time point are included in the labeling in the plots. The mean estimates in the plots are based on observed data at each given time point. In general, we can see that for all dropouts, the mean change from baseline values had a worsening trend, which is consistent with the majority of dropouts being due to lack of efficacy as shown in Table 1.

In real applications, several methods may be considered to handle the intermittent missing data. One of the proposed approaches in the literature (e.g., Carpenter, Roger, and Kenward 2013) is to do multiple imputation for the intermittent missing data under MAR assumption and then apply missing data analysis methods (e.g., selection model) and combine the results using the multiple imputation approach. With FB approach, the intermittent missing data can be incorporated in the posterior sampling automatically in some recent software such as Win-BUGS (Lunn et al. 2009), Stan (2012), and Proc MCMC in SAS/Stat 13.2 (2014).

2.2 Notations

With general notations, we assume that there are N patients in a clinical trial and \( Y_i = (Y_{i1}, \ldots, Y_{in}) \) is a vector of \( n \) repeated measurements to be collected for patient \( i \), and \( X_i = (X_{i1}, \ldots, X_{ip}) \) is a \( (p \times 1) \) vector of covariates. Let \( Y = \{Y_1, \ldots, Y_N\} \) and \( X = \{X_1, \ldots, X_N\} \). Likelihood-based method assumes a model for the distribution of \( Y \) given \( X \) with a collection of unknown parameters \( \gamma \), \( f(Y|X, \gamma) \). Assuming patients are

![Figure 2. Mean change from baseline for active vs. placebo by missing data pattern (number of patients given in parentheses).](image-url)
independent, then the full data (i.e., no missing data) likelihood function is given below,

\[ f (Y | X, \gamma) = \prod_{i=1}^{N} f(Y_i | X_i, \gamma). \]

When there are missing data, let \( R_{ij} \) be the missing-data indicator for \( Y_{ij} \), with value of 1 if \( Y_{ij} \) is observed and 0 if \( Y_{ij} \) is missing, and \( R = (R_1, \ldots, R_n) \). The vector \( Y^o \) denotes the set of observed values for patient \( i \), and \( Y^m \) the set of missing values.

With incomplete data, we need to specify a joint distribution of \( Y \) and \( R \) with density \( f(R, Y | X, \theta) \) and parameters \( \theta = (\gamma, \phi) ; \gamma \) contains the parameters for distribution of the full data response \( Y \), and \( \phi \) contains the parameters for the missing data mechanism. The complete-data likelihood would be

\[ L (\theta | R, Y, X) = \prod_{i=1}^{N} f(R_i, Y_i | X_i, \theta). \]

### 2.3 The Original Analysis

The primary analysis as specified in the protocol was based on mixed model for repeated measure (MMRM) (Mallinckrodt et al. 2008). The model included factors for baseline PANSS total score, treatment, week, country, treatment by week and baseline by week interactions, where week, country, and treatment are treated as categorical variables. An unstructured covariance matrix was used for modeling the intra-subject correlation over the repeated measures. To further simplify the analyses, we removed country from our analysis model as it was not a significant factor in the analysis. Specifically, in the MMRM model we assumed that the longitudinal data \( Y_i \) follow a multivariate normal distribution with unstructured covariance structure \( \Sigma \), that is, \( Y_i | X_i \sim N(\mu_i, \Sigma) \).

All data from three treatment groups were included in the original analysis model specified in the protocol. To make the analysis results more compatible with other sensitivity analysis models presented in the following sections, we analyzed the data in two separate models: one for active and placebo, and another for test drug and placebo. Specifically, the MMRM for test drug and placebo has the mean

\[ \mu_{ij} = \beta_j (1 - D_i) + \alpha_j D_i + \gamma_j B_i, \quad j = 1, \ldots, n. \]

The analysis results based on these separate MMRM models are given in the first column of Table 2. The results from the original MMRM model with all three treatment groups (results shown in Table 4) were very similar to that from these separate MMRM models except for some slight increase in variance of the mean estimate on the contrast between active and placebo group. The conclusions are that the tested drug showed no evidence of activity. The point estimate for active control showed some activity but the effect was not statistically significant.

One assumption for the MMRM method is that the missing data are MAR. This assumption cannot be verified from the observed data. Therefore, sensitivity analysis is suggested for evaluating the robustness of the analysis results against the missing data and MAR assumption. It will be interesting to see whether different models/approaches would generate “similar” results.
Table 2. Estimated mean different (SE) based on frequentist and Bayesian methods

| Difference | MMRM All Data | MMRM Complers | Selection Model | Shared Parameter | PMM ACMV |
|------------|---------------|---------------|----------------|------------------|----------|
| Active vs. Pbo | $-4.25 (3.87)$ | $-3.17 (3.34)$ | $-7.50 (4.26)$ | $-5.69 (4.36)$ | $-5.99 (3.76)$ |
| Test vs. Pbo  | $-0.42 (3.28)$ | $-2.97 (2.78)$ | $-0.46 (3.69)$ | $1.10 (3.71)$  | $-1.97 (3.43)$ |

Bayesian Methods

| Difference | MMRM All Data | MMRM Complers | Selection Model | Shared Parameter | PMM ACMV |
|------------|---------------|---------------|----------------|------------------|----------|
| Active vs. Pbo | $-4.27 (3.89)$ | $-3.27 (3.37)$ | $-6.53 (4.33)$ | $-6.43 (4.72)$ | $-5.91 (4.21)$ |
| Test vs. Pbo  | $-0.45 (3.32)$ | $-2.97 (2.79)$ | $-1.29 (3.74)$ | $0.04 (3.86)$  | $-1.91 (3.87)$ |

NOTES:

$a$The analysis models are run separately for Active vs. Placebo, and for Test vs. Placebo. The frequentist methods for selection and shared parameter models are based on SAS macros developed by DIA Missing Data Working Group.

$^b$The imputation is done by treatment group. The frequentist model is fit by SAS PROC MI with 500 imputations.

Table 3. Results for PMM with ACMV, NCMV, and CCMV

| Difference | PMM ACMV | PMM NCMV | PMM CCMV |
|------------|----------|----------|----------|
| Active vs. Pbo | $-4.37 (3.62)$ | $-4.75 (6.26)$ | $-3.97 (3.39)$ |
| Test vs. Pbo  | $-0.56 (3.43)$ | $4.10 (6.64)$ | $-1.17 (3.18)$ |

NOTE: The imputation is done with treatment as a covariables using SAS PROC MI with 500 imputations.

3. Some Frequentist Approaches for Sensitivity Analysis

We first applied a few conventional methods to handle missing data and analyze the case study. The methods are frequentist based and include complete case analysis, selection models, shared parameter model, and pattern mixture model. Some brief descriptions of these methods are given in this section. More details can be found in the recent paper by DIA missing data working group (Mallinckrodt et al. 2013), and references therein.

3.1 Complete-Case Analysis

With the complete-case analysis, the observations for patients who dropped out are ignored. This model is valid under that the missing data are missing completely at random. The analysis consists of 147 completers. The results from the separate MMRM models for the completers are shown in the second column of Table 2. It is clear that the results are quite different from that of the separate MMRM analysis with all available data. Especially for the test drug, the completers had larger treatment effects compared with that from the separate MMRM analysis with all available data. This implied that the missing data may not be missing completely at random. In fact, majority of the dropouts were discontinued due to lack of efficacy. Therefore, the probability of dropout is likely depending on the previously observed outcomes and/or potentially the missing outcomes. Figures 2 and 3 plot the mean change from baseline by treatment group and missing data pattern as defined by time of dropout. It can be seen that for those who dropped out early the observed mean change values at time of dropout are always larger (worse) than the previous mean changes, and are larger than 0 (worse than baseline) for most of the cases. For the completers, the test drug showed no treatment difference

Table 4. Summary of results for Bayesian selection model with modified dropout probability

| Parameters | Conventional MMRM | Bayesian Selection Models |
|------------|--------------------|--------------------------|
| Active vs. Pbo | $-4.37 (3.75)$ | $-4.34 (3.75)$ | $-4.23 (3.67)$ | $-4.27 (3.72)$ |
| Test vs. Pbo  | $-0.48 (3.28)$ | $-0.47 (3.31)$ | $-0.71 (3.32)$ | $-0.79 (3.29)$ |
| $\gamma_0$ | N/A | $-0.03 (0.06)$ | $-0.03 (0.06)$ | $-0.03 (0.06)$ |
| $\gamma_1$ | $0.14 (0.02)$ | $0.17 (0.07)$ | $0.22 (0.06)$ |
| $\gamma_2$ | N/A | N/A | $-0.04 (0.04)$ |
| $\gamma_3$ | N/A | N/A | $-0.08 (0.05)$ |
| $\gamma_4$ | N/A | N/A | $-0.03 (0.07)$ |
| $\zeta_2$ | $-3.49 (0.40)$ | $-3.73 (0.56)$ | $-3.97 (0.56)$ |
| $\zeta_3$ | $-2.53 (0.31)$ | $-2.81 (0.50)$ | $-2.85 (0.49)$ |
| $\zeta_4$ | $-1.73 (0.29)$ | $-1.92 (0.42)$ | $-1.98 (0.40)$ |
| $\zeta_5$ | $-1.75 (0.37)$ | $-1.96 (0.51)$ | $-2.45 (0.54)$ |
A Comparison of Frequentist and Bayesian Model Based Approaches

except for some small difference at last time point. Therefore, the analysis for completers had a positive treatment effect at last time point. In the MMRM based on all available data, the test drug had a smaller treatment effect because for those who dropped out early, the test drug showed the negative effect (worse than placebo) as seen in Figure 3.

3.2 Selection Models

Selection models specify the joint distribution of \( R_i \) and \( Y_i \) through models for the marginal distribution of \( Y_i \) and the conditional distribution of \( R_i \) given \( Y_i \),
\[
f(R_i, Y_i | X_i, \gamma, \phi) = f_X(Y_i | X_i, \gamma) f_{R|Y}(R_i | Y_i, X_i, \phi),
\]
where \( \gamma \) and \( \phi \) are distinct parameters for the full data response and the missing data mechanism, respectively. We consider the parametric selection model as specified in Diggle and Kenward (1994), in which the first part of the likelihood \( f_Y(Y_i | X_i, \gamma) \), is the same as in the MMRM model. For this case study, we used the SAS logit used for the missing data mechanism.

\[
y_{ij} = \beta_1(1-D_i) + \alpha_j D_i + \gamma_j B_i.
\]

For the second part of the likelihood, \( f_{R|Y}(R_i | Y_i, X_i, \phi) \), the following logistic model is used for the missing data mechanism.
\[
R_{ij} \sim \text{Bernoulli}(q_{ij}),
\]
\[
\text{logit}(q_{ij}) = \phi_1(1-D_i) + \phi_2 D_i + \phi_3(1-D_i)y_{j-1} \]
\[
+ \phi_4 D_i y_{j-1} + \phi_5 (1-D_i)y_j + \phi_6 D_i y_j, \ j = 2, \ldots, v_i
\]
where \( D_i \) is the indicator for the drug group, \( q_{ij} \) is the probability of patient \( i \) dropout at time \( j \), \( v_i \) is equal to \( n \) for completers, and is equal to the first visit number with a missing response for patients who dropped out from the study.

When \( \phi_3 = \phi_6 = 0 \), this model implies missing data only depends on previous observed data (i.e., MAR). When both models are specified with parametric distributions, all the parameters including \( \phi_3 \) and \( \phi_6 \) can typically be identified (Diggle and Kenward 1994). Using the DIA macro, the results for fitting these selection models are given in the third column of Table 2. Compared to the analysis based on MMRM, the selection model result for test versus placebo is similar but that for active versus placebo is somewhat different although the conclusion is similar. This difference may be resulted from the logistic model specification for missing data. We will discuss more details in Section 4.3.

3.3 Shared Parameter Models

With shared parameter model, the outcome process \( Y_i \) and the dropout process \( R_i \) depend on a shared latent variable \( U_i \). Given \( U_i, R_i, \) and \( Y_i \) are assumed to be independent:
\[
f(R_i, Y_i, U_i | X_i, \gamma, \phi) = f_U(U_i | X_i, \gamma)
\times f_{Y|U}(Y_i | X_i, U_i, \gamma) f_{R|YU}(R_i | Y_i, U_i, X_i, \phi).
\]

To fit this model, parametric forms are assumed for the response and dropout models. For example, linear and quadratic functions are considered in the DIA macro (see http://missingdata.lshtm.ac.uk). As in the selection model, the DIA macro can only fit data with two treatment groups at a time. Here, we consider a quadratic model to allow a potential nonlinear response profile for this case study. Specifically, the model between drug and placebo is:
\[
y_{ij} = (b_0 + b_1 w_j + b_2 w_j^2)(1-D_i)
+ (b_3 + b_4 w_j + b_5 w_j^2) D_i + b_6 B_i
+ U_{i0} + U_{i1} w_j + U_{i2} w_j^2 + \varepsilon_{ij}
\]
and
\[
R_{ij} \sim \text{Bernoulli}(q_{ij}),
\]
\[
\text{logit}(q_{ij}) = \xi_j + a \text{trt}_i + (a_0 + b_0 D_i) U_{i0}
+ (a_1 + b_1 D_i) U_{i1} + (a_2 + b_2 D_i) U_{i2}, \ j = 2, \ldots, v_i
\]
where \( w_j \) is the time in weeks for \( j \)th visit, \( (U_{i0}, U_{i1}, U_{i2}) \sim N(0, \Sigma) \) are shared random effects, \( \varepsilon_{ij} \sim N(0, \sigma^2) \). For this case study, \( w_j \) = 0.57, 1, 2, 3, and 4 for visit 1 through 5, respectively.

The results from fitting these quadratic shared random effects models are provided in the fourth column of Table 2. Compared to the analysis based on MMRM, the shared parameter model results are somewhat different for both active versus placebo and test versus placebo comparisons although the conclusions are similar. The difference may be due to the quadratic functional form used for the response and shared random effects models. Note that in the MMRM, a cell-mean model was used for the response so there is no restriction posted on the mean parameters over time.

3.4 Pattern Mixture Models

Pattern mixture models were first proposed by Little (1993). This approach has more appeal to many scientists
as it is transparent on how the observed data and missing data are modeled. For this case study, all patients had observations at Day 4. We define the dropout pattern based on the maximum visit number before a patient dropped out. For example, pattern \( s = 1 \) contains patients who dropped out after Week 1, pattern \( s = 2 \) contains patients who dropped out after Week 2, etc., and finally pattern \( s = 5 \) contains all completers.

To identify the parameters in the model, we consider the following 3 commonly used constraints in building the imputation models for missing data:

- **Complete case missing value (CCMV),**
  \[
  f(y_j | s = k, y_1, \ldots, y_{j-1}) = f(y_j | s = 5, y_1, \ldots, y_{j-1}), \quad k < j,
  \]
  which only uses data from completers to fit imputation models.

- **Neighboring case missing value (NCMV),**
  \[
  f(y_j | s = k, y_1, \ldots, y_{j-1}) = f(y_j | s = j, y_1, \ldots, y_{j-1}), \quad k < j,
  \]
  which imputes missing data \( y_j \) using a model built from data of pattern \( s = j \), that is, all patients with \( y_j \) observed and dropped out after visit \( j \).

- **All case missing value (ACMV),**
  \[
  f(y_j | s = k, y_1, \ldots, y_{j-1}) = f(y_j | s \geq j, y_1, \ldots, y_{j-1}), \quad k < j,
  \]
  which uses all available data to fit imputation model. For example, the imputation model for missing data \( y_j \) is built from data of all patterns \( s \geq j \), that is, all patients with \( y_j \) observed regardless of time of drop out.

ACMV is equivalent to MAR under monotone missingness (Molenberghs et al. 1998). For potential MNAR data, CCMV and NCMV can provide some sensitivity analysis in the sense that the imputation models are built using completers and neighboring cases only, respectively. In this case study, we apply the ACMV, NCMV, and CCMV approaches. For each approach, the missing data are imputed under the given constraint using regression models with factors of baseline PANSS total score, treatment, and previous observed PANSS change from baseline total scores.

The multiple imputed datasets are analyzed using a regression model for the response at last time point (i.e., week 4) with factors of baseline PANSS total score, dropout pattern, treatment, and pattern by treatment interaction. The effects and differences between treatment groups weighted by sample size across the patterns are obtained (e.g., using OM option in LSESTIMATE statement in SAS PROC Mixed procedure). Because the standard error from the regression model treated the proportion of patients in each pattern of missing data as fixed, we also adjusted the variance using the approach as proposed by Hedeker and Gibbons (2006).

In the case study analysis, the first two dropout patterns are combined because there were only eight patients who dropped at the second visit. The analysis results using the pattern mixture model approach are given in Table 3. The results from ACMV are very close to that from the MMRM analysis, which confirms that ACMV is equivalent to MAR under monotone missingness. In CCMV, because the imputation models for all missing data are built from the completers, the estimated treatment differences from CCMV are between that of MMRM with all data and that of MMRM based on completers. It is noticed that the NCMV result between test drug and placebo was quite different from that of the MMRM or completers. This is because the neighboring case missing value approach may impute higher values (worse) for test drug than placebo. The standard error values from the NCMV analyses were also much larger than those from the ACMV or CCMV analyses because the imputation models in NCMV were built from neighboring cases that had much smaller sample sizes.

In all the above analysis models, the variance covariance matrix was assumed to be the same between two treatment groups in the analysis. To assess the sensitivity on this assumption, we also run another PMM analysis with ACMV, where the imputation is done for each treatment group separately. As such, different variance covariance matrix was obtained for each treatment group in the imputation model. The results are somewhat different from the MMRM and are included in the last column of Table 2 for comparisons with the results from the Bayesian approach (to be described in the next section). In fact, these results were similar to that from the MMRM analyses (results not shown) when we used different covariance matrix for each treatment group.

4. Bayesian Approaches for Sensitivity Analysis

In this section, we provide Bayesian analogues to the frequentist models including MMRM models under MAR, selection models, shared parameter models, and pattern mixture models under MNAR. We also illustrate
flexibility of using the Bayesian method in a selection model with different parameterization.

In all Bayesian analyses, the missing data are treated as random parameters and will be sampled from the specified joint distribution of the repeated measures. We used SAS Proc MCMC (SAS/Stat 13.2) or Stan for the analyses in which the missing data will be automatically sampled from the specified distribution. Stan uses a variant of Hamiltonian Monte Carlo. For complex models, Stan usually converges more quickly than software using simple MCMC algorithms such as the random-walk Metropolis (Proc MCMC) and Gibbs sampling (WinBUGS) (Han et al. 2014).

4.1 Bayesian MMRM and Complete-Case Analysis

We first used Bayesian approach to compute the posterior distribution of the parameters in the MMRM model and complete-case analysis. The MMRM approach assumes MAR (or MCAR for complete-case analysis). Here, we are using Bayesian approach as a computational tool to get parameter estimates. Hence, we assume vague priors (or noninformative priors) on all the model parameters to confirm that Bayesian approach will produce similar results as the likelihood methods. Specifically, we assume the following priors for the parameters in the MMRM specified in Section 2.3,

\[
\alpha_j, \eta_j, \beta_j, \gamma_j \sim N \left(0, 100^2\right), j = 1, 2, \ldots, n, \\
\Sigma \sim \text{invWishart} \left(I_n, n\right),
\]

where \(n\) is the number of visits and \(I_n\) is the identity matrix of size \(n\).

The posterior summaries are shown in the bottom section of the first two columns in Table 2. For both models, the results are fairly close to those from the corresponding frequentist methods. For a study with moderate sample size, this is as expected because the Bayesian models use noninformative priors so that the posterior distribution is similar to the likelihood function from the frequentist method.

4.2 Bayesian Selection, Shared Parameter, and PMM Models

Bayesian approach can also be used to fit the selection and shared parameter models specified in Sections 3.2 and 3.3. For the selection model, we assume vague prior, \(N(0, 100^2)\), for all regression parameters \(\alpha_j, \beta_j, \gamma_j\) in the MMRM model, and \(\Sigma \sim \text{invWishart} \left(I_n, n\right)\) where \(n\) is the number of visits and \(I_n\) is the identity matrix of size \(n\). For missing data mechanism model, we use \(\phi_1, \phi_2 \sim \text{logistic}(0, 1)\) and \(\phi_3, \phi_4, \phi_5, \phi_6 \sim N(0, 5^2)\). These vague prior distributions are chosen so that they bear minimal weight to the posterior distributions. For particular situations, some appropriate modifications might be made to these prior specifications to speed up convergence without impacting the resulting posterior distribution. For example, with our case study, we also tried \(\phi_1, \phi_4, \phi_5, \phi_6 \sim N(0, 0.8^2)\). The posterior sample means and standard deviations for the analysis parameters are largely unchanged (data not shown). To ensure that MCMC sampling converges to the target posterior sample, multiple independent MCM chains may be carried out. The convergence can be effectively examined by visual inspections of trace-plot (see some selected trace plots in the online supplemental materials), but the diagnosis might be aided by statistical measures such as Gelman and Rubin diagnostics (Gelman and Rubin 1992).

Similarly, for the shared parameter model, we assume \(\beta_0, \beta_1, \ldots, \beta_6 \sim N(0, 100^2)\) for regression parameters in data model and \(\Sigma_j \sim \text{invWishart} \left(I_j, 3\right)\). For missing mechanism model, we assume \(\xi_j \sim \text{logistic}(0, 1)\) and \(\alpha, \alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2 \sim N(0, 5^2)\).

The results for the selection and shared parameter models are presented in the third and fourth columns of Table 2. The posterior means (and standard deviation) of the treatment effect of test drug versus placebo and for active drug versus placebo are somewhat different from those of the corresponding frequentist models. For this case study, however, the conclusions are the same with respect to statistical significance for testing the treatment effect between groups.

We also use Bayesian method to fit the PMM with ACMV restriction. Similar to the frequentist PMM analysis, the first two patterns are combined. The details of the model specification are given in Appendix A. We assumed that the pattern indicator follows a multinomial distribution, that is, \(S \sim \text{Mult}(\phi)\), \(\phi = (\phi_1, \phi_2, \phi_3, \phi_4)\) with \(\phi_s = P(S = s)\) for \(s \in \{1, 2, \ldots, 4\}\) and \(\sum_s \phi_s = 1\). The prior for \(\phi\) follows a Dirichlet distribution with all parameters set to one. For all standard deviations, we assign a uniform prior with support of \((0, 50)\). For other regression parameters, a normal prior with mean of zero and standard deviation of 100 is assumed. Separate models are specified for different treatment groups. The results are shown in the last column of Table 2 and are very similar to the corresponding PMM ACMV analysis based on the frequentist multiple imputation approach (where the imputation is done for each treatment group separately).

4.3 Bayesian Selection Model With a Modified Dropout Probability

In the selection model described in Section 3.2, the parameters for the probability of dropout do not depend on visit \(j\) but depend on the treatment group. This may pose
some restriction on the parameters because the proportion of patients who dropped out can be different by visit. It is also noticed that there were only eight patients dropped out from the active group. This may produce unstable estimates for the six parameters in the missing data probability for active group in the selection model as specified in Section 3.2. In a double-blinded clinical trial, we can expect that the dropout probability may mostly depend on efficacy outcome (observed or unobserved). Conditioning on efficacy outcome, the dropout probability may be less dependent on treatment groups (i.e., dropout may not depend on treatment after adjusting for efficacy outcomes as the treatment group is blinded during the study). Therefore, we modify the dropout probability model as follows.

\[
\text{logit} \left[ q_{ji} \right] = \zeta_j + \sum_{k=0}^{j} \gamma_k y_{i(k-j)}, \ j = 2, 3, 4, 5.
\]

In this model, the different probabilities of dropout by visit may be captured by intercept parameters \( \zeta_j \), and the parameters \( \gamma_k \) will estimate the dependence of dropout on the efficacy outcome at \( k \)-visit before the time of dropout. When \( \gamma_0 = 0 \), the model becomes MAR. Depending on the level of dependence of the dropout probability and the previous outcomes, we may assume some of the high level \( \gamma_k = 0 \). For example, assuming \( \gamma_k = 0 \) for \( k > 1 \) corresponds to the dropout probability not depending on the previous outcomes except for the one just prior to the time of dropout. In this one-level dependence model, there are only six parameters in the missing data model, which may help to estimate the parameters from the total of 53 patients who dropped out early (see Figures 2 and 3 for the number of dropout patients by visit). We use SAS PROC MCMC to fit this model and assume vague priors as above for location and covariance parameters (see, e.g., Chen 2013). The MCMC process converges well for these models (some selected trace and posterior density plots are provided in the online supplemental materials). The results from this Bayesian selection model are summarized in Table 4. The results from the primary analysis MMRM model are also included for easy comparisons.

From these analysis models, we can see a few interesting results:

1. The results for estimated treatment difference are fairly similar across the methods from conventional MMRM model to different Bayesian selection models.

2. In the all level selection model, only the coefficient \( \gamma_1 \) is significantly different from 0, which implies that the higher level dependence of missing probability on previous outcomes that are more than one-level away may be ignorable. A positive value for \( \gamma_1 \) implies that a higher outcome (worsening of disease) may lead to higher probability of dropout, which is consistent with the observation that the lack of efficacy is the main cause for discontinuation in this trial.

3. The estimated value for \( \gamma_0 \) is very close to 0 and insignificant, which implies that the dropout probability is not likely dependent on the missing data in this selection model. Therefore, the results are quite similar to that from the MMRM, which is obtained under MAR assumption.

4. The estimated intercept parameters (i.e., \( \zeta_2, \ldots, \zeta_5 \)) are all significantly different from 0. The values are also decreasing from \( \zeta_2 \) to \( \zeta_3 \), and from \( \zeta_3 \) to \( \zeta_4 \); which implies that the dropout probabilities may be different over time. Therefore, the modified model which allows dropout probability depend on time may be more appropriate.

5. Discussions

Although Bayesian methods for missing data exist in the statistical literature for quite a while, their applications in real clinical trials are still not common. This is partially due to the software to fit the Bayesian models that are not available in commonly used packages such as SAS, which is one of the most widely used statistical analysis packages that are considered as validated and accepted by regulatory agencies. Recent advancements in computational algorithms and Bayesian analysis tools such as MCMC with missing data handling become available in new versions of SAS software (SAS/STAT 12.3 or later). This brings opportunities for pharmaceutical statisticians to start using Bayesian methods in analyzing data from clinical trials with missing data.

In this article, we use a real clinical trial as a case study to explore the applications of Bayesian methods in analysis of continuous repeated measures obtained in a longitudinal trial with missing data, specifically in the area of different sensitivity analysis under missing not at random scenarios. We have considered some common sensitivity analysis approaches from frequentist methods, including selection models, shared parameter models, and pattern mixture models. Except for the MMRM models for all available data or complete-case analysis, these models allow for additional specifications for missing data under missing not at random. For each of those methods, we applied a corresponding Bayesian approach for the analysis. In most of the cases, the results from Bayesian methods with vague (or noninformative) prior were similar to those from the corresponding frequentist methods. This is helpful for applied statisticians to consider Bayesian
A Comparison of Frequentist and Bayesian Model Based Approaches

approach with vague (or noninformative) prior as a computational tool to fit complicated statistical models.

Most of the frequentist sensitivity analysis methods require special software to implement. For selection models and shared parameter models, we have used the SAS macros developed by DIA missing data working group (available at http://missingdata.lshtm.ac.uk). For pattern mixture models, we applied the new multiple imputation procedure in SAS 9.4. In general, these available macros or software have some restrictions that may preclude applied statisticians to apply the methods in real clinical trials. For example, the SAS macros developed by DIA missing data working group can only handle two treatment groups. This may produce different results as modeling all data together in a study with more than two treatment groups. The case study illustrated that Bayesian approach can be an alternative to fit these sensitivity analysis models. Under the noninformative (or vague conjugate) prior, the results would be similar to the corresponding frequentist methods. One advantage to use Bayesian approach is that it can be more flexible for users to implement different models by using Markov chain Monte Carlo and obtain the posterior distribution for the parameters of interest. Some of the sample analysis codes are made available in the online supplemental materials.

For this case study, the conclusions from different methods were consistent although the estimated treatment effects could be different. The consistent conclusions demonstrate some robustness of the analysis results for this study. However, we should notice some limitations of this case study dataset. Because of relatively small sample sizes, both test drug and active groups showed insignificant results compared to placebo across the methods. Therefore, we assess the sensitivity of the analysis results from checking the differences of the estimated treatment effect and corresponding standard error. Compared to the MMRM analysis, the selection model, shared parameter model, and pattern mixture model with ACMV all produced somewhat different results. Each of the sensitivity analysis models made some specific assumptions on the missing data. Based on the plots of mean response by treatment group and missing data pattern (i.e., time of dropout), we would anticipate that MCAR is unlikely and a linear or quadratic functional form may not fit the profile well. Hence, we are less concerned about the different results from the completers or shared parameter models. Among the selection models, the model with different parameters for treatment groups but no difference across visit may pose some unrealistic restriction. The small number of dropout patients in some visits may also cause excessive variation for the estimates from NCMV. To overcome some of these concerns, we considered a selection model with modified parameterization which allows different parameters over time but the same across treatment groups. The estimated treatment effect and its standard error from this modified selection model were similar to that of the original planned mixed model analyses. The analysis indicated that a worsening of response might have led to a higher probability of dropout, which is consistent with the observation that the lack of efficacy was the main cause for discontinuation in this trial.

Although software for Bayesian methods allows more flexibility for model specification, it is always important to understand the assumptions and interpretation for each parameterization. The convergence of MCMC algorithm is another important aspect that needs attention since models like selection models may involve many parameters that can cause slow mixture in the MCMC process. It is essential to check convergence using trace plots and/or some diagnostic measures.

In this article, we only consider the Bayesian application on a few common sensitivity analysis approaches using the case study dataset. There are other sensitivity analyses methods proposed in the recent publications, including control-based imputation models and tipping point analysis. Readers may refer to recent papers by Carpenter and Kenward (2007), Carpenter, Roger, and Kenward (2013), Ratitch, O’Kelly, and Tosiello (2013), Mallinckrodt et al. (2013), and Teshome et al. (2014). Their corresponding Bayesian approaches could be topics for future research. In the analysis for this case study dataset, we mostly used noninformation or vague prior for all model parameters. Further discussions of sensitivity analysis and using informative priors in Bayesian models are out of scope for this article and can refer to Daniels and Hogan (2008) and missing data handbook (2014).

Appendix: Bayesian Model Specification for Pattern Mixture Model Under ACMV

For the case study, the first two patterns are combined due to very few patients in first pattern so that we only have four patterns. The first pattern contains patients who dropped out at visit 2 or 3, second and third patterns contain patients who dropped out at visit 4 or 5, respectively; and pattern 4 are completers. For each pattern, there will be nonidentifiable elements (i.e., model parameters for missing observations). The nonidentified components are marked by *. The complete model is specified as follows for each treatment:

\[ Y_1|S = k \sim N(\mu_k + \gamma_{1,bl}, \sigma^{(k)}_1), \quad k = 1, \ldots, 4 \]

\[ *Y_2|Y_1, S = 1 \sim N(\alpha_0^{(1)} + \alpha_1^{(1)}Y_1 + \gamma_{2,bl}, \tau_2^{(1)}) \]
\[ Y_2|Y_1, S = k \sim N \left( \alpha_{0}^{(k)} + \alpha_{1}^{(k)} Y_1 + \gamma_2 b_l, \tau_2^{(k)} \right), \]
\[ k = 2, \ldots, 4 \]
\[ *Y_3|Y_1, Y_2, S = 1 \sim N \left( \beta_0^{(1)} + \beta_1^{(1)} Y_1 + \beta_2^{(1)} Y_2 + \gamma_3 b_l, \tau_3^{(1)} \right), k = 1 \]
\[ Y_3|Y_1, Y_2, S = k \sim N \left( \beta_0^{(k)} + \beta_1^{(k)} Y_1 + \beta_2^{(k)} Y_2 + \gamma_3 b_l, \tau_3^{(k)} \right), k = 2, \ldots, 4 \]
\[ *Y_4|Y_1, Y_2, Y_3 = 1 \sim N \left( \eta^{(1)}_0 + \eta^{(1)}_1 Y_1 + \eta^{(1)}_2 Y_2 + \eta^{(1)}_3 Y_3 + \gamma_4 b_l, \tau_4^{(1)} \right), k = 1, 2 \]
\[ Y_4|Y_1, Y_2, Y_3, S = k \sim N \left( \eta^{(k)}_0 + \eta^{(k)}_1 Y_1 + \eta^{(k)}_2 Y_2 + \eta^{(k)}_3 Y_3 + \gamma_4 b_l, \tau_4^{(k)} \right), k = 3, 4 \]
\[ *Y_5|Y_1, Y_2, Y_3 Y_4 = 4 \sim N \left( \zeta_0^{(4)} + \zeta_1^{(4)} Y_1 + \zeta_2^{(4)} Y_2 + \zeta_3^{(4)} Y_3 + \zeta_4^{(4)} Y_4 + \gamma_5 b_l, \tau_5^{(4)} \right) \]
\[ S \sim \text{Mult}(\phi). \]

The multinomial parameter is \( \phi = (\phi_1, \phi_2, \phi_3, \phi_4) \) with \( \phi_s = P(S = s) \) for \( s \in \{1, 2, \ldots, 4\} \) and \( \sum \phi_s = 1 \). With ACMV, we set the following identifying restrictions:

\[ \alpha^{(1)}_0 = \alpha^{(2)}_0, \quad \alpha^{(1)}_1 = \alpha^{(2)}_1, \quad \beta^{(1)}_0 = \beta^{(2)}_0, \quad \beta^{(1)}_1 = \beta^{(2)}_1, \quad \eta^{(1)}_l = \eta^{(2)}_l, \quad l = 0, \ldots, 3, k = 1, 2 \]
\[ \zeta^{(1)}_l = \zeta^{(2)}_l, \quad l = 0, \ldots, 4, k = 1, 2, 3 \]
\[ \tau_2^{(1)} = \tau_2^{(2)}, \quad \tau_3^{(1)} = \tau_3^{(2)}, \quad \tau_4^{(1)} = \tau_4^{(2)}, \quad \tau_5^{(1)} = \tau_5^{(2)}, \quad k = 1, 2, 3 \]

**Supplementary Materials**

The supplementary materials contain sample codes, and selected trace and density plots.

**Acknowledgments**

We would like to thank Prof. Michael Daniels for discussions and providing valuable inputs during the development of this article. We also thank a referee for providing many helpful comments that led us to improve the contents and presentation of the article.

[Received November 2014. Revised May 2015]

**References**

Best, N. G., Spiegelhalter, D. J., Thomas, A., and Brayne, C. E. G. (1996), “Bayesian Analysis of Realistically Complex Models.” *Journal of the Royal Statistical Society, Series A*, 159, 323–342. [117]

Carpenter, J. R., and Kenward, M. G. (2007), *Missing Data in Randomised Controlled Trials—A Practical Guide*. Available at [http://missingdata.lshtm.ac.uk/downloads/rm04_jh17_mk.pdf](http://missingdata.lshtm.ac.uk/downloads/rm04_jh17_mk.pdf). [116,125]

Carpenter, J. R., Pocock, S., and Lamm, C. J. (2002), “Coping with Missing Data in Clinical Trials: A Model-Based Approach Applied to Asthma Trials,” *Statistics in Medicine*, 21, 1043–1066. [117]

Carpenter, J. R., Roger, J., and Kenward, M. (2013), “Analysis of Longitudinal Trials With Protocol Deviation: A Framework for Relevant, Accessible Assumptions, and Inference Via Multiple Imputation,” *Journal of Biopharmaceutical Statistics*, 23, 1352–1371. [118,125]

Chen, F. (2013), “Missing No More: Using the MCMC Procedure to Model Missing Data,” *Proceedings for SAS Global Forum*. Available at [http://support.sas.com/resources/papers/proceedings13/436-2013.pdf](http://support.sas.com/resources/papers/proceedings13/436-2013.pdf). [124]

Chen, Q., and Ibrahim, J. G. (2013), “A Note on the Relationships Between Multiple Imputation, Maximum Likelihood and Fully Bayesian Methods for Missing Responses in Linear Regression Models,” *Statistics and Its Interface*, 6, 315–324. [117]

Daniels, M. J., and Hogan, J. W. (2008), *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*, Boca Raton, FL: Chapman & Hall/CRC. [117,125]

Diggle, P., and Kenward, M. G. (1994), “Informative Dropout in Longitudinal Data Analysis” (with discussion), *Applied Statistics*, 43, 49–94. [121]

European Medicines Agency (2010), *Guideline on Missing Data in Confirmatory Clinical Trials*, EMA/CPMP/EWP/1776/99 Rev. 1, July 2010. [116]

Han, B., Yu, M., Dignam, J. J., and Rathouz, P. J. (2014), “Bayesian Approach for Flexible Modeling of Semicompeting Risks Data,” *Statistics in Medicine*, 33, 5111–5125. [123]

Gelman, A., and Rubin, D. B. (1992), “Inference From Iterative Simulation Using Multiple Sequences,” *Statistical Science*, 7, 457–472. [123]

Hedeker, D., and Gibbons, R. D. (2006), *Longitudinal Data Analysis*, New York: Wiley. [122]

Hogan, J. W., Daniels, M. J., and Hu, L. (2014), “Bayesian Sensitivity Analysis,” in *Handbook of Missing Data Methodology*, eds. G. Molenberghs, G. Fitzmaurice, M. Kenward, A. Tsiatis, and G. Verbeke, New York: Chapman and Hall/CRC. [117]

Ibrahim, J. G., Chen, M. H., Lipsitz, S. R., and Herring, A. H. (2005), “Missing-Data Methods for Generalized Linear Models: A Comparative Review,” *Journal of the American Statistical Association*, 100, 332–346. [117]

Little, R. J. (1993), “Pattern-Mixture Models for Multivariate Incomplete Data,” *Journal of the American Statistical Association*, 88, 125–134. [121]
A Comparison of Frequentist and Bayesian Model Based Approaches

Little, R. J., and Rubin, D. B. (2002), Statistical Analysis with Missing Data (2nd ed.), New York: Wiley. [117]

Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. (2009), “The BUGS Project: Evolution, Critique and Future Directions,” Statistics in Medicine, 28, 3049–3067. [118]

Mallinckrodt, C., Lane, P., Schnell, D., Peng, Y., and Mancuso, J. (2008), “Recommendations for the Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials,” Drug Information Journal, 42, 303–319. [119]

Mallinckrodt, C., Roger, J., Chuang-Stein, C., et al. (2013), “Missing Data: Turning Guidance Into Action,” Statistics in Biopharmaceutical Research, 5, 369–382. [120,125]

Molenberghs, G., and Kenward, M. G. (2007), Missing Data in Clinical Studies (1st ed.), New York: Wiley. [117]

Molenberghs, G., Michiels, B., Kenward, M. G., and Diggle, P. J. (1998), “Missing Data Mechanisms and Pattern-Mixture Models,” Statistica Neerlandica, 52, 153–161. [122]

National Research Council (2010), The Prevention and Treatment of Missing Data in Clinical Trials, Panel on Handling Missing Data in Clinical Trials, Committee on National Statistics, Division of Behavioral and Social Sciences and Education, Washington, DC: The National Academies Press. [116]

Ratitch, B., O’Kelly, M., and Tosiello, R. (2013), “Missing Data in Clinical Trials: From Clinical Assumptions to Statistical Analysis Using Pattern Mixture Models,” Pharmaceutical Statistics, 12, 337–347. [125]

SAS/STAT 13.2 (2014), User’s Guide, Cary, NC: SAS Institute Inc. [118]

Stan Development Team (2012), A C++ Library for Probability and Sampling, Version 1.0. http://mc-stan.org/ 2012. [118]

Sterne, J. A. C., Carlin, J. B., Spratt, M., Kenward, M. G., Wood, A. M., and Carpenter, J. R. (2009), “Multiple Imputation for Missing Data in Epidemiological and Clinical Research: Potential and Pitfalls,” British Medical Journal, 338, 157–160. [117]

Teshome, A. B., Lipkovich, I., Molenberghs, G., and Mallinckrodt, C. H. (2014). A Multiple-Imputation-Based Approach to Sensitivity Analyses and Effectiveness Assessments in Longitudinal Clinical Trials,” Journal of Biopharma Statistics, 24, 211–228. [125]

About the Author

G. Frank Liu, corresponding author, Merck Sharp & Dohme Corp., UG1CD-38, P.O. Box 1000 351 Sumneytown Pike, North Wales, PA 19454 (E-mail: guanghan_frank_liu@merck.com). Baoguang Han, Biogen, Cambridge, MA 02142 (E-mail: baoguang.han@biogenidec.com). Xin Zhao is Director of Biostatistics, J & J Janssen, San Francisco, CA 94080 (E-mail: xzhao2@its.jnj.com). Qun Lin, Eli Lilly and Company, Indianapolis, IN 46285 (E-mail: lin.qun@lilly.com).