Global Sensitivity Analysis of Randomized Trials with Missing Data

Daniel O. Scharfstein and Aidan McDermott, Johns Hopkins University

Missing Data Matters

During almost 30 years of review experience, the issue of missing data in ... clinical trials has been a major concern because of the potential impact on the inferences that can be drawn .... when data are missing .... the analysis and interpretation of the study pose a challenge and the conclusions become more tenuous as the extent of 'missingness' increases. *Robert Temple and Bob O'Neil (FDA)*

Crux: Inference about treatment effects relies on *unverifiable* assumptions.

Need for Sensitivity Analysis

Sensitivity analyses should be part of the primary reporting of findings from clinical trials. Examining sensitivity to the assumptions about the missing data mechanism should be a mandatory component of reporting. *2010 NRC Report*

... it is important to evaluate the robustness of the results to various limitations of the data, assumptions, and analytic approaches to data analysis *ICH Guidance - E9*

In all submissions with non-negligible amounts of missing data sensitivity analyses should be presented as support to the main analysis. *EMEA 2009 Draft Guidance*

Sensitivity Analysis Approaches

Set of possible assumptions is too large to be fully explored. Types of sensitivity analysis:

- Ad-hoc
  - Specify a few different analytic methods (e.g., LOCF, BOCF, MMRM) and see if the results are consistent.
- Local
  - Specify a reasonable benchmark assumption and evaluate the robustness within a small neighborhood of this assumption.
- Global (or less local)
  - Specify a reasonable benchmark assumption and evaluate the robustness within a large neighborhood of this assumption.
  - Can see how far one needs to deviate from the benchmark assumption in order for inferences to change.
  - If the assumptions under which the inferences change are judged to be sufficiently far from the benchmark assumption, then greater credibility is lent to the benchmark analysis; if not, the benchmark analysis can be considered to be fragile.
  - Operates like "stress testing" in reliability engineering or "tipping point" analysis in business strategy.

Global Sensitivity Analysis

- Randomized study designs where measurements of an outcome are scheduled to be taken at fixed time-points after randomization.
- Focus on monotone missing data pattern.
- Comparison of treatment arm means at the last scheduled visit (in a world with no missing data).

Assumptions

Inference requires two types of assumptions:

(i) *unverifiable* assumptions about the distribution of outcomes among those with missing data

(ii) additional testable assumptions that serve to increase the efficiency of estimation.

Type (i) Assumptions

- stratify individuals with missing outcomes according to the data that we were able to collect on them and the occasions at which the data were collected
- separately for each stratum, hypothesize a connection (or link) between the distribution of the missing outcome(s) with the distribution of the outcome(s) among those with the observed outcome(s) and who share the same recorded data.

Common Benchmark Assumption: Missing at Random (Type (i) Assumption)

- For subjects last seen at visit k with a given recorded history of outcomes through that visit, the distribution of outcomes after visit k is the same as the distribution of outcomes for subjects who remain on-study after visit k and who share the same recorded history.
- Equivalent to assuming that the conditional hazard of dropping out between visits k and $k+1$ only depends on the history of recorded outcomes through visit k.

Building a Class of Assumptions Around Missing at Random

- For subjects last seen at visit k with a given recorded history of outcomes through visit k and unobserved outcome at visit $k+1$, the distribution of outcomes after visit $k+1$ is the same as the distribution outcomes for subjects who remain on-study after visit k and who share the same recorded history through visit k and have the same outcome at visit $k+1$.
- Equivalent to assuming that the conditional hazard of dropping out between visits k and $k+1$ only depends on the history of recorded outcomes through visit k and the outcome at visit $k+1$.
- For subjects last seen at visit k with a given recorded history of outcomes through visit k, use "exponential tilting" to link the distribution of the missing outcome at visit $k+1$ to the distribution of the outcome at this timepoint among those who remain on-study after visit k and who share the same recorded history through visit k, where there is a parameter (α) that governs "differences" between the distributions being linked.
- $\alpha = 0$ indicates no differences and yields the missing at random assumption, $\alpha > 0$ (< 0) indicates that distribution of the missing outcomes is more heavily weighted toward higher (lower) values that those with observed outcomes.
- Equivalent to assuming a logistic regression model for the conditional hazard of dropping out between visits k and $k+1$, where α quantifies the influence of the outcome at visit $k+1$, above and beyond the recorded history of outcomes through visit k.

Type (ii) Assumptions

- The stratum of people who share the same recorded data will typically be small - need to draw strength across strata by imposing type (ii) "smoothing" assumptions.
- We model the distribution of the observed data and smooth by assuming
  - the conditional distribution of the outcome at visit k given on-study at visit k and the past history of outcomes only depends on the outcome at visit $k-1$
  - the conditional probability of dropping out between visits k and $k+1$ given on-study at visit k and the history of outcomes only depends on the outcome at visit k

  and by using non-parametric covariate smoothing techniques.

Estimation and Inference

For given α,

- Generalized Newton-Raphson estimator (plug-in plus average of estimated influence functions)
- Confidence intervals via double bootstrap

Case Study: SCA-3004

- Randomized trial designed to evaluate the efficacy/safety of once-monthly, injectable paliperidone palmitate (PP1M) relative to placebo (PBO) in delaying the time to relapse in subjects with schizoaffective disorder.
- Open-label phase with a flexible-dose, lead-in period and a fixed-dose, stabilization period.
- Stable subjects entered a 15-month relapse-prevention phase and were randomized to receive PP1M or placebo injections at baseline (Visit 0) and every 28 days (Visits 1-15).
- Additional clinic visit (Visit 16) scheduled for 28 days after the last scheduled injection.
- 170 and 164 subjects were randomized to the PBO and PP1M arms.
- Question: Are functional outcomes better in patients with schizoaffective disorder better maintained if they continue on treatment or are withdrawn from treatment and given placebo instead?
- Study design required that patients who had signs of relapse be discontinued from the study.
- Some patients discontinued due to adverse events, withdrew consent or were lost.
- 38% and 60% of patients in the PBO and PP1M arms were followed through Visit 16.
- Patient function was measured by the Personal and Social Performance (PSP) scale.
- The PSP scale is scored from 1 to 100 with higher scores indicating better functioning.
- By treatment, estimate mean PSP at Visit 16 in a world in which all patients who are followed to Visit 16.
- The mean PSP score among completers was 76.05 and 76.96 in the PBO and PP1M arms; the estimated difference is -0.91 (95%: -3.98:2.15).

Case Study: SCA-3004

- Under missing at random, the estimated means are 69.60 and 74.37 for the PBO and PP1M arms.
- The estimated treatment difference is -4.77 (95% CI: -10.89 to 0.09).
- Relative to the complete-case analysis, the missing at random analysis corrects for bias in a direction and relative magntitude that is anticipated.
- As a consequence, the estimated treatment effect is more favorable to PP1M, although the 95% CI still includes 0.

Case Study: SCA-3004

Figure : *Treatment-specific mean PSP at Visit 16 as a function of α, with 95% pointwise confidence intervals.*

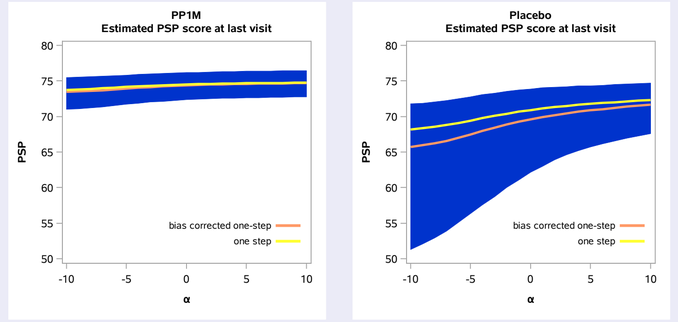
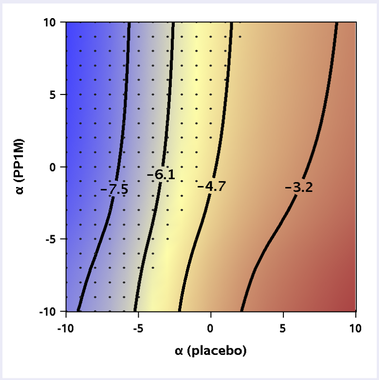


Figure : *Contour plot of the estimated differences between mean PSP at Visit 16 for PBO vs. PP1M for various treatment-specific combinations of α. The point (0,0) corresponds to the missing at random assumption in both treatment arms. Dots indicate whether the treatment difference is statistically significant at the 0.05 level.*



- When it is plausibly assumed that in both treatment groups $\alpha < 0$ (i.e., the lower left quadrant of the figure), the magnitude of the estimated treatment effect generally increases in favor of PP1M, with most combinations of the treatment-specific sensitivity analysis parameters yielding statistically significant results.

Software

- An R-based software package called SAMON is available at www.missingdatamatters.org.
- A SAS version is currently under development and will be ready this summer.

Extensions

- Intermittent missing data
- Incorporation of auxiliary covariates

Conclusion

- The methods and software we have developed will help sponsors and regulators more comprehensively assess the robustness of trials with missing data, as per recommendations of the NRC Report and regulatory guidance documents.

Funding

- FDA
- PCORI