Global Sensitivity Analysis of Randomized Trials with Missing Data: A Frequentist Perspective

FDA/CTP Statistics Workshop

Daniel Scharfstein
Johns Hopkins University
dscharf@jhu.edu

November 6, 2015

Schizophrenia Clinical Trial

- Multi-center, randomized clinical trial to assess the safety and efficacy of a test drug (81 subjects) relative to placebo (78 subjects) for individuals suffering from acute schizophrenia.
- The primary instrument used to assess the severity of symptoms was the positive and negative syndrome scale (PANSS).
- Measurements were scheduled to be collected at baseline, day 4 after baseline, and weeks 1, 2, 3, and 4 after baseline.
- One goal was to compare the two treatment groups with respect the mean PANSS score at week 4.

Problem: Missing Data

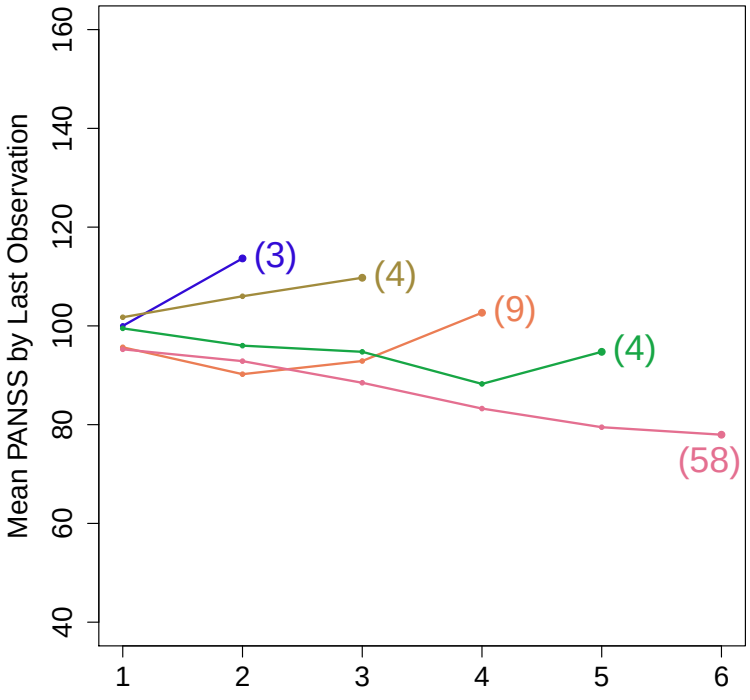


Figure: Placebo

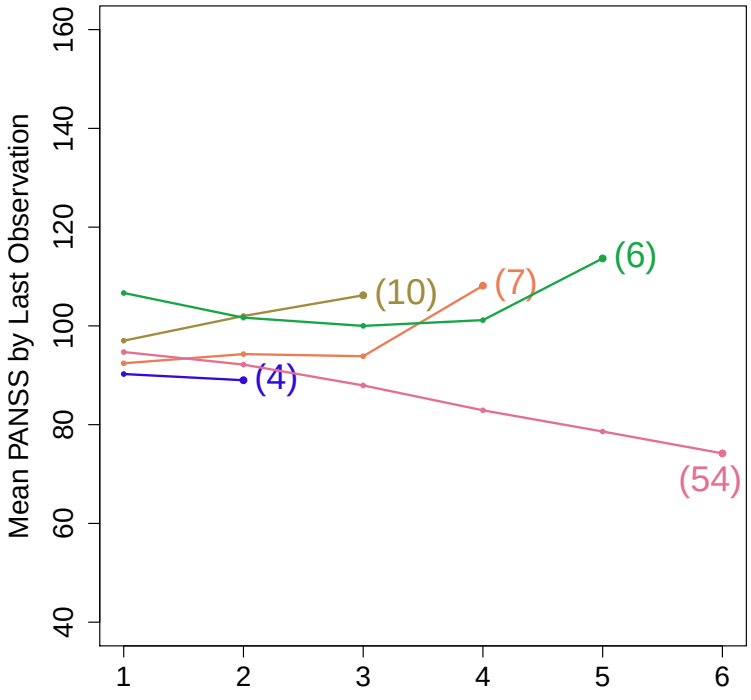


Figure: Test

Fundamental Issue

- Even with infinite data, we cannot learn about the treatment-specific mean PANSS score at week 4.
- We don't know the distribution of PANSS scores for individuals who have dropped out prior to week 4.
- **Need to make assumptions**!

Sensitivity Analysis

The set of possible assumptions about the missing data mechanism is very large and cannot be fully explored. There are different approaches to sensitivity analysis:

- Ad-hoc
- Local
- Global

Ad-hoc Sensitivity Analysis

- Analyzing data using a few different analytic methods, such as last or baseline observation carried forward, complete or available-case analysis, mixed models or multiple imputation, and evaluate whether the resulting inferences are consistent.

Local Sensitivity Analysis

- Specify a reasonable benchmark assumption (e.g., missing at random) and evaluate the robustness of the results within a small neighborhood of this assumption.

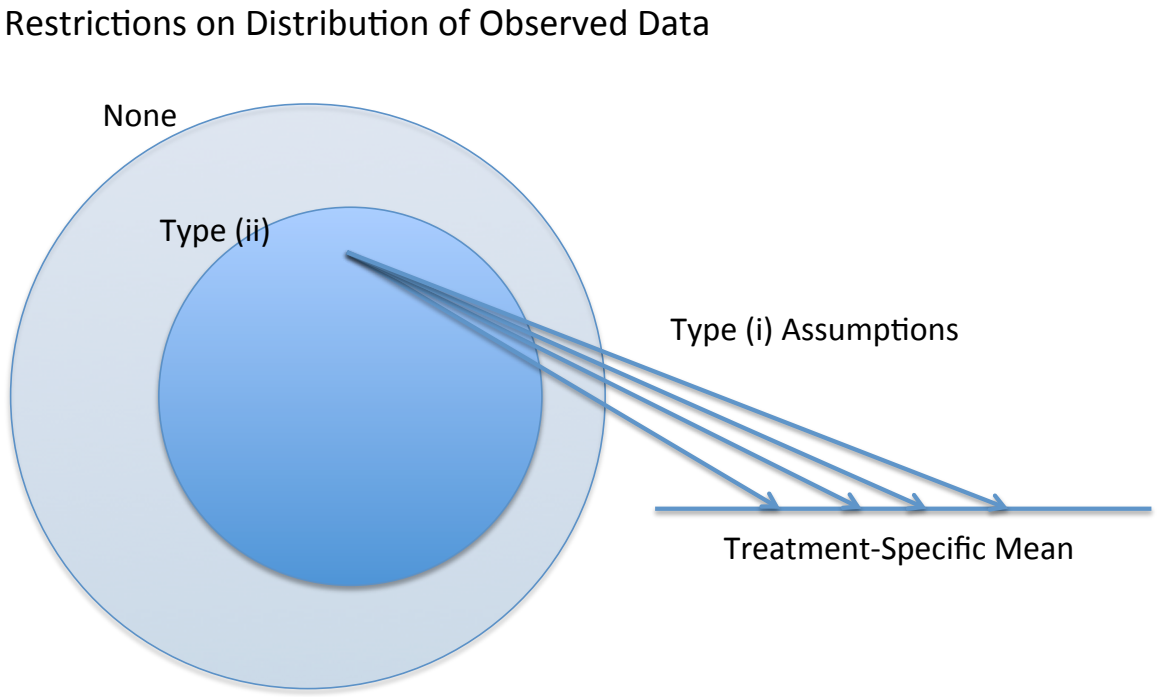
Global Sensitivity Analysis

- Evaluate robustness of results across a much broader range of assumptions that include a reasonable benchmark assumption
- Allows one to see how far one needs to deviate from the benchmark assumption in order for inferences to change.
- "Tipping point" analysis
- 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.

Global Sensitivity Analysis

- Inference about the treatment arm means requires two types of assumptions:
  - (i) *unverifiable* assumptions about the distribution of outcomes among those with missing data and
  - (ii) additional testable assumptions that serve to increase the efficiency of estimation.

Global Sensitivity Analysis



Notation

- K scheduled post-baseline assessments.
- There are $(K+1)$ patterns representing each of the visits an individual might last be seen, i.e., $0, \ldots, K$.
- The $(K+1)^{st}$ pattern represents individuals who complete the study.
- Let Y_k be the outcome scheduled to be measured at visit k, with visit 0 denoting the baseline measure (assumed to be observed).
- Let $Y_k^- = (Y_0, \ldots, Y_k)$ and $Y_k^+ = (Y_{k+1}, \ldots, Y_K)$.

Notation

- Let R_k be the indicator of being on study at visit k
- $R_0 = 1$; $R_k = 1$ implies that $R_{k-1} = 1$.
- Let C be the last visit that the patient is on-study.
- We focus inference separately for each treatment arm.
- The observed data for an individual is $O = (C, Y_C^-)$.
- We want to estimate $\mu^* = E[Y_K]$.

Missing at Random (MAR)

- For patients on study at visit k with observed history Y_k^-, the distribution of outcomes after visit k (Y_k^+) is the same for
  - those are last seen at visit k and
  - those who remain on-study
- Among those on study at visit k, the decision to drop-out before visit $k+1$ only depends on the observed history Y_k^-.
- MAR is a type (i) assumption. It is "unverifiable."
- Inference will rely on models for either
  - $f(Y_{k+1}|R_{k+1}=1, Y_k^-)$
  - $P(R_{k+1}=0 \mid R_k=1, Y_k^-)$

Missing Not at Random (MNAR)

$$\text{logit } P[R_{k+1} = 0 | R_k = 1, Y_K^-] = h_{k+1}(Y_k^-) + \alpha r(Y_{k+1})$$

where

$$\begin{aligned} h_{k+1}(Y_k^-) &= \text{logit } P[R_{k+1} = 0 | R_k = 1, Y_k^-] - \\ &\quad \log\{E[\exp\{\alpha r(Y_{k+1})\} | R_{k+1} = 1, Y_k^-]\} \end{aligned}$$

- $r(Y_{k+1})$ is a specified function of Y_{k+1}
- α is a sensitivity analysis parameter
- Each α is type (i) assumption.

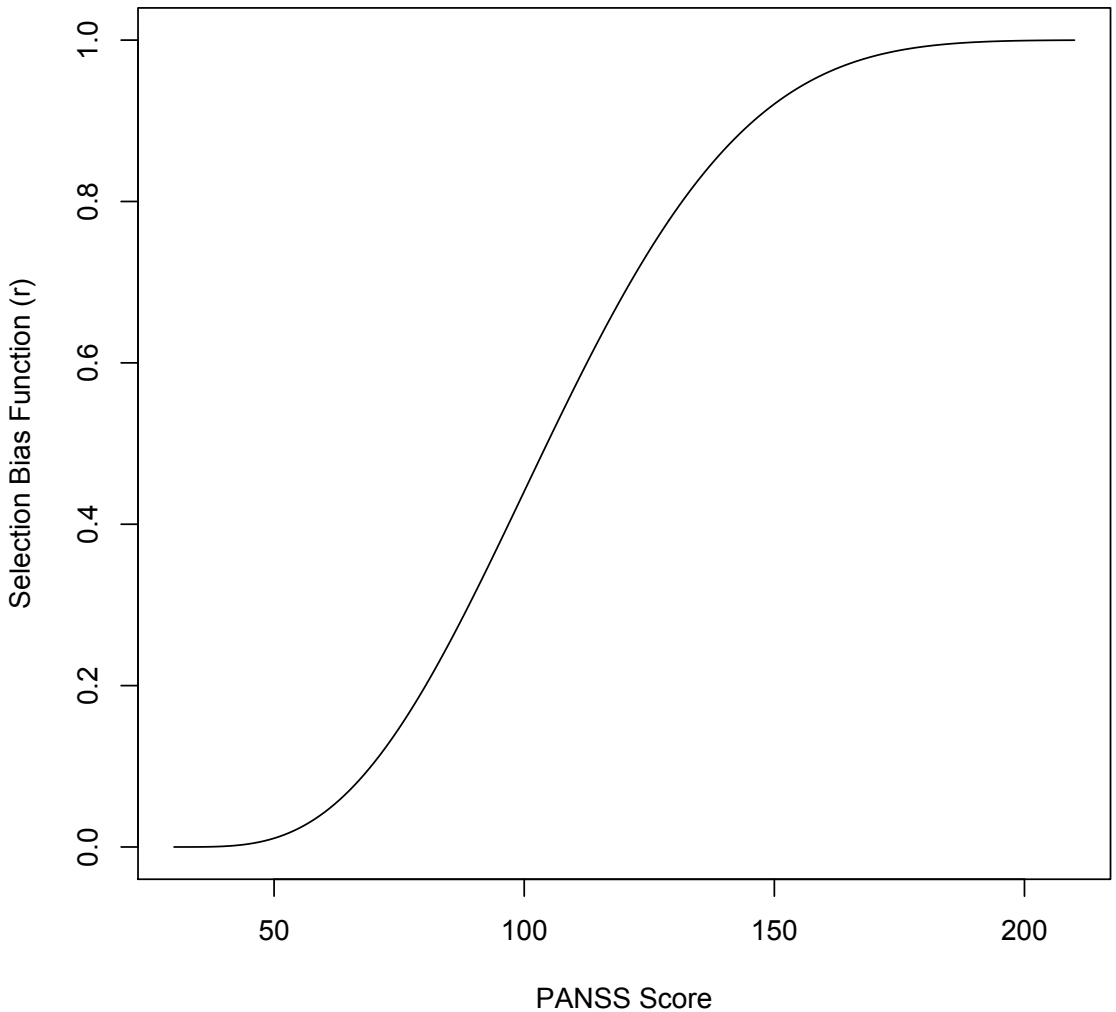
Inference

- Inference will rely on models for either
  - $f(Y_{k+1}|R_{k+1} = 1, Y_k^-)$
  - $P(R_{k+1} = 0 \mid R_k = 1, Y_k^-)$
- Impose first-order Markov assumption (Type (ii) assumption)
- Non-parametric smoothing using cross-validation
- Corrected plug-in estimator
- Confidence intervals using t-based bootstrap

Analysis

	Placebo	Test	Difference
Observed	77.97	74.19	-3.78
LOCF	84.68	84.73	0.05
MAR	83.19	80.44	-2.75

Analysis


Selection Bias Function (r)
0.0
0.2
0.4
0.6
0.8
1.0
50
100
150
200
PANSS Score

Analysis

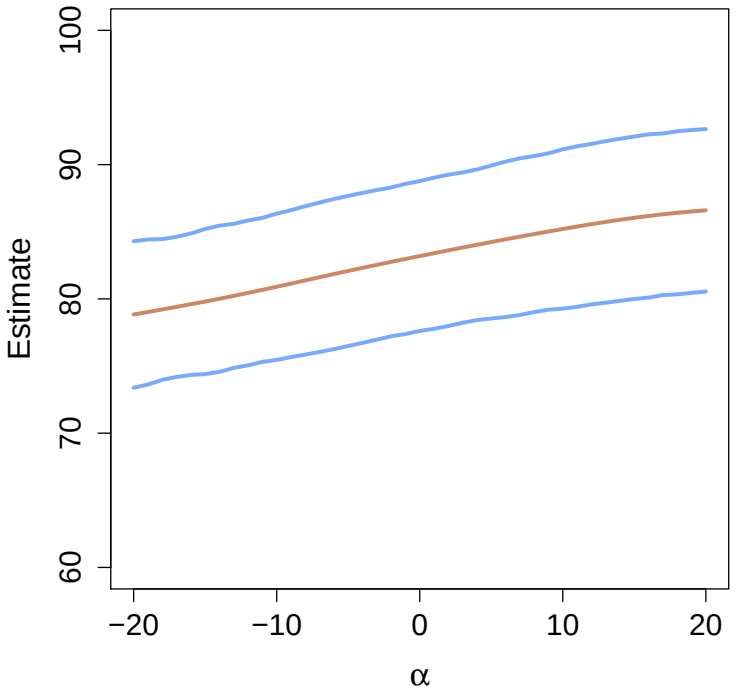


Figure: Placebo

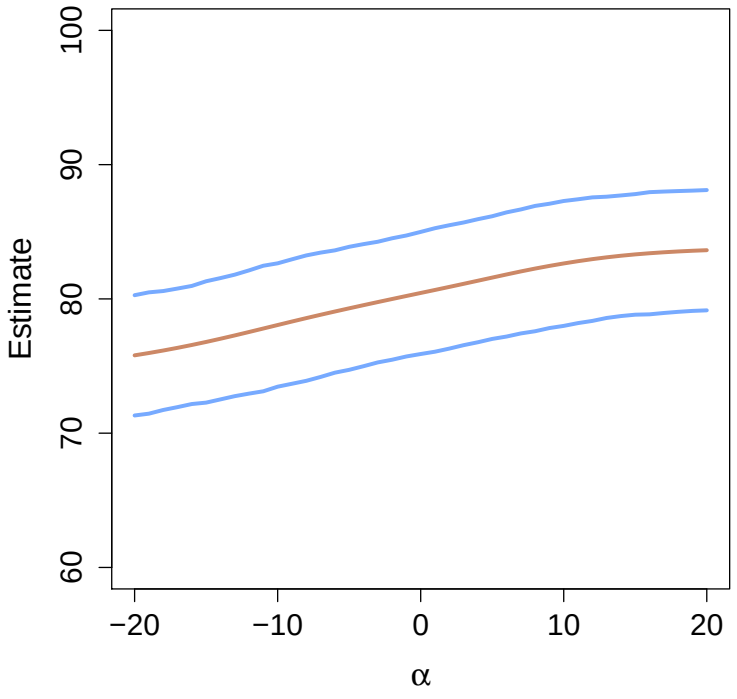
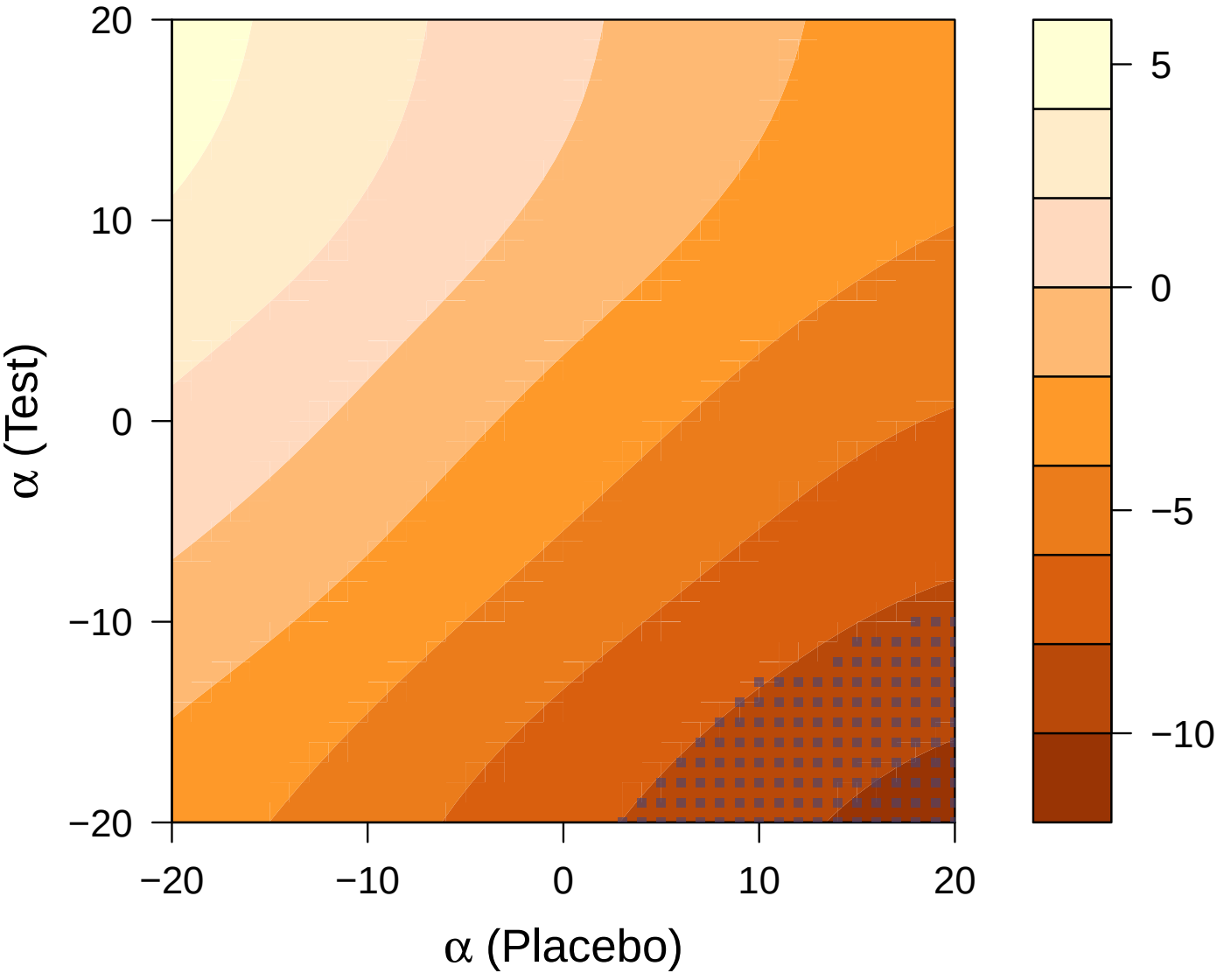


Figure: Test

Analysis



More Information

Software, Papers, Presentations

www.missingdatamatters.org