Assessing missing data impact in a clinical trial prior to unblinding using a parametric bootstrap approach

Jianing Di* Qing Liu† Keith M. Gregg*

Abstract

The problem of missing data is frequently encountered in clinical studies. The potential impact of missing data ranges from estimation inefficiency to estimation bias/invalidity. In practice, to assess the robustness of the primary efficacy analysis method to missing data, sensitivity analyses are conducted after data unblinding. This paper discusses an alternative simulation-based framework that can be used to assess the missing data impact by applying the primary method to data generated with different characteristics. The proposed approach can be used prior to data unblinding to evaluate the missing data impact on any metrics or statistical methods. An example of using such framework to assess the type I error rate of the mixed model for repeated measures in a parallel-group study is used to illustrate the methodology.

Key Words: missing data, type I error rate, parametric bootstrap, mixed model for repeated measures, simulation

1. Introduction

Missing data is a frequently encountered issue in clinical studies. Depending on the nature of the disease and the design of the study, missing data may exhibit various characteristics. A large number of factors may contribute to the generation of missing data, and their routes of contribution are often complicated.

Missing data can impact the analysis results in several ways. At a minimum, they result in a loss of estimation efficiency due to sample size reduction. Additionally, in a clinical study, missing data may affect the validity of the data analysis method. For instance, missing data may cause the observed distribution to deviate from the underlying data generating process. This may happen if dropouts are more likely in one tail of the distribution (eg, dropout due to "lack of efficacy") so the observed distribution becomes skewed, or if dropouts tend to occur more frequently in a particular subgroup so the observed distribution becomes an altered mixture. Also, if dropouts occur at different rates across treatment groups, the observed data may exhibit a false treatment difference. There is particular concern with clinical study result interpretation when such bias, often referred as *selection bias*, favors the active treatment over placebo.

The nature of the loss of validity of inference depends on the analytical method and its underlying assumptions. For example, as noted above, the loss of information will increase the estimation variability, but, in addition, most likelihood-based approaches rely on the asymptotic distribution of estimators, and thus their validity will be impaired by a reduction in sample size. More generally, the missing data impact increases as the deviation from the assumed missing data mechanism (Rubin 1976, Little and Rubin 1987, Little 1995) increases. For example, in order to completely ignore the missing mechanism, generalized estimating equation (GEE) regression generally requires missing data to be missing completely at random (MCAR) (Liang and Zeger 1986), and likelihood-based approaches such as the mixed model for repeated measures (MMRM) require missing data to be MCAR or missing at random (MAR) (Verbeke and Molenberghs 2000).

*Janssen Alzheimer Immunotherapy R&D, 700 Gateway Blvd., South San Francisco, CA 94080
†Janssen R&D, 920 Route 202, P.O.Box 300, Raritan, NJ 08869

Because the data characteristics are unobservable after dropout and there is no general way to ascertain the mechanism of missing data, it is difficult to assess the real missing data impact on the primary efficacy analysis results. Therefore, it is often preferred that these results can be confirmed via sensitivity analyses, where the observed study dataset is processed and analyzed using alternative approaches with the hope that the results are consistent. The simplest method is imputation (eg, multiple imputation, Rubin 1996, Schafer 1999), where the missing portion of the study dataset is filled using various approaches. Because the imputation is often made based on observed data, the imputation-based methods assume missing data to be MAR. Methods that allow MNAR, such as the pattern mixture model (PMM) and the selection model (SM) (Verbeke and Molenberghs 2000), are also popular. In practice, the results of these methods are compared with the primary analysis results qualitatively and without a well-defined framework, meaning that the primary results may be considered robust to missing data if the sensitivity analysis results do not seem to be contradictory. Also, since these sensitivity analyses are performed after unblinding, even if the results suggest a large missing data impact, there remain very few (if any) options to salvage the primary analysis.

Motivated by common regulatory requests to evaluate the actual bias of MMRM under MNAR, this paper describes a bootstrap-based framework for sensitivity analysis. In contrast with the previously mentioned procedures, which analyze the same study dataset via alternative approaches, the proposed procedure tests the primary analysis method using simulated datasets with different but plausible characteristics. Therefore, compared with the usual approaches for sensitivity analysis, the proposed approach has the advantages that

- It can be applied prior to data unblinding and offers an opportunity to further refine this assessment after data unblinding.
- It provides a full spectrum of plausible missing data characteristics under which the primary analysis method can be evaluated.
- It can be used to quantitatively evaluate various kinds of output from the primary analysis method.

The paper is organized as follows: in Section 2 the proposed approach is introduced with details; in Section 3 the proposed approach is illustrated via an example of assessing the missing data impact on type I error rate in a hypothetical double-blind parallel-group clinical study in Alzheimer's disease (AD); Section 4 concludes and also offers some discussion.

2. Assessing Missing Data Impacts via Parametric Bootstrap

The difficulty in assessing missing data impacts arises from the unknown true mechanism of missing data. Therefore, in the proposed procedure, the primary analysis method will be tested using simulated data with known characteristics. As data with different properties are generated, the primary analysis method can be tested under all scenarios. The proposed framework is established via two main steps:

1. Construct a set of missing data generating models, which can be used to produce missing data with different characteristics.
2. Assess the selection bias suggested by each missing data generating model, and evaluate the primary analysis method under each of the data generating models, thereby establishing the relationship between selection bias and the missing data impact.

After the establishment of the relationship of missing data impact and selection bias, the evaluation of missing data impact on a given statistical method translates to evaluation of selection bias given the available data.

2.1 Dropout Models

In order to fundamentally understand the missing data impact, it is essential to understand the missing data generation process, which, for the purposes of simulation, is captured by a parametric function that is often referred to as the *dropout model*. A dropout model mathematically aggregates the factors that contribute to subjects' chance of dropout. In other words, given a complete dataset (ie, without missing data), a dropout model assigns a dropout probability to each data point, and then in simulations these data points are deleted randomly according to the assigned probability. Examples of dropout models can be found in several research papers (Little 1995, Wu 2001, Yoo 2009, Siddiqui et al. 2009, Chen et al. 2011). These dropout models can be used in creating missing data with different mechanisms. For example, Siddiqui et al. (2009) proposed that

> "For the MCAR mechanism, certain percentages of missing data are generated randomly at each visit and all subsequent visits. Similarly, for the MAR mechanism, missing data at visit i and the subsequent visits are assumed to be dependent on the values of outcome measure at visit $i-1$. For the MNAR mechanism, if the value of the outcome measure is higher at visit i, then the subject will have missing data at ith visit and the subsequent visits".

Our approach mixes these characteristics and others into a general model. Within certain constraints, the model is sufficiently flexible that the model parameters can be tuned so a portion of the dropouts are not related to efficacy measurements, a portion of the dropouts are related to unobserved efficacy measurements, and a certain level of selection bias can be generated. To satisfy the first requirement, the dropout models include terms that generate dropout probabilities regardless of the actual efficacy measurements. To satisfy the second requirement, the dropout models include terms that are directly related to the efficacy measurements. To satisfy the third requirement, the dropout models include terms that are treatment (exposure) specific.

2.1.1 The Structure of the Dropout Models

For incorporation into dropout models, dropouts are classified into two categories: efficacy-related and efficacy-unrelated. It should be noted that efficacy-relatedness is not a simple determination of the cause of dropout, but rather of the presence or absence of an effect of dropout on the unobserved efficacy data. For example, suppose a subject dropped out due to an adverse event (AE). Despite being a safety issue, the AE could have impacted efficacy if the subject chose to continue the study. A simple example is a back pain that distracts the subject from concentrating on the cognitive test. Therefore, dropouts due to AE are generally considered efficacy-related. In practice, depending on the design of the case report form, other dropout reasons could also be considered efficacy-related. For example, often investigators are required to provide more details if the classified dropout reasons are those such as "other" or "withdrawal of consent". In this case, the classification of efficacy-relatedness should be done by considering the additional information. For instance, the dropout of a subject who early terminated due to "withdrawal of consent" with the detailed description of "I feel so frustrated during the cognitive test so I don't want to participate in the study anymore" should be considered efficacy-related. However, the classification exercise should not be viewed as an attempt to provide a deterministic

mapping between the collected dropout reasons and the potential impact on efficacy, but rather to give a plausible assessment of the proportion of dropouts that could potentially be related to the unobserved efficacy measure. In summary, because the details of dropout reasons are not always available and the underlying relationship between the dropout reasons and the missing efficacy measurements is not directly observable, this classification should be made based on clinical judgment and the assumptions that investigators' determination of subjects' dropout reasons is correct. During the course of the study, when additional information becomes available, if it is determined that the proportion of efficacy-related dropouts is different, then the overall methodology remains unchanged but the model calibration needs to be modified. For a related discussion in classifying the dropouts using early termination reasons, see Heyting et al. (1992).

The classification of efficacy-relatedness suggests an additive structure to the dropouts: the probability of dropout has an efficacy-related component and an efficacy-unrelated component. As will be shown in Section 2.1.2, the models for both efficacy-related dropout and efficacy-unrelated dropout share a similar structure and several common factors. On the other hand, the breakdown was constructed to allow the dropout model to capture the real causes of the dropouts. Thus, in building the dropout model, efficacy-related dropouts are further classified as either exposure-related or exposure-unrelated. For example, an AE such as hospitalization due to an automobile accident can lead to an efficacy-related dropout but is unlikely to be related to the active treatment. This classification is not identical to the treatment-relatedness as is typically assessed for AEs by (blinded) investigators. In addition, it should also be noted that due to the complex interactions among all the model factors, separation of dropouts into efficacy-unrelated or efficacy-related subgroups does not imply that the dropout models characterize MCAR/MAR/MNAR data separately.

2.1.2 Dropout Model Mathematical Formulation

To begin, the overall dropout probability (probability that a data value for a specified subject at a specified time point and all later time points is missing) at time point t is given by an additive structure

$$P_{Overall}(t) = P_{\overline{Eff}}(t) + P_{Eff}(t)\ ,\ t = 1, 2, ..., m\ , \tag{1}$$

where $P_{\overline{Eff}}$ is the probability of efficacy-unrelated dropout, and P_{Eff} is the probability of efficacy-related dropout. In practice t usually indexes one of m scheduled study visits, with $t = 0$ representing the baseline assessment. To allow the model to incorporate a differential probability of dropout for active treatment subjects compared to placebo subjects, P_{Eff} can be further decomposed to be the sum of exposure-related dropout and exposure-unrelated dropout. Therefore the overall dropout probability can be written as

$$P_{Overall}(t) = P_{\overline{Eff}}(t) + P_{Eff,\overline{Exp}}(t) + P_{Eff,Exp}(t) \cdot I_{Active}\ , \tag{2}$$

where I_{Active} is the indicator function that equals 1 for the active treatment group and 0 otherwise. (We assume a two-group trial comparing active treatment to placebo, but extensions to multiple treatment arms is straightforward.)

Let y_t be the efficacy measurement at time point t, $\{x_1, ..., x_p\}$ be a set of variables, continuous or categorical (indicator functions). Then each of the three components in Equation (2) is given in the form of a logistic function as follows:

$$P_{\overline{Eff}}(t) = \text{logit}^{-1}\Big(c_1 + \sum_{i=1}^{p} \alpha_{1,i} x_i + \beta_{1,t}\Big) \tag{3}$$

$$P_{Eff,\overline{Exp}}(t) = \text{logit}^{-1}\Big(c_2 + \sum_{i=1}^{p} \alpha_{2,i} x_i + \beta_{2,t} + \gamma_2 (y_t - y_0)\Big) \quad (4)$$

$$P_{Eff,Exp}(t) = \text{logit}^{-1}\Big(c_3 + \sum_{i=1}^{p} \alpha_{3,i} x_i + \beta_{3,t} + \gamma_3 (y_t - y_0)\Big) \quad (5)$$

where $c.$, $\alpha.$, $\beta.$, and $\gamma.$ are constant coefficients of the corresponding variables.

2.1.3 Parameter Selection and Tuning in Dropout Models

While the fundamental causes of dropouts are often at least partially unknown, a large part of the data collected in a trial is potentially related to dropout. For instance, dropout could be related to factors such as treatment (exposure), time (visit), outcome measures, subject characteristics, background therapy, age, etc. But given the potential correlation among these factors, usually only a subset of these terms should be included into dropout modeling to avoid over-fitting.

Because the true process that governs the dropouts in a study is unobservable, multiple (a family of) dropout models need to be constructed based on the structure given by Equations (3 - 5). This can be accomplished by starting with a base model with a fixed structure and then tuning the parameter values in the base model. Several restrictions should be noted when creating the dropout model family. First, only a subset of the parameters may be modified, because the discrete coefficients $\beta_{2,t}$ and $\beta_{3,t}$ make the overall model over-parameterized in the sense that different combinations of the parameter values may imply the same dropout pattern ($P_{Overall}$ at each time point). Second, the models should be tuned such that they imply a dropout pattern that is consistent with the blinded observations. Third, the models should be tuned such that they imply different levels of dropout-measurement sensitivity (eg, by increasing γ_2 and/or γ_3) and selection bias (eg, by increasing γ_3). Fourth, the models should be tuned in a scientifically plausible manner. For example, the models should not be tuned such that the proportion of MNAR dropouts is extremely large (ie, large γ_2 and/or γ_3), because several research efforts have concluded that missing data in clinical trials are mostly MAR (Siddiqui et al. 2009, Little and Rubin 2002, Verbeke and Molenberghs 2000, Mallinckrodt et al. 2001).

Additional restrictions may be applied to simplify the tuning process. For instance, different rules could be considered to form subgroups (subfamilies) within the dropout model family. One possibility is to restrict the models within the same subgroup to have the same value for $g(\gamma_2, \gamma_3)$ for some function g but different values of γ_2 and γ_3. The simplest example is $g(\gamma_2, \gamma_3) = 2\gamma_2 + \gamma_3$, where the constant 2 comes from the fact that in $P_{Overall}$ the parameter γ_2 will be applied to both placebo and active groups while γ_3 will only be applied to the active group. The effect of such a constraint is to keep the contribution of y_t roughly constant within the model subgroup. See Section 3 for an example of dropout model tuning.

2.2 Missing Data Impacts via Selection Bias

Informative missing data can impact the results of a statistical method in many ways, one of the most important of which is by creating a selection bias, defined as the difference between the observed treatment difference and the true treatment difference. For example, under the null hypothesis, the selection bias at the study end is the expected difference between study completers. Selection bias often result in biased statistical estimation, which may inflate the type I error rate or falsely change the statistical power. Therefore, given a metric to evaluate the performance of an analysis method, it is important to first understand the relationship between that metric and the selection bias created by missing data.

Consider an example that evaluates the missing data impact on type I error rate. The relationship between the α-level and selection bias, suggested by a family of dropout models, can be established via a parametric bootstrap, as follows.

1. Generate a complete dataset (ie, without missing data), and randomly split it into placebo and active groups (ie, the null hypothesis holds). Within each treatment group, randomly delete data from the complete dataset according to probabilities implied by the dropout model.

2. Calculate the selection bias based on this incomplete dataset by subtracting the observed mean efficacy of the placebo group from that of the active group.

3. Ascertain the test result based on this incomplete dataset by analyzing it using the method of interest and recording the test result (significant or not).

4. Repeat steps 1 - 3 a large number of times (eg, 50,000) to generate multiple selection biases and test results. Use the average of the selection biases as the estimated selection bias implied by this dropout model, and use the proportion of significant tests as the estimated α-level of the statistical method under this dropout model.

5. Repeat steps 1 - 4 for all other dropout models in the family, obtain the selection bias and type I error rate implied by each model, and establish the relationship between selection bias and α-level as suggested by the dropout model family.

Different approaches may be used to generate the complete dataset. For example, complete data could be generated under the setup of a parametric bootstrap using a multivariate Gaussian distribution for the value

$$Y = \Big(y_0, y_1 - y_0, ..., y_m - y_0\Big)' , \tag{6}$$

(expressed as such because, as is commonly done, the efficacy variable is analyzed in terms of change from baseline, although this is not required by the method) where the parameters of the distribution are estimated from the data. In addition, in case the study randomization is stratified by a set of factors, the data should be generated within the subgroups and then combined together according to the observed proportions. If allowed by the computational power, multiple distributions should be considered to minimize the impact of the data generating bias, especially when dropouts are believed to have caused the empirical distribution to deviate from the real one. But regardless of the approach, the generated data should be intended to reflect the characteristics of the observed blinded data.

In the next section, the proposed method is illustrated via a hypothetical clinical study.

3. An Example

Compared with trials in most other disease areas, the missing data issue in AD trials is much more pronounced, due to several reasons. First, the overall dropout rate in AD studies is high due to several factors such as long study duration and elderly patient population. Second, the reasons for dropout in AD studies are complicated. A typical example is the dropout due to caregiver issues such as illness, caregiver burden, etc. Third, the clinical measurements in AD studies are often subjective and variable, and therefore the potential relationship between the missing measure and dropout is difficult to characterize.

Consider a hypothetical double-blind, placebo-controlled, longitudinal AD clinical trial, in which subjects will be randomized (1:1 ratio) to receive placebo or the experimental drug. The study duration is approximately 2 years and during the study each subject will

receive a baseline and seven post-baseline cognitive tests (eg, Alzheimers Disease Assessment Scale Cognitive [ADAS-Cog]). Subjects may early terminate from the trial due to any of the seven reasons listed in Figure 1 (left panel). At the end of the study, approximately 30% of subjects (typically observed in AD trials, see Winblad et al. 2008, Green et al. 2009, and Salloway et al. 2012) had early terminated from the study, and the overall dropout pattern is illustrated in Figure 1 (left panel). The treatment difference in the change from

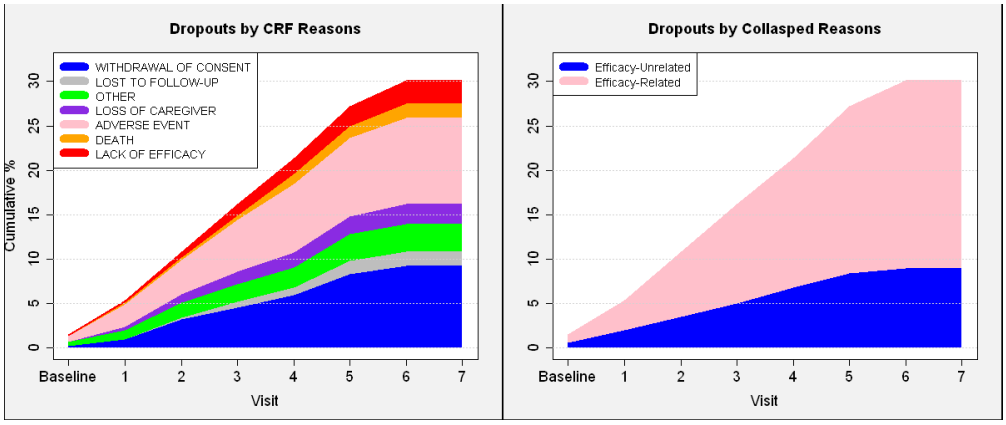

Figure 1: Proportion of early dropout subjects with last visit at a particular visit

baseline at the 7th post-baseline visit will be estimated based on a prespecified MMRM. The proposed approach will be used to evaluate the potential impact of missing data on the MMRM type I error rate.

3.1 Classification of Dropouts

In this example, the proportion of efficacy-related dropouts was set equal to the sum of early terminations due to "lack of efficacy", "death", "adverse event", and a certain portion of "withdrawal of consent" and "other" (see Section 2.1.1), and efficacy-unrelated dropouts include the rest of the early terminations. The dropout pattern based on these collapsed categories is shown in Figure 1 (right panel).

3.2 Dropout Model Construction and Parameter Tuning

A total of 33 different dropout models are constructed. To generate MNAR data that create selection biases in favor of the active treatment, ie, poor cognitive measurements are more likely to be dropped out in the active group, γ_3 in Equation (5) is set to be positive (as an increase in ADAS-Cog score reflects disease progression). In addition, to avoid overfitting (see Section 2.1.3), the following restrictions are applied

- In Equation (5) set $\alpha_{3,i} = \beta_{3,t} = 0$.
- Once the base model is established by fitting a logistic regression model based on Equations (3-5) to blinded data (indicators of missingness at each visit for each subject, as displayed in Figure 1 right panel), the tuning will only be applied to γ_2 and γ_3, and correspondingly the constants c_2 and c_3.
- Based on γ_2 and γ_3 the models are further classified into three groups.

- Group 1 has the weakest overall dropout-to-efficacy sensitivity (small $2\gamma_2 + \gamma_3$), Group 2 has moderate sensitivity (moderate $2\gamma_2 + \gamma_3$), and Group 3 has the strongest sensitivity (large $2\gamma_2 + \gamma_3$).

- Within each group, models with higher numbers were designed to suggest more efficacy-related dropouts from the active treatment group and thus larger selection bias. This is accomplished by splitting the total efficacy-related dropouts between placebo and active based on 11 different ratios γ_2/γ_3 such that γ_2/γ_3 decrease from Model 0 to Model 10.

3.3 Characteristics Implied by the Dropout Models

Simulations based on the parametric bootstrap (see Section 2.2) are executed to assess the characteristics (dropout pattern, selection bias, MMRM α-level, and MMRM estimation bias) of the constructed dropout models. Figure 2 displays the dropout pattern implied by some selected models. As expected, these dropout patterns are very similar, as the models are formulated to align with the blinded data. In addition, Figure 3 displays the

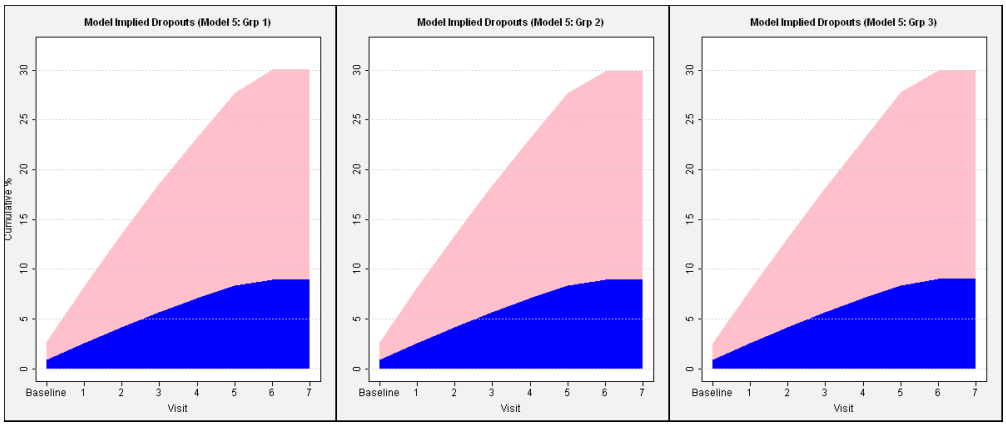

Figure 2: Example dropout patterns implied by dropout models

selection bias implied by all the 33 models as well as the estimation bias of the MMRM under them. Although it is understandable that there is an upward trend in α when the underlying selection bias is increasing in magnitude, it worthwhile to note that the MMRM estimation bias is generally smaller than the selection bias. This is supported by the fact that MMRM estimation is only adversely impacted by MNAR data, which account for only a portion of the missing data generated by the dropout models. Finally, Figure 4 displays the relationship between the simulated α and the selection bias at the 7th post-baseline visit. To allow better visualization, a LOESS (tricubic weight, bandwidth of 0.75, polynomials of degree 2) curve is superimposed.

Several points should be noted. First, if the planned test is 2-sided, then a selection bias in either direction (in favor of active drug or placebo) may inflate the type I error rate. However, because only selection bias in favor of the active drug is of interest, the dropout models should be constructed to only generate negative selection biases. Second, although Figure 4 connects the α-level with selection bias, it does not suggest a general one-to-one mapping between them. Instead, this connection only reflects the relationship across a particular set of dropout models, ignoring other determinants of α-level.

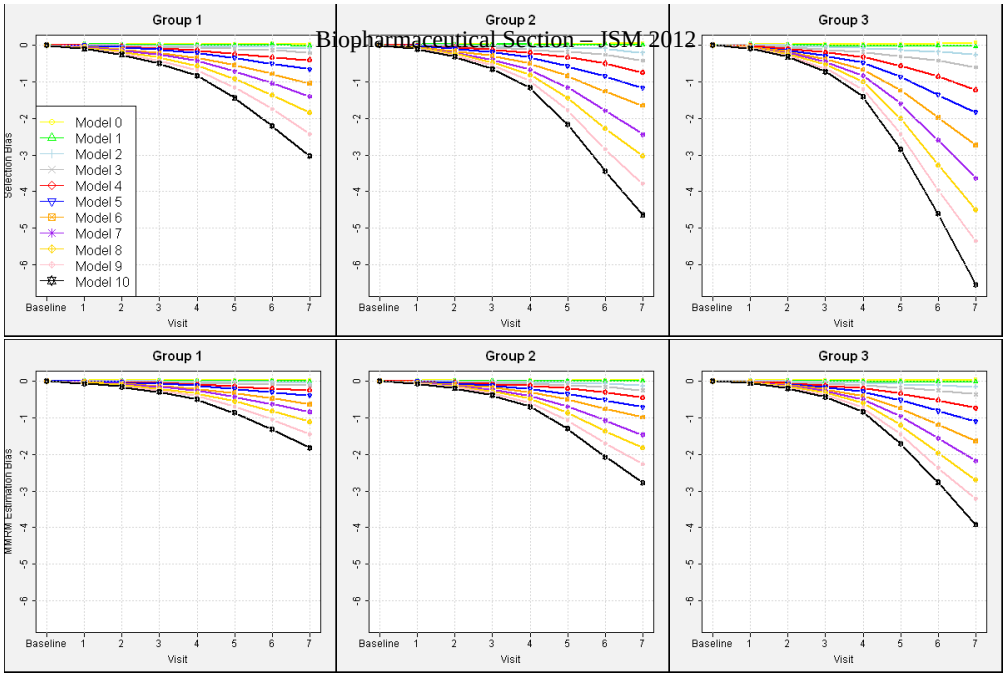

Figure 3: Selection bias and estimation bias by visit implied dropout models

4. Discussion

In this paper a parametric bootstrap-based approach that evaluates the missing data impact prior to data unblinding is discussed. The proposed method offers a framework that does not make specific assumptions about the characteristics of the actual data, and therefore can be used to provide a general and complete assessment of missing data impact.

First, all plausible missing data generation processes (mechanisms) can be considered. The dropout models can be constructed to include all factors that potentially contribute to the chance of dropout. The factor list can be extended such that the dropout model is not designed for a specific endpoint. In addition, once the model structure is determined, the coefficients of the models can be tuned such that all realistic dropout characteristics (eg, pattern, mechanism, etc) can be created.

Second, all statistical methods can be evaluated. Because the proposed framework does not require additional assumptions for its validity, the assessment of missing data impact under such framework will not create potential conflict with the primary method used to analyze the data, in contrast to the usual approach to sensitivity analyses.

Third, all metrics used to evaluate the statistical methods can be considered. For example, if during the parametric bootstrap a measurement difference is added between the treatment groups, then the same method can be used to evaluate the missing data impact on statistical power and estimation bias. Similarly, the proposed method can be used to evaluate other aspects of the study such as the choice of endpoint and the study design.

There are, however, practical considerations that need to be taken when assessing the missing data impact under this framework. First, effort should be made to identify a set of true driving factors for incorporation into the dropout models. This may require meta analyses that combine multiple datasets. Second, because the proposed approach uses observed blinded data, which is often a mixture of placebo and active treatment, extra care should be taken when using the observed data to infer properties of the true data distribution un-

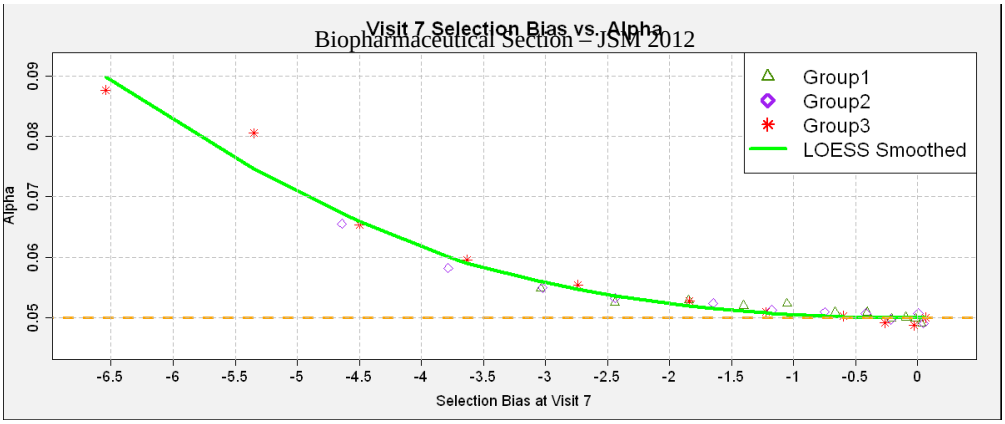

Figure 4: Corresponding selection bias and α implied by dropout models

der placebo. Third, because the result depends on accurate and informative determination of dropout reasons, dropout reasons should be closely monitored throughout the study via good case report form design. Also, patient followups should be performed to ensure all relevant information for dropout had been captured. Fourth, there needs to be a systematic classification of efficacy-related and non-efficacy-related dropouts in the construction of the model. Fourth, the proposed method links the missing data impacts with selection bias, which needs to be estimated based on the blinded data. Generally speaking, the selection bias could be related to many factors such as the actual disposition category (complete or early dropouts of different kinds), the efficacy profile of each disposition category, the likelihood that each subject falls into each of the disposition category, all conditioned on the treatment. A method to describe the relationship of these factors to selection bias and the corresponding assumptions needs to be further investigated.

REFERENCES

Laird, N. (1988), "Missing data in longitudinal studies," *Statistics in Medicine*, 7, 305-315.

Little, R. (1995), "Modeling the drop-out mechanism in repeated-measures studies," *Journal of American Statistical Association*, 90, 431:1112-1121.

Verbeke, G., and Molenberghs, G. (2000), "Linear mixed models for longitudinal data," New York: Springer.

Wu, M., Albert, P., and Wu, B. (2001), "Adjusting for drop-out in clinical trials with repeated measures: design and analysis issues," *Statistics in Medicine*, 20, 93-108.

Yoo, B. (2010), "Impact of missing data on type 1 error rates in non-inferiority trials," *Pharmaceutical Statistics*, 9, 87-99.

Siddiqui, O., Hung, H., and O'Neill, R. (2009), "MMRM vs LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets," *Journal of Biopharmaceutical Statistics*, 19, 227-246.

Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G. (2008), "Longitudinal data analysis," New York: Chapman & Hall/CRC.

Rubin, B. (1976), "Inference and missing data," *Biometrika*, 63, 581-592.

Liang, K., and Zeger, S. (1986), "Longitudinal data analysis using generalized linear models," *Biometrika*, 73, 13-22.

Little, R., and Rubin, D (2002), "Statistical analysis with missing data, second edition," New York: Wiley.

Schafer, J. (1999), "Multiple imputation: a primer," *Statistical Methods in Medical Research*, 8, 3-15.

Rubin, D (1996), "Multiple imputation after 18+ years," *Journal of the American Statistical Association*, 91, 473-489.

Chen, Y., Yang, Y., Hung, J., and Wang, S. (2011), "Evaluation of performance of some enrichment designs dealing with high placebo response in psychiatric clinical trials," *Contemporary Clinical Trials*, 32, 592-604.

Heyting, A., Tolboom, J., and Essers, J. (1992), "Statistical handling of drop-outs in longitudinal clinical trials," *Statistics in Medicine*, 11: 20432061.

Mallinckrodt, H., Clark, S., and David, R. (2001), "Accounting for dropout bias using mixed-effects models," *Journal of Biopharmaceutical Statistics*, 11: 921.

Winblad, B., Gauthier, S., Scinto, L., Feldman, H., Wilcock, K., Truyen, L., Mayorga, J., Wang, D., Brashear, R., and Nye, S. (2008), "Safety and efficacy of galantamine in subjects with mild cognitive impairment," *Neurology*, 70 (22): 2024-2035.

Green, C., Schneider, S., Amato, A., Beelen, P., Wilcock, G., Swabb, A., and Zavitz, H. (2009). "Effect of tarenflurbil on cognitive decline and activities of daily living in patients with mild Alzheimer disease: a randomized controlled trial." *Journal of the American Medical Association*, 302 (23): 2557-2564.

Salloway, S., Sperline, R., Honig, L., Porsteinsson, A., Sabbagh, M., Liu, E., Yuen, E., Lull, J., Miloslavsky, M., Reichert, M., Ketter, N., Grundman, M., Black, R., and Brashear, R. (2012). "A randomized, double-blind, placebo-controlled clinical trial of intravenous bapineuzumab in patients with mild to moderate Alzheimer's disease who are apolipoprotein E ε4 non-carriers," 16th Congress of the European Federation of Neurological Societies, Stockholm, Sweden, http://efns2012.efns.org/.