Missing Data and Sensitivity Analyses in Randomized Studies

Daniel Scharfstein
Johns Hopkins University
dscharf@jhu.edu

November 12, 2013

Missing Data Matters

- Missing outcome data are a widespread problem in randomized trials
- We reviewed all randomized trials reporting five major patient-reported outcomes published in five leading general medical journals between January 1, 2008 and March 14, 2015
  - 83.5% reported percentages greater than 10%,
  - 46.1% reported percentages greater than 20%
  - 23.1% reported percentages greater than 30%.

Missing Data Matters

- While unbiased estimates of treatment effects can be obtained from trials with no missing data, this is no longer true when data are missing on some patients.
- The essential problem is that inference about treatment effects relies on *unverifiable* assumptions about the nature of the mechanism that generates the missing data.
- While we usually know the reasons for missing data, we do not know the distribution of outcomes for patients with missing data, how it compares to that of patients with observed data and whether differences in these distributions can be explained by the observed data.

Robert Temple and Bob O'Neil (FDA)

- *"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."*

Outline

- Recommendations from 2010 NRC Report
- Sensitivity Analysis

2010 NRC Report

- The report, commissioned by the FDA, provides 18 recommendations

Recommendation 1

- The trial protocol should explicitly define (a) the objective(s) of the trial; (b) the associated primary outcome or outcomes; (c) how, when, and on whom the outcome or outcomes will be measured; and (d) the measures of intervention effects, that is, the causal estimands of primary interest.
- These measures should be meaningful for all study participants, and estimable with minimal assumptions.
- The protocol should address the potential impact and treatment of missing data.

Estimands

- 1. (Difference in) Outcome Improvement for all Randomized Patients
  - ITT
  - Interpreted as a treatment policy.
  - Parallel group, randomized trial in which outcome is collected on all patients, regardless of treatment adherence.
- 2. (Difference in) Outcome Improvement in Tolerators
  - Active treatment run-in phase, followed by placebo washout, followed by randomization
  - Outcome data collected on all patients.

Estimands

- 3. (Difference in) Outcome Improvement If All Patients Tolerated and Adhered
  - Parallel group, randomized trial in which all patients are provided adjunctive or supportive care to insure tolerability and adherence.
  - Outcome data collected on all patients.
- 4. (Difference in) Area Under the Outcome Curve During Adherence
  - Simultaneously quantifies the effect of treatment on both the outcome and the duration of tolerability or adherence in all patients.
  - No need to collect outcome data after treatment discontinuation.

Estimands

- 5. (Difference in) Outcome Improvement During Adherence to Treatment
  - Simultaneously quantifies the effect of treatment on both the outcome and the duration of tolerability or adherence in all patients.
  - No need to collect outcome data after treatment discontinuation.

Estimands

- Estimands 1, 4 and 5 may be influenced by both pharmacological efficacy and tolerance and adherence. They have the potential to be be misinterpreted.
- Estimand 5 does not distinguish between highly effective but toxic treatments from a non-toxic treatment with gradual improvement over time.

Recommendation 2

Investigators, sponsors, and regulators should design clinical trials consistent with the goal of maximizing the number of participants who are maintained on the protocol-specified intervention until the outcome data are collected.

Ideas to Limit Missing Data

- Target a population which is not adequately served by current treatments, and hence has an incentive to remain in the study.
- Include a run-in period where all patients are assigned to the active treatment, after which only individuals who tolerated and adhered to therapy are randomized to a treatment.
- Allow flexible dosing that accommodates individual differences in efficacy and tolerability, reducing the frequency of dropout for lack of efficacy or tolerability.

Ideas to Limit Missing Data

- Consider add-on designs, where the study treatment (or placebo) is added to an existing treatment, typically with a different mechanism of action known from previous studies to be effective.
- Shorten the follow-up period for the primary outcome.
- Allow rescue medications, designated as components of a treatment regimen in the study protocol.
- For assessing long-term efficacy, where dropouts are likely, consider randomized withdrawal designs so only participants who have remained on therapy are randomized (to continue or withdraw to placebo)

Ideas to Limit Missing Data

- Avoid outcome measures that are likely to lead to substantial missing data; in some cases it may be appropriate to consider time to use of rescue treatment as an outcome measure, or discontinuation of study treatment as a form of treatment failure.

Recommendation 3

Trial sponsors should continue to collect information on key outcomes on participants who discontinue their protocol-specified intervention in the course of the study, except in those cases for which a compelling cost-benefit analysis argues otherwise, and this information should be recorded and used in the analysis.

Recommendation 9

Statistical methods for handling missing data should be specified by clinical trial sponsors in study protocols, and their associated assumptions stated in a way that can be understood by clinicians.

Recommendation 10

Single imputation methods like last observation carried forward and baseline observation carried forward should not be used as the primary approach to the treatment of missing data unless the assumptions that underlie them are scientifically justified.

Recommendation 11

Parametric models in general, and random effects models in particular, should be used with caution, with all their assumptions clearly spelled out and justified. Models relying on parametric assumptions should be accompanied by goodness-of-fit procedures.

Recommendation 12

It is important that the primary analysis of the data from a clinical trial should account for the uncertainty attributable to missing data, so that under the stated missing data assumptions the associated significance tests have valid type I error rates and the confidence intervals have the nominal coverage properties.

- For inverse probability weighting and maximum likelihood methods, this can be accomplished by appropriate computation of standard errors, using either asymptotic results or the bootstrap.
- For imputation, it s necessary to use appropriate rules for multiply imputing missing responses and combining results across imputed datasets because single imputation does not account for all sources of variability.

Recommendation 13

Weighted generalized estimating equations methods should be more widely used in settings when missing at random can be well justified and a stable weight model can be determined, as a possibly useful alternative to parametric modeling.

Recommendation 14

- When substantial missing data are anticipated, auxiliary information should be collected that is believed to be associated with reasons for missing values and with the outcomes of interest. This could improve the primary analysis through use of a more appropriate missing at random model or help to carry out sensitivity analyses to assess the impact of missing data on estimates of treatment differences.
- Investigators should seriously consider following up all or a random sample of trial dropouts, who have not withdrawn consent, to ask them to indicate why they dropped out of the study, and, if they are willing, to collect outcome measurements from them.

Recommendation 15

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.

ICH, EMEA and Sensitivity Analysis

- 1998 International Conference of Harmonization (ICH) Guidance document (E9) entitled "Statistical Principles in Clinical Trials" states: "*it is important to evaluate the robustness of the results to various limitations of the data, assumptions, and analytic approaches to data analysis*"
- European Medicines Agency 2009 draft "Guideline on Missing Data in Confirmatory Clinical Trials" states "*[i]n all submissions with non-negligible amounts of missing data sensitivity analyses should be presented as support to the main analysis.*"

PCORI and Sensitivity Analysis

- In 2012, Li *et al.* issued the report "Minimal Standards in the Prevention and Handling of Missing Data in Observational and Experimental Patient Centered Outcomes Research"
- This report, commissioned by PCORI, provides 10 standards targeted at (1) design, (2) conduct, (3) analysis and (4) reporting.
- Standard 8 echoes the NRC report, stating
  - *Examining sensitivity to the assumptions about the missing data mechanism (i.e., sensitivity analysis) should be a mandatory component of the study protocol, analysis, and reporting.*

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). Higher scores worse.
- 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 (6th timepoint).

Problem: Missing Data

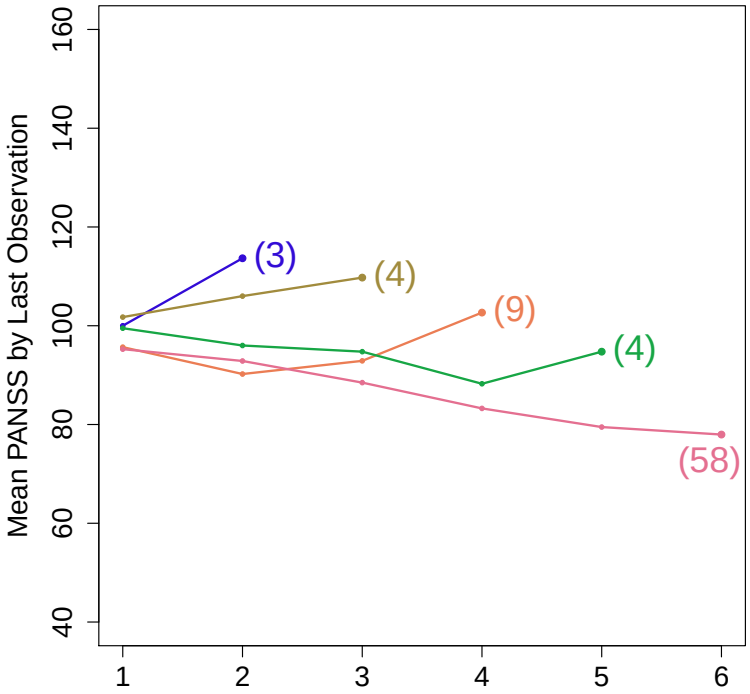


Figure: Placebo

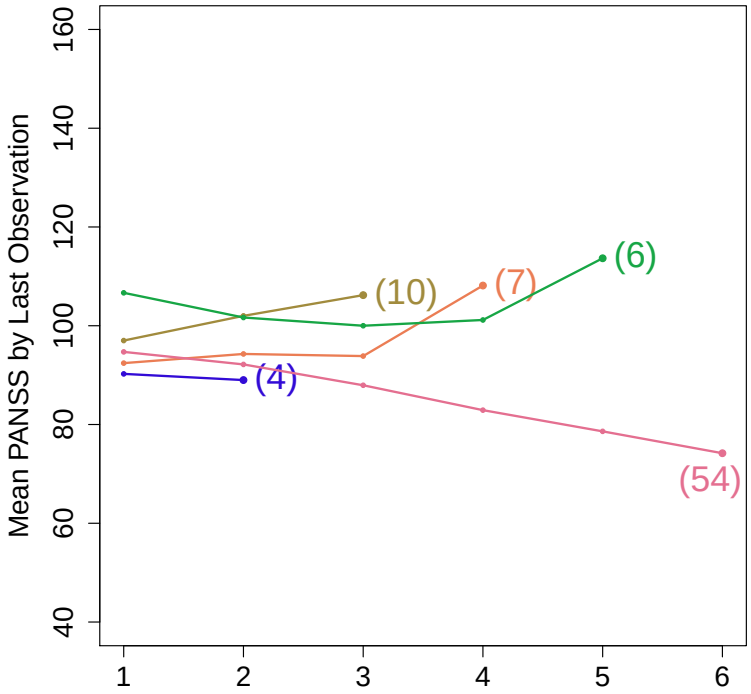


Figure: Test

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.

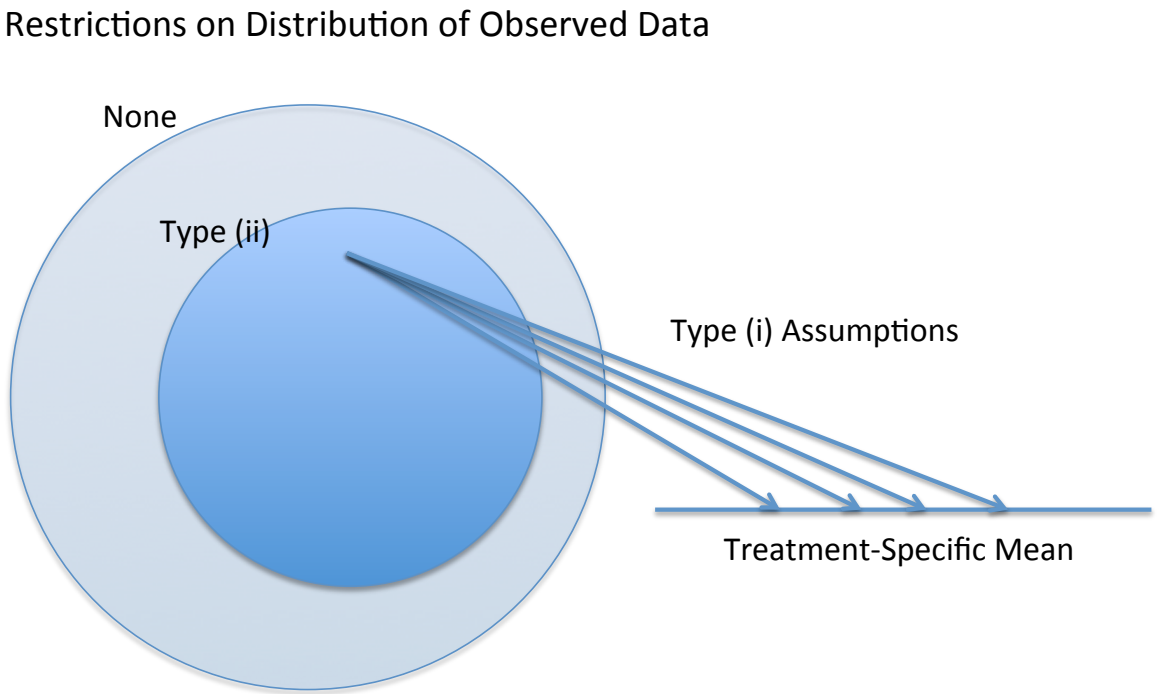
Global Sensitivity Analysis

- 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.

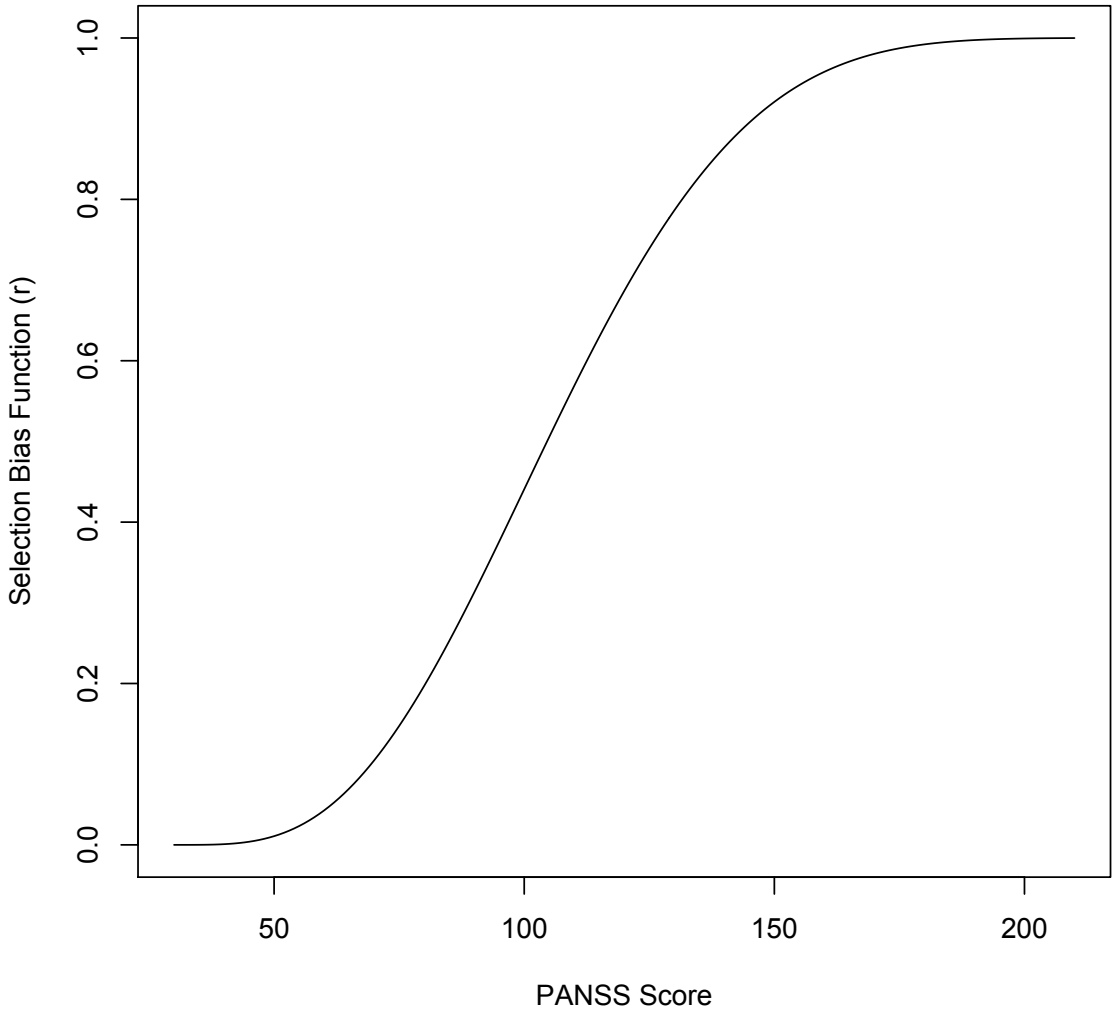
Inference

- 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

y^*_{k+1}	y_{k+1}	Log Odds Ratio
50	30	α0.02
60	40	α0.07
80	60	α0.22
100	80	α0.30
120	100	α0.24
140	120	α0.12
160	140	α0.04
180	160	α0.01
200	180	α0.00

Analysis

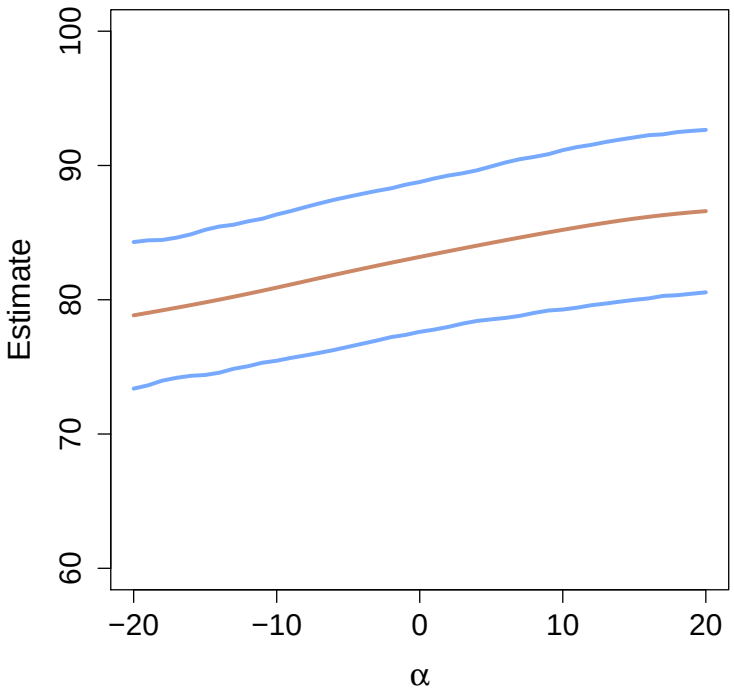


Figure: Placebo

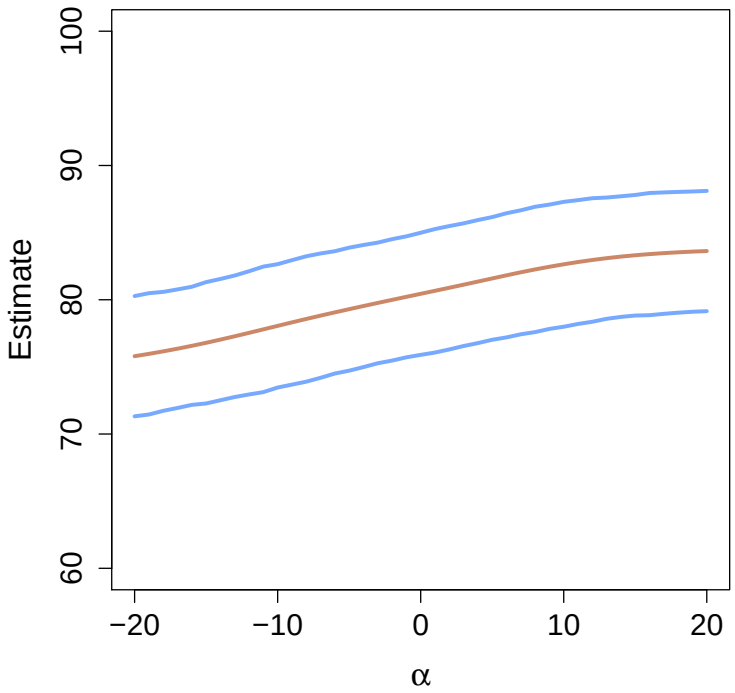
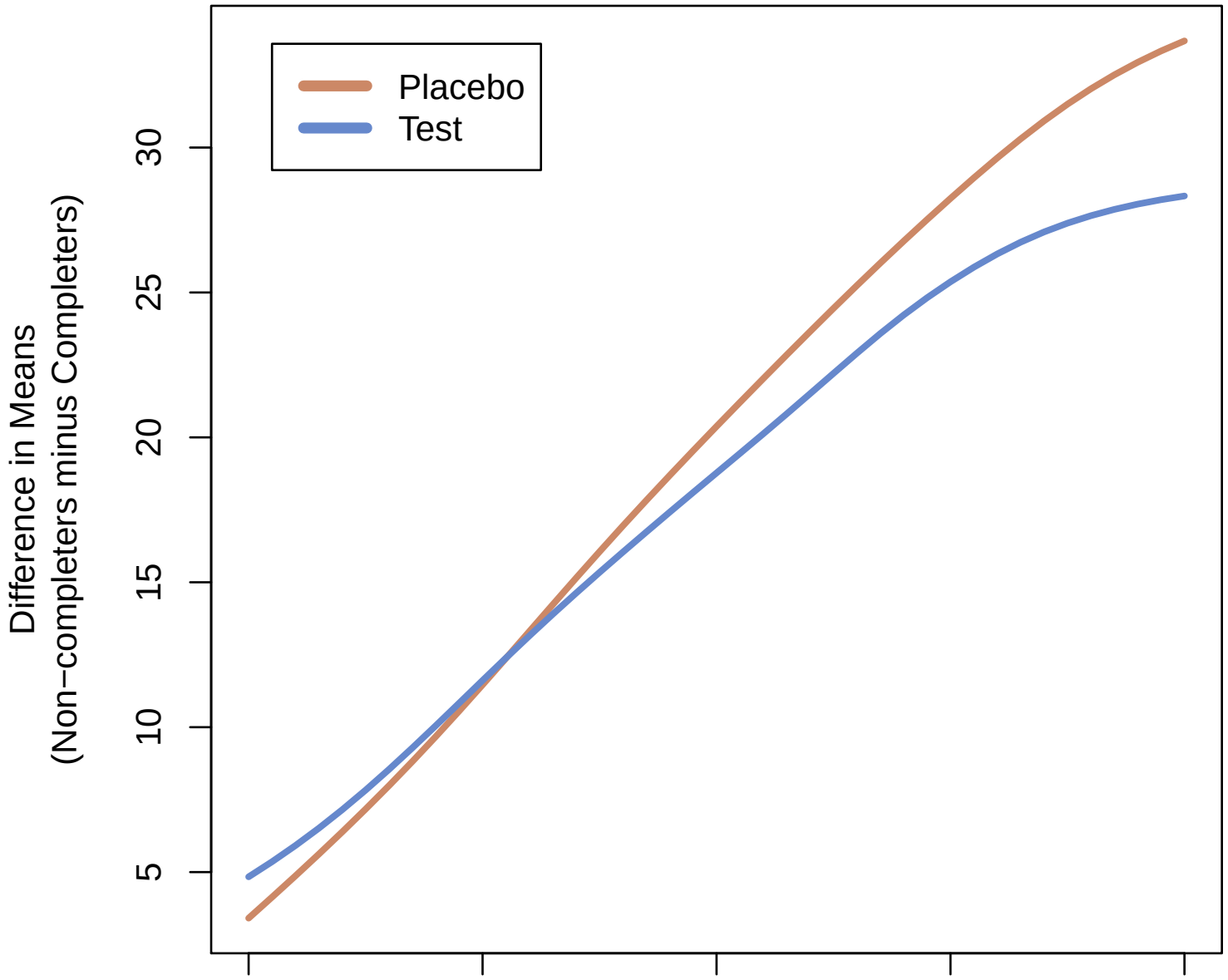
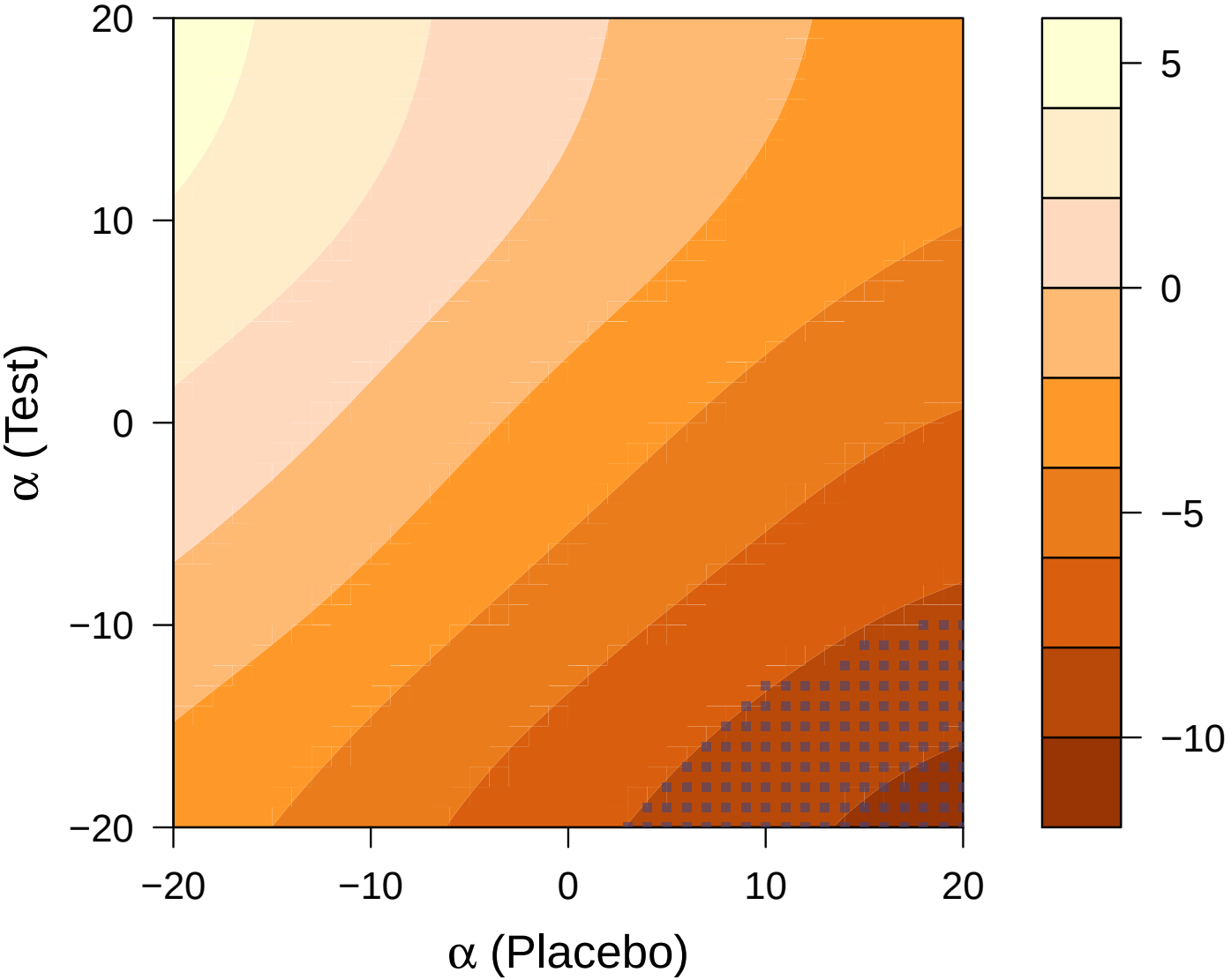


Figure: Test

Analysis


Placebo
Test
Difference in Means
(Non−completers minus Completers)
5
10
15
20
25
30

Analysis



Summary

- Missing data is a widespread problem in clinical trials
- Study design and study procedures can be employed to minimize missing data
- Sensitivity analysis

More Information

Software, Papers, Presentations

www.missingdatamatters.org

- Funded by FDA and PCORI