

Applications of Simulation for Missing Data Issues in Longitudinal Clinical Trials

G. Frank Liu and James Kost

Abstract Missing data handling in longitudinal clinical trials has gained considerable interest in recent years. Although a lot of research has been devoted to statistical methods for missing data, there is no universally best approach for analysis. It is often recommended to perform sensitivity analyses under different assumptions to assess the robustness of the analysis results from a clinical trial. To evaluate and implement statistical analysis models for missing data, Monte-Carlo simulations are often used. In this chapter, we present a few simulation-based approaches related to missing data issues in longitudinal clinical trials. First, a simulation-based approach is developed for generating monotone missing data under a variety of missing data mechanism, which allows users to specify the expected proportion of missing data at each longitudinal time point. Secondly, we consider a few simulation-based approaches to implement some recently proposed sensitivity analysis methods such as control-based imputation and tipping point analysis. Specifically, we apply a delta-adjustment approach to account for the potential difference in the estimated treatment effects between the mixed model (typically used as the primary model in clinical trials) and the multiple imputation model used to facilitate the tipping point analysis. We also present a Bayesian Markov chain Monte-Carlo method for control-based imputation which provides a more appropriate variance estimate than conventional multiple imputation. Computation programs for these methods are implemented and available in SAS.

1 Introduction

Handling missing data in longitudinal clinical trials has gained considerable interest in recent years among academic, industry, and regulatory statisticians alike. Although substantial research has been devoted to statistical methods for missing data, no single analytic approach has been accepted as universally optimal. A common recommendation is to simply conduct sensitivity analyses under different assumptions to assess

G.F. Liu (✉) · J. Kost
Merck & Co. Inc., 351 N. Sumneytown Pike, North Wales, PA, USA
e-mail: Guanghan_frank_liu@merck.com

© Springer Nature Singapore Pte Ltd. 2017
D.-G. Chen and J.D. Chen (eds.), *Monte-Carlo Simulation-Based Statistical Modeling*, ICSA Book Series in Statistics, DOI 10.1007/978-981-10-3307-0_11

211

the robustness of the analysis results from a clinical trial. For decades, the gold standard for longitudinal clinical trials has been using mixed-models for repeated measures (MMRM, see e.g., Mallinckrodt et al. 2008). However, the MMRM analysis requires the assumption that all missing data are missing at random (MAR), an assumption which cannot be verified, and might even be considered as unlikely for some study designs and populations. The current expectation is that regulatory agencies will require sensitivity analyses to be conducted to evaluate the robustness of the analytic results to different missing data assumptions (European Medicines Agency 2010; National Academy of Sciences 2010). Clinical trial statisticians are thus well-advised to understand the ramifications that various missing data mechanisms (MDMs, see e.g., Little and Rubin 1987) have on their proposed analyses, most notably on bias, type I error control, and power. To that end, Monte-Carlo simulations are often used to conduct trial simulations under different MDMs for evaluating statistical analysis models.

In this chapter, we present three simulation-based approaches related to missing data issues in longitudinal clinical trials. First, a simulation-based approach is developed for generating monotone missing multivariate normal (MVN) data under a variety of MDMs (i.e., missing-completely-at-random [MCAR], MAR, and missing-not-at-random [MNAR]), which allows users to specify the expected proportion of missing data at each longitudinal time point. Second, a simulation-based approach is used to implement a recently proposed “tipping-point” sensitivity analysis method. Specifically, a delta-adjustment is applied to account for the potential difference in the estimated treatment effects between the mixed model (typically used as the primary model in clinical trials) and the multiple imputation model used to facilitate the tipping point analysis. Last, a Bayesian Markov chain Monte-Carlo (MCMC) method for control-based imputation is considered that provides a more appropriate variance estimate than conventional multiple imputation. Computation programs for some of these methods are made available in SAS.

In practice, there are two types of missing data. The first type is *intermittent missing* where the missing data of a subject is followed by at least one timepoint at which data are observed. The second type is *monotone missing*, which is typically caused by study attrition (i.e., early drop out of a subject). Common reasons for intermittent missing data include missed subject-visits, data collection errors, or data processing (e.g., laboratory) errors. Because these errors are unlikely to be related to the value of the data itself (had that value been observed), an assumption that the intermittent missing data are MAR, or even MCAR, is often appropriate. Therefore, it may be considered reasonable to first impute the intermittent missing data under MAR before performing the analysis (e.g., Chap.4, O’Kelly and Ratitch 2014). With this consideration, the discussion that follows focuses primarily on monotone missing data. Although caution should be exercised because intermittent data might be MNAR for some studies in which the disease condition is expected to fluctuate over time.

2 Generation of Study Data with a Specified MDM and Cumulative Drop-Out Rates

Whether the clinical trial statistician wants to investigate novel approaches in the analysis of missing data or simply wants to compute power for an upcoming study, it is often useful to generate MVN data (given mean μ and covariance matrix Σ) under a specific MDM, with specified expected cumulative drop-out rates at each longitudinal timepoint. This section presents a method for generating monotone missing data, with a simple process outlining how to add intermittent (i.e., nonmonotone) missing data provided at the end of this section.

It is of interest to generate longitudinal MVN data, given μ and Σ , with expected cumulative drop-out rates (CDRs) over time under a given monotone MDM. For the MCAR MDM, this step is easily accomplished by first generating a subject-specific $U(0,1)$ random-variate, and then comparing that variate to the target CDR at each timepoint. Starting with the first postdose timepoint, if the random variate is smaller than the target CDR, then that subject can be considered as having dropped out, with the data at that timepoint and all subsequent timepoints set to missing.

The approach for the MAR and MNAR MDMs is more complicated because the missingness for these MDMs depends on the data itself. Analytic or closed-form solutions are not yet available for a general MDM specification. As opposed to the unconditional approach used for the MCAR MDM, the subjects who have already dropped out need to be accounted for. Specifically, the conditional probability needs to be calculated for each individual who will drop out at Time t given that the subject is still in the study at Time $t-1$. Defining CDR_t , $t = 1$ to T , as the desired postdose expected cumulative droprate at Time t , this conditional probability is expressed as $(CDR_t - CDR_{t-1}) / (1 - CDR_{t-1})$. For the purposes of this chapter, the baseline time point is assumed to be nonmissing (i.e., $CDR_0 = 0$).

Let $Y_{ijk} \sim N(\mu_k, \Sigma)$, with $t = 0$ to T , $j = 1$ to n , and $k = 1$ to K , for T total timepoints, n total observations, and K total groups (e.g., treatment arms). As noted, the baseline measurement, Y_{0jk} , is assumed to be nonmissing. Let p_{ij} represent the estimated conditional probability of dropping out at postdose time t for subject j , conditioned on subject j not having already dropped out. Finally, let Ψ represent one or more tuning parameters governing the effect of Y values on p_{ij} , with the designation of a positive (negative) value of Ψ indicating that higher (lower) values of Y are more likely to result in drop out.

Specifically, a logistic model, $\text{logit}(p_{ij}) = f(Y_j, \alpha, \Psi)$ is considered, with the tuning parameter(s) Ψ pre-specified by users based on the desired MDM. The vector $\alpha = (\alpha_1, \dots, \alpha_t)$ is then estimated based on Monte-Carlo simulations such that the resulting missing data are sufficiently close to the specified CDRs (per the user-defined tolerance parameter ϵ). Without loss of generality, consider the following MDM, which follows a simple MAR process, where missingness at a given timepoint is solely a function of the observation at the previous timepoint (conditioned on the subject having not already dropped out). To simplify notation, the subject indicator j for p_{ij} and y_{ij} is suppressed in the following formulas:

$$\text{logit}(p_t) = \sum_{i=1}^t \alpha_i + \psi y_{t-1}, \quad t = 1, 2, \dots, T. \tag{1}$$

The α_i are solved in a stepwise manner, first solving for α_1 using $\text{logit}(p_1) = \alpha_1 + \psi y_0$, such that p_1 is sufficiently close to CDR_1 .

Next, solve for α_2 using $\text{logit}(p_2) = \hat{\alpha}_1 + \alpha_2 + \psi y_1$, such that p_2 is sufficiently close to $(CDR_2 - CDR_1)/(1 - CDR_1)$, where $\hat{\alpha}_1$ was estimated from the previous step. Prior to iteratively solving for α_2 , it is required to identify and exclude data from those subjects in the simulated dataset that have dropped out. This step is accomplished by comparing a subject-specific (and timepoint-specific) $U(0, 1)$ variate with the subject-specific value of p_1 (which is, in part, a function of the recently solved $\hat{\alpha}_1$). This process is continued through to time T .

Each α_t is solved using a bisectional approach in conjunction with a large random sample drawn from the specified MVN distribution, with convergence for α_t declared when

$$|\hat{p}_t - (CDR_t - CDR_{t-1})/(1 - CDR_{t-1})| < \epsilon,$$

where ϵ is a user-defined convergence criterion and \hat{p}_t is a function of Y, Ψ , and $\hat{\alpha}_t$. The detailed steps for calculating these $\hat{\alpha}_t$ are provided in the section that follows.

General Algorithm to Solve for α

A bisectional approach is used to solve for each α_t sequentially as follows:

0. Generate a large dataset of observations (e.g., 100000), Y , comprised of $Y_k \sim N(\mu_k, \Sigma)$, with the proportion of observations following the distribution of Y_k equal to π_k , as determined by the treatment ratio per the study design.

Do Steps 1–9 for $t = 1, \dots, T$:

1. Initialize $\alpha_L = -10000, \alpha_C = 0, \alpha_U = 10000, DONE = 0, COUNTER = 0$
2. If $t \geq 2$, then simulate the missingness of observations at earlier timepoints by using the previously computed α . Delete any subjects who are simulated as having dropped out.
3. For each remaining observation in Y , compute an estimate of $f(\widehat{\alpha_{ij}})$ ($= f(Y, \alpha, \Psi)$), which is a function of the previously estimated α, α_C , and some function of the y and Ψ (depending on the MDM model).
4. Compute $\hat{p}_{ij} = (1 + \exp(-f(\widehat{\alpha_{ij}})))^{-1}$ for each remaining observation in Y .
5. Compute \hat{p}_t as the mean of the \hat{p}_{ij} .
6. Compute $DIFF = \hat{p}_t - (CDR_t - CDR_{t-1})/(1 - CDR_{t-1})$, a measure of how accurate this guess at $\alpha_t (= \alpha_C)$ is.
7. (a) If $|DIFF| < \epsilon$ then $DONE = 1$ (We are satisfied with α_C).
 (b) else if $DIFF > 0$ then $\alpha_U = \alpha_C$ and $\alpha_C = (\alpha_L + \alpha_C) / 2$ (i.e., search lower).
 (c) else if $DIFF < 0$ then $\alpha_L = \alpha_C$ and $\alpha_C = (\alpha_U + \alpha_C) / 2$ (i.e., search higher).
8. $COUNTER = COUNTER + 1$; If $COUNTER = 50$ then $DONE = 1$. (the value of $COUNTER$ may be adjusted by users to avoid an endless loop due to nonconvergence, though $COUNTER = 50$ is likely sufficient.)
9. If Not $DONE$ then Go to Step 3; if $DONE$, set $\hat{\alpha}_t = \alpha_C$

After the α vector has been estimated, the actual missing data of interest can be simulated by randomly generating the complete MVN dataset (Y), followed by the determination of missingness. Based on α , Ψ , the MDM model, and the randomly generated complete dataset Y , subject-specific cutpoints \hat{p}_{ij} are computed. These cutpoints represent the probability that Subject j will drop out of the study at timepoint t (conditioned on not already having dropped out). For each timepoint, a uniform variate is generated and compared to the appropriate cutpoint to determine whether the subject drops out at that timepoint. The process starts at the first postdose timepoint, and proceeds sequentially up to last time point T . As noted above, this process is actually required in the stepwise generation of the α values themselves (Step 2 in the algorithm above).

We note that the proposed algorithm is set up to handle a mixed distribution, with the CDRs at each timepoint defined over all treatment arms. Of course, the different treatment arms will presumably still have different CDRs as a function of the μ_k . If different defined CDRs are desired for each treatment arm (as opposed to defining the CDRs over all treatment arms and letting the μ_k provide differentiation between the treatment drop rates), then the algorithm would need to be run once for each such arm (or group of arms), yielding a different α vector for each.

In the absence of historical treatment-specific drop-out information, one could take a two-step approach in specifying the CDRs. The first step involves running the algorithm assuming a CDR (perhaps corresponding to placebo data from literature) over all treatment arms and letting the assumed efficacy (i.e., μ_k) drive the treatment-specific drop-out rates. One could then use available safety information on the drug (and placebo) to fine tune those treatment-specific CDRs, thus generating a separate α vector for each treatment group.

The process described above can also be used to generate study data with MNAR MDM. For example, the model $\text{logit}(p_t) = \sum_{i=1}^t \alpha_i + \psi y_{t-1}$ can be replaced with two options: (a) $\text{logit}(p_t) = \sum_{i=1}^t \alpha_i + \psi y_t$, in which the missing probability depends on the missing data; or (b) $\text{logit}(p_t) = \sum_{i=1}^t \alpha_i + \psi_1 y_{t-1} + \psi_2 y_t$, in which p_t depends on both observed y_{t-1} and the missing data y_t . In all these models, the tuning parameters ψ , ψ_1 , and ψ_2 are prespecified by users.

One might also wish to add intermittent (i.e., nonmonotone) missing data. This can be accomplished by generating a vector (one for each postdose timepoint) of independent uniform variates for each subject and then comparing that vector to cutpoints (timepoint-specific, as desired) corresponding to the missing probability at each timepoint (e.g., 0.01). Presumably, this process would be applied in conjunction with the monotone process, with the monotone process meant to emulate missingness due to subject-drop out and the intermittent process meant to emulate an MCAR process. In such a case, the process used to generate the intermittent missing data is applied independently of the process used to generate monotone missing data, with a given observation designated as missing if either process determines that observation is missing. The overall probability of missing would clearly be higher than the level specified for either process on its own, thus requiring some downward adjustment of the defined drop-out probabilities of one or both processes.

Example: Power and Bias Evaluation for a Longitudinal Study with Missing Data

For sample size and power calculations, analytic approaches are available when missing data are MAR (e.g., Lu et al. 2008, 2009). In general, power loss stemming from missing data in a longitudinal trial depends on the proportion and timing of the missing data; that is, the cumulative drop-out rates (CDRs), as a function of the different effective information yielded from the observations over time. For example, one would expect that a study with drop outs occurring gradually over time would have less power than a study in which all of the drop outs occurred between the second-to-last and the last (presumably primary) timepoint.

Despite the available analytic approaches, power calculations for longitudinal clinical trials are often conducted via simulations, given the extreme flexibility that simulations afford. The use of simulations is especially common in trials with complicating factors such as (a) interim analyses for futility or for overwhelming efficacy, (b) multiplicity approaches covering multiple timepoints/endpoints, or (c) adaptations built into the designs (e.g., dropping an arm or adjusting the randomization ratio as a function of the accruing data). Of course, power calculations can also be simulated for relatively straightforward clinical trial designs.

The following simulation study investigates the effect that different methods of generating random MVN data, primarily with respect to MDMs, have on both the bias of the parameter estimates and the corresponding power calculations. The assumed parameters are based on data obtained from an actual clinical trial. The results from 10,000 simulation runs are summarized in Table 1.

The simulations used the following assumptions (four postdose timepoints):

$$\alpha = 0.050;$$

$$N/\text{per arm} = 120.$$

$$\mu_{\text{Pbo}} = (0.00, 0.46, 0.92, 1.37, 1.83); \mu_{\text{Act}} = (0.00, 0.23, 0.46, 0.69, 0.92);$$

(Higher means represent lower efficacy).

A functional form was used for the variance-covariance matrix, with $\sigma_i = c^i * \sigma_0$, with $\sigma_0 = 0.860$, and $c = 1.288$; $i = 1, \dots, 4$; and $\rho_{ij} = r * b^{|j-i|-1}$, with $r = 0.748$ and $b = 0.832$; $i, j = 0, \dots, 4$.

Notably $\sigma_0^2 = 0.740$, $\sigma_4^2 = 5.616$, and $\rho_{0,4} = 0.431$, with $\text{Var}(Y_4 - Y_0) = 4.6$.

CDR = (0.086, 0.165, 0.236, 0.300), with a 30% CDR at Timepoint 4 (T4).

The following missing data patterns (MDPs) were considered (with $\Psi = \Psi_1 = \Psi_2 = 0.5$).

- MDP0: No missing data.
- MDP1: MCAR
- MDP2: Data are MCAR but only the baseline and last timepoint values are included in the analysis (Completers Analysis).
- MDP3: MAR with $\text{logit}(p_t) = \sum_{i=1}^t \alpha_i + \psi y_{t-1}$
- MDP4: MNAR with $\text{logit}(p_t) = \sum_{i=1}^t \alpha_i + \psi y_t$
- MDP5: Mixture of MAR and MNAR with $\text{logit}(p_t) = \sum_{i=1}^t \alpha_i + \psi_1 y_{t-1} + \psi_2 y_t$

Table 1 Summary of missing data, power, and estimated treatment effects from MMRM at time 4 (10,000 simulations)

MDP	Estimated bias for mean at T4										Power (%)	
	Percentage of missing at T4			Completers			MMRM					
	Drug	Pbo		Drug	Pbo		Drug-Pbo	Drug	Pbo	Drug-Pbo		
MDP0	0.0	0.0		0.00	0.00		0.00	0.00	0.00	0.00	0.00	91.0
MDP1	30.1	30.0		0.00	0.00		0.00	0.00	0.00	0.00	0.00	81.1
MDP2	30.0	30.0		0.00	0.00		0.00	0.00	0.00	0.00	0.00	78.4
MDP3	28.1	31.9		-0.27	-0.33		0.06	0.00	0.00	0.00	0.00	81.1
MDP4	27.0	33.0		-0.41	-0.53		0.08	-0.18	-0.23	0.05	0.05	79.4
MDP5	25.9	33.9		-0.53	-0.73		0.20	-0.17	-0.24	0.07	0.07	79.0

Note MDP = missing data pattern. T4 = time point 4. MMRM = mixed-methods for repeated measures

Although the simulations were conducted by defining a 30% drop-out rate over the two treatment arms, readers should note that the two treatment arms still have different drop-out rates for the MAR and MNAR MDMs. This difference is a result of higher efficacy over time in the drug test group as compared with the placebo group. In the data simulations, we use $\Psi = \Psi_1 = \Psi_2 = 0.5$ such that a higher value (worse) of observed (in MDP3) or unobserved (in MDP4 and MDP5) response will result in a high probability of drop out. This simulates common drop out in clinical trials due to lack of efficacy.

In going from MDP2 to MDP1, a modest gain in power is observed as a result of using the partial data from subjects that dropped out prior to T4. This gain underscores the importance of using the full longitudinal dataset when calculating power, as opposed to considering only the final timepoint.

Focusing on the completers only, clear bias can be observed for the MAR and MNAR scenarios (MDP3 to MDP5); this fact is important to note when powering a study based on simple summary statistics from completers, as is often done when using results from the literature. As expected, the MMRM is unbiased for all of the MCAR scenarios (MDP0 to MDP2), as well as for the MAR scenario (MDP3), since MMRMs assume that all missing data is MAR. Conversely, bias is present in the MMRM analysis for both of the MNAR scenarios (MDP4 and MDP5). Since this bias is to the detriment of the drug under study (i.e., leads to a diluted estimate of the treatment effect), these scenarios result in roughly 2% lower power. Compared with the results from completers, the MMRM analysis had relatively smaller bias.

3 Tipping Point Analysis to Assess the Robustness of MMRM Analyses

As mentioned in Sect. 1, the MMRM, which is often used as the primary analysis model, assumes that all missing data are MAR—an assumption that cannot be verified. Both the clinical trial sponsor and government regulatory agencies are interested in assessing the robustness of any conclusions coming from an MMRM analyses against deviations from the MAR assumption. Given this interest, many methods for sensitivity analysis have been proposed and developed (see e.g., NRC, 2010 and references therein). Some of the more notable proposed methods include selection models, pattern-mixture models, and controlled-imputation models (see, e.g., Carpenter et al. 2013; Mallinckrodt et al. 2013; O’Kelly and Ratitch 2014). Another method, which has recently gained attention is the so-called *tipping point* approach. At a high level, a tipping point analysis varies the imputed values for the missing data (usually for the treatment arms only) by the exact amount needed to make a significant result turn nonsignificant.

Ratitch et al. (2013) have proposed three variations of tipping point analyses using pattern imputation with a delta adjustment. Our discussion considers the variation in which standard multiple imputation is performed first, and then a delta-adjustment

(δ) is applied simultaneously to all imputed values in the treatment group. The goal is to find the smallest δ that will turn the significant p -value (as calculated from the primary MMRM model) to a nonsignificant value. In addition to being relatively straightforward to interpret, this approach has the attractive quality of returning a quantitative result that is directly comparable on the scale of interest, which can then be put into clinical context. The following steps provide details for a bisectional procedure to solve for this tipping point δ .

General Algorithm to Solve for Tipping Point δ

Note the definitions of the following algorithm variables:

- m : the number of imputations to be used in the multiple imputation procedure (a value should be prespecified in the study protocol).
- d : the difference between the maximum and minimum values for the variable/endpoint under investigation (i.e., the maximum allowable shift)
- df : degrees of freedom
- p_{target} : Target probability (e.g., type-I error)
- t_{target} : t -value corresponding to p_{target} . If lower values of Y represent higher levels of efficacy, then this value must be negated in the search algorithm because the target t -value needs to be negative. (Note that the corresponding degrees of freedom (df) are actually a function of the data, as defined in Step 5 below).
- ϵ : A tolerance level, on the t -scale, under which convergence can be declared (e.g. 0.001).
- p_{prim} : p -value from the primary model

Given a dataset with intermittent missing data, the basic algorithm to conduct the tipping-point analysis is outlined below; instructions for procedures conducted in SAS refer to SAS version 9.3 or later.

0. Initialize $\delta_L = -d$, $\delta_C = 0$, $\delta_U = d$, $DONE = 0$, $COUNTER = 0$.
1. Using a Markov Chain Monte-Carlo method (see e.g., Schafer 1997), make the observed dataset monotone-missing. This step can be accomplished for each treatment group using `proc mi` within SAS by using the option `mcmc chain=multiple impute=monotone` in conjunction with all covariates (excluding treatment) included in the primary analysis model. This step will generate m monotone-missing datasets. Note that the study protocol should specify the random seed used in this step.
2. Applying parametric regression to the m monotone-missing datasets, impute the missing values in a stepwise fashion starting with the first postdose timepoint. This step can be accomplished for each treatment group using `proc mi` in SAS using the `monotone reg` option in conjunction with all covariates (excluding treatment) included in the primary analysis model. This step will generate m complete datasets (one imputed dataset for each of the m monotone-missing datasets).

Do Steps 3–9 while Not DONE

3. Subtract δ_C from each of the imputed values of the test drug treatment arms (to the detriment of test drug).
4. Analyze each of the m post-imputation complete datasets using the primary model, obtaining point estimates for the parameter of interest (e.g., mean change-from-baseline treatment difference at the last timepoint) and the associated variance.
5. Using the `proc mianalyze` procedure in SAS, combine the m means and variances from the m analyses to obtain the final test statistic and p -value, t_{δ_C} and p_{δ_C} , respectively (Rubin 1987). The final test statistic $\hat{Q} / (V^{(1/2)})$ is approximately distributed as t_ν , where \hat{Q} is the sample mean of the m point estimates, $V = \hat{U} + (m + 1) (B/m)$, \hat{U} is the sample mean of the m variance estimates, and B is the sample variance of the m point estimates. The degrees of freedom, ν , are computed as follows (Barnard and Rubin 1999): $\nu = [(\nu_1)^{-1} + (\nu_2)^{-1}]^{-1}$, where $\nu_1 = (m - 1) [1 + (\hat{U}/(1 + m^{-1}) B)]^2$ and $\nu_2 = (1 - \gamma) \nu_0(\nu_0 + 1) / (\nu_0 + 3)$, with $\gamma = (1 + m^{-1}) B/V$ and where ν_0 represents the complete-data degrees of freedom.
6. Compute $\text{DIFF} = t_{\delta_C} - t_{\text{target}}$.
7. (a) if $|\text{DIFF}| < \epsilon$ then $\text{DONE} = 1$ (We are satisfied with δ_C)
 - (b) else if $(\text{DIFF} > 0)$ then $\delta_L = \delta_C$; $\delta_C = (\delta_C + \delta_U)/2$; (subtract larger δ)
 - (c) else if $(\text{DIFF} < 0)$ then $\delta_U = \delta_C$; $\delta_C = (\delta_C + \delta_L)/2$; (subtract smaller δ)
8. If Not DONE, then $\text{COUNTER} = \text{COUNTER} + 1$;
9. If $\text{COUNTER} = 50$, then $\text{DONE} = 1$; (guard against endless loop due to non-convergence)

The final δ can be interpreted as the detrimental offset needed to apply to each imputed observation to change a significant result to a nonsignificant result. Confidence in the primary results stem from a large value of δ , relative to (a) the assumed treatment difference, (b) the observed treatment difference per the primary model, and/or (c) a widely accepted clinically meaningful difference. For example, in a trial of an anti-depressant drug in which a clinically meaningful difference might be around 2–3 points, a trial result could be considered as robust if we were to subtract $\delta \geq 3$ points from every imputed value in the treatment arm and still maintain a statistically significant result.

Conventionally, MMRM analysis is based on restricted maximum likelihood while the tipping point methodology is implemented using multiple imputation (MI) analysis. Ideally, applying $\delta = 0$ to the MI analysis would yield the p -value from the MMRM analysis (p_{prim}). More important, setting $p_{\text{target}} = p_{\text{prim}}$ would ideally yield a solution of $\delta = 0$. If the δ obtained does not equal 0, then the value of δ obtained when setting $p_{\text{target}} = \alpha$ will be biased, per the intended interpretation. Unfortunately, simulation results indicate that the above method will not always yield a δ value equal to 0 when setting $p_{\text{target}} = p_{\text{prim}}$. This inconsistency might be due to additional variation in the analysis of multiple imputation as compared to the restricted maximum

likelihood analysis. To overcome this discrepancy, we advise running the above algorithm twice and in the following order: first, with the setting $p_{\text{target}} = p_{\text{prim}}$, and then with the setting $p_{\text{target}} = \alpha$, yielding δ_{prim} and δ_{α} . The final δ is then computed as $\delta = \delta_{\alpha} - \delta_{\text{prim}}$. The value δ_{prim} can be thought of as a calibration factor in going from the MMRM to the MI model, accounting for the methodological differences between the two, as well as for the inherent randomness in the MI process. Simulation results indicate that t -values and p -values arising from the MMRM and MI models ($\delta = 0$) are highly similar, providing reassurance that the δ (i.e., $\delta_{\alpha} - \delta_{\text{prim}}$) obtained using the MI model translates well to the MMRM model.

As might be expected, higher values of m will yield results with greater stability. This stability applies not only to the estimates produced by the MI approach, but also to the adjusted degrees of freedom (df). The adjustment for the df was first proposed by Barnard and Rubin (1999), and has subsequently gained widespread use (e.g., adopted in SAS). The major impetus for the adjustment, as compared to the initial proposal for a df adjustment as cited in Rubin (1987), was to guard against the possibility that the df used for the MI approach would exceed the df present in the original MMRM for the complete data. However, this df adjustment might be very conservative in certain situations, particularly for smaller sample sizes when low numbers of imputations are used. This characteristic of the df adjustment might have the effect of producing abnormally large δ values since the respective t -values from the original MMRM and from the MI approach will be based on different t -distributions. A simple fix is to ensure that a sufficient number of imputations are used in the MI.

The following simulation study investigates (a) the variation of the df for the MI approach at $\delta = 0$ for different values of m , (b) the differences between the MMRM and the MI approach (at $\delta = 0$) for the t -values and p -values, (c) the variation of δ_{prim} for different values of m , and (d) the variation of $\delta = \delta_{\alpha} - \delta_{\text{prim}}$ for various treatment effect sizes and CDRs.

Unless otherwise noted, the assumptions used in the simulation study for Sect. 2 were also used for all simulations in Sect. 3. For ease of interpretability, a simple MAR mechanism (MDP3 from Sect. 2) was assumed for the missing data.

Assessing the variation of the df at $\delta = 0$

Moderate-to-large differences in df between the original MMRM and the MI model could cause convergence issues or unreliable results when attempting to solve for the tipping point δ . Table 2 shows the summary of df s from the MMRM and MI model using 1,000 simulations. For the case of MDP3, the trial had about 80% power with about 30% missing data at Time 4. The tipping point analysis was performed for about 800 simulated cases for which the MMRM results were significant. The df s from the MMRM analyses varied between 150 and 204. However, the df s for MI varied from 7 to 232 for $m = 5$, and from 47 to 199 for $m = 20$. Since the adjusted df are a direct function of the data itself, it is challenging to provide absolute general guidance as to how many imputations, m , are enough. For the scenario considered in this section, it appears as though $m = 100$ is sufficient to have relatively small variation for the df under consideration.

Table 2 Summary of Variation for the df under MI approach (at $\delta = 0$) by m (based on 1,000 simulations)

m	# Sig	df			
		MMRM		MI	
		Mean	p00, p25, p50, p75, p100	Mean	p00, p25, p50, p75, p100
5	800	179	150, 174, 179, 184, 201	63	7, 27, 46, 84, 232
20	800	179	154, 173, 179, 184, 204	113	47, 95, 111, 129, 199
50	811	179	153, 174, 179, 184, 197	144	91, 133, 144, 155, 195
100	825	179	151, 173, 179, 185, 203	158	122, 150, 158, 166, 194
250	777	179	154, 174, 179, 184, 201	168	137, 162, 168, 174, 192

Note MMRM = mixed-methods for repeated measures. MI = multiple imputation

Table 3 Summary of differences between the MMRM and the MI approach (at $\delta = 0$) for the t -values and p -values by m (based on 1,000 simulations)

m	# sig	t -value difference (MI-MMRM)				p -value difference (MI-MMRM)			
		Mean	p00, p25, p50, p75, p100	Mean	p00, p25, p50, p75, p100				
5	800	-0.05	-1.13, -0.30, -0.06, 0.19, 1.23	0.005	-0.034, 0.000, 0.000, 0.004, 0.114				
20	800	-0.03	-0.59, -0.14, -0.03, 0.08, 0.58	0.000	-0.018, 0.000, 0.000, 0.000, 0.037				
50	811	-0.01	-0.53, -0.07, -0.01, 0.06, 0.37	0.000	-0.015, 0.000, 0.000, 0.000, 0.027				
100	825	0.01	-0.26, -0.04, 0.01, 0.06, 0.30	0.000	-0.010, 0.000, 0.000, 0.000, 0.017				
250	777	0.01	-0.19, -0.02, 0.01, 0.04, 0.17	0.000	-0.011, 0.000, 0.000, 0.000, 0.000				

Note MI = multiple imputation. MMRM = mixed-methods for repeated measures

Assessing the variation of δ_{prim}

As discussed above, an adjustment (δ_{prim}) is needed to account for the differences between the primary MMRM and the MI analysis used in the tipping point procedure. The following simulation study examines the variation of δ_{prim} as well as the differences of the t -values and p -values between the primary MMRM and the MI approach with $\delta = 0$ for various values of m . The tolerance level for convergence of the t -values was set at $\epsilon = 0.005$.

The simulation results in Table 3 indicate that the differences of both the t -values and the p -values between the MMRM and the MI model at $\delta = 0$ are typically small, particularly for $m \geq 100$. This finding provides general confidence that the MI model adequately approximates the MMRM.

Due to the extensive computation required to estimate δ_{prim} , only 100 simulations were conducted to investigate the variation of δ_{prim} as a function of m . Table 4 indicates that the variation of δ_{prim} generally decreases as m increases, with the percentiles generally shrinking to values closer to 0 as m increases. However, this trend cannot be expected to continue as $m \rightarrow \infty$, because some differences due to methodology will persist. For the examined scenario, no clear improvement was seen moving from $m = 100$ to $m = 250$.

Table 4 Summary of δ_{prim} by m (based on 100 simulations)

m	# sig	δ_{prim}	
		Mean	p00, p25, p50, p75, p100
5	81	0.48	-0.93, -0.26, 0.30, 0.95, 3.67
20	79	0.20	-0.34, -0.08, 0.19, 0.46, 1.02
50	79	0.17	-0.30, -0.09, 0.21, 0.40, 0.79
100	81	0.14	-0.16, -0.05, 0.16, 0.32, 0.56
250	85	0.18	-0.16, -0.04, 0.21, 0.37, 0.79

Assessing the Variation of the Final Tipping Point $\delta = \delta_{\alpha} - \delta_{\text{prim}}$

The effect of various treatment differences and CDRs on the distribution of δ was examined using the same simulation assumptions as before, but fixing $m = 100$. Note that $\mu_{\text{Pbo}} = (0.00, 0.46, 0.92, 1.37, 1.83)$ is held constant, while μ_{Act} is set equal to $(1 - \theta)\mu_{\text{Pbo}}$, $\theta = 0.35$ and 0.50 , with larger values of θ resulting in larger efficacy (since higher values of μ represent lower efficacy). Cumulative drop-out rates of (0.054, 0.106, 0.154, 0.200) and (0.086, 0.165, 0.236, 0.300) were considered.

As shown in Table 4, the offset δ_{prim} needed to align the results between the MMRM and the MI model is non-ignorable across the scenarios. Focusing on the scenario with a 30% CDR at T4 and $\theta = 0.50$, we note that the mean value of δ_{prim} needed to calibrate the two models was estimated as 0.14, with observed values ranging from -0.16 to 0.56. As a frame of reference, the true treatment difference at T4 is $(1 - \theta)\mu_{\text{Pbo}} - \mu_{\text{Pbo}} = -0.5(1.83) = -0.92$.

Staying with the same scenario, the mean value of δ is equal to -1.34. That is, on average, all of the missing values in the treatment arm would need to be detrimentally adjusted 1.34 points in order to make the significant p -value obtained become non-significant (i.e., equal to 0.05). Assuming these results were obtained for a single study, and in the context of an observed (or assumed) treatment difference of -0.92, such a δ can be considered as evidence of a fairly robust treatment effect.

Conclusions across scenarios are best drawn by focusing on the estimated mean and quartiles (as opposed to the more variable quantities of the simulated minimum and maximum values). As expected, Table 5 demonstrates that larger detrimental values have to be applied to the imputed data from the treatment arms as (a) the drop-out rate goes down and (b) the true treatment effect goes up.

Without going into great detail, one technical point bears mentioning. When applying δ to the imputed values, it seems reasonable to not allow adjusted values past the minimum or maximum allowable value of the endpoint. However, this restriction might need to be relaxed when applying the convergence algorithm.

Table 5 Summary of Variation of δ_{prim} and Final Tipping Point δ (based on 100 simulations)

CDR at T4 (%)	θ	Number sig.	δ_{prim}		δ	
			Mean	p00, p25, p50, p75, p100	Mean	p00, p25, p50, p75, p100
20	0.35	57	0.11	-0.19, -0.05, -0.03, 0.28, 1.02	-1.41	-4.41, -1.85, -1.13, -0.70, -0.01
20	0.50	83	0.12	-0.18, -0.05, -0.01, 0.35, 0.56	-2.21	-6.68, -3.24, -1.97, -1.19, -0.11
30	0.35	52	0.16	-0.17, -0.04, 0.22, 0.30, 0.70	-1.00	-4.34, -1.53, -0.80, -0.36, -0.01
30	0.50	81	0.14	-0.16, -0.05, 0.16, 0.32, 0.56	-1.34	-3.69, -1.88, -1.32, -0.69, -0.02

Note CDR = cumulative drop-out rates

4 Monte-Carlo Approaches for Control-Based Imputation Analysis

Control-based imputation (CBI) has recently been proposed as an approach for sensitivity analysis (Carpenter et al. 2013), in which different imputation methods are used for the treatment and control groups. The missing data in the control group are imputed under the assumption of MAR, while the missing data in the treatment group are imputed using the imputation model built from the control group. One of the primary assumptions in this CBI approach is that the true post-discontinuation efficacy response in the test drug group is similar to the efficacy response of those subjects continuing in the trial in the control group. This control-based imputation model might be reasonable when no rescue or other active medications are taken by patients who drop out (Mallinckrodt et al. 2013). In general, this CBI can provide a conservative estimate of the treatment effect in superiority trials. Recently, these methods have become more attractive because the assumptions are transparent and understandable for clinical trial scientists. The methods address an attributable treatment effect (estimand) under the intent-to-treat principle but exclude the potential confounding effect of rescue medications (Mallinckrodt et al. 2013). Thus, the estimand captures the causal-effect outcomes for the test therapy.

The three most commonly used CBI methods (Carpenter et al. 2013) are defined by specifying the mean profile after drop out in the treatment group using the profile in the control group as follows:

- I. **Copy Increments in Reference (CIR)**: The increment mean change from the time of drop out for a patient in the treatment group will be the same as the increment mean change for a patient in the control group. Namely, the mean profile after drop out for the treatment group will be parallel to the mean profile of the control group.
- II. **Jump to Reference (J2R)**: The mean profile after drop out for the test drug group will equal the mean profile of the control group. That is, the mean profile for the test drug group has a ‘jump’ from the mean of test drug before drop out to the mean of control after drop out.
- III. **Copy Reference (CR)**: the mean profile for a drop-out patient in test drug group will equal the mean profile for the control group for all time points, including the time points before drop out.

These CBI approaches can be implemented using multiple imputation. Several SAS macros to implement this methodology have been developed by the Drug Information Association (DIA) Missing Data Working Group (macros are available at www.missingdata.org.uk).

Consider a response vector for patient i , $Y_i = \{Y_{ij}, j = 1, \dots, t\}$, and assume

$$Y_i | X_i \sim N(\mu_i, \Sigma).$$

Let μ_{ij} represent the mean for patient i at time j , with the MMRM specified as

$$\mu_{ij} = \alpha_j + \beta_j D_i + \gamma_j' X_{i.}, \tag{2}$$

where β_j is the mean treatment difference from control at time j after adjusting for the covariates X_i , D_i is an indicator for treatment (1 for treatment and 0 for control), and γ_j is a vector of coefficients for the covariates. The following steps can be used to implement the CBI analysis,

1. Fit the MMRM thus yielding the estimates $\hat{\alpha}_j, \hat{\beta}_j, \hat{\gamma}_j$ and $\hat{\Sigma}$;
2. Assume non-informative priors for the parameters, and draw a sample for these parameters from their posterior distribution, denoted by $\alpha_j, \beta_j, \gamma_j$ and Σ . Note that the DIA Missing Data Working Group macros used SAS PROC MCMC to fit the MMRM model and draw these parameters.
3. For a patient who dropped out at time j , draw a sample from the conditional distribution to impute the missing vector, i.e.,
- 4.

$$y_{\text{mis}} | y_{\text{obs}}, X, \mu, \Sigma \sim N(\mu_m + \Sigma_{\text{mo}} \Sigma_{\text{oo}}^{-1} (y_{\text{obs}} - \mu_o), \Sigma_{\text{mm}} - \Sigma_{\text{mo}} \Sigma_{\text{oo}}^{-1} \Sigma_{\text{om}}) \tag{3}$$

where

$$\mu = \begin{pmatrix} \mu_o \\ \mu_m \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{\text{oo}} & \Sigma_{\text{om}} \\ \Sigma_{\text{mo}} & \Sigma_{\text{mm}} \end{pmatrix}$$

are split into sub-vectors and block matrices with dimensions corresponding to the observed (indicated with ‘o’) and missing data (indicated with m) portions of the response vector. The patient and time indicators i and j are omitted in the formulas for simplicity. To implement the CBI, if a patient is in placebo group, the μ_o and μ_m for the placebo group will be used. Otherwise, the means will be modified as specified per the chosen CBI approach. Specifically, a patient in the treatment group who dropped out after time j will have

a.

$$\mu_m^{\text{d}} = \begin{cases} \mu_m^{\text{p}} + \mu_j^{\text{d}} - \mu_j^{\text{p}} & \text{for CIR} \\ \mu_m^{\text{p}} & \text{for JR} \\ \mu_m^{\text{p}} + \Sigma_{\text{mo}} \Sigma_{\text{oo}}^{-1} (\mu_o^{\text{d}} - \mu_o^{\text{p}}) & \text{for CR} \end{cases} \tag{4}$$

b. where the superscripts d and p are used to indicate the mean vector for drug or placebo.

5. Repeat Steps 2 and 3 for the number of imputed datasets;
6. Analyze each imputed dataset using the primary model (e.g., ANCOVA model at last visit) to get estimated treatment difference and its standard error;
7. Combine the results using Rubin’s rule for final statistical inference (Rubin 1987).

Simulation studies show the estimated variances for the treatment differences using the regular MI techniques are always larger than the corresponding empirical variances. This phenomenon has been noticed for the copy-reference imputation method by Lu (2014) and Ayele et al. (2014). Lu (2014) proposed an analytical approach to get the correct variance estimate using the delta method. However, that approach is available only for copy-reference imputation and requires special programming for each specific analysis model.

Liu and Pang (2015) proposed methods to get more appropriate variances for the CBI estimates. One of their approaches is a Bayesian MCMC-based method that accounts for the pattern of missing data and obtains the estimates for the treatment difference and its variance from empirical MCMC samples. Based on the mean profile specified in equation (3), the overall treatment difference at the last time point under CBI can be written as a weighted average over the missing data patterns, that is

$$\theta^{CBI} = \sum_{j=1}^t \pi_j \mu_{ij}^d - \mu_i^p$$

where μ_{ij}^d is the mean at last time point t under missing data pattern j as given in (3) and the $\{\pi_j, j = 1, \dots, t\}$ are the proportions of patients in the missing data patterns for the drug group. As $\sum_{j=1}^t \pi_j = 1$, we have

$$\theta^{CBI} = \begin{cases} \sum_{j=1}^t \pi_j (\mu_j^d - \mu_j^p) & \text{for CIR} \\ \pi_t (\mu_t^d - \mu_t^p) & \text{for J2R} \\ \sum_{j=1}^t \pi_j (\mu_{ij}^d - \mu_i^p) & \text{for CR} \end{cases} \tag{5}$$

where μ_j^d and μ_j^p are the means at time j for drug and placebo, respectively. Therefore, the treatment effect under CBI can be expressed as a linear combination of the parameters of MMRM and the proportions of patients in each pattern of missing data. We note this approach is a special pattern-mixture model (PMM) where the missing data are handled differently by the pattern of missing data only for the treatment group. The missing data in the placebo group are all sampled assuming a MAR process.

To account for the uncertainty of the proportions of missing data $\{\pi_j, j = 1, \dots, t\}$, random proportions are also drawn from a Dirichlet distribution in the Bayesian MCMC process, which corresponds to a posterior distribution for the proportions with a Jefferys prior. The empirical distribution and statistical inference for θ^{CBI} are obtained from the MCMC samples. Specifically, the following steps are implemented in the process:

1. Specify flat priors for $\alpha_j, \beta_j, \gamma_j$ and Σ , for example, use $N(0, \sigma, 2 = 10000)$ for α_j, β_j , and each of the element of γ_j , and an inverse Wishart distribution $IW(t + I, I)$, where I is an identity matrix of dimension t . We used SAS PROC MCMC to fit the MMRM model, specifically:

- a. Use conjugate sampling to draw a sample for Σ ;
 - b. Use random walk Metropolis to draw samples for $\alpha_j, \beta_j, \gamma_j$;
 - c. For a patient who dropped out at time j , PROC MCMC will draw a sample for the missing vector \mathbf{y}_{mis} from conditional distribution as specified in (2) with the parameters from above;
 - d. Draw $\{\pi_j, j = 1, \dots, t\}$ from the Dirichlet $(m_1 + 0.5, \dots, m_t + 0.5)$, where m_j is the number of drop-out patients at time $j+1$ in treatment group, and m_t is the number of completers;
 - e. Evaluate θ^{CBI} with the formula (4).
2. The above process will be run with a burn-in, then repeat Steps a through e to obtain an empirical posterior distribution for θ^{CBI} .

Note that this Bayesian MCMC process is a simulation based approach. It is important to check the convergence of the MCMC samples. Usually, the trace-plot can be examined visually, or some statistical measures can be checked such as Geweke or Raftery-Lewis that are provided by SAS PROC MCMC procedure.

We apply the regular MI analysis and Bayesian MCMC approach to an antidepressant drug trial dataset created by the DIA Missing Data Working Group (Mallinckrodt et al. 2013). The dataset was constructed from an actual clinical trial and made available by the Working Group (see www.missingdata.org.uk). The dataset contains 172 patients (84 in the treatment arm, 88 in the placebo control arm). Repeated measures for Hamilton Depression 17-item total scores were taken at baseline and Weeks 1, 2, 4, and 6, post-randomization. The Week 6 measurements were completed by about 76% of the treatment group patients and about 74% of the the control group patients. The analysis dataset included one patient record with intermittent missing data; in all analysis methods, the missing data for this patient were imputed under the assumption of MAR. The monotone missing data were imputed under CBI methods of CIR, CR, or J2R.

In the analysis of this dataset, we noticed that the MCMC sampling had high autocorrelation. To increase the stability of the results, we used 200 imputations in the conventional MI analysis, and used 2,000 iterations for turning, 2,000 iterations for burn-in, and 200,000 in the main sampling, keeping one from every 10 samples (with option THIN=10 in PROC MCMC) to get a total of 20,000 samples for the posterior mean and standard deviation. Table 6 shows the analysis results. Compared with the mixed model analysis, the Bayesian MCMC under MAR produced very similar results. As compared to the results from the mixed model, the CBI analyses based on regular MI are conservative. With CBI, the point estimates are shrunk toward 0 but the standard errors (SEs) are very similar to the SEs from the MAR analysis. As such, the CBI analyses with regular MI have large p -values compared to the primary analysis under MAR. In fact, the result of the J2R analysis becomes insignificant. With the Bayesian MCMC approach, the CBI analysis results have similar point estimates as the CBI analysis with regular MI but have smaller SEs. As a result, the p -values from CIR, CR, and J2R all maintained significance.

Table 6 Primary and sensitivity analysis results for an anti-depressant trial

Method	$\hat{\theta}$ (SE)	95 % CI	p-value (%)
Primary analysis under MAR			
Mixed Model	-2.80(1.12)	[-5.01, -0.60]	1.3
Bayesian MCMC	-2.80(1.13)	[-5.03, -0.57]	1.4
CBI using regular MI			
CIR	-2.46(1.10)	[-4.65, -0.28]	2.7
CR	-2.38(1.11)	[-4.57, -0.20]	3.3
J2R	-2.12(1.13)	[-4.36, 0.12]	6.3
CBI using MCMC approach			
CIR w MCMC	-2.43(1.01)	[-4.39, -0.44]	1.7
CR w MCMC	-2.34(0.99)	[-4.25, -0.53]	1.9
J2R w MCMC	-2.10(0.86)	[-3.77, -0.42]	1.5

Note CI = confidence intervals for mixed model and regular multiple imputation (MI), credible interval for Bayesian MCMC (Markov chain Monte-Carlo method) approach. MAR = missing at random. CBI = control-based imputation. CIR = copy increments in reference. CR = copy reference. J2R = jump to reference

To check the convergence of the MCMC sampling, Fig. 1 shows the diagnostics plots for both the Bayesian MCMC samples for the primary analysis under MAR and the CR analysis. The trace-plots for both parameters show good mixing and stabilization. With the option of THIN=10, the autocorrelation decreases quickly. The posterior density curves are estimated well from the 20,000 samples.

5 Discussions and Remarks

In many clinical trials, missing data might be unavoidable. We have illustrated some applications of Monte-Carlo simulation methods for handling of missing data issues for longitudinal clinical trials. Simulation-based approaches to dealing with missing data can be extremely useful in the conduct of clinical trials; most notably in the design stage to calculate needed sample size and power, as well as in the final analysis stage to conduct sensitivity analyses. We described a method to generate MVN longitudinal data under different assumed MDMs with specified CDRs. As a sensitivity analysis, we applied a δ -adjustment approach to account for the potential difference between the MMRM (typically used as the primary model in clinical trials) and the MI model used to facilitate the tipping point analysis, and to propose an adjustment to the final tipping point calculation. Depending on the number of imputations used, the inferential statistics produced by the MMRM and MI models can differ, due in part to differences in the approximated degrees of freedom. A sufficient number of imputations should be used to reduce this variation. The appropriate number of imputations to be used should be confirmed via simulation

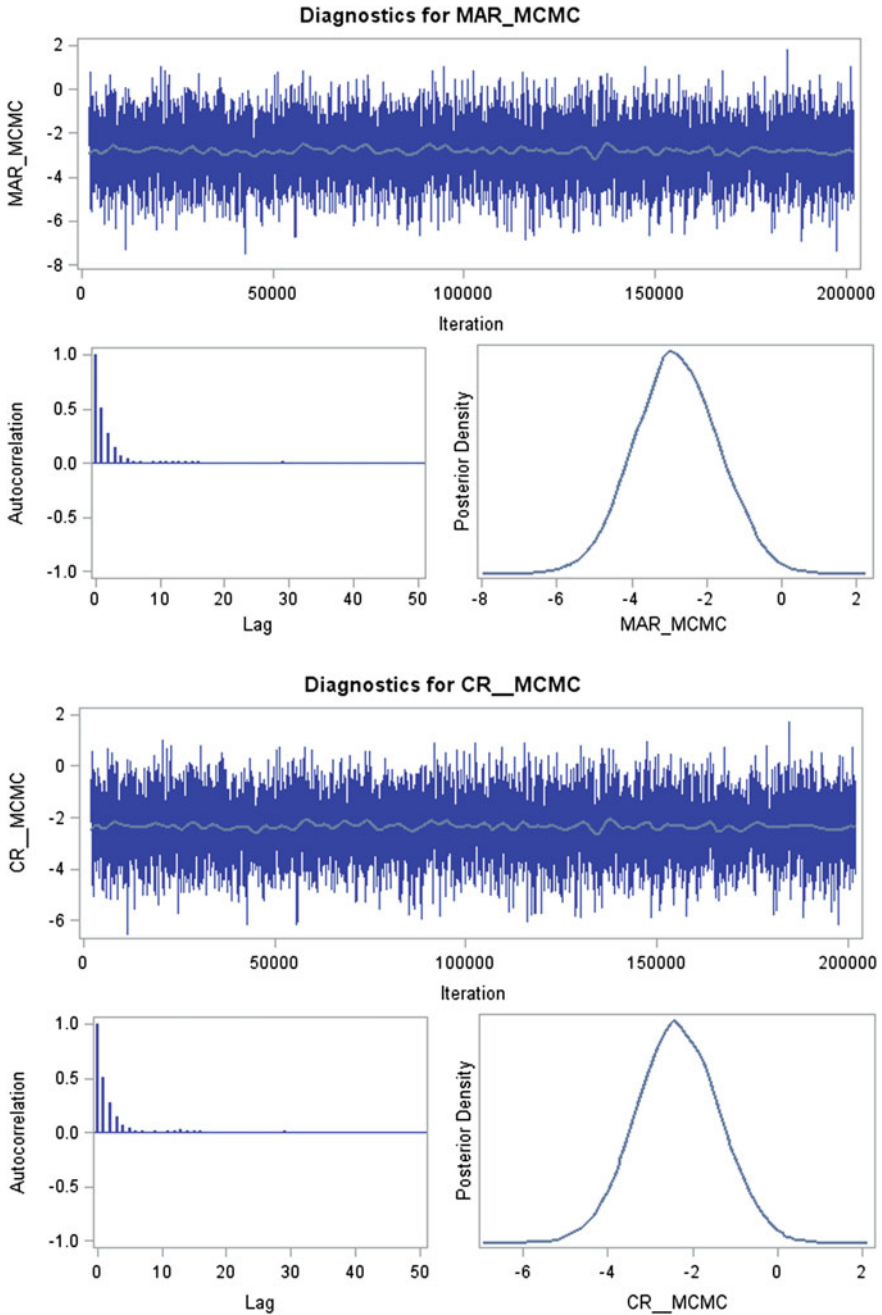


Fig. 1 Diagnostics plots for Bayesian MCMC under MAR and for Copy Reference Imputation Method

by the statistician during the analysis planning stage. We also presented a Bayesian MCMC method for CBI that provides a more appropriate variance estimate than regular multiple imputation.

The methods presented are only a few applications of simulation methods for missing data issues. Of course, many other simulation-based methods are available that can be used for missing data. For example, we considered only a logistic model for the MDM, noting that other models such as a probit model can also be used. In addition, the missing probabilities defined by the example MDMs depended only on the current time point and/or the next time point. Other MDMs may be defined allowing for the incorporation of additional time points. In the CBI approaches, we considered a Bayesian MCMC approach, although other avenues might also be used such as bootstrapping to obtain the appropriate variance for CBI methods. Although we considered only continuous endpoints, simulation-based methods can also be very useful in dealing with missing data for other types of endpoints such as binary, categorical, or time-to-event data.

One drawback for simulation-based methods is the random variation from the simulations. It is critical to assess the potential variation and/or monitor the convergence. When using simulation-based method for analysis of clinical trials, the analysis plan should pre-specify all the algorithms, software packages, and random seeds for the computation. Naturally, the analysis should use a sufficient number of imputations or replications in order to reduce the random variation. Of course, simulations examine only the statistical properties under the assumptions used in those simulations. Whenever possible, theoretical or analytic methods should be considered over simulations.

References

- Ayele, B. T., Lipkovich, I., Molenberghs, G., & Mallinckrodt, C. H. (2014). A multiple imputation based approach to sensitivity analysis and effectiveness assessments in longitudinal clinical trials. *Journal of Biopharmaceutical Statistics*, 24, 211–228. doi:10.1080/10543406.2013.859148.
- Barnard, J., & Rubin, D. B. (1999). Small-sample degrees of freedom with multiple imputation. *Biometrika*, 86, 948–955. doi:10.1093/biomet/86.4.948.
- Carpenter, J. R., Roger, J. H., & Kenward, M. G. (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. doi:10.1080/10543406.2013.834911.
- European Medicines Agency. (2010). *Guideline on missing data in confirmatory clinical trials*. Retrieved from http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf.
- Little, R. J., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: John Wiley.
- Liu, G. F., & Pang, L. (2015). On analysis of longitudinal clinical trials with missing data using reference-based imputation. *Journal of Biopharmaceutical Statistics*. Advance online publication. <http://dx.doi.org/10.1080/10543406.2015.1094810>.

- Lu, K., Luo, X., & Chen, P.-Y. (2008). Sample size estimation for repeated measures analysis in randomized clinical trials with missing data. *International Journal of Biostatistics*, 4, 1–16. doi:10.2202/1557-4679.1098.
- Lu, K., Mehrotra, D. V., & Liu, G. F. (2009). Sample size determination for constrained longitudinal data analysis. *Statistics in Medicine*, 28, 679–699. doi:10.1002/sim.3507.
- Lu, K. (2014). An analytic method for the placebo-based pattern mixture model. *Statistics in Medicine*, 33, 1134–1145.
- Mallinckrodt, C. H., Lane, P. W., Schnell, D., Peng, Y., & Mancuso, J. P. (2008). Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. *Drug Information Journal*, 42, 303–319. doi:10.1177/009286150804200402.
- Mallinckrodt, C., Roger, J., Chuang-Stein, C., Molenberghs, G., Lane, P. W., O’Kelly, M., et al. (2013). Missing data: Turning guidance into action. *Statistics in Biopharmaceutical Research*, 5, 369–382. doi:10.1080/19466315.2013.848822.
- National Academy of Sciences. (2010). *The prevention and treatment of missing data in clinical trials. Panel on handling missing data in clinical trials* [Prepared by the Committee on National Statistics, National Research Council]. Washington, DC: National Academics Press. Retrieved from <http://www.nap.edu/catalog/12955/the-prevention-and-treatment-of-missing-data-in-clinical-trials>.
- O’Kelly, M., & Ratitch, B. (2014). *Clinical trials with missing data: A guide for practitioners*. West Sussex: Wiley. doi:10.1002/9781118762516.
- Ratitch, B., O’Kelly, M., & Tosiello, R. (2013). Missing data in clinical trials: From clinical assumptions to statistical analysis using pattern mixture models. *Pharmaceutical Statistics*, 12, 337–347. doi:10.1002/pst.1549.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley. doi:10.1002/9780470316696.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman and Hall. doi:10.1201/9781439821862.