GEMS PRACTICE POSTER

Addressing the Problem of Data Missingness: Intent to Attend Plus Scale

Ibrahim Turkoz[1], Elena Polverejan[1], Jiyeong Jang[2], Hakan Demirtas[2], Larry Alphs[3]

[1]Janssen Research and Development, LLC, Titusville, NJ, USA; [2]Department of Biostatistics, UIC, Chicago, IL, USA; [3]Newron Pharmaceutical, S.p.A., Morristown, NJ, USA

Introduction

- Missing data are unavoidable in most trials, especially when measurements are taken repeatedly. Subject attrition rates from neuroscience clinical trials range from 20% to more than 50%, leading to data missingness.
- Statistical techniques alone cannot solve the key issues of internal and external validity raised by the attrition process. If assumptions about the missing data are not accurate, crude statistical analyses are biased and can lead to false inferences.
- Data missingness can decrease the statistical power, precision, and generalizability of study results. Failing to measure predictors of missing data further hinders modeling the missing data process sufficiently in sensitivity analyses.
- Recent ICH-E9 (R1)[1] guideline - Estimands and Sensitivity Analysis in Clinical Trials – has revisited missing data related issues in defining estimands and provided number of examples.
- The National Research Council guidance on The Prevention and Treatment of Missing Data in Clinical Trials[2] encourages the development of mechanisms to predict, prevent and adjust for patient dropout. For this purpose, the Janssen Missing Data Working Group has developed a new scale Intent to Attend (ITA)-Plus based on the ITA scale proposed in the Leon and colleagues (2007)[3].
- The ITA-Plus is a 2-item instrument designed to assess patients' self-rated risk of study dropout at baseline and subsequent visits.
  - At baseline: likelihood of subjects completing the clinical trial (**Table 1**).
  - Throughout the trial: likelihood that subjects will attend their next study visit (**Table 2**).
  - If the dropout risk is imminent, follow-up questions can be asked to help identify reasons for dropout or hurdles to study completion
- The ITA-Plus has both logistical and statistical value.
- On the logistical side, ITA-Plus is a tool designed to evaluate potential risks of dropout and allows the investigators and sites to understand them and be proactive towards the prevention of study dropouts.
  - Provides clear anchor points to the likelihood of completing the study or making the next visit (for patients).
  - Include a standardized collection of reasons for dropout or failure to make the next visit (for patients).
  - To increase the chance of study completion, document responses that may address patient-identified reasons for study dropout or making the next visit (for study staff).
  - Provide an instruction manual to explain how to administer the ITA (for sponsor).
- On the statistical side, ITA-Plus could be used in statistical models for missing data (e.g. as a time-varying covariate in a multiple imputation model) as it is expected to be correlated with the study dropout and outcome.
  - In practice, both the outcome and dropout mechanism are expected to be dependent on ITA-Plus scores. It was assumed that adjusting for ITA-Plus in the analysis model should improve bias and overall performance in addition to creating of richer MAR models

Table 1: CRF

Intent-to-Attend (ITA) Plus

Baseline

Please answer the following question:

1. How likely is it that you will complete this study given what you know about it now? (Please consider all the things in your life that may prevent you from completing it.

0	1	2	3	4	5	6
Definitely Not	Very Unlikely	Unlikely	Possibly	Likely	Very Likely	Definitely

Note: If the response is 0 or less, ask the following the question:

2. "What might prevent you from completing the study?" (check all that apply)

_____ Transportation considerations
_____ Treatment side effects
_____ Job or school considerations
_____ Family considerations
_____ Length and complexity of study
_____ Planned move to another area
_____ Treatment no longer needed
_____ Vacations
_____ Other medical conditions
_____ Other study concerns (please specify)
_____ Other personal concerns (please specify)

† Clinician: For each item identified by the subject as an impediment to completing the study, provide a brief summary of your key responses to address the subject's concerns.

Note: This item is optional and should be considered based on study characteristics and disease state. Based on study need, determine whether randomization should occur.

Table 2: CRF

Intent-to-Attend (ITA) Plus

Every Post-Baseline Visit

Please answer the following question:

1. Given what you know about this study, how likely will you attend the next scheduled visit? (Please consider all the things in your life that may prevent you from keeping your next appointment.)

0	1	2	3	4	5	6
Definitely Not	Very Unlikely	Unlikely	Possibly	Likely	Very Likely	Definitely

Note: If the response is 0 or less, ask the following the question:

2. What might prevent you from making your next study visit? (check all that apply)

_____ Transportation considerations
_____ Treatment side effects
_____ Job or school considerations
_____ Family considerations
_____ Length and complexity of study
_____ Planned move to another area
_____ Treatment no longer needed
_____ Vacations
_____ Other medical conditions
_____ Other study concerns (please specify)
_____ Other personal concerns (please specify)

† Clinician: For each item identified by the subject as an impediment to completing the study, provide a brief summary of your key responses to address the subject's concerns.

Note: This item is optional and should be considered based on study characteristics and disease state.

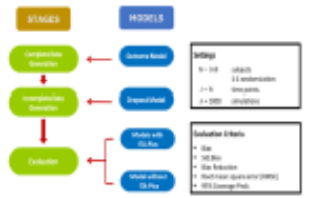
Objective

- To evaluate the performance of Intent to Attend (ITA)-Plus scale in an simulation study.
- A simulation study was carried out to quantify the operating characteristics of statistical missing data models that adjust or not for the ITA-Plus collected values under various scenarios. This investigation was extended to include influence of the treatment by ITA-Plus interaction and of the ITA-Plus use as a time-varying covariate.

Simulation Study

- Separate data generation and data analytic models were considered, see figure 1.
  - For the first stage, complete data generation, generated data sets for 160 subjects under 1:1 randomization with the outcome model. Each subject had 6 visits after the baseline.
  - In the second stage, incomplete data generation, imposed missing on outcome based on the dropout model.
  - At the last stage, evaluation, compared the performance of models with or without ITA-Plus score using simulations for each scenario.

Figure 1: Simulation Methods



Complete Data Generation Model

- Mixed model for continuous outcome y_{ij}

$$y_{ij} = \beta_0 + \beta_1 G_i + \beta_2 T_j + \beta_3 G_i T_j + \beta_4 A_{ij-1} + \beta_5 G_i A_{ij-1} + v_i + \epsilon_{ij}$$

where $i=1,\ldots,N$ subjects ($N=160$), $j=1,\ldots,J$ repeated measurements ($J=6$)

G_i: Group (0 for control, 1 for treatment)

T_j: Time for visit, $T_j=j$

A_{ij-1}: Time varying ITA score before the next visit j

v_i: Subject-specific random intercept

ϵ_{ij}: Residual error

- Mixed model for the baseline continuous outcome y_{i0}

$$y_{i0} = \beta_0 + v_i + \epsilon_{i0}$$

- Assumed higher values indicate better outcomes.
- ITA-Plus scales were generated from a left skewed ordinal distribution from 0 to 6.

$$A_{ij-1} \sim Ord(0.05, 0.10, 0.10, 0.17, 0.18, 0.20, 0.20)$$

- Underlying parameter values to create plausible scenarios
  - Coefficients: $\beta_0 = 0$, $\beta_1 = 0.05$, $\beta_2 = 0.05$, $\beta_3 = 0.10$, $\beta_4 = 0.5$, $\beta_{05} = 0/0.15/0.5$
  - Additional scenarios are considered with: $\beta_1 = 0$, $\beta_{05} = 0$
  - The outcome is assumed to be higher in the treatment group at the baseline, and it is assumed to slightly increase over time in both groups.
  - With the positive interaction of Group by Time, we assumed symptoms are getting better in the treatment group than in the control group.
  - Assumed higher ITA scores provide better symptom scores with beta.4 equal to 0.5.
  - Differentiated impact of ITA in the treatment group, by varying the interaction of Group by ITA from 0 to 0.5. As the interaction increases, the impact of ITA on the outcome is greater in the treatment group.
  - Examined the scenarios where there is no impact of ITA on outcome with 0 coefficients for both beta.4 and beta.6A.

Incomplete Data Generation Model

- A latent dropout score d_{ij} was generated under the following dropout model to impose missingness.

$$d_{ij} = \alpha_0 + \alpha_1 G_i + \alpha_2 T_j + \alpha_{3A} G_i T_j + \alpha_4 A'_{ij-1} + \alpha_{5A} G_i A'_{ij-1} + \alpha_{MAR} y_{ij-1} + \alpha_{MNAR} y_{ij} + \epsilon^*_{ij}$$

where ϵ^*_{ij} ~ standard logistic distribution

y_{ij} is missing when dropout criteria met ($d_{ij} >$ threshold)

- Alpha coefficients are interpreted as the log of odds ratio of dropping out according to one unit increase of corresponding covariate.
- A'_{ij-1} will=A_{ij-1}, **intent to drop-out.**
- When dropout score is larger than the chosen threshold value, the outcome y_{ij} is set to missing using Rubin's[4] hierarchy mechanism.

Coefficients:

$\alpha_0 = 0.05$, $\alpha_1 = \log(0.98)$, $\alpha_2 = \log(1.1)$, $\alpha_{3A} = \log(1.1)$, $\alpha_{4A} = \log(1.7)$, $\alpha_{5A} = -\log(1.4)$

- MAR (Missing at random): $\alpha_{MAR} = -\log(1.2)$, $\alpha_{MNAR} = 0$
- MNAR (Missing not at random): $\alpha_{MAR} = 0$, $\alpha_{MNAR} = -\log(1.2)$
- Dropout criteria: $d_{ij} \geq 4.1$
- Assumed monotone missing data pattern
- MAR and MNAR scenarios fit separately. MAR dropout depends on the variable that is always observed, so here under MAR, dropout score depends on the prior outcome y_{ij-1}. MNAR dropout depends on the variable that is subject to be missing, dropout is the function of the current outcome y_{ij}.
- Used negative values for α_{MAR} and α_{MNAR} to force higher scores lead to less dropouts.

Data Analytic Model

- A random-intercept linear regression model examined efficacy. This model accounts for the time-varying intent to Dropout by including the term, A_{ij-1}; this model was labelled as LDH to refer original Leon paper. To examine the model misspecification, a naïve model that excluded A_{ij-1} was fitted, this was labelled as Naive. Full model includes both ITA-Plus term A_{ij-1} and the interaction term between A_{ij-1} and treatment group.

Analysis Models

- Naïve model without ITA-Plus terms

$$y_{ij} = \beta_0 + \beta_1 G_i + \beta_2 T_j + \beta_{3A} G_i T_j + v_i + \epsilon_{ij}$$

- LDH model listed as Naive + ITA-Plus score

$$y_{ij} = \beta_0 + \beta_1 G_i + \beta_2 T_j + \beta_{3A} G_i T_j + \beta_4 A_{ij-1} + v_i + \epsilon_{ij}$$

- Full model listed Naive + ITA-Plus + ITA-Plus*Group

$$y_{ij} = \beta_0 + \beta_1 G_i + \beta_2 T_j + \beta_{3A} G_i T_j + \beta_4 A_{ij-1} + \beta_{5A} G_i A_{ij-1} + v_i + \epsilon_{ij}$$

- Models were compared on several criteria including bias (the absolute value of the difference between the specified coefficient and the parameter estimate), standardized bias (ratio of the bias to the empirically estimated standard deviation of the parameter estimate), root mean squared error (reflects combination of accuracy and precision since it includes both bias and variance of the parameter estimate), bias reduction (proportion of bias in adjusted models that is reduced with the adjustment for the ITA-Plus), and 95% probability coverage (the proportion of the estimated 95% confidence intervals that include the specified coefficient), see Table 3.

Table 3: Evaluation Criteria

Criteria	Formula
Bias	$\hat{\theta}(\hat{\alpha}) - \hat{\alpha}_0$
Standardized bias	$\hat{\theta}(\hat{\alpha}) - \hat{\alpha}_0 / \sqrt{\hat{\theta}(\hat{\alpha})}$
Root Mean Square Error (RMSE)	$\sqrt{\hat{\theta}(\hat{\alpha}) - \hat{\alpha}_0)^2} + \sqrt{\widehat{Var}(\hat{\alpha})}$
Bias reduction (%)	$\frac{(\text{bias}\hat{\alpha}_0 \text{ in the naive model} - \text{bias}\hat{\alpha}_0 \text{ in the analytic model})}{\text{bias}\hat{\alpha}_0 \text{ in the naive model}} \times 100$
95% Coverage	Estimated 95% interval: $(\hat{\alpha}_0 - t_{crit}\sqrt{\hat{\theta}(\hat{\alpha})}, \hat{\alpha}_0 + t_{crit}\sqrt{\hat{\theta}(\hat{\alpha})})$; $\frac{\text{Number of estimated 95\% intervals that include the true } \alpha_0}{\text{Total number of simulations}}$

Results

Table 4: Using the Evaluation Criteria for β_{3A} when (β_4 = 0.5) and Correlation Between ITA's and Groups: $r_{bet} = 0$

Figure 2: Under MAR, Performance of Models: Bias and RMSE using the Evaluation Criteria for β_{3A} when (β_4 = 0.5)

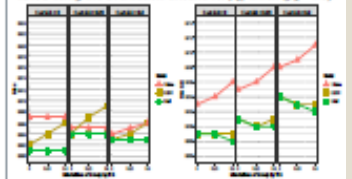
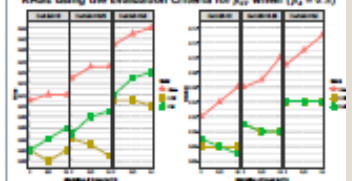


Figure 3: Under MNAR, Performance of Models: Bias and RMSE using the Evaluation Criteria for β_{3A} when (β_4 = 0.5)



- The Full Model RMSE is lower or the same as the LDH Model under both MAR and MNAR. The Full Model have the best performance in terms of BIA and RMSE.

Conclusions

Methodological advantages

- In practice, both the outcome and dropout mechanism are expected to be dependent on ITA-Plus scales.
- Adding ITA-Plus as a covariate in the model improves bias and overall performance.
- Inclusion of ITA-Plus allows the creation of richer MAR models.
- Even when there is no assumed relationship between ITA-Plus and outcome & dropout mechanism, inclusion of ITA-Plus as a covariate does not adversely affect bias and RMSE.

Operational advantages

- Allows identification of subjects at risk of dropping out.
- Includes collection of standardized reasons of dropping out at next visits.
- Allows sponsors to proactively address subjects' concerns prior to dropping out.
- Adds minimal burden to implement.
- May increase empathy between subjects and clinicians.

References

1.https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/E9-R1EWG_Step2_Guideline_2017_0616.pdf
2.National Research Council. The prevention and treatment of missing data in clinical trials. Washington, DC: National Academies Press, 2010.
3.Leon AC et al. Clin Trials. 2007;4:540-547.
4.Rubin DB. Inference and missing data. Biometrika 1976; 63(3):581-