Global Sensitivity Analysis of Randomized Trials with Missing Data

Harvard Shortcourse

Daniel Scharfstein

Johns Hopkins University

dscharf@jhu.edu

September 24, 2019

Funding Acknowledgments

Missing Data Matters

- While unbiased estimates of treatment effects can be obtained from randomized 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.*"

NRC Report and Sensitivity Analysis

- In 2010, the National Research Council (NRC) issued a reported entitled "The Prevention and Treatment of Missing Data in Clinical Trials."
- This report, commissioned by the FDA, provides 18 recommendations targeted at (1) trial design and conduct, (2) analysis and (3) directions for future research.
- Recommendation 15 states
  - *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*"

- The recent draft Addendum to ICH-E9 confirms the importance of sensitivity 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.*

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.

- The problem with this approach is that the assumptions that underlie these methods are very strong and for many of these methods unreasonable.

- More importantly, just because the inferences are consistent does not mean that there are no other reasonable assumptions under which the inference about the treatment effect is different.

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.
- What if there are assumptions outside the local neighborhood which are plausible?

Global Sensitivity Analysis

- Evaluate robustness of results across a much broader range of assumptions that include a reasonable benchmark assumption and a collection of additional assumptions that trend toward best and worst case assumptions.
- Emphasized in Chapter 5 of the NRC report.
- This approach is substantially more informative because it operates like "stress testing" in reliability engineering, where a product is systematically subjected to increasingly exaggerated forces/conditions in order to determine its breaking point.

Global Sensitivity Analysis

- In the missing data setting, global sensitivity analysis allows one to see how far one needs to deviate from the benchmark assumption in order for inferences to change.
- "Tipping point" analysis (Yan, Lee and Li, 2009; Campbell, Pennello and Yue, 2011)
- 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

- Restrict consideration to follow-up randomized study designs that prescribe that measurements of an outcome of interest are to be taken on each study participant at fixed time-points.
- First part of course will focus on monotone missing data pattern. Second part will address how to handle intermittent missing data patterns.
- Consider the case where interest is focused on a comparison of treatment arm means at the last scheduled visit in a counterfactual world without missingness.

Case Study: Quetiapine Bipolar Trial

- Patients with bipolar disorder randomized equally to one of three treatment arms: placebo, Quetiapine 300 mg/day or Quetiapine 600 mg/day (Calabrese *et al.*, 2005).
- Randomization was stratified by type of bipolar disorder.
- Short-form version of the Quality of Life Enjoyment Satisfaction Questionnaire (QLESSF, Endicott *et al.*, 1993), was scheduled to be measured at baseline, week 4 and week 8.
- Focus on the subset of 234 patients with bipolar 1 disorder who were randomized to either the placebo (n=116) or 600 mg/day (n=118) arms.

Quetiapine Bipolar Trial

- 600 mg/day dose was titrated to achieve target by Day 8.
- In each treatment group, a dose reduction of 100 mg was allowed to improve tolerability.
- At discretion of the investigator, patients could be discontinued from study treatment and assessments at any time.
- Patients were free to discontinue their participation in the study at any time.
- Use of psychoactive drugs, except lorazepam and zolpidem tartrate during the first 3 weeks, was prohibited. Investigators were allowed to prescribe other medications for the safety and well-being of the participant.

- Only 65 patients (56%) in placebo arm and 68 patients (58%) in the 600mg/day arm had a complete set of QLESSF scores.
- Patients with complete data tend to have higher average QLESSF scores, suggesting that a complete-case analysis could be biased.

Observed Data

Figure: Treatment-specific (left: placebo; right: 600 mg/day Quetiapine) trajectories of mean QLESSF scores, stratified by last available measurement.
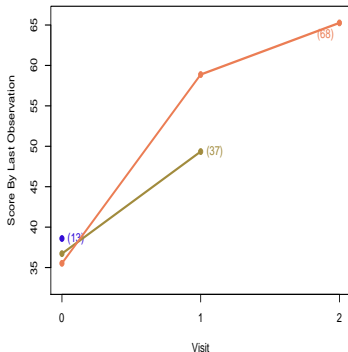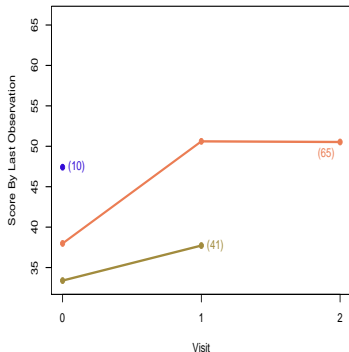
What is the difference in the mean QLESSF score at week 8 between Quetiapine 600 mg/day and placebo in the counterfactual world in which all patients were followed to that week?

Imagination

- Validity of assumptions will depend on what is *imagined* about treatments that patients receive off-study.
- *Not* imagining the continuation of assigned treatment after occurrence of intolerable side effects or lack of efficacy.
- Imagining that patients receive treatment as close to the assigned treatment as ethically possible.
- The difference of the treatment-specific mean QLESSF outcomes at week 8 under this imaginary, yet plausible, treatment scenario is the target *estimand* of interest.

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

- Type (i) assumptions are necessary to identify the treatment-specific means.
- By *identification*, we mean that we can write it as a function that depends only on the distribution of the observed data.
- When a parameter is identified we can hope to estimate it as precisely as we desire with a sufficiently large sample size,
- In the absence of identification, statistical inference is fruitless as we would be unable to learn about the true parameter value even if the sample size were infinite.

Global Sensitivity Analysis

- To address the identifiability issue, it is essential to conduct a sensitivity analysis, whereby the data analysis is repeated under different type (i) assumptions, so as to investigate the extent to which the conclusions of the trial are dependent on these subjective, unverifiable assumptions.

- The usefulness of a sensitivity analysis ultimately depends on the plausibility of the unverifiable assumptions.

- It is key that any sensitivity analysis methodology allow the formulation of these assumptions in a transparent and easy to communicate manner.

Global Sensitivity Analysis

- There are an infinite number of ways of positing type (i) assumptions.
- Ultimately, however, these assumptions prescribe how missing outcomes should be "imputed."
- A reasonable way to posit these assumptions is to
  - 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 with the distribution of the outcome among those with the observed outcome and who share the same recorded data.

Global Sensitivity Analysis

- Type (i) assumptions will not suffice when the repeated outcomes are continuous or categorical with many levels. This is because of *data sparsity*.

- For example, the stratum of people who share the same recorded data will typically be small. As a result, it is necessary to draw strength across strata by "smoothing."

- Without smoothing, the data analysis will rarely be informative because the uncertainty concerning the treatment arm means will often be too large to be of substantive use.

- As a result, it is necessary to impose type (ii) smoothing assumptions.

- Type (ii) assumptions should be scrutinized with standard model checking techniques.

Global Sensitivity Analysis

- The global sensitivity framework proceeds by parameterizing (i.e., indexing) the connections (i.e., type (i) assumptions) via sensitivity analysis parameters.
- The parameterization is configured so that a specific value of the sensitivity analysis parameters (typically set to zero) corresponds to a benchmark connection that is considered reasonably plausible and sensitivity analysis parameters further from the benchmark value represent more extreme departures from the benchmark connection.

Global Sensitivity Analysis

- The global sensitivity analysis strategy that we propose is focused on separate inferences for each treatment arm, which are then combined to evaluate treatment effects.
- Until later, our focus will be on estimation of the mean outcome at week 8 (in a world without missing outcomes) for one of the treatment groups and we will suppress reference to treatment assignment.

Notation: Quetiapine Bipolar Trial

- Y_0, Y_1, Y_2: QLESSF scores scheduled to be collected at baseline, week 4 and week 8.
- Let R_k be the indicator that Y_k is observed.
- We assume $R_0 = 1$ and that $R_k = 0$ implies $R_{k+1} = 0$ (i.e., missingness is monotone).
- Patient is on-study at visit k if $R_k = 1$
- Patient discontinued prior to visit k if $R_k = 0$
- Patient last seen at visit $k - 1$ if $R_{k-1} = 1$ and $R_k = 0$.
- Y_k^{obs} equals to Y_k if $R_k = 1$ and equals to *nil* if $R_k = 0$.

- The observed data for an individual are

$$O = (Y_0, R_1, Y_1^{obs}, R_2, Y_2^{obs}),$$

which has some distribution P^* contained within a set of distributions $\mathcal{M}$ (type (ii) assumptions discussed later).

- The superscript $*$ will be used to denote the true value of the quantity to which it is appended.

- Any distribution $P \in \mathcal{M}$ can be represented in terms of the following distributions:

  - $f(Y_0)$
  - $P(R_1 = 1 | Y_0)$
  - $f(Y_1 | R_1 = 1, Y_0)$
  - $P(R_2 = 1 | R_1 = 1, Y_1, Y_0)$
  - $f(Y_2 | R_2 = 1, Y_1, Y_0)$.

- We assume that n independent and identically distributed copies of O are observed.
- The goal is to use these data to draw inference about $\mu^* = E^*[Y_2]$.
- When necessary, we will use the subscript i to denote data for individual i.

Benchmark Assumption (Missing at Random)

- $A_0(y_0)$: patients last seen at visit 0 ($R_0 = 1, R_1 = 0$) with $Y_0 = y_0$.
- $B_1(y_0)$: patients on-study at visit 1 ($R_1 = 1$) with $Y_0 = y_0$.
- $A_1(y_0, y_1)$: patients last seen at visit 1 ($R_1 = 1, R_2 = 0$) with $Y_0 = y_0$ and $Y_1 = y_1$.
- $B_2(y_0, y_1)$: patients who complete study ($R_2 = 1$) with $Y_0 = y_0$ $Y_1 = y_1$.

Missing at random posits the following type (i) "linking" assumptions:

- For each y_0, the distribution of Y_1 and Y_2 is the same for those in stratum $A_0(y_0)$ as those in stratum $B_1(y_0)$.
- For each y_0, y_1, the distribution of Y_2 is the same for those in stratum $A_1(y_0, y_1)$ as those in stratum $B_2(y_0, y_1)$.

Mathematically, we can express these assumptions as follows:

$$f^*(Y_1, Y_2|A_0(y_0)) = f^*(Y_1, Y_2|B_1(y_0)) \text{ for all } y_0 \qquad (1)$$

and

$$f^*(Y_2|A_1(y_0, y_1)) = f^*(Y_2|B_2(y_0, y_1)) \text{ for all } y_0, y_1 \qquad (2)$$

Benchmark Assumption (Missing at Random)

Using Bayes' rule, we can re-write these expressions as:

$$P^*(R_1 = 0 | R_0 = 1, Y_0 = y_0, Y_1 = y_1, Y_2 = y_2)$$
$$= P^*(R_1 = 0 | R_0 = 1, Y_0 = y_0)$$

and

$$P^*(R_2 = 0 | R_1 = 1, Y_0 = y_0, Y_1 = y_1, Y_2 = y_2)$$
$$= P^*(R_2 = 0 | R_1 = 1, Y_0 = y_0, Y_1 = y_1)$$

Missing at random implies:

- The decision to discontinue the study before visit 1 is like the flip of a coin with probability depending on the value of the outcome at visit 0.
- For those on-study at visit 1, the decision to discontinue the study before visit 2 is like the flip of a coin with probability depending on the value of the outcomes at visits 1 and 0.

Benchmark Assumption (Missing at Random)

- MAR is a type (i) assumption. It is "unverifiable."

- For patients last seen at visit k, we cannot learn from the observed data about the conditional (on observed history) distribution of outcomes after visit k.

- For patients last seen at visit k, any assumption that we make about the conditional (on observed history) distribution of the outcomes after visit k will be unverifiable from the data available to us.

- For patients last seen at visit k, the assumption that the conditional (on observed history) distribution of outcomes after visit k is the same as those who remain on-study after visit k is unverifiable.

Under MAR, μ^* is identified. That is, it can be expressed as a function of the distribution of the observed data. Specifically,

$$\mu^* = \mu(P^*) = \int_{y_0} \int_{y_1} \int_{y_2} y_2 \, dF_2^*(y_2|y_1, y_0) \, dF_1^*(y_1|y_0) \, dF_0^*(y_0)$$

where

- $F_2^*(y_2|y_1, y_0) = P^*(Y_2 \leq y_2 | B_2(y_1, y_0))$
- $F_1^*(y_1|y_0) = P^*(Y_1 \leq y_1 | B_1(y_0))$
- $F_0^*(y_0) = P^*(Y_0 \leq y_0)$.

The MAR assumption is not the only one that is (a) unverifiable and (b) allows identification of the mean of Y_2.

Missing Not at Random (MNAR)

The first part of the MAR assumption (see (1) above) is

$$f^*(Y_1, Y_2|A_0(y_0)) = f^*(Y_1, Y_2|B_1(y_0)) \text{ for all } y_0$$

It is equivalent to

$$f^*(Y_2|A_0(y_0), Y_1 = y_1)$$
$$= f^*(Y_2|B_1(y_0), Y_1 = y_1) \text{ for all } y_0, y_1 \quad (3)$$

and

$$f^*(Y_1|A_0(y_0)) = f^*(Y_1|B_1(y_0)) \text{ for all } y_0 \quad (4)$$

Missing Not at Random (MNAR)

In building a class of MNAR models, we will retain (3):

- For all y_0, y_1, the distribution of Y_2 for patients in stratum $A_0(y_0)$ with $Y_1 = y_1$ is the same as the distribution of Y_2 for patients in stratum $B_1(y_0)$ with $Y_1 = y_1$.

- The decision to discontinue the study before visit 1 is independent of Y_2 (i.e., the future outcome) after conditioning on the Y_0 (i.e., the past outcome) and Y_1 (i.e., the most recent outcome).

- *Non-future dependence* (Diggle and Kenward, 1994)

Missing Not at Random (MNAR)

Generalizing (4) using Exponential Tilting

$$f^*(Y_1|A_0(y_0))$$
$$\propto f^*(Y_1|B_1(y_0)) \exp\{\alpha r(Y_1)\} \text{ for all } y_0 \quad (5)$$

Generalizing (2) using Exponential Tilting

$$f^*(Y_2|A_1(y_0, y_1))$$
$$\propto f^*(Y_2|B_2(y_0, y_1)) \exp\{\alpha r(Y_2)\} \text{ for all } y_0, y_1 \quad (6)$$

- $r(y)$ is a specified increasing function; α is a sensitivity analysis parameter.
- $\alpha = 0$ is MAR.

Missing Not at Random (MNAR)

When $\alpha > 0 \; (< 0)$

- For each y_0, the distribution of Y_1 for patients in stratum $A_0(y_0)$ is weighted more heavily to higher (lower) values than the distribution of Y_1 for patients in stratum $B_1(y_0)$.

- For each y_0, y_1, the distribution of Y_2 for patients in stratum $A_1(y_0, y_1)$ is weighted more heavily to higher (lower) values than the distribution of Y_2 for patients in stratum $B_2(y_0, y_1)$.

The amount of "tilting" increases with the magnitude of α.

Missing Not at Random (MNAR)

Using Bayes' rule, we can re-write (3), (5) and (6) as:

$$\text{logit } P^*(R_1 = 0 | R_0 = 1, Y_0 = y_0, Y_1 = y_1, Y_2 = y_2)$$
$$= l_1^*(y_0; \alpha) + \alpha r(y_1)$$

and

$$\text{logit } P^*(R_2 = 0 | R_1 = 1, Y_0 = y_0, Y_1 = y_1, Y_2 = y_2)$$
$$= l_2^*(y_0, y_1; \alpha) + \alpha r(y_2)$$

where

$$
\begin{aligned}
l_1^*(y_0; \alpha) &= \text{logit } P^*(R_1 = 0 | R_0 = 1, Y_0 = y_0) - \\
&\qquad \log E^*(\exp\{\alpha r(Y_1)\} | B_1(y_0))
\end{aligned}
$$

and

$$
\begin{aligned}
l_2^*(y_1, y_0; \alpha) &= \text{logit } P^*(R_2 = 0 | R_1 = 1, Y_0 = y_0, Y_1 = y_1) - \\
&\qquad \log E^*(\exp\{\alpha r(Y_2)\} | B_2(y_1, y_0))
\end{aligned}
$$

Missing Not at Random (MNAR)

Written in this way:

- The decision to discontinue the study before visit 1 is like the flip of a coin with probability depending on the value of the outcome at visit 0 *and (in a specified way)* the value of the outcome at visit 1.

- For those on-study at visit 1, the decision to discontinue the study before visit 2 is like the flip of a coin with probability depending on the value of the outcomes at visits 0 and 1 *and (in a specified way)* the value of the outcome at visit 2.

Exponential Tilting Explained

$$f(Y|R = 0) \propto f(Y|R = 1) \exp\{\alpha r(Y)\}$$

- If $[Y|R = 1] \sim N(\mu, \sigma^2)$ and $r(Y) = Y$,
  $[Y|R = 0] \sim N(\mu + \alpha\sigma^2, \sigma^2)$
- If $[Y|R = 1] \sim Beta(a, b)$ and $r(Y) = \log(Y)$,
  $[Y|R = 0] \sim Beta(a + \alpha, b)$, $\alpha > -a$.
- If $[Y|R = 1] \sim Gamma(a, b)$ and $r(Y) = \log(Y)$,
  $[Y|R = 0] \sim Gamma(a + \alpha, b)$, $\alpha > -a$.
- If $[Y|R = 1] \sim Gamma(a, b)$ and $r(Y) = Y$,
  $[Y|R = 0] \sim Gamma(a, b - \alpha)$, $\alpha < b$.
- If $[Y|R = 1] \sim Bernoulli(p)$ and $r(Y) = Y$,
  $[Y|R = 0] \sim Bernoulli\left(\frac{p\exp(\alpha)}{p\exp(\alpha)+1-p}\right)$.

Beta
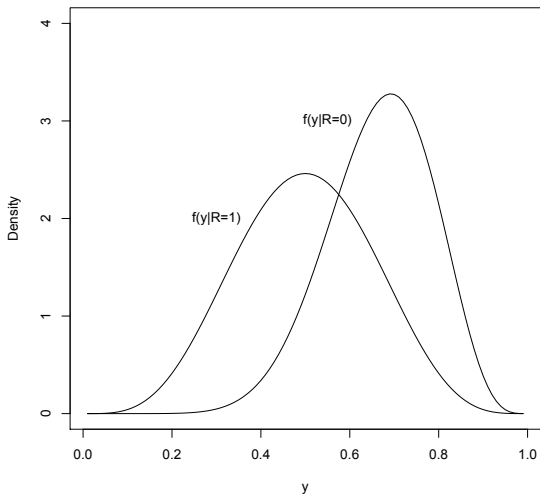
Gamma
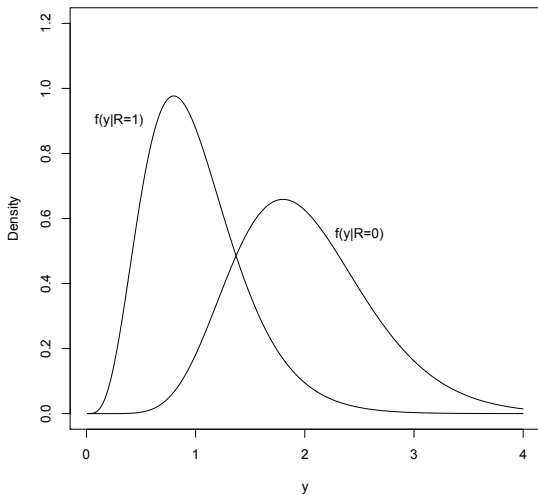
Missing Not at Random (MNAR)

For given α, μ^* is identified. Specifically, $\mu^* = \mu(P^*; \alpha)$ equals

$$\int_{y_0} \int_{y_1} \int_{y_2} y_2 \left\{ dF_2^*(y_2|y_1, y_0)\{1 - H_2^*(y_1, y_0)\} + \frac{dF_2^*(y_2|y_1, y_0)\exp\{\alpha r(y_2)\}}{\int_{y_2'} dF_2^*(y_2'|y_1, y_0)\exp\{\alpha r(y_2')\}} H_2^*(y_1, y_0) \right\} \times$$

$$\left\{ dF_1^*(y_1|y_0)\{1 - H_1^*(y_0)\} + \frac{dF_1^*(y_1|y_0)\exp\{\alpha r(y_1)\}}{\int_{y_1'} dF_1^*(y_1'|y_0)\exp\{\alpha r(y_1')\}} H_1^*(y_0) \right\} dF_0^*(y_0)$$

where

$$H_2^*(y_1, y_0) = P^*(R_2 = 0|R_1 = 1, Y_1 = y_1, Y_0 = y_0)$$

and

$$H_1^*(y_0) = P^*(R_1 = 0|R_0 = 1, Y_0 = y_0)$$

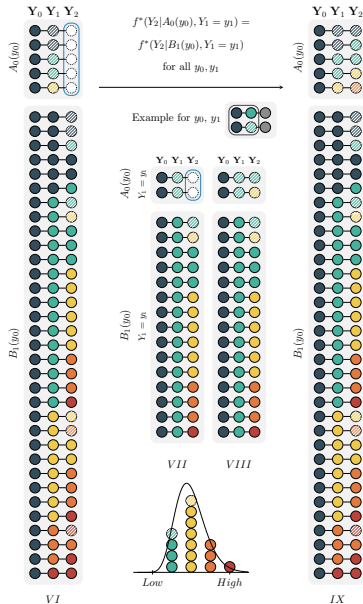- μ^* is written as a function of the distribution of the observed data (depending on α).

Inference
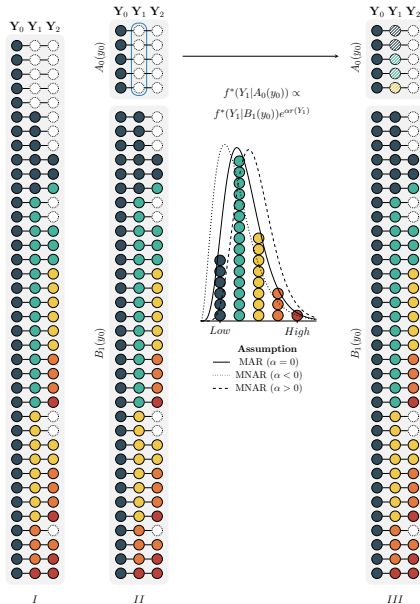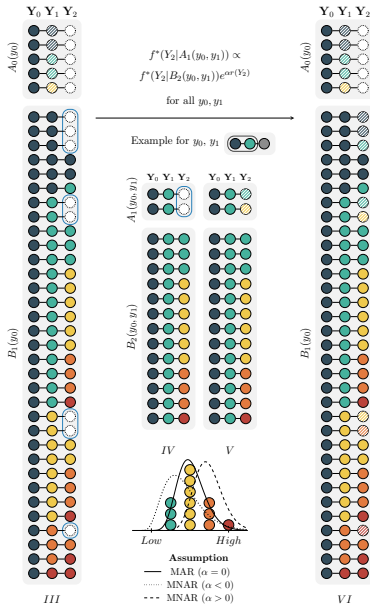
For given α, the above formula shows that μ^* depends on

- $F_2^*(y_2|y_1, y_0) = P^*(Y_2 \leq y_2|B_2(y_1, y_0))$
- $F_1^*(y_1|y_0) = P^*(Y_1 \leq y_1|B_1(y_0))$
- $H_2^*(y_1, y_0) = P^*(R_2 = 0|R_1 = 1, Y_1 = y_1, Y_0 = y_0)$
- $H_1^*(y_0) = P^*(R_1 = 0|R_0 = 1, Y_0 = y_0)$.

It is natural to consider estimating μ^* by "plugging in" estimators of these quantities.

How can we estimate these latter quantities? With the exception of $F_0^*(y_0)$, it is tempting to think that we can use non-parametric procedures to estimate these quantities.

Inference

A non-parametric estimate of $F_2^*(y_2|y_1, y_0)$ would take the form:

$$\widehat{F}_2(y_2|y_1, y_0) = \frac{\sum_{i=1}^n R_{2,i} I(Y_{2,i} \leq y_2) I(Y_{1,i} = y_1, Y_{0,i} = y_0)}{\sum_{i=1}^n R_{2,i} I(Y_{1,i} = y_1, Y_{0,i} = y_0)}$$

- This estimator will perform very poorly (i.e., have high levels of uncertainty in moderate sample sizes) because the number of subjects who complete the study (i.e., $R_2 = 1$) and are observed to have outcomes at visits 1 and 0 exactly equal to y_1 and y_0 will be very small and can only be expected to grow very slowly as the sample size increases.

- As a result, a a plug-in estimator of μ^* that uses such non-parametric estimators will perform poorly.

We make the estimation task slightly easier by assuming that

$$F_2^*(y_2|y_1, y_0) = F_2^*(y_2|y_1) \qquad (7)$$

and

$$H_2^*(y_1, y_0) = H_2^*(y_1) \qquad (8)$$

Inference - Kernel Smoothing

Estimate $F_2^*(y_2|y_1)$, $F_1^*(y_1|y_0)$, $H_2^*(y_1)$ and $H_1^*(y_0)$ using kernel smoothing techniques.

To motivate this idea, consider the following non-parametric estimate of $F_2^*(y_2|y_1)$

$$\widehat{F}_2(y_2|y_1) = \frac{\sum_{i=1}^n R_{2,i} I(Y_{2,i} \leq y_2) I(Y_{1,i} = y_1)}{\sum_{i=1}^n R_{2,i} I(Y_{1,i} = y_1)}$$

- This estimator will still perform poorly, although better than $\widehat{F}_2(y_2|y_1, y_0)$.
- Replace $I(Y_{1,i} = y_1)$ by $\phi\left(\frac{Y_{1,i} - y_1}{\sigma_{F_2}}\right)$, where $\phi(\cdot)$ is standard normal density and σ_{F_2} is a tuning parameter.

$$\widehat{F}_2(y_2|y_1; \sigma_{F_2}) = \frac{\sum_{i=1}^n R_{2,i} I(Y_{2,i} \leq y_2) \phi\left(\frac{Y_{1,i} - y_1}{\sigma_{F_2}}\right)}{\sum_{i=1}^n R_{2,i} \phi\left(\frac{Y_{1,i} - y_1}{\sigma_{F_2}}\right)}$$

Inference - Kernel Smoothing

- This estimator allows *all* completers to contribute, not just those with Y_1 values equal to y_1
- It assigns weight to completers according to how far their Y_1 values are from y_1, with closer values assigned more weight.
- The larger σ_{F_2}, the larger the influence of values of Y_1 further from y_1 on the estimator.
- As $\sigma_{F_2} \to \infty$, the contribution of each completer to the estimator becomes equal, yielding bias but low variance.
- As $\sigma_{F_2} \to 0$, only completers with Y_1 values equal to y_1 contribute, yielding low bias but high variance.

Inference - Cross-Validation

To address the bias-variance trade-off, cross validation is typically used to select σ_{F_2}.

- Randomly divide dataset into J (typically, 10) approximately equal sized validation sets.
- Let V_j be the indices of the patients in jth validation set.
- Let n_j be the associated number of subjects.
- Let $\widehat{F}_2^{(j)}(y_2|y_1; \sigma_{F_2})$ be the estimator of $F_2^*(y_2|y_1)$ based on the dataset that excludes the jth validation set.
- If σ_{F_2} is a good choice then one would expect

$$CV_{F_2^*(\cdot|\cdot)}(\sigma_{F_2}) = \frac{1}{J} \sum_{j=1}^{J} \left\{ \frac{1}{n_j} \sum_{i \in V_j} R_{2,i} \underbrace{\int \left\{ I(Y_{2,i} \leq y_2) - \widehat{F}_2^{(j)}(y_2|Y_{1,i}; \sigma_{F_2}) \right\}^2 d\widehat{F}_2^\circ(y_2)}_{\text{Distance for } i \in V_j} \right\}$$

will be small, where $\widehat{F}_2^\circ(y_2)$ is the empirical distribution of Y_2 among subjects on-study at visit 2.

Inference - Cross-Validation

- For each individual i in the jth validation set with an observed outcome at visit 2, we measure, by the quantity above the horizontal brace, the distance (or loss) between the collection of indicator variables $\{I(Y_{2,i} \leq y_2) : d\widehat{F}_2^{\circ}(y_2) > 0\}$ and the corresponding collection of predicted values $\{\widehat{F}_2^{(j)}(y_2|Y_{1,i}; \sigma_{F_2}) : d\widehat{F}_2^{\circ}(y_2) > 0\}$.

- The distances for each of these individuals are then summed and divided by the number of subjects in the jth validation set.

- An average across the J validation/training sets is computed.

- We can then estimate $F_2^*(y_2|y_1)$ by $\widehat{F}_2(y_2|y_1; \widehat{\sigma}_{F_2})$, where $\widehat{\sigma}_{F_2} = \text{argmin } CV_{F_2^*(\cdot|\cdot)}(\sigma_{F_2})$.

We use similar ideas to estimate

- $F_1^*(y_1|y_0)$
- $H_2^*(y_1)$
- $H_1^*(y_0)$

In our software, we set $\sigma_{F_2} = \sigma_{F_1} = \sigma_F$ and minimize a single CV function. The software refers to this smoothing parameter as σ_Q.

In our software, we set $\sigma_{H_2} = \sigma_{H_1} = \sigma_H$ and minimize a single CV function. The software refers to this smoothing parameter as σ_P.

Inference - Potential Problem

- The cross-validation procedure for selecting tuning parameters achieves optimal finite-sample bias-variance trade-off for the quantities requiring smoothing.
- This optimal trade-off is usually not optimal for estimating μ^*.
- The plug-in estimator of μ^* could possibly suffer from excessive and asymptotically non-negligible bias due to inadequate tuning.
- This may prevent the plug-in estimator from having regular asymptotic behavior, upon which statistical inference is generally based.
- The resulting estimator may have a slow rate of convergence, and common methods for constructing confidence intervals, such as the Wald and bootstrap intervals, can have poor coverage properties.

Inference - Correction Procedure

- Let $\mathcal{M}$ be the class of distributions for the observed data O that satisfy constraints (7) and (8).
- For $P \in \mathcal{M}$, it can be shown that

$$\mu(P; \alpha) - \mu(P^*; \alpha)$$
$$= -E^*[\psi_P(O; \alpha) - \psi_{P^*}(O; \alpha)] + \text{Rem}(P, P^*; \alpha), \quad (9)$$

  where $\psi_P(O; \alpha)$ is a "derivative" of $\mu(\cdot; \alpha)$ at P and $\text{Rem}(P, P^*; \alpha)$ is a "second-order" remainder term which converges to zero as P tends to P^*.
- The derivative is used to quantify the change in $\mu(P; \alpha)$ resulting from small perturbations in P; it also has mean zero (i.e., $E^*[\psi_{P^*}(O; \alpha)] = 0$).
- The remainder term is second order in the sense that it can be written as or bounded by the product of terms involving differences between (functionals of) P and P^*.

Inference - Correction Procedure

Equation (9) plus some simple algebraic manipulation teaches us that

$$\underbrace{\mu(\widehat{P};\alpha)}_{\text{Plug-in}} - \mu(P^*;\alpha)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\psi_{P^*}(O_i;\alpha) - \frac{1}{n}\sum_{i=1}^{n}\psi_{\widehat{P}}(O_i;\alpha) \tag{10}$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\{\psi_{\widehat{P}}(O_i;\alpha) - \psi_{P^*}(O_i;\alpha) - E^*[\psi_{\widehat{P}}(O;\alpha) - \psi_{P^*}(O;\alpha)]\} \tag{11}$$

$$+ \text{Rem}(\widehat{P},P^*;\alpha) \tag{12}$$

where $\widehat{P}$ is the estimated distribution of P^* discussed in the previous section.

Inference - Correction Procedure

- Under smoothness and boundedness conditions, term (11) will be $o_{P^*}(n^{-1/2})$ (i.e., will converge in probabity to zero even when it is multipled by $\sqrt{n}$).

- Provided $\widehat{P}$ converges to P^* at a reasonably fast rate, term (12) will also be $o_{P^*}(n^{-1/2})$.

- The second term in (10) prevents us from concluding that the plug-in estimator can be essentially represented as an average of i.i.d terms plus $o_{P^*}(n^{-1/2})$ terms.

- By adding the second term in (10) to the plug-in estimator, we can construct a "corrected" estimator that does have this representation.

Inference - Correction Procedure

The corrected estimator is

$$\tilde{\mu}_\alpha = \underbrace{\mu(\widehat{P}; \alpha)}_{\text{Plug-in}} + \frac{1}{n} \sum_{i=1}^{n} \psi_{\widehat{P}}(O_i; \alpha)$$

The practical implication is that $\tilde{\mu}_\alpha$ converges in probability to μ^* and

$$\sqrt{n}(\tilde{\mu}_\alpha - \mu^*) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{P^*}(O_i; \alpha) + o_{P^*}(1)$$

With this representation, we see that $\psi_{P^*}(O; \alpha)$ is the so-called influence function.

Inference - Correction Procedure

- By the central limit theorem, we then know that $\sqrt{n}\,(\tilde{\mu}_\alpha - \mu^*)$ converges to a normal random variable with mean 0 and variance $\sigma_\alpha^2 = E^*[\psi_{P^*}(O; \alpha)^2]$.
- The asymptotic variance can be estimated by $\tilde{\sigma}_\alpha^2 = \frac{1}{n}\sum_{i=1}^n \psi_{\hat{P}}(O_i; \alpha)^2$.
- A $(1-\gamma)\%$ Wald-based confidence interval for μ^* can be constructed as $\tilde{\mu}_\alpha \pm z_{1-\gamma/2}\tilde{\sigma}_\alpha/\sqrt{n}$, where z_q is the qth quantile of a standard normal random variable.

Let

$$\pi^*(y_0, y_1, y_2; \alpha)^{-1} = (1 + \exp\{l_1^*(y_0; \alpha) + \alpha r(y_1)\}) \times$$
$$(1 + \exp\{l_2^*(y_1; \alpha) + \alpha r(y_2)\})$$

$$w_1^*(y_0; \alpha) = E^* \left[\exp\{\alpha r(Y_1)\} \mid R_1 = 1, Y_0 = y_0 \right],$$

$$w_2^*(y_1; \alpha) = E^* \left[\exp\{\alpha r(Y_2)\} \mid R_2 = 1, Y_1 = y_1 \right],$$

$$g_1^*(y_0, y_1; \alpha) = \{1 - H_1^*(y_0)\} w_1^*(y_0; \alpha) + \exp\{\alpha r(y_1)\} H_1^*(y_0).$$

$$g_2^*(y_1, y_2; \alpha) = \{1 - H_2^*(y_1)\} w_2^*(y_1; \alpha) + \exp\{\alpha r(y_2)\} H_2^*(y_1).$$

$$
\begin{aligned}
\psi_{P^*}(O; \alpha) \ :=\ & a_0^*(Y_0; \alpha) + \\
& R_1 b_1^*(Y_0, Y_1; \alpha) + \\
& R_2 b_2^*(Y_1, Y_2; \alpha) + \\
& \{1 - R_1 - H_1^*(Y_0)\} c_1^*(Y_0; \alpha) + \\
& R_1 \{1 - R_2 - H_2^*(Y_1)\} c_2^*(Y_1; \alpha)
\end{aligned}
$$

where

$$a_0^*(Y_0) = E^* \left[\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \,\middle|\, Y_0 \right] - \mu(P^*; \alpha)$$

$$b_1^*(Y_0, Y_1; \alpha) = E^* \left[\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \,\middle|\, R_1 = 1, Y_1, Y_0 \right] - E^* \left[\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \,\middle|\, R_1 = 1, Y_0 \right]$$

$$+ E^* \left[\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \left[\frac{\exp\{\alpha r(Y_1)\}}{g_1^*(Y_0, Y_1; \alpha)} \right] \,\middle|\, R_1 = 1, Y_0 \right] H_1^*(Y_0) \left\{ 1 - \frac{\exp\{\alpha r(Y_1)\}}{w_1^*(Y_0; \alpha)} \right\}$$

$$b_2^*(Y_1, Y_2; \alpha) = E^* \left[\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \,\middle|\, R_2 = 1, Y_2, Y_1 \right] - E^* \left[\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \,\middle|\, R_2 = 1, Y_1 \right]$$

$$+ E^* \left[\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \left[\frac{\exp\{\alpha r(Y_2)\}}{g_2^*(Y_1, Y_2; \alpha)} \right] \,\middle|\, R_2 = 1, Y_1 \right] H_2^*(Y_1) \left\{ 1 - \frac{\exp\{\alpha r(Y_2)\}}{w_2^*(Y_1; \alpha)} \right\}$$

$$c_1^*(Y_0) = E^* \left[\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \left[\frac{\exp\{\alpha r(Y_1)\}}{g_1^*(Y_0, Y_1; \alpha)} \right] \,\middle|\, Y_0 \right]$$

$$- E^* \left[\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \left[\frac{1}{g_1^*(Y_0, Y_1; \alpha)} \right] \,\middle|\, Y_0 \right] w_1^*(Y_0; \alpha)$$

$$c_2^*(Y_1) = E^* \left[\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \left[\frac{\exp\{\alpha r(Y_2)\}}{g_2^*(Y_1, Y_2; \alpha)} \right] \,\middle|\, R_1 = 1, Y_1 \right]$$

$$- E^* \left[\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \left[\frac{1}{g_2^*(Y_1, Y_2; \alpha)} \right] \,\middle|\, R_1 = 1, Y_1 \right] w_2^*(Y_1; \alpha)$$

Inference - Uncertainty

- Wald-based confidence intervals don't always have adequate coverage properties is finite samples.
- In equal-tailed studentized bootstrap, the confidence interval takes the form $[\widehat{\mu} - t_{0.975}\widehat{se}(\widehat{\mu}), \widehat{\mu} - t_{0.025}\widehat{se}(\widehat{\mu})]$, where t_q is the qth quantile of $\left\{ \frac{\widehat{\mu}^{(b)} - \widehat{\mu}}{\widehat{se}(\widehat{\mu}^{(b)})} : b = 1, \ldots, B \right\}$
- In symmetric studentized bootstrap, the confidence interval takes the form $[\widehat{\mu} - t_{0.95}^*\widehat{se}(\widehat{\mu}), \widehat{\mu} + t_{0.95}^*\widehat{se}(\widehat{\mu})]$, where $t_{0.95}^*$ is selected so that 95% of the distribution of $\left\{ \frac{\widehat{\mu}^{(b)} - \widehat{\mu}}{\widehat{se}(\widehat{\mu}^{(b)})} : b = 1, \ldots, B \right\}$ falls between $-t_{0.95}^*$ and $t_{0.95}^*$.
- Useful to replace influence-function based standard error estimator with jackknife standard error estimator.

Quetiapine Bipolar Trial - Fit

- Estimated smoothing parameters for the drop-out model are 11.54 and 9.82 for the placebo and 600 mg arms.
- Estimated smoothing parameters for the outcome model are 6.34 and 8.05 for the placebo and 600 mg arms.
- In the placebo arm, the observed percentages of last being seen at visits 0 and 1 among those at risk at these visits are 8.62% and 38.68%. Model-based estimates are 7.99% and 38.19%.
- For the 600 mg arm, the observed percentages are 11.02% and 35.24% and the model-based estimates are 11.70% and 35.08%.

Quetiapine Bipolar Trial - Fit

- In the placebo arm, the Kolmogorov-Smirnov distances between the empirical distribution of the observed outcomes and the model-based estimates of the distribution of outcomes among those on-study at visits 1 and 2 are 0.013 and 0.033.
- In the 600 mg arm, these distances are 0.013 and 0.022.
- These results suggest that our model for the observed data fits the observed data well.

Quetiapine Bipolar Trial - MAR

- Under MAR, the estimated values of μ^* are 46.45 (95% CI: 42.35,50.54) and 62.87 (95% CI: 58.60,67.14) for the placebo and 600 mg arms.
- The estimated difference between 600 mg and placebo is 16.42 (95% 10.34, 22.51)
- Statistically and clinically significant improvement in quality of life in favor of Quetiapine.

Quetiapine Bipolar Trial - Sensitivity Analysis

- We set $r(y) = y$ and ranged the sensitivity analysis parameter from -10 and 10 in each treatment arm.
- According to experts, there is no evidence to suggest that there is a differential effect of a unit change in QLESSF on the hazard of drop-out based on its location on the scale.

Quetiapine Bipolar Trial - Sensitivity Analysis

Figure: Treatment-specific (left: placebo; right: 600 mg/day Quetiapine) estimates (along with 95% pointwise confidence intervals) of μ^* as a function of α.

Figure: Treatment-specific differences between the estimated mean QLESSF at Visit 2 among non-completers and the estimated mean among completers, as a function of α.
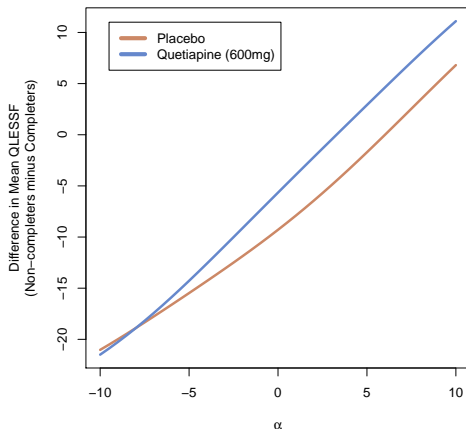
Quetiapine Bipolar Trial - Sensitivity Analysis

Figure: Contour plot of the estimated differences between mean QLESSF at Visit 2 for Quetiapine vs. placebo for various treatment-specific combinations of the sensitivity analysis parameters.
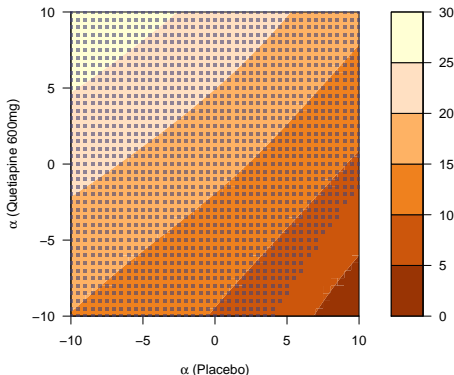
- Only when the sensitivity analysis are highly differential (e.g., $\alpha(\text{placebo}) = 8$ and $\alpha(\text{Quetiapine}) = -8$) are the differences no longer statistically significant.
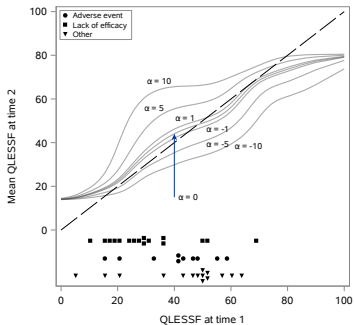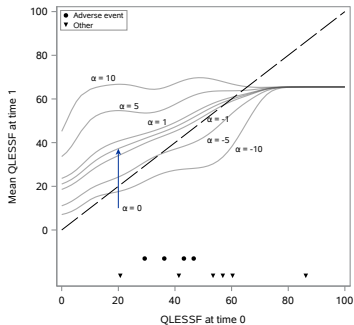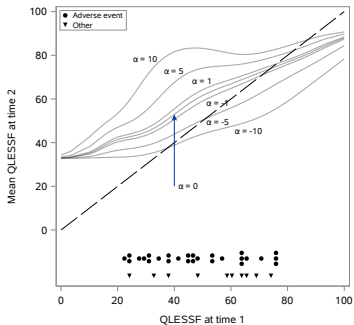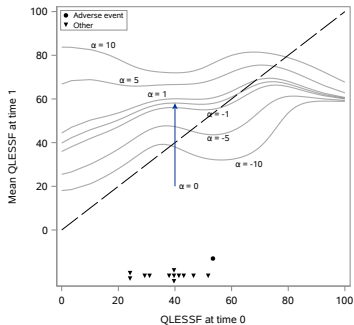
- Conclusions under MAR are highly robust.

Simulation Study

- Generated 2500 placebo and Quetiapine datasets using the estimated distributions of the observed data from the Quentiapine study as the true data generating mechanisms.
- For given treatment-specific α, these true data generating mechanisms can be mapped to a true value of μ^*.
- For each dataset, the sample size was to set to 116 and 118 in the placebo and Quetiapine arms, respectively.

Simulation Study - Bias/MSE

α	Estimator	Placebo			Quetiapine		
		μ^*	Bias	MSE	μ^*	Bias	MSE
-10	Plug-in	40.85	0.02	4.43	56.07	0.40	4.69
	Corrected		0.43	4.56		0.42	4.72
-5	Plug-in	43.45	0.05	4.29	59.29	0.34	4.55
	Corrected		0.27	4.26		0.24	4.35
-1	Plug-in	46.02	0.28	4.34	62.58	0.50	4.39
	Corrected		0.18	4.22		0.14	4.00
0	Plug-in	46.73	0.36	4.44	63.42	0.55	4.36
	Corrected		0.17	4.27		0.14	3.95
1	Plug-in	47.45	0.43	4.57	64.25	0.59	4.32
	Corrected		0.16	4.36		0.15	3.92
5	Plug-in	50.48	0.66	5.33	67.34	0.59	4.20
	Corrected		0.14	5.11		0.19	4.15
10	Plug-in	54.07	0.51	5.78	70.51	0.07	4.02
	Corrected		0.04	6.30		-0.05	4.66

Simulation Study - Coverage

α	Procedure	Placebo Coverage	Quetiapine Coverage
-10	Wald-IF	91.5%	90.5%
	Wald-JK	95.0%	94.6%
	Bootstrap-IF-ET	94.3%	93.8%
	Bootstap-JK-ET	94.4%	93.4%
	Bootstap-IF-S	95.2%	94.6%
	Bootstap-JK-S	95.0%	94.6%
-5	Wald-IF	93.5%	92.9%
	Wald-JK	95.0%	94.8%
	Bootstrap-IF-ET	95.2%	94.6%
	Bootstap-JK-ET	94.8%	94.6%
	Bootstap-IF-S	95.4%	95.2%
	Bootstap-JK-S	95.1%	95.2%
-1	Wald-IF	93.9%	94.2%
	Wald-JK	94.9%	95.4%
	Bootstrap-IF-ET	95.1%	94.8%
	Bootstap-JK-ET	95.1%	94.6%
	Bootstap-IF-S	95.3%	96.4%
	Bootstap-JK-S	95.1%	96.3%
0	Wald-IF	93.8%	94.0%
	Wald-JK	95.0%	95.4%
	Bootstrap-IF-ET	94.6%	94.5%
	Bootstap-JK-ET	94.6%	94.6%
	Bootstap-IF-S	95.5%	96.6%
	Bootstap-JK-S	95.2%	96.7%

Simulation Study - Coverage

α	Procedure	Placebo Coverage	Quetiapine Coverage
1	Wald-IF	93.3%	93.7%
	Wald-JK	95.1%	95.5%
	Bootstrap-IF-ET	94.6%	94.6%
	Bootstap-JK-ET	94.6%	94.6%
	Bootstap-IF-S	95.5%	96.5%
	Bootstap-JK-S	95.2%	96.5%
5	Wald-IF	90.8%	91.3%
	Wald-JK	95.3%	95.7%
	Bootstrap-IF-ET	93.2%	91.6%
	Bootstap-JK-ET	93.8%	93.0%
	Bootstap-IF-S	95.5%	95.4%
	Bootstap-JK-S	95.8%	96.4%
10	Wald-IF	85.4%	87.8%
	Wald-JK	94.9%	94.5%
	Bootstrap-IF-ET	88.2%	87.0%
	Bootstap-JK-ET	92.2%	89.7%
	Bootstap-IF-S	94.6%	93.9%
	Bootstap-JK-S	95.5%	95.1%

1. R

   `samon` library
   functions with pass to C code

2. SAS

   procedures and macros

The Quet1/Quet2 Dataset

```
proc print data = quet1;
run;

    Obs    V1    V2    V3

     1     36    56    57
     2     32    32    31
     3     25    42    27
     4     48    52    54
     5     27    40    40
     6     38    43    .
     7     35    44    39
     8     24    26    28
     9     25    30    .
    10     25    28    .
    11     40    45    45
    12     14    16    .
    13     34    37    .
    14     27    25    .
    15     26    25    .
    16     38    .     .
    17     44    .     .
    18     35    31    .
    19     25    45    49
    20     45    53    56
    21     32    38    48
    22     46    45    .
    23     30    .     .
    24     43    56    43
    25     42    44    41
    26     34    45    47
    27     34    .     .
```

SAS: `samonDataCheck` macro

- `samonDataCheck` macro can be used to check the missing data pattern in the data.

%samonDataCheck	
(	
data =	input dataset
vars =	variable list (in time order)
out =	output data
stats =	output statistics dataset
mpattern =	missing pattern counts dataset
);	

```
%samonDataCheck(
   data     =  quet1,
   vars     =  v1 v2 v3,
   out      =  check1,
   stats    =  stats1t,
   mpattern =  pattern1);
```

R: samonDataCheck function

In R, the samonDataCheck function can be used for the same purpose:

```
samonDataCheck( quet1 )
```

```
> # R version of samonDataCheck is a
> # function of the same name
>
> # Check data
> chk1  <- samonDataCheck( quet1 )
>
> chk2  <- samonDataCheck( quet2 )
```

```
Samon Data Check:
-------------------------------------------------
Number of time-points:                        3
Number of subjects:                         116
Minimum observed value:                      14
Maximum observed value:                      63
Average number of timepoints on study:     2.47
Total number of observed values:            287
Subjects observed at final timepoint:        65
Subjects observed at all timepoint:          65

Missing Patterns:
                   N     proportion
*__  :            10        0.0862
**_  :            41        0.3534
***  :            65        0.5603
```

samonDataCheck

```
Samon Data Check:
-------------------------------------------------
Number of time-points:                          3
Number of subjects:                           118
Minimum observed value:                        15
Maximum observed value:                        67
Average number of timepoints on study:       2.47
Total number of observed values:              291
Subjects observed at final timepoint:          68
Subjects observed at all timepoint:            68

Missing Patterns:
                N       proportion
*__  :         13          0.1102
**_  :         37          0.3136
***  :         68          0.5763
```
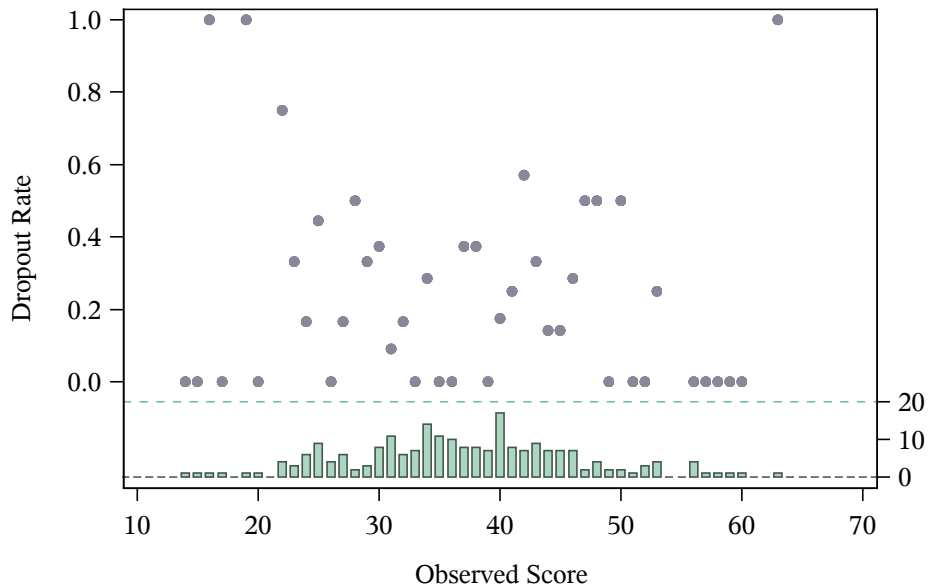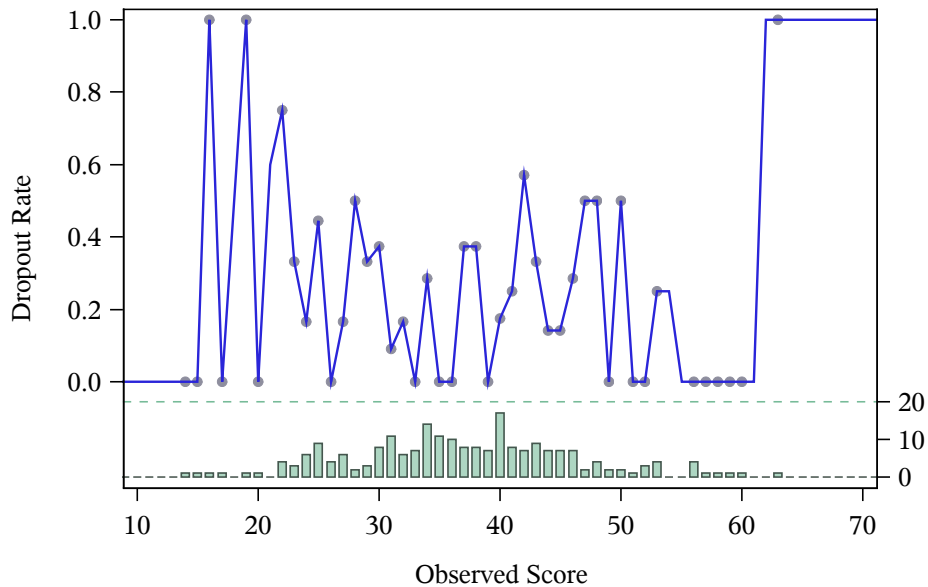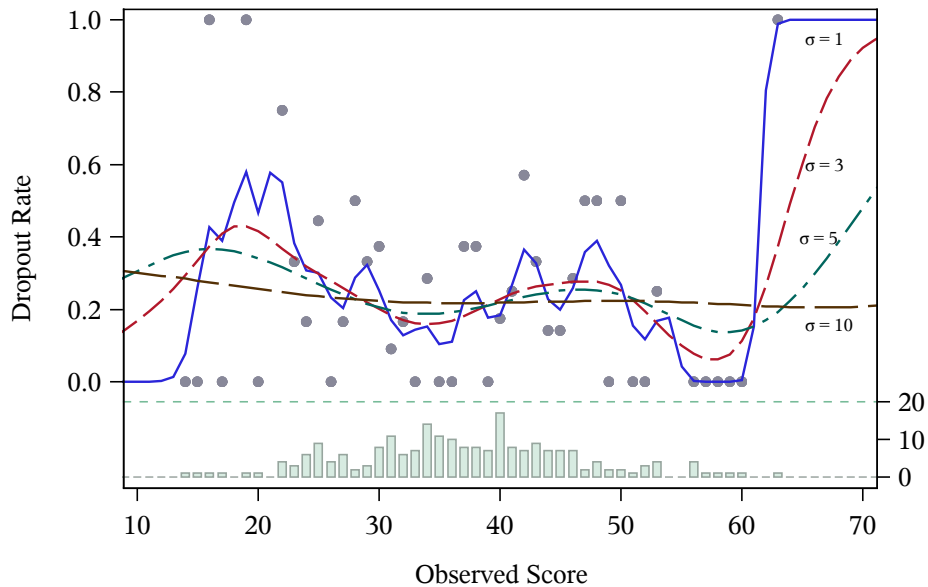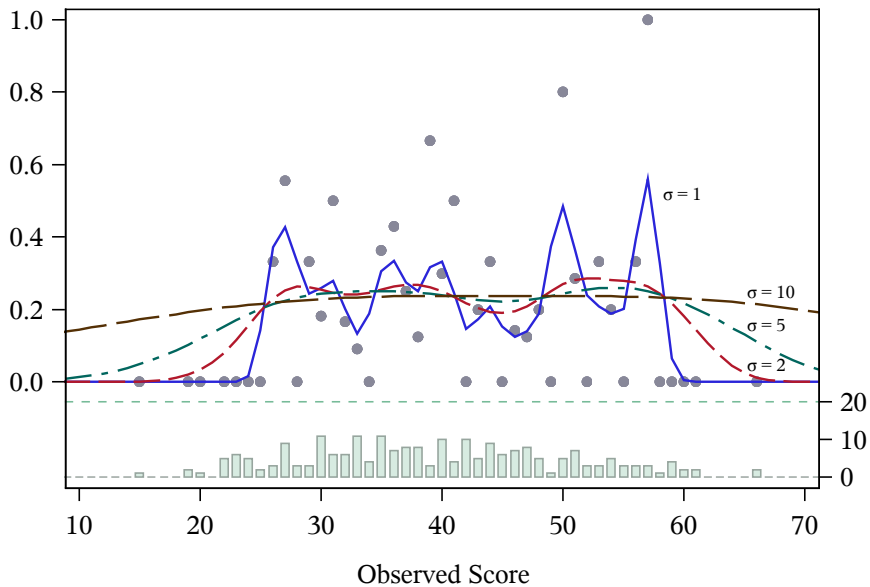
R: samoneval function

- samoneval function can be used to compute the loss function for a range of σ.
- Takes four arguments:

  samoneval(

  mat = , # input matrix to evaluate

  Npart = 10, # number of partitions

  sigmaList = c(0,1), # vector of sigma values

  type = "both" # compute the loss function for σ_H, σ_F, or both σ_H and σ_F

)

- Returns a matrix containing σ_H and its corresponding loss function value and σ_F and its corresponding loss function value.

R: samoneval function

```
> library(samon, lib.loc="../samlib")
>
> Results1 <- samoneval(
+     mat             = quet1,
+     Npart           = 10,
+     sigmaList       = seq(0.2,30.0,by=0.1),
+     type            = "both"
+                     )
```

R: samoneval function

```
> Results2 <- samoneval(
+      mat            = quet2,
+      Npart          = 10,
+      sigmaList       = seq(0.2,30.0,by=0.1),
+      type            = "both"
+                       )
> ResultsH <- cbind(Results1[,1:2],
+                   Results2[,1:2])
> HFPlot(ResultsH, "H.pdf", 4.2, 4.2,
+        "Loss function (H)",  c(0,30),
+        c( 2.5, 4.5), c(15,4.5) )
```

samonev procedure

The `samonev` procedure computes the loss function for a range of σ.

samonev	
data =	Input dataset
out =	Output dataset
npart =	Number of partitions
var varlist	list of variables in time order
sigma sigmalist	list of sigmas

```
proc samonev
   data      = quet1
   out       = ev1
   Npart     = 10;

   var v1 - v3;
   sigma 0.2 to 30 by 0.1;
run;
```
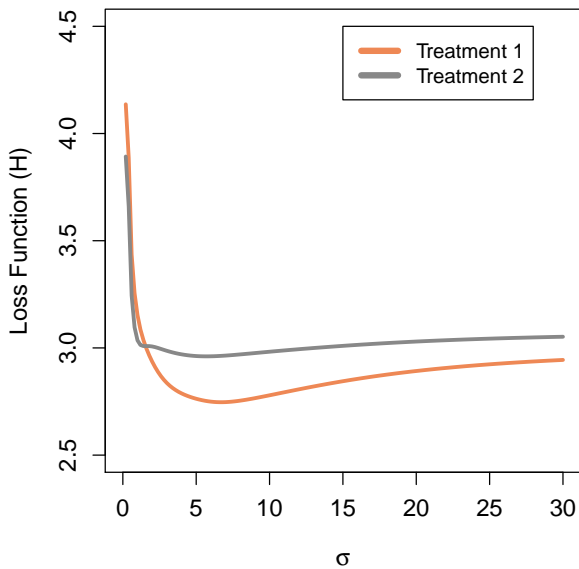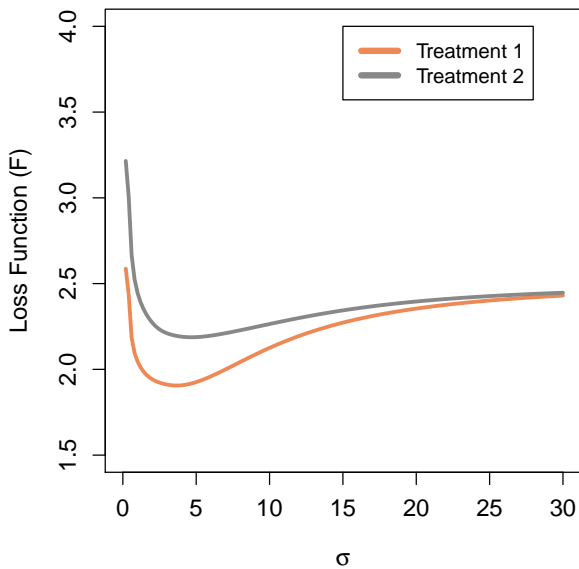
The samon function and the samon procedure can be used to find optimal values of σ_H and σ_F. Like many optimization techniques providing good initial estimates can improve the efficiency of convergence of the optimization. Within samon we also provide an upper bound for σ_H and σ_F. Should the algorithm begin to converge to an optimal value greater than the upper bound, samon returns the upper bound itself rather than search for an optimal value above this upper bound. This is to reflect the fact that larger values of σ_H or σ_F result in little change in the smoothing.

R: samon function

```
# Example1.R
# Finding optimal Sigma_h and Sigma_f.
# ----------------------------------------
library(samon, lib.loc="../samlib")

samonResults <- samon(
    mat             = quet1,
    Npart           = 10,

    InitialSigmaH   = 6.0,
    HighSigmaH      = 50.0,

    InitialSigmaF   = 4.0,
    HighSigmaF      = 50.0

    )

# print the output
print(samonResults)
```

```
* Finding optimal Sigma_h and Sigma_f.;
* ----------------------------------------;
proc samon data   = quet1
           out    = samon1
           Npart  = 10
           Hinit  = 6.0
           Hhigh  = 50.0
           Finit  = 4.0
           Fhigh  = 50
           Hout   = HM1
           Fout   = FM1;
  var v1 - v3;
run;
```

Treatment 1

```
proc print data = HM1 noobs;
run;

  rc    Niter    Sigma      loss
   2        3    6.6918    2.7468


proc print data = FM1 noobs;
run;

  rc    Niter    Sigma      loss
   2        6    3.6771    1.9057
```

```
proc print data = HM2 noobs;
run;

  rc    Niter    Sigma      loss
   2        3    5.6938    2.9607


proc print data = FM2 noobs;
run;

  rc    Niter    Sigma      loss
   2        3    4.6704    2.1872
```

Sensitivity Analysis

- Within samon the sensitivity bias function is the cumulative function of the beta distribution, a flexible function with bounded support.
- This together with the sensitivity analysis parameter α provides the mechanism by which we measure the sensitivity of the results to informative drop-out.

Sensitivity Analysis

- The cumulative beta function is defined on the interval (0,1) and in order to use it as the sensitivity bias function we need to map the range of our data into (0,1).
- In the case of QLESSF scores the data are limited to the range 13 and 71.
- We take the parameters for the cumulative beta function ζ_1 and ζ_2 to be 1.

$(\zeta_1, \zeta_2) = (1,4)$

$(\zeta_1, \zeta_2) = (1,1)$

$(\zeta_1, \zeta_2) = (8,2)$

QLESSF
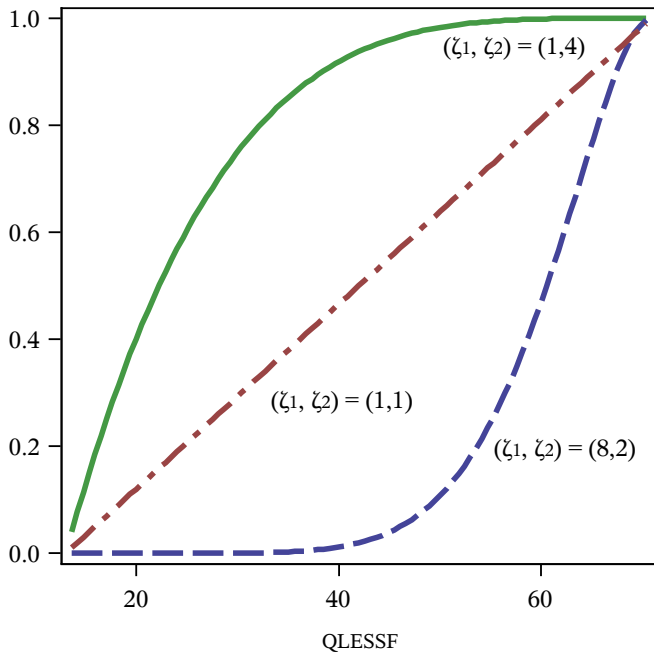
samon procedure

samon	
data =	Input dataset
IFout =	Influence function estimates
npart =	Number of partitions
Hinit =	initial value for smoothing parameter sigma H
Hhigh =	Highest value for smoothing parameter sigma H
Finit =	initial value for smoothing parameter sigma F
Fhigh =	Highest value for smoothing parameter sigma F
lb =	lower bound of data
ub =	upper bound of data
zeta1 =	parameter for cumulative beta distribution
zeta2 =	parameter for cumulative beta distribution
nsamples =	Number of bootstrap samples
seed0 =	Seed to pass to random number generator
var varlist	list of variables in time order
sensp senslist	list of sensitivity parameters

```
proc samon  data = quet1 IFout = IFM1
    Npart = 10
    Hinit =  6.0    HHigh = 50.0
    Finit =  4.0    FHigh = 50.0

    lb   =   13    ub   =  71
    zeta1 =  1.0    zeta2 = 1.0
    nomj            nsamples = 0 ;

  var v1 - v3;
  sensp -10 to 10 by 1;
run;
proc print data = IFM1 noobs;
  var alpha AEst AVar IFEst IFVar;
run;
```

```
# Produce one-step influence function estimates
# ---------------------------------------------
library(samon, lib.loc="../samlib")

Results1 <- samon(
    mat              = quet1,
    Npart            = 10,

    # initial value and upper bound for sigmaH
    InitialSigmaH    = 6.0,
    HighSigmaH       = 50.0,

    # initial value and upper bound for sigmaF
    InitialSigmaF    = 4.0,
    HighSigmaF       = 50.0,

    AlphaList        = -10:10,      # alphas

    lb        = 13,    ub       = 71,
    zeta1     = 1.0,   zeta2    = 1.0
```

alpha	AEst	AVar	IFEst	IFVar
-10	36.6909	0.08754	36.9454	1.3374
-9	36.9556	0.09050	37.2108	1.3438
-8	37.2402	0.09354	37.4881	1.3497
-7	37.5435	0.09653	37.7740	1.3547
-6	37.8641	0.09934	38.0654	1.3585
-5	38.2007	0.10185	38.3611	1.3609
-4	38.5526	0.10395	38.6609	1.3625
-3	38.9189	0.10556	38.9660	1.3641
-2	39.2993	0.10660	39.2787	1.3668
-1	39.6935	0.10701	39.6020	1.3715
0	40.1010	0.10678	39.9386	1.3792
1	40.5210	0.10590	40.2911	1.3898
2	40.9517	0.10436	40.6609	1.4026
3	41.3907	0.10222	41.0484	1.4160
4	41.8343	0.09950	41.4525	1.4278
5	42.2785	0.09626	41.8710	1.4356
6	42.7186	0.09252	42.3006	1.4374
7	43.1501	0.08832	42.7372	1.4320
8	43.5690	0.08370	43.1760	1.4190
9	43.9724	0.07871	43.6127	1.3989
10	44.3586	0.07345	44.0435	1.3729
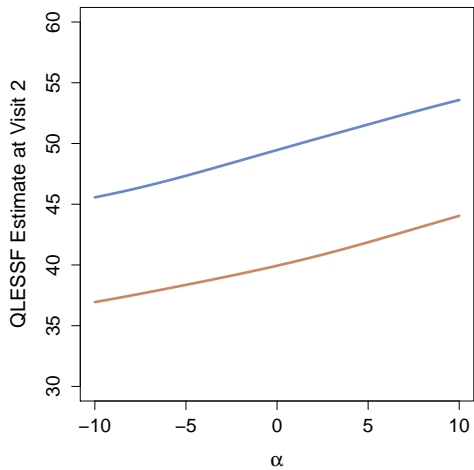
Confidence intervals

- Use bootstrap with jackknife to compute confidence intervals.
- The NSamples argument controls the number of bootstraps to make.
- The flags mj and sj control whether jackknifes are performed on the main (input) data and the bootstrap samples respectively.
- For a small dataset with 100 individuals, 1,000 bootstraps each with bootstrap estimates on 50 sensitivity parameters gives rise to 50 x 100 x 1000 = 5 million estimates.

```
proc samon data = quet1 out = samon1
    Npart   = 10

    Hinit   =  6.0    HHigh = 50.0
    FInit   =  4.0    FHigh = 50.0
    lb      = 13      ub    = 71
    zeta1   = 1.0     zeta2 = 1.0
    NSamples = 500    seed0 = 81881
    sj;
  var v1-v3;
  sensp -10 to 10 by 1;
run;
```

```
%samonSummary(
  data    = results1,
  out     = data.Summary1,
  sampout = data.sampSummary1
  );
proc print data=data.Summary1;
  var alpha IFEst lb ub;
run;
```

alpha	IFEst	lb	ub
-10	36.9454	34.2645	39.6263
-9	37.2108	34.5635	39.8580
-8	37.4881	34.9005	40.0757
-7	37.7740	35.2496	40.2984
-6	38.0654	35.5840	40.5468
-5	38.3611	35.9267	40.7955
-4	38.6609	36.2627	41.0590
-3	38.9660	36.6038	41.3281
-2	39.2787	36.9218	41.6356
-1	39.6020	37.2546	41.9493
0	39.9386	37.5628	42.3144
1	40.2911	37.8957	42.6866
2	40.6609	38.2211	43.1007
3	41.0484	38.5820	43.5148
4	41.4525	38.9005	44.0045
5	41.8710	39.2378	44.5043
6	42.3006	39.5695	45.0318
7	42.7372	39.9254	45.5490
8	43.1760	40.2880	46.0641
9	43.6127	40.5862	46.6392
10	44.0435	40.8705	47.2165

We use the `samonCrossSummary` function to compute the difference in estimates for each pair of alpha.

```
%samonCrossSummary(
    IFM1      = data.Summary1,
    sampIF1   = data.sampSummary1,
    IFM2      = data.Summary2,
    sampIF2   = data.sampSummary2,
    out       = data.Cross
    );
```

macro	description
`samonCombine`	Combines results from multiple runs of proc `samon`
`samonSummary`	Summarizes samon results. Combines bootstrap and jackknife results to produce confidence intervals
`samonDifferenceSummary`	Computes treatment effect differences and confidence intervals from a pair of `samonSummary` objects.
`samonCrossSummary`	Computes treatment effect differences and confidence intervals for each pair of sensitivity parameters α.
`samonECompleterStatus`	Computes the difference in the expected value of non-completers and completers

samonCombine macro

samonCombine	combines samon results into one dataset
(	
inlib =	input libref
stem = results	file name stem
connect = _	name connector
partfrom = 1	parts start at 1
partto = 100	to 100
partform = z5	format to use on partno
outlib =	output libref
)	

samonSummary macro

samonSummary	computes summary of samon object
(	
data =	input dataset to summarize
out =	summary of main data
sampSummary =	summary of parametric bootstrap samples
)	

samonDifferenceSummary macro

samonDifferenceSummary	Treatment-specific differences
(	
IFM1 =	main results from samonSummary for trt 1
sampIF1 =	sample results from samonSummary for trt 1
IFM2 =	main results from samonSummary for trt 2
sampIF2 =	sample results from samonSummary for trt 2
out =	summary of difference
)	

samonCrossSummary macro

samonCrossSummary	Treatment-specific differences for all pairs of sensitivity parameter
(	
IFM1 =	main results from samonSummary for trt 1
sampIF1 =	sample results from samonSummary for trt 1
IFM2 =	main results from samonSummary for trt 2
sampIF2 =	sample results from samonSummary for trt 2
out =	summary of difference
)	

R: `samon` function

function	description
`samonCombine`	combines the outputs from `samon` into one `samonMat` object. The results are stored as `.Rds` files. `samonCombine` takes a list of such files and combines them.
`samonDiff`	Takes two `samonMat` objects and produces a `samonMat` object for the difference in influence function estimates
`samonBiasCorrection`	Takes a `samonMat` object and produces corrected influence function estimates
`samonXBiasCorrection`	Takes two `samonMat` objects (one from each treatment groups) and for each pair of alphas produces the difference in influence function estimates.

Generalization

- Y_k: outcome scheduled to be measured at assessment k.
- R_k: indicator that individual is on-study at assessment k.
- All individuals are present at baseline, i.e., $R_0 = 1$.
- Monotone missing data: $R_{k+1} = 1$ implies $R_k = 1$.
- $C = \max\{k : R_k = 1\}$, $C = K$ implies that the individual completed the study.
- For any given vector $z = (z_1, z_2, \ldots, z_K)$,
    - $\overline{z}_k = (z_0, z_1, \ldots, z_k)$
    - $\underline{z}_k = (z_{k+1}, z_{k+2}, \ldots, z_K)$.
- For each individual, the data unit $O = (C, \overline{Y}_C)$ is drawn from some distribution P^* contained in the non-parametric model $\mathscr{M}$ of distributions.
- The observed data consist of n independent draws $O_1, O_2, \ldots, O_n$ from P^*.

Generalization

By factorizing the distribution of O in terms of chronologically ordered conditional distributions, any distribution $P \in \mathcal{M}$ can be represented by

- $F_0(y_0) := P(Y_0 \leq y_0)$;

- $F_{k+1}(y_{k+1} \mid \overline{y}_k) := P(Y_{k+1} \leq y_{k+1} \mid R_{k+1} = 1, \overline{Y}_k = \overline{y}_k)$,
  $k = 0, 1, \ldots, K - 1$;

- $H_{k+1}(\overline{y}_k) := P(R_{k+1} = 0 \mid R_k = 1, \overline{Y}_k = \overline{y}_k)$,
  $k = 0, 1, \ldots, K - 1$.

The main objective is to draw inference about $\mu^* := E^*(Y_K)$, the true mean outcome at visit K in a hypothetical world in which all patients are followed to that visit.

For every $\overline{y}_k$, define the following strata:

- $A_k(\overline{y}_k)$: patients last seen at visit k (i.e., $R_k = 1, R_{k+1} = 0$) with $\overline{Y}_k = \overline{y}_k$.
- $B_{k+1}(\overline{y}_k)$: patients on-study at visit $k+1$ (i.e., $R_{k+1} = 1$) with $\overline{Y}_k = \overline{y}_k$.

> For all $\overline{y}_k$, the distribution of $\underline{Y}_k$ for patients in stratum $A_k(\overline{y}_k)$ is the same as the distribution of $\underline{Y}_k$ for patients in stratum $B_{k+1}(\overline{y}_k)$

Mathematically, we can express these assumptions as follows:

$$f^*(\underline{Y}_k | A_k(\overline{y}_k)) = f^*(\underline{Y}_k | B_{k+1}(\overline{y}_k)) \text{ for all } \overline{y}_k \qquad (13)$$

Missing at Random

Using Bayes' rule, we can re-write these expressions as:

$$P^*(R_{k+1} = 0 | R_k = 1, \overline{Y}_K = \overline{y}_K)$$
$$= P^*(R_{k+1} = 0 | R_k = 1, \overline{Y}_k = \overline{y}_k) \text{ for all } \overline{y}_K$$

Written in this way, missing at random implies that the drop-out process is stochastic with the following interpretation:

Among those on study at visit k, the decision to discontinue the study before the next visit is like the flip of a coin with probability depending only on the observable history of outcomes through visit k (i.e., no outcomes after visit k).

Under missing at random, μ^* is identified. That is, it can be expressed as a functional of the distribution of the observed data. Specifically, $\mu^* = \mu(P^*)$ is

$$\int_{y_0} \cdots \int_{y_K} y_K \left\{ \prod_{k=0}^{K-1} dF_{k+1}^*(y_{k+1}|\overline{y}_k) \right\} dF_0^*(y_0)$$

Missing Not at Random

Equation (13) is equivalent to the following two assumptions:

$$f^*(\underline{Y}_{k+1}|A_k(\overline{y}_k), Y_{k+1} = y_{k+1})$$
$$= f^*(\underline{Y}_{k+1}|B_{k+1}(\overline{y}_k), Y_{k+1} = y_{k+1}) \text{ for all } \overline{y}_{k+1} \tag{14}$$

and

$$f^*(Y_{k+1}|A_k(\overline{y}_k)) = f^*(Y_{k+1}|B_{k+1}(\overline{y}_k)) \text{ for all } \overline{y}_k \tag{15}$$

Equation (14) posits the following "linking" assumption:

> For all $\overline{y}_{k+1}$, the distribution of $\underline{Y}_{k+1}$ for patients in stratum $A_k(\overline{y}_k)$ with $Y_{k+1} = y_{k+1}$ is the same as the distribution of $\underline{Y}_{k+1}$ for patients in stratum $B_{k+1}(\overline{y}_k)$ with $Y_{k+1} = y_{k+1}$.

Missing Not at Random

Using Bayes' rule, this assumption can be re-written as:

$$P^*(R_{k+1} = 0 | R_k = 1, \overline{Y}_K = \overline{y}_K)$$
$$= P^*(R_{k+1} = 0 | R_k = 1, \overline{Y}_{k+1} = \overline{y}_{k+1}) \text{ for all } \overline{y}_K$$
$$(16)$$

This assumption has been referred to as the "non-future" dependence assumption (Diggle and Kenward, 1994) because it has the following interpretation:

> Among those on study at visit k, the decision to discontinue the study before the next visit is like the flip of a coin with probability depending only on the observable history of outcomes through visit k and the potentially unobserved outcome at visit $k + 1$ (i.e., no outcomes after visit $k + 1$).

We will retain this assumption.

Next, we generalize (15) and impose the following exponential tilting "linking" assumptions:

$$f^*(Y_{k+1}|A_k(\overline{y}_k)) \propto f^*(Y_{k+1}|B_{k+1}(\overline{y}_k)) \exp(\alpha r(Y_{k+1})) \text{ for all } \overline{y}_k \tag{17}$$

where $r(\cdot)$ is a specified function which we will assume to be an increasing function of its argument and α is a sensitivity analysis parameter.

Missing Not at Random

- The missing not at random class of assumptions that we propose involves Equations (14) and (17), where $r(\cdot)$ is considered fixed and α is a sensitivity analysis parameter that serves as the class index.

- (17) reduces to (15) when $\alpha = 0$. Thus, when $\alpha = 0$, the missing at random assumption is obtained.

- When $\alpha > 0$ (< 0), (17) implies:

  > For all $\overline{y}_k$, the distribution of Y_{k+1} for patients in stratum $A_k(\overline{y}_k)$ is weighted more heavily (i.e., tilted) to higher (lower) values than the distribution of Y_{k+1} for patients in stratum $B_{k+1}(\overline{y}_k)$.

  The amount of "tilting" increases with magnitude of α.

Inference

Three steps:

1. Assume
   - $F_{k+1}^*(y_{k+1} \mid \bar{y}_k) = F_{k+1}^*(y_{k+1} \mid y_k)$
   - $H_{k+1}^*(\bar{y}_k) = H_{k+1}^*(y_k)$

2. Estimate $F_{k+1}^*(y_{k+1} \mid y_k)$ and $H_{k+1}^*(\bar{y}_k) = H_{k+1}^*(y_k)$ using non-parametric smoothing with tuning parameters selected by cross-validation.

3. Use plug-in + average of estimated influence functions.

4. Use alternatives to Wald-based confidence intervals.

Case Study: SCA-3004

- Randomized trial designed to evaluate the efficacy and 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 consisting of 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.

- Research 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?
- An ideal study would follow all randomized subjects through Visit 16 while maintaining them on their randomized treatment and examine symptomatic and functional outcomes at that time point.
- Since clinical relapse can have a major negative impact, the study design required that patients who had signs of relapse were discontinued from the study.
- In addition, some patients discontinued due to adverse events, withdrew consent or were lost to follow-up.
- 38% and 60% of patients in the PBO and PP1M arms were followed through Visit 16 (p=0.0001).
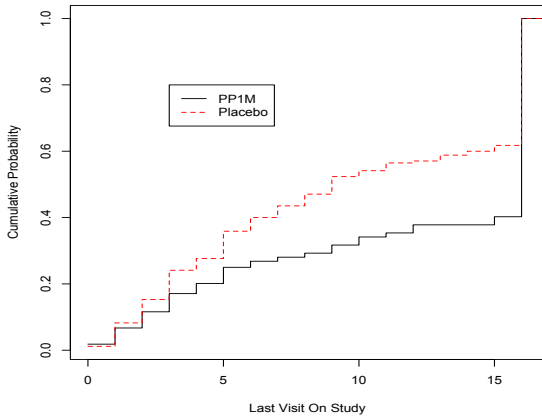
Case Study: SCA-3004

- Focus: Patient function as measured by the Personal and Social Performance (PSP) scale.
- The PSP scale is scored from 1 to 100 with higher scores indicating better functioning based on evaluation of 4 domains (socially useful activities, personal/social relationships, self-care, and disturbing/aggressive behaviors).
- Estimate treatment-specific mean PSP at Visit 16 in the counterfactual 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).
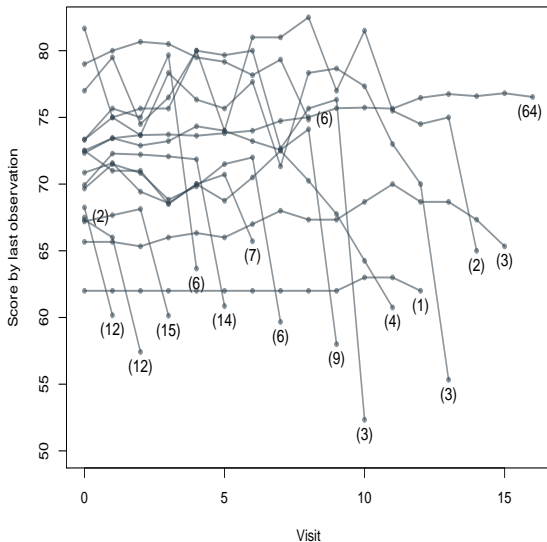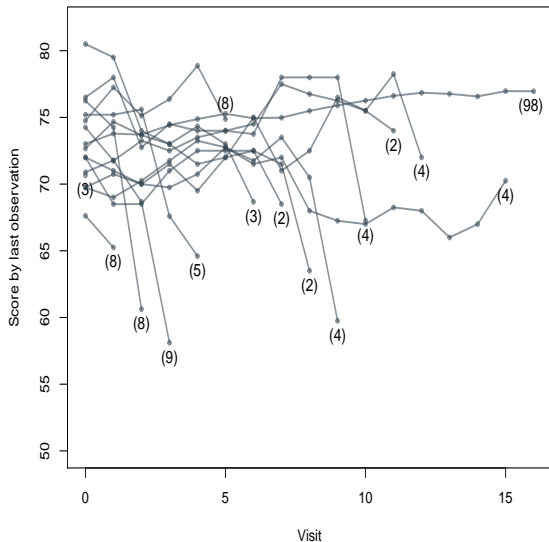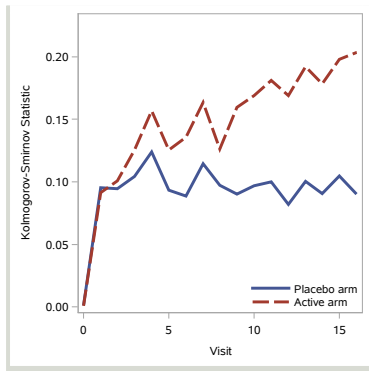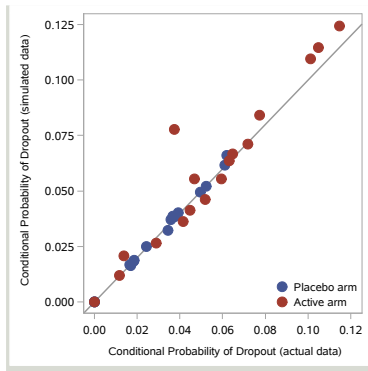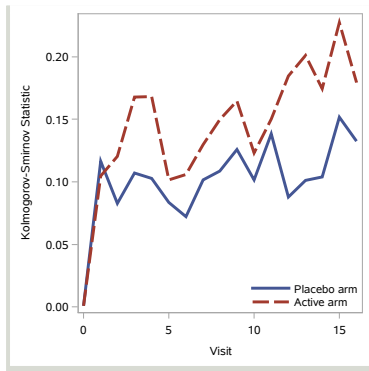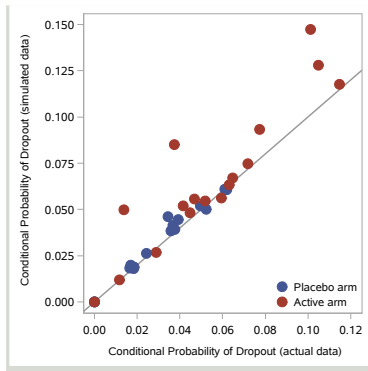
Case Study: SCA-3004
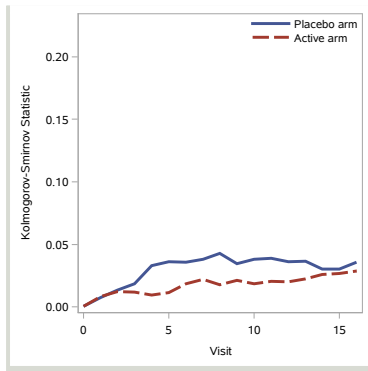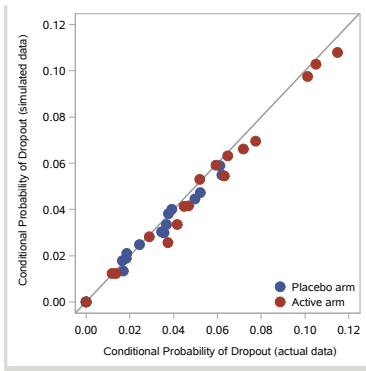
- Under MAR (i.e., $\alpha = 0$), the estimated means of interest are 69.60 and 74.37 for the PBO and PP1M arms.
- The estimated treatment difference is -4.77 (95% CI: -10.89 to 0.09).
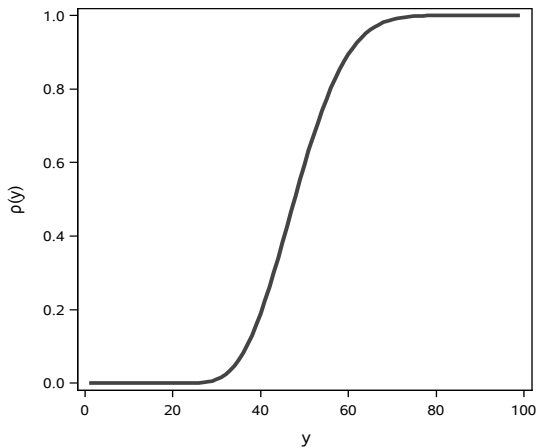
y_{k+1}	y_{k+1}^*	Log Odds Ratio
30	20	$\alpha \times 0.01$
40	30	$\alpha \times 0.18$
50	40	$\alpha \times 0.40$
60	50	$\alpha \times 0.30$
70	60	$\alpha \times 0.09$
80	700	$\alpha \times 0.01$
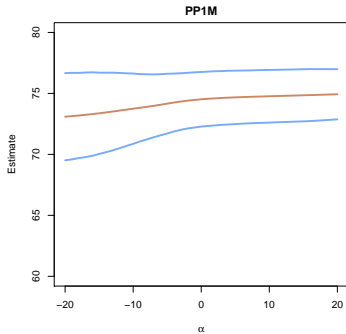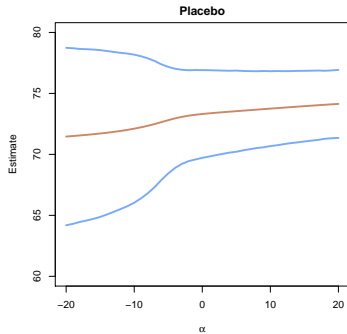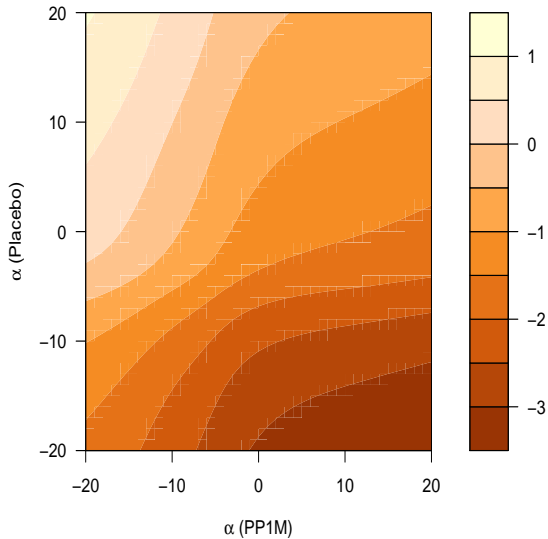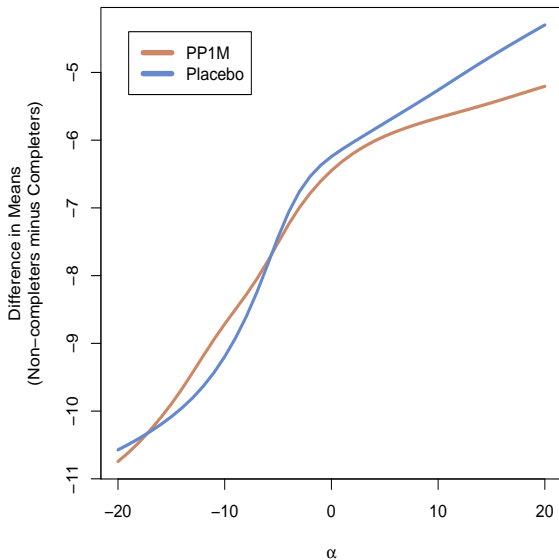
Simulation Study

α	Estimator	PBO			PP1M		
		μ^*	Bias	MSE	μ^*	Bias	MSE
-10	$\mu(\widehat{P})$	72.89	0.76	1.75	73.76	0.41	1.36
	$\widehat{\mu}$		0.50	1.58		0.31	1.26
-5	$\mu(\widehat{P})$	73.38	0.52	1.42	74.25	0.26	1.14
	$\widehat{\mu}$		0.31	1.32		0.16	1.05
-1	$\mu(\widehat{P})$	73.74	0.38	1.23	74.59	0.17	1.02
	$\widehat{\mu}$		0.19	1.18		0.06	0.95
0	$\mu(\widehat{P})$	73.80	0.36	1.21	74.63	0.16	1.01
	$\widehat{\mu}$		0.18	1.17		0.08	0.95
1	$\mu(\widehat{P})$	73.84	0.35	1.19	74.67	0.18	1.01
	$\widehat{\mu}$		0.17	1.15		0.05	0.94
5	$\mu(\widehat{P})$	74.00	0.30	1.13	74.67	0.16	1.00
	$\widehat{\mu}$		0.13	1.11		0.04	0.93
10	$\mu(\widehat{P})$	74.15	0.24	1.08	74.84	0.15	0.97
	$\widehat{\mu}$		0.10	1.08		0.06	0.91

Simulation Study

α	Procedure	PBO Coverage	PP1M Coverage
-10	Normal-IF	86.1%	88.6%
	Normal-JK	92.1%	92.6%
	Bootstrap-IF-ET	90.2%	91.9%
	Bootstap-JK-ET	92.4%	93.7%
	Bootstap-IF-S	92.3%	92.7%
	Bootstap-JK-S	93.9%	94.3%
-5	Normal-IF	89.0%	91.7%
	Normal-JK	94.1%	94.2%
	Bootstrap-IF-ET	91.7%	92.6%
	Bootstrap-JK-ET	93.6%	94.9%
	Bootstap-IF-S	94.1%	94.2%
	Bootstap-JK-S	95.1%	95.1%
-1	Normal-IF	90.8%	93.4%
	Normal-JK	94.9%	94.8%
	Bootstrap-IF-ET	91.0%	94.0%
	Bootstrap-JK-ET	92.8%	94.9%
	Bootstap-IF-S	94.4%	94.7%
	Bootstap-JK-S	95.0%	95.3%
0	Normal-IF	90.7%	93.5%
	Normal-JK	95.0%	94.9%
	Bootstrap-IF-ET	92.8%	93.9%
	Bootstrap-JK-ET	94.3%	95.0%
	Bootstap-IF-S	95.3%	94.7%
	Bootstap-JK-S	96.0%	95.1%

Simulation Study

α	Procedure	PBO Coverage	PP1M Coverage
1	Normal-IF	90.9%	93.5%
	Normal-JK	94.9%	94.8%
	Bootstrap-IF-ET	92.8%	93.5%
	Bootstap-JK-ET	94.2%	95.0%
	Bootstrap-IF-S	95.3%	94.6%
	Bootstap-JK-S	96.0%	95.2%
5	Normal-IF	91.5%	93.7%
	Normal-JK	94.6%	95.1%
	Bootstrap-IF-ET	92.6%	93.8%
	Bootstrap-JK-ET	93.8%	94.7%
	Bootstrap-IF-S	94.9%	95.1%
	Bootstap-JK-S	96.0%	95.5%
10	Normal-IF	92.1%	93.4%
	Normal-JK	94.8%	95.0%
	Bootstrap-IF-ET	92.9%	93.8%
	Bootstrap-JK-ET	93.9%	94.8%
	Bootstrap-IF-S	94.7%	95.0%
	Bootstap-JK-S	95.6%	95.4%

Intermittent Missing Data

- Propose a method for multiply imputing missing data prior to the last study visit in order to create a monotone missing data structure.
- Previous methods are applied to the monotonized datasets.
- Results are averaged across imputed datasets.
- Confidence intervals computed using methods that properly accounting for uncertainty due to imputation.

Data Structure

- M_k: indicator that Y_k is unobserved at time k.
- $M_0 = 0$ and $M_C = 0$.
- $M_k = 1$ if $R_k = 0$.
- $O_k = (M_k, Y_k : M_k = 0)$.
- Observed data for an individual are $\overline{O}_K$.
- $O_0 = Y_0$ and C can be computed from $\overline{O}_K$ as $\max\{k : M_k = 0\}$.

Assumption

For $0 < k < C$,

$$M_k \perp Y_k \,\big|\, \overline{Y}_{k-1}, \underline{O}_k \qquad (18)$$

- While on-study, the probability of providing outcome data at time k can depend on previous outcomes (observed or not) and observed data after time k.
- Imagine a stratum of individuals who share the same history of outcomes prior to time k and same observed data after time k.
- Imagine splitting the stratum into two sets: those who provide outcome data at time k (stratum B) and those who do not (stratum A).
- The distribution of the outcome at time k is the same for these two strata.

Assumption

For $0 < k < C$,

$$f^*(Y_k | \underbrace{M_k = 1, \overline{Y}_{k-1}, \underline{O}_k}_{\text{Stratum A}}) = f^*(Y_k | \underbrace{M_k = 0, \overline{Y}_{k-1}, \underline{O}_k}_{\text{Stratum B}}) \quad (19)$$

Using Bayes' rule, (19) can be written as follows:

$$P^*(M_k = 1 \mid \overline{Y}_k, \underline{O}_k) = P^*(M_k = 1 \mid \overline{Y}_{k-1}, \underline{O}_k): \quad 0 < k < C. \quad (20)$$

In our imputation algorithm, we will use the following fact:

$$M_k \perp Y_k \mid \rho_k^*(\overline{Y}_{k-1}, \underline{O}_k) : \quad 0 < k < C \qquad (21)$$

where

$$\rho_k^*(\overline{Y}_{k-1}, \underline{O}_k) = P^*(M_k = 1 \mid \overline{Y}_{k-1}, \underline{O}_k) \qquad (22)$$

Under assumption (18), the joint distribution of $(C, \overline{Y}_C)$ (i.e., the monotonized data) is identified by a recursive algorithm.

Illustration
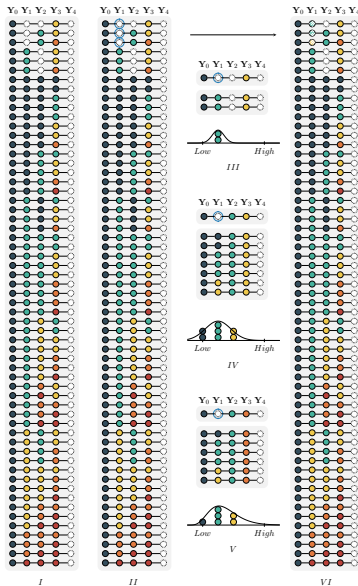
Imputation
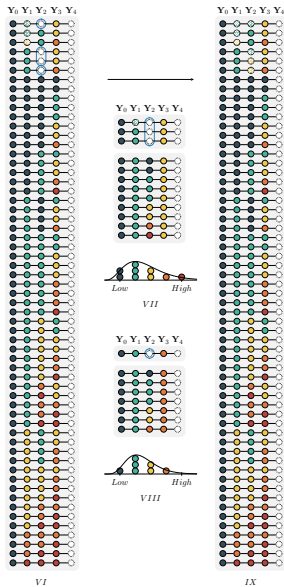
- The number of individuals contributing to the histograms that form of the basis of the imputation strategy may be quite small.
- Rather than matching on the past outcomes and future observed data, we plan to use (21) and match on estimates of $\rho_k^*(\overline{Y}_{k-1}, \underline{O}_k)$.

$$\text{logit}\{\rho_k^*(\overline{Y}_{k-1}, \underline{O}_k)\} = w_k(\overline{Y}_{k-1}, \underline{O}_k; \nu_k^*); \quad k = 1, \ldots, K-1 \tag{23}$$

where $w_k(\overline{Y}_{k-1}, \underline{O}_k; \nu_k)$ is a specified function of its arguments and ν_k is a finite-dimensional parameter with true value ν_k^*.

Simultaneous Estimation/Imputation

The parameters ν_k^* $(k = 1, \ldots, K-1)$ can be estimated and the intermittent missingness can be imputed using the following sequential procedure:

1. Set $k = 1$.
2. Estimate ν_k^* by $\widehat{\nu}_k$ as the solution to:

$$\sum_{i=1}^{n} R_{k,i} d_k(\overline{Y}_{k-1,i}, \underline{O}_{k,i}; \nu_k) \left(M_{k,i} - \text{expit}\{w_k(\overline{Y}_{k-1,i}, \underline{O}_{k,i}; \nu_k)\} \right) = 0,$$

where

$$d_k(\overline{Y}_{k-1}, \underline{O}_k; \nu_k^*) = \frac{\partial w_k(\overline{Y}_{k-1}, \underline{O}_k; \nu_k)}{\partial \nu_k}$$

Simultaneous Estimation/Imputation

3. For each individual i with $R_{k,i} = 1$, compute

$$\widehat{\rho}_k(\overline{Y}_{k-1,i}, \underline{O}_{k,i}) = \text{expit}\{w_k(\overline{Y}_{k-1,i}, \underline{O}_{k,i}; \widehat{\nu}_k)\}.$$

Let

$$\mathcal{J}_k = \{i : R_{k,i} = 1, M_{k,i} = 0\}$$
$$\mathcal{J}'_k = \{i : R_{k,i} = 1, M_{k,i} = 1\}.$$

For each individual $i \in \mathcal{J}'_k$, impute $Y_{k,i}$ by randomly selecting an element from the set

$$\big\{ Y_{k,l} : l \in \mathcal{J}_k, \widehat{\rho}_k(\overline{Y}_{k-1,l}, \underline{O}_{k,l}) \text{ is "near" } \widehat{\rho}_k(\overline{Y}_{k-1,i}, \underline{O}_{k,i}) \big\}$$

4. Set $k = k + 1$. If $k = K$ then stop. Otherwise, return to Step 2.

Inference

- Use algorithm to create to M monotone missing datasets.
- Apply monotone missing data methods to each of these datasets.
- Overall point estimates are obtained by averaging across imputed datasets.

$$\tilde{\mu}_\alpha = \frac{1}{M} \sum_{m=1}^{M} \tilde{\mu}_{\alpha,m},$$

where $\tilde{\mu}_{\alpha,m}$ is the corrrected estimator of μ^* based on the mth imputed dataset.

Confidence Intervals

When $M > 1$, we can replace $\tilde{\sigma}_\alpha^2$ with Rubin's (1987) multiple imputation variance estimator, i.e.,

$$\tilde{\sigma}_\alpha^2 = \frac{1}{M} \sum_{m=1}^{M} \tilde{\sigma}_{\alpha,m}^2 + \left(1 + \frac{1}{M}\right) \frac{1}{M-1} \sum_{m=1}^{M} (\tilde{\mu}_{\alpha,m} - \tilde{\mu}_\alpha)^2 \quad (24)$$

- In simulations, we have found success using (24) coupled with symmteric bootstrap to form confidence intervals.

Bootstrap

Let $\mathcal{D}$ be the observed dataset. To create a bootstrap dataset $\mathcal{D}^{(b)}$, use the following procedure:

1. Use $\mathcal{D}$ to estimate the $\widehat{\nu}_k$'s and impute a monotonized dataset $\mathcal{D}^{\dagger}$.

2. Using $\mathcal{D}^{\dagger}$, estimate of $F_0^*(y_0)$, $F_{k+1}^*(y_{k+1}|y_k)$ and $H_{k+1}^*(y_k)$ and simulate a new monotonized dataset $\mathcal{D}^{\ddagger}$.

3. Use $\mathcal{D}^{\ddagger}$ and the $\widehat{\nu}_k$'s from Step 1 to create a non-monotone dataset $\mathcal{D}^{(b)}$.

Bootstrap

In Step 3, we create a non-monotone dataset by applying the following procedure to each patient i with $C_i > 1$:

1. Set $k = C_i - 1$.
2. Generate $U \sim \text{Uniform}(0, 1)$. If $U < \widehat{\rho}_k(\overline{Y}_{k-1,i}, \underline{O}_{k,i})$, set $M_{k,i} = 1$ and delete $Y_{k,i}$; otherwise set $M_{k,i} = 0$ and retain $Y_{k,i}$.
3. Set $k = k - 1$. If $k = 0$ then stop; otherwise go to step 2.

Diabetic Peripheral Polyneuropathy

- Peripheral neuropathy is a common complication of diabetes.
- Diabetic peripheral polyneuropathy is characterized by damage to small-diameter sensory fibers in distal areas of the peripheral nervous system.
- This condition commonly manifests itself by painful tingling or burning sensations in the hands and feet.
- This pain can be so severe that it compromises day-to-day actitivities and quality of life.

Topiramate

- Topiramate is an approved medication for the treatment of epileptic seizures.
- It operates by dampening neuronal hyperexcitability in the central nervous system.
- It was hypothesized that topiramate might also dampen the excitability of nerves in peripheral nervous system.
- Small studies were conducted that showed that topiramate reduced the pain associated with peripheral neuropathies, including diabetic peripheral neuropathy.
- Based on these data, three placebo-controlled randomized trials to evaluate the efficacy of different doses of topiramate in reducing pain in patients with diabetic peripheral polyneuropathy (Thienel *et al.*, 2004).

NP 001 and 002

- Two these studies had nearly identical designs and will form the basis of our second case study.
- In Studies NP 001 and 002, there were baseline and double-blind phases.
- Eligibility was determined during the baseline phase that lasted up to 28 days.
- At least 7 days before randomization, subjects must have been tapered off all background medications being used to treat neuropathic pain.
- During the baseline phase, all subjects were to have their diabetes controlled on a stable regimen of oral hypoglycemics, insulin, or diet alone.
- The double-blind phase included 2 periods: a 10 week titration period and a 12 week maintenance peiod.

- The primary efficacy variable was the pain score measured on a 100-mm Visual Analog Scale (VAS), where higher levels of VAS indicate worse pain.
- VAS scores were scheduled on day 1 of the baseline phase, every two weeks during titration, and then monthly during the maintenance phase.
- Treatment effects were based on the difference in the mean VAS scores at the final scheduled follow-up visit.
- Adverse events and use of rescue medications was also scheduled to be monitored throughout the double-blind phase.
- The trials were not designed to follow patients after they discontinued their assigned therapy.

- In NP 001, 531 subjects were randomized to one of four study arms: placebo ($n = 137$), 100 mg/day ($n = 129$), 200 mg/day ($n = 132$), and 400 mg/day ($n = 133$).
- In NP 002, 370 subjects were randomized to one of three study arms: placebo ($n = 123$), 200 mg/day ($n = 118$), and 400 mg/day ($n = 129$).
- Seven subjects in NP 001 and six subjects NP 002 did not have at least one follow-up visit and were not considered part of the intent-to-treat (ITT) population.

- In our analysis, we merge the data from the two studies.
- We focus our analysis on a comparison of the placebo versus 400 mg/day arms.
- One individual from the 400 mg/day arm was excluded because of undue influence on the analysis.
- The sample sizes are 255 and 256 in the placebo and 400 mg/day arms, respectively.

```
              Placebo
Monotone:
                N      %
*_____  :    5    0.0196
**_____  :    5    0.0196
***_____  :   10    0.0392
****_____  :    3    0.0118
*****____  :   19    0.0745
******___  :   12    0.0471
*******__  :   12    0.0471
********_  :    5    0.0196
*********  :   81    0.3176

Intermittent:
                N      %
*_*_*****  :   14    0.0549
*_*******  :   13    0.0510
****_****  :    7    0.0275
***_*****  :    6    0.0235
******_**  :    5    0.0196
Other      :   47    0.1843
```

```
              400 mg/day
Monotone:
                  N      %
*_____  :    4    0.0156
**_____  :   14    0.0547
***_____  :   19    0.0742
****_____  :    7    0.0273
*****____  :   19    0.0742
******___  :   10    0.0391
*******__  :    9    0.0352
********_  :    2    0.0078
*********  :   67    0.2617

Intermittent:
                  N      %
*_*_____  :   15    0.0586
*_*_*****  :    9    0.0352
*_*******  :    8    0.0312
***_*****  :    7    0.0273
*_*_***__  :    5    0.0195
Other      :   56    0.2188
```

What is the difference in the mean VAS scores at the end of the double blind phase between topiramate at a specified dose level vs. placebo in the counterfactual world in which there is no missing data at that visit?

Observed Data

		Placebo			
		Number			
	On-	Last	Obs.		Std.
k	Study	Seen	Value	Mean	Dev.
0	255	5	255	58.902	19.196
1	250	5	188	53.202	23.048
2	245	14	238	48.899	24.888
3	231	5	186	45.849	23.928
4	226	27	203	42.291	25.338
5	199	24	192	38.896	25.117
6	175	15	162	37.549	25.827
7	160	10	150	35.047	26.313
8	150	150	150	35.613	26.446

Observed Data

<table>
<thead>
<tr><th colspan="7" align="center">400 mg/day</th></tr>
<tr><th></th><th colspan="3" align="center">Number</th><th></th><th></th></tr>
<tr><th></th><th>On-</th><th>Last</th><th>Obs.</th><th></th><th>Std.</th></tr>
<tr><th>k</th><th>Study</th><th>Seen</th><th>Value</th><th>Mean</th><th>Dev.</th></tr>
</thead>
<tbody>
<tr><td>0</td><td>256</td><td>4</td><td>256</td><td>58.305</td><td>19.958</td></tr>
<tr><td>1</td><td>252</td><td>14</td><td>192</td><td>51.297</td><td>22.605</td></tr>
<tr><td>2</td><td>238</td><td>34</td><td>223</td><td>47.466</td><td>25.268</td></tr>
<tr><td>3</td><td>204</td><td>12</td><td>162</td><td>44.228</td><td>22.956</td></tr>
<tr><td>4</td><td>192</td><td>28</td><td>174</td><td>41.879</td><td>23.851</td></tr>
<tr><td>5</td><td>164</td><td>26</td><td>159</td><td>36.528</td><td>24.101</td></tr>
<tr><td>6</td><td>138</td><td>20</td><td>133</td><td>36.211</td><td>24.334</td></tr>
<tr><td>7</td><td>118</td><td>6</td><td>109</td><td>33.138</td><td>21.842</td></tr>
<tr><td>8</td><td>112</td><td>112</td><td>112</td><td>31.482</td><td>22.149</td></tr>
</tbody>
</table>

Observed Data
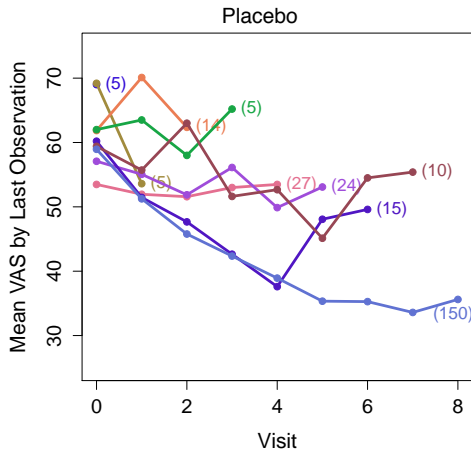
Placebo

Observed Data
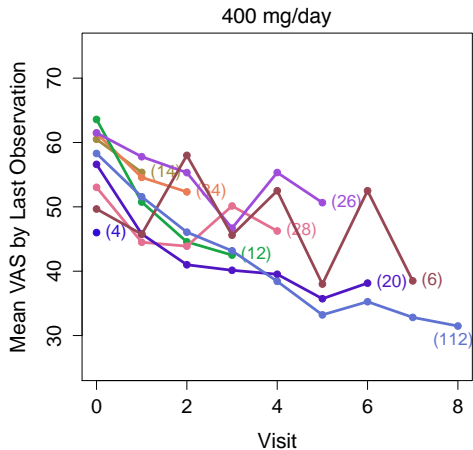
400 mg/day

Observed Data

- In the placebo arm, 59.6% of individuals have a monotone missing data pattern, with only 31.8% having complete data.

- In the 400 mg/day arm, these numbers are 59.0% and 26.2%.

- There is a statistically significant difference in the proportion of individuals who completed the study in the placebo versus 400/day arms (58.8% vs. 42.8%; $p < 0.001$).

- The primary reason for premature discontinuation of the study differed by treatment arm.

- The most common reason for placebo patients was lack of efficacy and, for 400/mg day patients, it was adverse events.

Observed Data

- In both treatment arms, there is a decline in the average observed VAS scores through time.
- The mean of the observed VAS scores at time $K = 8$ is 35.6 and 31.48 in the placebo versus 400/day arms, respectively.
- A naive t-test based of the observed outcomes at time $K = 8$ does not suggest a statistical difference between the treatment arms ($p = 0.17$).
- Patients who prematurely discontinue the study tend to have higher VAS scores at their penultimate visit than those who complete the study. This is true for both treatment arms, although the differences appear somewhat larger in the placebo group.

Observed Data

- Using last observation carried forward, the means at time $K = 8$ are 43.8 and 40.6 in the placebo versus 400/day arms, respectively. The estimated treatment difference between 400 mg/day and placebo of -3.3.
- A t-test based on LOCF also does not suggest a statistical difference between the treatment arms ($p = 0.18$).
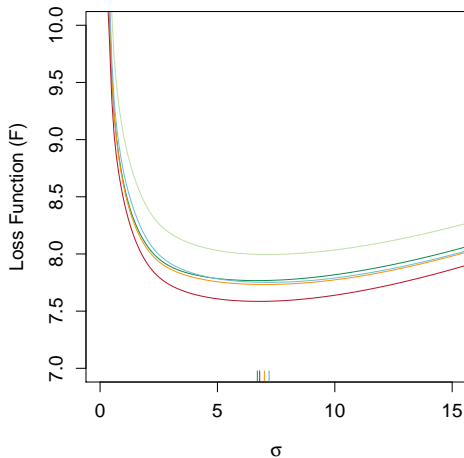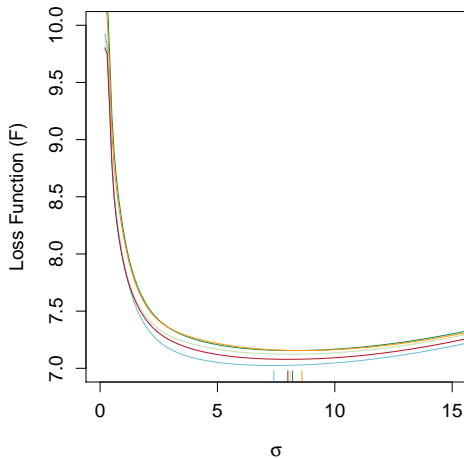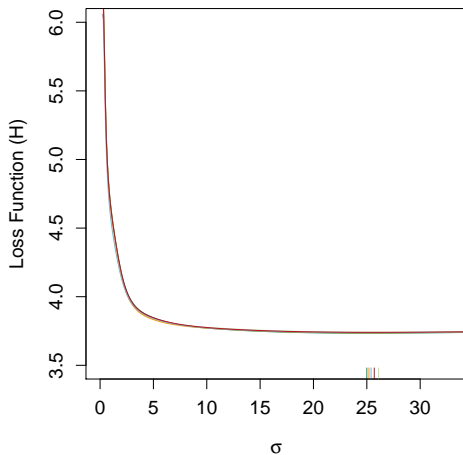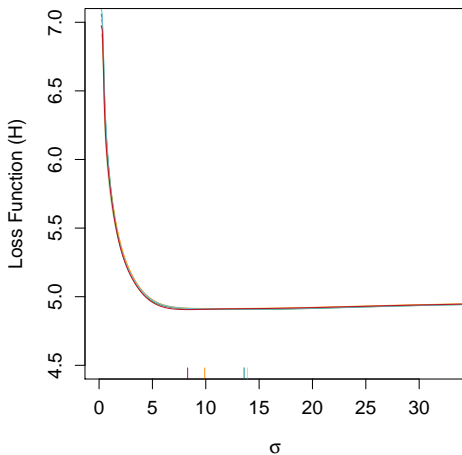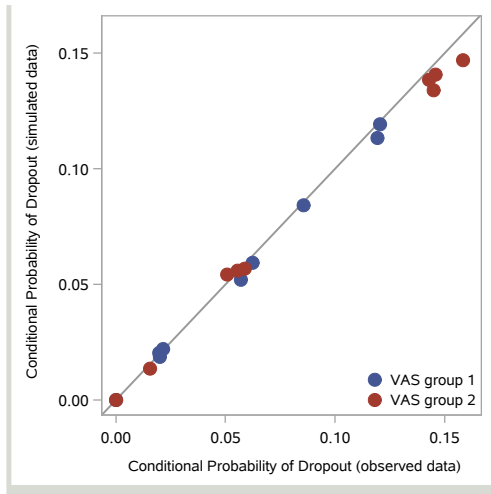
Goodness of Fit
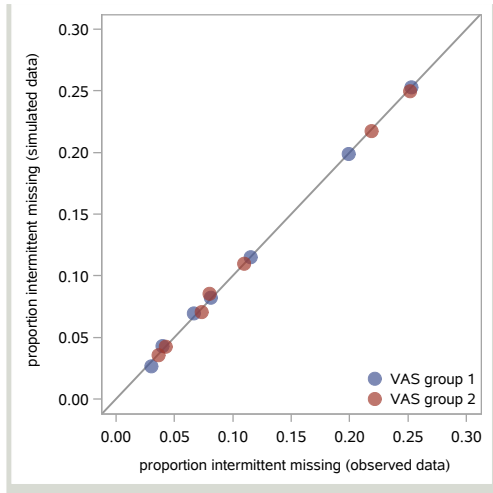
Goodness of Fit

Goodness of Fit
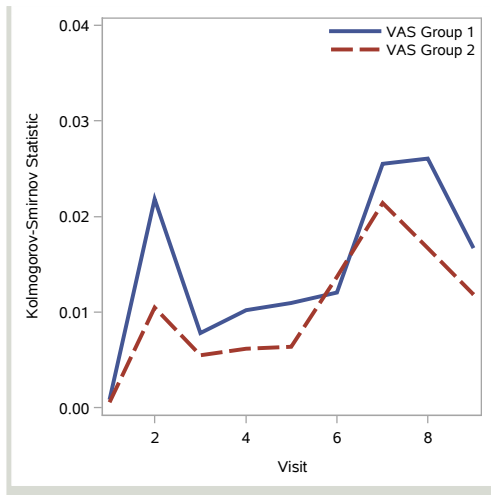
MAR Analysis

- The estimates of μ^* are 39.07 (95% CI: 34.19 to 43.95) and 33.06 (95% CI: 28.33 to 37.78) in the placebo and 400 mg/day arms, respectively.
- These estimates correct for the fact that individuals with higher VAS scores appear to be dropping out of the study.
- The correction is bigger for placebo versus 400 mg/day arm.
- The estimated difference in means between the arms is -6.01 (95% CI: -11.70, -0.329), indicating a statistically significant difference in favor of the 400 mg/day arm. This is a different inference than the naive inferences reported above.
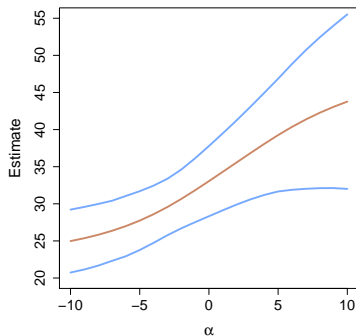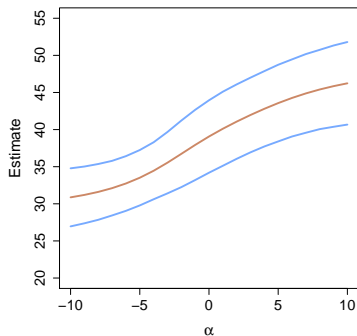
Sensitivity Analysis
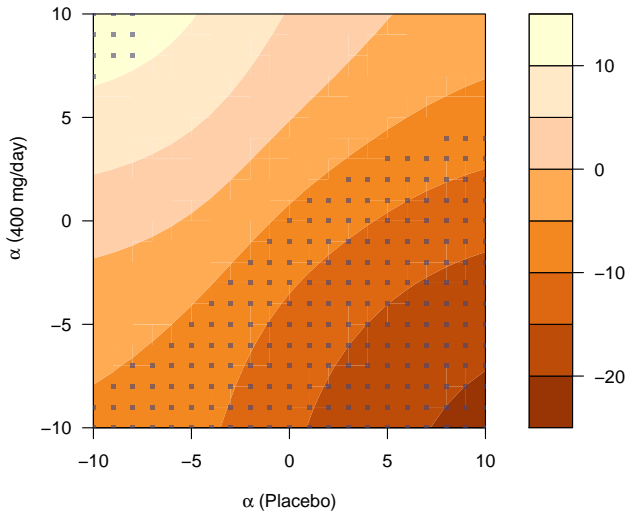
Sensitivity Analysis
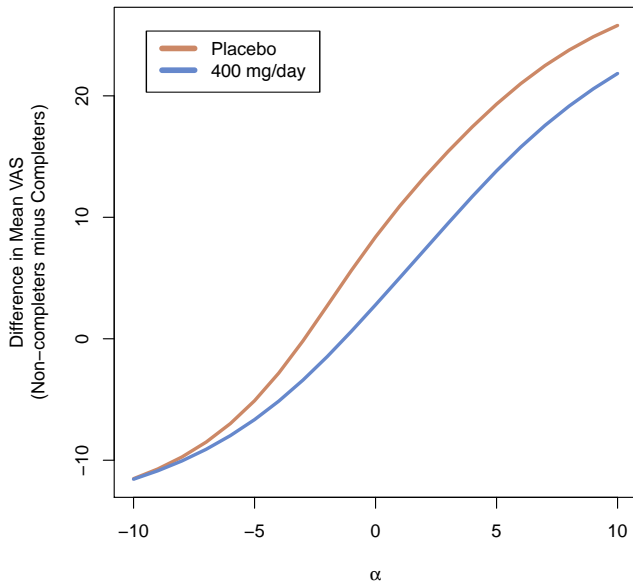
Simulation Study - Five Imputes

α	Estimator	PBO μ^*	Bias	MSE	400 mg/day μ^*	Bias	MSE
-10	$\mu(\widehat{P})$	31.69	0.53	4.13	25.88	1.05	4.75
	$\widehat{\mu}$		0.10	3.76		0.60	3.75
-5	$\mu(\widehat{P})$	33.60	0.03	4.03	28.20	0.37	3.89
	$\widehat{\mu}$		-0.02	3.93		0.26	3.55
-1	$\mu(\widehat{P})$	37.10	-0.70	4.93	31.40	-0.30	3.99
	$\widehat{\mu}$		-0.32	4.41		-0.08	3.66
0	$\mu(\widehat{P})$	38.12	-0.82	5.23	32.35	-0.45	4.17
	$\widehat{\mu}$		-0.36	4.53		-0.16	3.77
1	$\mu(\widehat{P})$	39.10	-0.89	5.47	33.32	-0.59	4.39
	$\widehat{\mu}$		-0.37	4.65		-0.22	3.92
5	$\mu(\widehat{P})$	42.75	-1.06	6.43	37.32	-1.32	6.26
	$\widehat{\mu}$		-0.49	5.35		-0.62	5.11
10	$\mu(\widehat{P})$	45.59	-1.43	8.10	41.07	-2.48	11.39
	$\widehat{\mu}$		-0.70	6.50		-1.38	8.11

Simulation Study - Five Imputes

α	Procedure	PBO Coverage	400 mg/day Coverage
-10	Normal-IF(Rubin)	94.6%	93.6%
	Normal-BootstrapSE	94.4%	93.8%
	Bootstrap-Percentile	92.6%	86.6%
	Bootstap-IF(Rubin)-ET	93.8%	94.5%
	Bootstap-IF(Rubin)-S	95.4%	95.5%
-5	Normal-IF(Rubin)	93.9%	94.5%
	Normal-BootstrapSE	94.3%	94.6%
	Bootstrap-Percentile	93.8%	93.3%
	Bootstap-IF(Rubin)-ET	94.3%	95.8%
	Bootstap-IF(Rubin)-S	94.7%	95.2%
-1	Normal-IF(Rubin)	92.8%	94.8%
	Normal-BootstrapSE	92.8%	94.7%
	Bootstrap-Percentile	92.0%	93.4%
	Bootstap-IF(Rubin)-ET	94.0%	96.0%
	Bootstap-IF(Rubin)-S	94.8%	95.2%
0	Normal-IF(Rubin)	92.7%	95.1%
	Normal-BootstrapSE	92.9%	95.0%
	Bootstrap-Percentile	91.3%	92.5%
	Bootstap-IF(Rubin)-ET	93.8%	95.9%
	Bootstap-IF(Rubin)-S	94.9%	96.1%

Simulation Study - Five Imputes

α	Procedure	PBO Coverage	400 mg/day Coverage
1	Normal-IF(Rubin)	93.1%	94.6%
	Normal-BootstrapSE	92.9%	94.4%
	Bootstrap-Percentile	90.7%	91.8%
	Bootstap-IF(Rubin)-ET	94.3%	95.7%
	Bootstap-IF(Rubin)-S	95.0%	96.2%
5	Normal-IF(Rubin)	93.3%	93.2%
	Normal-BootstrapSE	93.4%	93.2%
	Bootstrap-Percentile	89.8%	84.5%
	Bootstap-IF(Rubin)-ET	94.3%	93.3%
	Bootstap-IF(Rubin)-S	95.2%	96.0%
10	Normal-IF(Rubin)	94.0%	88.6%
	Normal-BootstrapSE	93.5%	88.4%
	Bootstrap-Percentile	86.8%	70.6%
	Bootstap-IF(Rubin)-ET	93.6%	89.7%
	Bootstap-IF(Rubin)-S	96.4%	94.8%

Honey-do List

- Develop data adaptive technique for handling outliers
- Incorporate auxiliary covariates

Missing Data Matters

- No substitute for better trial design and procedures to minimize missing data.
- Global sensitivity analysis should be a mandatory component of trial reporting.
- Visit us at `www.missingdatamatters.org` or email me at dscharf@jhu.edu

Inference in Randomized Trials with Death and Missingness

Harvard Shortcourse

Daniel Scharfstein

Johns Hopkins University

dscharf@jhu.edu

September 24, 2019

HT-ANAM 302 Study

- Anamorelin is a drug developed for the treatment of cancer cachexia and anorexia.
- HT-ANAM 302 was a randomized, double-blind, placebo-controlled Phase III study designed to evaluate the efficacy of anamorelin in patients with advanced non-small cell lung cancer.
- Lean body mass (LBM) was scheduled to be measured at baseline (Y_0), 6 weeks (Y_1) and 12 weeks (Y_2)
- Primary functional endpoint: $Z = \frac{(Y_2 + Y_1)}{2} - Y_0$

	Placebo $n = 157$	Anamorelin $n = 322$
Died Prior to Wk 12	24 (15.3%)	54 (16.8%)
Survivors with complete data	93 (59.2%)	185 (57.5%)
Survivors missing only Wk 6	3 (1.9%)	17 (5.3%)
Survivors missing only Wk 12	17 (10.8%)	31 (9.6%)
Survivors missing both Wks 6, 12	20 (12.7%)	35 (10.9%)

How should data from studies like HT-ANAM 302 be analyzed
to evaluate the effect of treatment on the functional outcome?

Key Issue

- Distinction between missing data and data truncated by death
  - Missing data: exist but not collected
  - Data truncated by death: does not exist and undefined
- Can't just treat as a missing data problem.

Common Approaches

1. Evaluate treatment effect on functional outcome conditional on survival
   - Conditioning on post-baseline factor
2. Joint modeling survival and functional outcomes
   - Allows extrapolation of outcomes after death
3. Principal stratification
   - Applies to a subset of patients who are not identifiable at baseline
4. Composite endpoint combining survival and functional outcomes
   - May be hard to separate effect on function.

NO PERFECT SOLUTIONS

Not a fan of Approaches 1 and 2.

To construct a composite endpoint approach that handles both death and missing data

Notation

- $T = 0, 1$: treatment assignment
- X vector baseline covariates
- Y_0: baseline functional measure at t_0
- $Y_1, \ldots, Y_K$: functional outcomes at $t_1, \ldots, t_K$
- L: survival time
- $A_k = I(L > t_k)$: survival status at t_k
- $Z = g(Y_0, \ldots, Y_K)$: primary functional endpoint
    - e.g. $K = 2$, $Z = (Y_2 + Y_1)/2 - Y_0$
    - only defined when $A_K = 1$

Composite Outcome

Finite-valued random variable U which assigns a score to each patient such that

- each patient who dies prior to t_K is assigned a score according to their survival time (L), with shorter survival times assigned lower scores
- each patient who survives past t_K is assigned a score (higher than those who died prior to t_K) according to their functional status (Z), with lower functional status assigned lower scores.

Only the ordering of U is important, not the actual score assignments.

Treatment Effect

Treatment effect (θ) is measured by the probability that the outcome for an individual with $T = 0$ is less than the outcome of an individual with $T = 1$ minus the probability that the outcome for an individual with $T = 0$ is greater than the outcome of an individual with $T = 1$

- $\theta = 0$ under the null
- $\theta > 0$ favors $T = 1$; $\theta < 0$ favors $T = 0$
- First part: Mann-Whitney
- Second part: needed to handle ties

Can also compare the treatment-specific quantiles of U.

In the absence of missing data,

$$\widehat{\theta} = \frac{1}{n_0 n_1} \sum_{i:\, T_i=0} \sum_{j:\, T_j=1} \{I(U_i < U_j) - I(U_i > U_j)\}$$

where $n_0 = \sum_i (1 - T_i)$ and $n_1 = \sum_i T_i$.

Missing Data

- R_k: missing data indicator (defined when $A_k = 1$)
- $S = (R_1, \ldots, R_K)$ (defined when $A_K = 1$)
  - $Y_{obs}^{(s)} = \{Y_k : R_k = 1, k \geq 1, S = s\}$
  - $Y_{mis}^{(s)} = \{Y_k : R_k = 0, k \geq 1, S = s\}$
  - Z is unobserved when $S \neq \mathbf{1}$.

To estimate θ, need to impute Z or equivalently $Y_{mis}^{(s)}$ for $s \neq \mathbf{1}$

Missing Data Assumptions

$$f(Y_{mis}^{(s)}|A_K = 1, Y_{obs}^{(s)}, Y_0, X, T, S = s)$$
$$\propto \exp(\beta_T Z) \underbrace{f(Y_{mis}^{(s)}|A_K = 1, Y_{obs}^{(s)}, Y_0, X, T, S = \mathbf{1})}_{\text{Reference Distribution}}$$

for all $s \neq \mathbf{1}$,

- β_T is a treatment-specific sensitivity parameter.
- $\beta_T = 0$ (i.e., benchmark assumption) reduces to the complete case missing value (CCMV) restrictions applied to the missing data patterns for patients alive at t_K.
- CCMV is different than missing at random (MAR) assumption.

- $K = 2$, $Z = (Y_1 + Y_2)/2 - Y_0$.
- $\beta'_T = 2\beta_T$

$$f(Y_2|A_2 = 1, Y_1, Y_0, X, T, S = (1, 0))$$
$$\propto \exp(\beta'_T Y_2) \underbrace{f(Y_2|A_2 = 1, Y_1, Y_0, X, T, S = \mathbf{1})}_{\text{Reference Distribution}}$$

For subjects alive at t_2, who are observed at time t_1, who share the same functional measure at t_1 and who share the same baseline factors, the distribution of Y_2 for those whose functional measure at t_2 is missing is, when $\beta'_T > 0$ (< 0), more heavily weighted toward higher (lower) values of Y_2 than those whose functional measure at t_2 is observed.

$$f(Y_1 | A_2 = 1, Y_2, Y_0, X, T, S = (0,1))$$
$$\propto \exp(\beta'_T Y_1) \underbrace{f(Y_1 | A_2 = 1, Y_2, Y_0, X, T, S = \mathbf{1})}_{\text{Reference Distribution}}$$

For subjects alive at t_2, who are observed at time t_2, who share the same functional measure at t_2 and who share the same baseline factors, the distribution of Y_1 for those whose functional measure at t_1 is missing is, when $\beta'_T > 0$ (< 0), more heavily weighted toward higher (lower) values of Y_1 than those whose functional measure at t_1 is observed.

$$f(Y_1, Y_2 | A_2 = 1, Y_0, X, T, S = (0, 0))$$
$$\propto \exp\left(\beta'_T(Y_1 + Y_2)\right) \underbrace{f(Y_1, Y_2 | A_2 = 1, Y_0, X, T, S = \mathbf{1})}_{\text{Reference Distribution}}$$

For subjects alive at t_2 and who share the same baseline factors, the joint distribution of Y_1 and Y_2 for those whose functional measures at t_1 and t_2 are missing is, when $\beta'_T > 0$ (< 0), more heavily weighted toward higher (lower) values of Y_1 and Y_2 than those whose measures are fully observed.

HT-ANAM 302 Study

- Ignore conditioning on Y_0 and X and suppose $f(Y_1, Y_2 | A_2 = 1, T, S = \mathbf{1})$ is multivariate normal with mean $(\mu_{T,1}, \mu_{T,2})$ and variance-covariance matrix

$$\Sigma_T = \begin{bmatrix} \sigma_{T,1}^2 & \rho_T \sigma_{T,1} \sigma_{T,2} \\ \rho_T \sigma_{T,1} \sigma_{T,2} & \sigma_{T,2}^2 \end{bmatrix}$$

- $f(Y_2 | A_2 = 1, Y_1, T, S = (1,0))$ is normal with mean $\mu_{T,2} + \beta_T' (1 - \rho_T^2) \sigma_{T,2}^2 + \rho_T \frac{\sigma_{T,2}}{\sigma_{T,1}} (Y_1 - \mu_{T,1})$ and variance $(1 - \rho_T^2) \sigma_{T,2}^2$

- $f(Y_1 | A_2 = 1, Y_2, T, S = (0,1))$ is normal with mean $\mu_{T,1} + \beta_T' (1 - \rho_T^2) \sigma_{T,1}^2 + \rho_T \frac{\sigma_{T,1}}{\sigma_{T,2}} (Y_2 - \mu_{T,2})$ and variance $(1 - \rho_T^2) \sigma_{T,1}^2$

- $f(Y_1, Y_2 | A_2 = 1, T, S = (0, 0))$ is multivariate normal with mean $(\mu_{T,1} + \beta'_T \sigma^2_{T,1} + \beta'_T \rho_T \sigma_{T,1} \sigma_{T,2}, \mu_{T,2} + \beta'_T \sigma^2_{T,2} + \beta'_T \rho_T \sigma_{T,1} \sigma_{T,2})$ and variance-covariance matrix Σ_T.
- If $\rho_T > 0$, then the means increase linearly in β'_T
- β'_T has no impact on the variances and covariances.
- $\beta'_T > 0$ ($\beta'_T < 0$) implies that the non-identified distributions have more (less) mass at higher values than their reference distributions.

Example: Exponential tilting
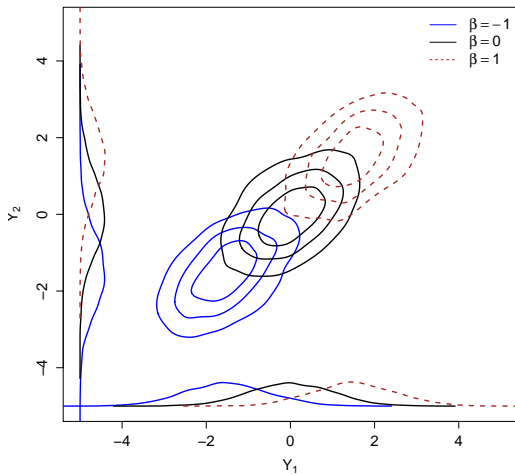
Modeling

Need to specify of a model for

$$f(\overline{Y}_K | A_K = 1, Y_0, X, T, S = \mathbf{1})$$

- To respect bounds, define

$$\phi(y_k) = \log \left\{ \frac{y_k - B_L}{B_U - y_k} \right\},$$

- $Y_k^\dagger = \phi(Y_k)$ and $\overline{Y}_k^\dagger = (Y_1^\dagger, \ldots, Y_k^\dagger)$.
- One-to-one mapping between

$$h(\overline{Y}_K^\dagger | A_K = 1, Y_0, X, T, S = \mathbf{1})$$

and

$$f(\overline{Y}_K | A_K = 1, Y_0, X, T, S = \mathbf{1})$$

.

Modeling

$$h(\overline{Y}_K^\dagger | A_K = 1, Y_0, X, T, S = \mathbf{1}) =$$
$$\prod_{k=1}^K h(Y_k^\dagger | A_K = 1, \overline{Y}_{k-1}^\dagger, Y_0, X, T, S = \mathbf{1})$$

- Posit a model for each component of the product.

Modeling

$$h(Y_k^{\dagger}|A_K = 1, \overline{Y}_{k-1}^{\dagger}, Y_0, X, T = t, S = \mathbf{1})$$
$$= h_{k,t}(Y_k^{\dagger} - \mu_{k,t}(\overline{Y}_{k-1}^{\dagger}, Y_0, X; \boldsymbol{\alpha}_{k,t}))$$

- $\mu_{k,t}(\overline{Y}_{k-1}^{\dagger}, Y_0, X; \boldsymbol{\alpha}_{k,t})$ is a specified function
- $\boldsymbol{\alpha}_{k,t}$ is an unknown parameter vector
- $h_{k,t}$ is an unspecified time/treatment-specific density function.

Estimation

- The parameter vectors $\boldsymbol{\alpha}_{k,t}$ can be estimated by minimizing the least squares objective function

$$\sum_{i=1}^{n} I(T_i = t) A_{K,i} \left(\prod_{k=1}^{K} R_{k,i} \right) \{ Y_{k,i}^{\dagger} - \mu_{k,t}(\overline{Y}_{k-1}^{\dagger}, Y_0, X; \boldsymbol{\alpha}_{k,t}) \}^2$$

- The density function $h_{k,t}$ can be estimated by kernel density estimation based on the residuals
  $\{ Y_{k,i}^{\dagger} - \mu_{k,t}(\overline{Y}_{k-1,i}^{\dagger}, Y_{0,i}, X_i; \widehat{\boldsymbol{\alpha}}_{k,t}) : T_i = t, A_{K,i} = 1, R_{1,i} = \dots, R_{K,i} = 1, i = 1, \dots, n \}$

- $f(\overline{Y}_K | A_K = 1, Y_0, X, T, S = \mathbf{1})$ is estimated by

$$\prod_{k=1}^{K} \widehat{h}_{k,t}(Y_k^{\dagger} - \mu_{k,t}(\overline{Y}_{k-1}^{\dagger}, Y_0, X; \widehat{\boldsymbol{\alpha}}_{k,t})) \left| \frac{d\phi(Y_k)}{dY_k} \right|.$$

Imputation/Estimation

- For each individual i alive at t_K and who is in a stratum $s \neq 1$, impute the missing functional outcomes by drawing (using rejection sampling technqiues) from the density that is proportional to

$$\exp(\beta_T Z)\widehat{f}(Y_{mis}^{(s)}|A_K = 1, Y_{obs}^{(s)} = Y_{obs,i}, Y_0 = Y_{0,i}, X = X_i, T = T_i, S = 1)$$

- Draw M copies of the missing functional outcomes to create M complete datasets.
- For each complete dataset m, estimate θ by $\widehat{\theta}_m$.
- Overall estimator of θ is $\tilde{\theta} = \frac{1}{M}\sum_{m=1}^{M}\widehat{\theta}_m$.
- Confidence intervals can be constructed by non-parametric bootstrap

HT-ANAM 302 Study

- Baseline covariates: ECOG performance status, age, gender, BMI, weight loss in prior 6 months
- LBM is bounded between 24 and 140
- 10 imputed datasets
- Under benchmark assumptions,
    - $\widehat{\theta} = 0.30$ (95% CI: 0.16 to 0.37, $p < 0.0001$)
    - Placebo: Median -0.98 kg (95% CI: -1.27 kg to -0.28 kg).
    - Anamorelin: Median 0.69 kg (95% CI: 0.43 kg to 0.93 kg).

HT-ANAM 302 Study

HT-ANAM 302 Study

(A)

(B)

(C)

Discussion
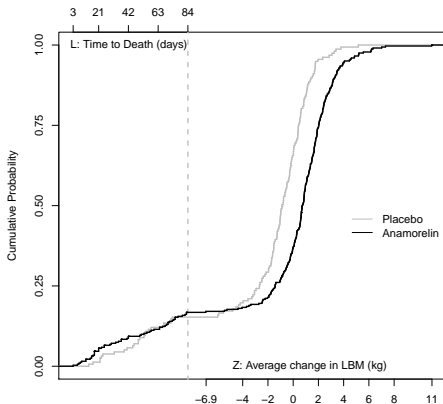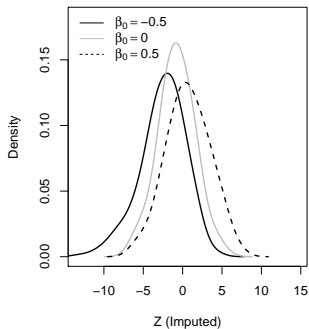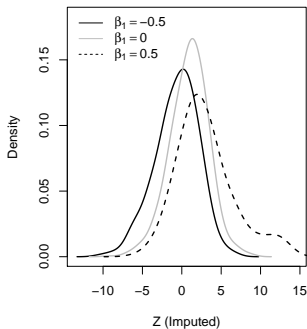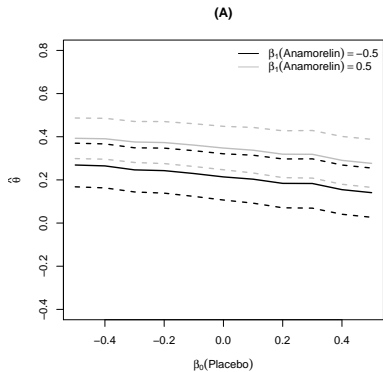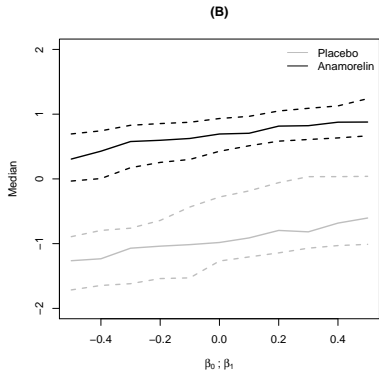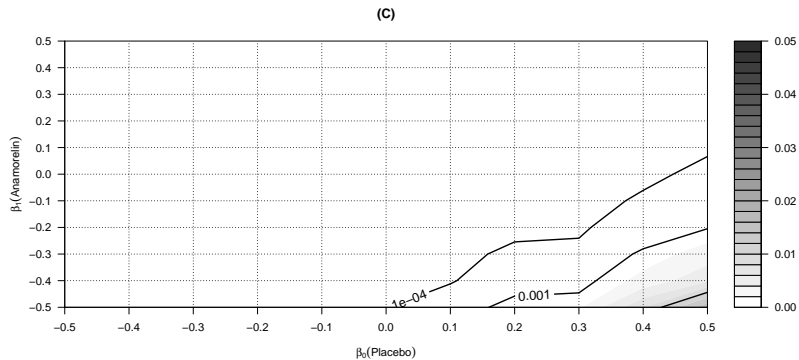
- Method presumes that death and the functional outcome can be ordered in a scientifically meaningful way.
- Use mixed methods to confirm that ordering is consistent with the health preferences of patient population.
- Ranking scheme is similar to 'untied worst-rank score analysis" for missing data of Lachin (1999).
- The "worst-rank score analysis" ranks all the patients who died ($A_K = 0$) the same and is also commonly used.
- CCMV is a strong benchmark assumption.
- Assumed survival time is always known, need to extend methods to handle censoring.
- `idem` software is available on CRAN.

idem: An **R** Package for Inferences in Clinical Trials with Death and Missingness

Chenguang Wang
Johns Hopkins University

Elizabeth Colantuoni
Johns Hopkins University

Andrew Leroux
Johns Hopkins University

Daniel O. Scharfstein
Johns Hopkins University

Abstract

In randomized controlled trials of seriously ill patients, death is common and often defined as the primary endpoint. Increasingly, non-mortality outcomes such as functional outcomes are co-primary or secondary endpoints. Functional outcomes are not defined for patients who die, referred to as "truncation due to death", and among survivors, functional outcomes are often unobserved due to missed clinic visits or loss to follow-up. It is well known that if the functional outcomes "truncated due to death" or missing are handled inappropriately, treatment effect estimation can be biased. In this paper, we describe the package **idem** that implements a procedure for comparing treatments that is based on a composite endpoint of mortality and the functional outcome among survivors. Among survivors, the procedure incorporates a missing data imputation procedure with a sensitivity analysis strategy. A web-based graphical user interface is provided in the **idem** package to facilitate users conducting the proposed analysis in an interactive and user-friendly manner. We demonstrate **idem** using data from a recent trial of sedation interruption among mechanically ventilated patients.

Keywords: Clinical trial, Truncation due to death, Composite endpoint, Imputation, Missing data, R, SACE, Sensitivity analysis, Shiny, STAN.

1. Introduction

In randomized clinical trials (RCTs) that evaluate medical interventions for patients at high risk of death, functional outcomes scheduled to be measured at pre-specified post-randomization time points may be pre-empted due to death. Furthermore, patients alive

at a pre-specified time may fail to be evaluated due to missed visits or withdrawal, yielding missing data. The distinction between the two types of unobserved functional outcomes is that data pre-empted due to death are considered to be undefined, whereas missing data exist but were not collected.

The so-called issue of "truncation due to death" is challenging even if there is no missing data among survivors. One method proposed for analyzing such data is to create a composite endpoint that combines mortality information among patients that die prior to the pre-specified time and the functional outcome among survivors (Diehr, Patrick, Spertus, Kiefe, Donell, and Fihn 2001; Lachin 1999; Joshua Chen, Gould, and Nessly 2005). In cases where patients can be ordered in a scientifically meaningful way , the simplicity of the composite outcome approach can be a useful way of globally assessing treatment effects that are causally interpretable.

Wang, Scharfstein, Colantuoni, Girard, and Yan (2017) integrated the composite endpoint definition based on Lachin (1999) with a missing data imputation approach for intermittent missing data. They proposed a ranking scheme that ranks all the patients who died before the end of the study according to their time of death (earlier times are worse than later times) lower than patients who survived past the end of the study and survivors are then ranked according to their functional outcome. The inference for treatment arm comparisons are based on comparing the distribution of ranks across the treatment arms, accounting for the possibility of ties. Their method considered the complete case missing value constraints (Little 1993) as the benchmark assumption for intermittent missing data imputation and suggested a global sensitivity analysis framework to further assess the robustness of the findings through exponential tilting.

In this paper, we describe the R package **idem** that implements the proposed method in Wang *et al.* (2017) for making inferences in randomized clinical trials with both intermittent missing data and deaths. Notably, there are several extensions and modifications in **idem** from the original paper. First, Wang *et al.* (2017) proposed a Metropolis-Hastings algorithm for imputing missing data from their target distributions. In contrast, the package **idem** implements a rejection sampling approach where the candidate samples are drawn by **rstan** (Carpenter, Gelman, Hoffman, Lee, Goodrich, Betancourt, Brubaker, Guo, Li, and Riddell 2017). Second, the package **idem** implements two alternative approaches to estimate and test for a treatment effect when data are "truncated due to death", the *survivors only* analysis and the *survivor average causal effect* (SACE) analysis (Chiba and VanderWeele 2011). Lastly, the package **idem** implements a web-based graphical user interface (GUI) where users can conduct the analysis in an interactive and user-friendly manner.

There are several software packages on the Comprehensive R Archive Network (CRAN) for analyzing death truncated data. The package **JM** (Rizopoulos 2010) applies shared parameter models for the joint modeling of longitudinal and survival data and the package **JMbayes** (Rizopoulos 2016) implements the shared parameter joint modeling approach under the Bayesian framework. The joint modeling approach implemented in the two packages introduces a shared set of latent random effects for modeling both the functional outcome and survival. In this approach, the model for the functional outcome allows trajectories of the functional outcome after death, which is not scientifically meaningful. The package **sensitivityPStrat** (Dupont and Shepherd 2014) applies the causal inference framework that addresses the problem in terms of counterfactuals and seeks to estimate the "principal stratum" causal effect (Frangakis and Rubin 2002; Hayden, Pauler, and Schoenfeld 2005; Chiba and VanderWeele 2011), e.g., the SACE. Although this approach is useful for understanding the mechanistic effect of

treatment on clinical outcomes, it requires strong assumptions to identify whether a patient is a member of the "principal stratum" at the time of the treatment decision. To the best of our knowledge, there is no statistical software package that handles both the "truncation due to death" problem and intermittent missing data among survivors for RCTs, let alone one with a graphical user interface (GUI).

In this paper, we demonstrate **idem** by using data from the Awakening and Breathing Controlled (ABC) trial (Girard, Kress, Fuchs, Thomason, Schweickert, Pun, Taichman, Dunn, Pohlman, Kinniry, Jackson, Canonico, Light, Shintani, Thompson, Gordon, Hall, Dittus, Bernard, and Ely 2008). The ABC trial randomized acute respiratory failure patients receiving mechanical ventilation 1:1 within each study site to management with a paired sedation plus ventilator weaning protocol involving daily interruption of sedatives through spontaneous awakening trials (SATs) and spontaneous breathing trials (SBTs) or sedation per usual care (UC) and SBTs (Girard *et al.* 2008). In a single-site substudy, cognitive, psychological and physical function was measured at 3 and 12-months post-randomization among $n = 94$ and $n = 93$ patients in the UC+SBT and SAT+SBT arms, respectively. We analyze a continuous measure of cognitive function where higher scores indicate better cognition.

The reminder of the paper is organized as follows. In Section 2, we briefly introduce the method proposed in Wang *et al.* (2017). We demonstrate the **idem** package in the R interactive mode using data from the ABC trial in Section 3. In Section 4, we describe the details of the **idem** GUI. In Section 5, we illustrate the **idem** GUI using the ABC trial. Section 6 is devoted to discussion.

2. Method

In this section, we briefly introduce the composite endpoint approach implemented in Wang *et al.* (2017), the *survivors only* analysis and the SACE analysis.

2.1. Notation

Consider a randomized study with K post-randomization assessment times $l_1, \ldots, l_K$. Let Y_k $(k = 1, \ldots, K)$ denote the functional outcome scheduled to be measured at time l_k. We use $\overline{Y}_k$ to denote $(Y_1, Y_2, \ldots, Y_k)$. Let X denote covariates measured at baseline, which may or may not include the functional outcome, Y_0. Let T define the treatment assignment. Let L denote the survival time and $A_k = I(L > l_k)$, an indicator that the patient survived past assessment time l_k. Let $Z = g(Y_0, \ldots, Y_K)$ be the study's functional endpoint (e.g., $Z = Y_K - Y_0$). Assume that higher values of Z denote better outcomes.

In the absence of missing data, patients i and j are ranked as follows:

- If $A_{K,i} = A_{K,j} = 1$, then patient i is ranked better than patient j if $Z_i > Z_j$ and ranked the same if $Z_i = Z_j$.

- If $A_{K,j} = 0$ and $A_{K,i} = 1$, then patient i is ranked better than patient j.

- If $A_{K,i} = A_{K,j} = 0$, then patient i is ranked better than patient j if $L_i > L_j$ and ranked the same if $L_i = L_j$.

More formally, let U be a function of (A_K, W) where $W = L$ if $A_K = 0$ and $W = Z$ if $A_K = 1$ with the ordering following the above ranking rules. Wang *et al.* (2017) argued that U is a

composite endpoint in the sense that it is univariate and contains information on survival and functional status.

When $A_k = 1$, define R_k to be the indicator that Y_k is observed. For patients alive at l_K (i.e., $A_K = 1$), let $\boldsymbol{S} = (R_1, \ldots, R_K)$ denote the missing data pattern; further, let $\boldsymbol{Y}_{obs}^{(s)} = \{Y_k : R_k = 1, k \geq 1, \boldsymbol{S} = \boldsymbol{s}\}$ and $\boldsymbol{Y}_{mis}^{(s)} = \{Y_k : R_k = 0, k \geq 1, \boldsymbol{S} = \boldsymbol{s}\}$ denote the observed and missing post-randomization functional outcomes. Note that Z is only observed when $\boldsymbol{S} = \boldsymbol{1}$, where $\boldsymbol{1}$ is a K-dimensional vector of 1's, if $g(\cdot)$ is a non-constant function of all Y_k's.

2.2. Missing data imputation

To impute the missing functional outcomes, $\boldsymbol{Y}_{mis}^{(s)}$, for patients alive at l_K, the following class of untestable assumptions are posited:

$$f(\boldsymbol{Y}_{mis}^{(s)}|A_K = 1, \boldsymbol{Y}_{obs}^{(s)}, Y_0, \boldsymbol{X}, T, \boldsymbol{S} = \boldsymbol{s})$$
$$\propto \exp(\Delta_T Z) f(\boldsymbol{Y}_{mis}^{(s)}|A_K = 1, \boldsymbol{Y}_{obs}^{(s)}, Y_0, \boldsymbol{X}, T, \boldsymbol{S} = \boldsymbol{1}) \quad (1)$$

for all $\boldsymbol{s} \neq \boldsymbol{1}$, where Δ_T is a treatment-specific sensitivity parameter. Note that the benchmark assumption in the class (i.e., $\Delta_T = 0$) is the complete case missing value (CCMV) restrictions (Little 1993).

To avoid non-sensical imputations that generate out-of-bound functional outcomes, Wang *et al.* (2017) suggested the following data transformation of Y_k ($k = 1, \ldots, K$):

$$\phi(y_k) = \log\left(\frac{y_k - B_L}{B_U - y_k}\right), \quad (2)$$

where (B_L, B_U) denote the lower and upper bound of the functional outcome. Let $Y_k^{\dagger} = \phi(Y_k)$ and $\overline{\boldsymbol{Y}}_k^{\dagger} = (Y_1^{\dagger}, \ldots, Y_k^{\dagger})$. Note that there is a one-to-one mapping between the conditional distributions $h(\overline{\boldsymbol{Y}}_K^{\dagger}|A_K = 1, Y_0, \boldsymbol{X}, T, \boldsymbol{S} = \boldsymbol{1})$ and $f(\overline{\boldsymbol{Y}}_K|A_K = 1, Y_0, \boldsymbol{X}, T, \boldsymbol{S} = \boldsymbol{1})$.

We first factorize $h(\overline{\boldsymbol{Y}}_K^{\dagger}|A_K = 1, Y_0, \boldsymbol{X}, T, \boldsymbol{S} = \boldsymbol{1})$ as follows

$$h(\overline{\boldsymbol{Y}}_K^{\dagger}|A_K = 1, Y_0, \boldsymbol{X}, T, \boldsymbol{S} = \boldsymbol{1}) = \prod_{k=1}^{K} h(Y_k^{\dagger}|A_K = 1, \overline{\boldsymbol{Y}}_{k-1}^{\dagger}, Y_0, \boldsymbol{X}, T, \boldsymbol{S} = \boldsymbol{1}) \quad (3)$$

and posit a model for each component of the product. Specifically, we consider models of the form:

$$h(Y_k^{\dagger}|A_K = 1, \overline{\boldsymbol{Y}}_{k-1}^{\dagger}, Y_0, \boldsymbol{X}, T = t, \boldsymbol{S} = \boldsymbol{1}) = h_{k,t}(Y_k^{\dagger} - \mu_{k,t}(\overline{\boldsymbol{Y}}_{k-1}^{\dagger}, Y_0, \boldsymbol{X}; \boldsymbol{\alpha}_{k,t})) \quad (4)$$

where $\mu_{k,t}(\overline{\boldsymbol{Y}}_{k-1}^{\dagger}, Y_0, \boldsymbol{X}; \boldsymbol{\alpha}_{k,t})$ is a specified conditional mean function of $\overline{\boldsymbol{Y}}_{k-1}^{\dagger}$, Y_0, $\boldsymbol{X}$ and $\boldsymbol{\alpha}_{k,t}$, $\boldsymbol{\alpha}_{k,t}$ is an unknown parameter vector and $h_{k,t}$ is an unspecified time and treatment-specific mean zero density function.

Let $\widehat{\boldsymbol{\alpha}}_{k,t}$ denote the least squares estimator of $\boldsymbol{\alpha}_{k,t}$. The density function $h_{k,t}$ can be estimated by kernel density estimation based on the residuals or estimated with parametric assumptions (e.g., normality) if the sample size is small. Let $\widehat{h}_{k,t}$ denote the kernel density estimator of $h_{k,t}$. We then estimate $f(\overline{\boldsymbol{Y}}_K|A_K = 1, Y_0, \boldsymbol{X}, T, \boldsymbol{S} = \boldsymbol{1})$ by

$$\widehat{f}(\overline{\boldsymbol{Y}}_K|A_K = 1, Y_0, \boldsymbol{X}, T, \boldsymbol{S} = \boldsymbol{1}) = \prod_{k=1}^{K} \widehat{h}_{k,t}(Y_k^{\dagger} - \mu_{k,t}(\overline{\boldsymbol{Y}}_{k-1}^{\dagger}, Y_0, \boldsymbol{X}; \widehat{\boldsymbol{\alpha}}_{k,t})) \left|\frac{d\phi(Y_k)}{dY_k}\right|. \quad (5)$$

2.3. Treatment effect quantification: composite endpoint approach

Let i and j be random individuals randomized to treatment $T = 0$ and $T = 1$, respectively. Wang *et al.* (2017) proposed to quantify the treatment effect, denoted θ, as

$$\theta = P(U_i < U_j) - P(U_i > U_j). \tag{6}$$

Values of $\theta > 0$ and $\theta < 0$ favor $T = 1$ and $T = 0$, respectively. Note that $\theta = 0$ under the null hypothesis of no treatment effect.

In the absence of missing data, θ can be estimated by

$$\widehat{\theta} = \frac{1}{n_0 n_1} \sum_{i:T_i=0} \sum_{j:T_j=1} \{\mathrm{I}(U_i < U_j) - \mathrm{I}(U_i > U_j)\}$$

where n_0 and n_1 are the sample size of treatment arm $T = 0$ and $T = 1$, respectively.

In addition to estimating θ, Wang *et al.* (2017) suggested reporting quantiles (e.g., median) of the treatment-specific distribution of the composite endpoint U to further help characterize the treatment effect.

2.4. Treatment effect quantification: alternative approaches

In the absence of missing data, several alternative approaches to quantify the effect of an intervention on the functional endpoint in the presence of mortality have also been proposed and utilized in the statistical and clinical literature.

The *survivors only* approach defines the treatment effect of the intervention on the functional endpoint as

$$\theta_{surv} = E(Z|T = 1, A_K = 1) - E(Z|T = 0, A_K = 1),$$

i.e., the difference in the mean functional endpoint comparing survivors receiving the intervention to survivors receiving the control. If survival is independent of the treatment assignment, then this treatment effect definition has a causal interpretation. However, in cases where the intervention affects mortality then this treatment effect definition does not define a causal effect and interpreting the estimated treatment effect can be misleading.

To remedy the potential bias in the *survivors only* approach, one may compare the functional endpoint within a special subset of patients, referred to as the principle stratum. This special subset of patients would survive to the end of the follow-up regardless of which intervention they receive. To define the *survivor average causal effect* (SACE), we define what would happen to patients (in terms of survival and functional endpoint) under both intervention and control. Let $A_K(t)$ be the indicator that the patient survives to time l_K under treatment $T = t$, and if $A_K(t) = 1$, define $Z(t)$ as the potential functional endpoint observed (otherwise, $Z(t)$ is not defined). Among patients who survive to time l_K regardless of which treatment they receive (i.e., $A_K(0) = A_K(1) = 1$), the SACE is defined as

$$\theta_{SACE} = E[Z(1)|A_K(0) = 1, A_K(1) = 1] - E[Z(0)|A_K(1) = 1, A_K(0) = 1].$$

Since the survival status and functional endpoint are only observed for the treatment that was received, additional assumptions are required to estimate SACE or obtain bounds. Under the monotonicity assumption $A_K(1) \geq A_K(0)$, that is, if a patient would survive to time l_K under

control, then the patient would survive to time l_K under intervention, Chiba and VanderWeele (2011) showed that

$$\theta_{SACE} = \theta_{surv} - \Delta_{SACE},$$

where Δ_{SACE} is the difference in the mean functional endpoint for surviving intervention arm patients and the mean functional endpoint if surviving control group patients had, contrary to fact, received the intervention. That is,

$$\Delta_{SACE} = E[Z|T = 1, A_K = 1] - E[Z(1)|A_K(0) = 1]. \tag{7}$$

Possible values for Δ_{SACE} should be elicited from expert opinions. In practice, an additional assumption is often made that the surviving control group patients are healthier than the surviving intervention group patients. Consequently, Δ_{SACE} is assumed to be non-positive if the healthier patients are expected to obtain a better functional outcome.

2.5. Inference

For individual i alive at l_K with missing functional outcomes, M copies of the missing functional outcomes can be drawn from the density that is proportional to $\exp(\Delta_T Z)\widehat{f}(\boldsymbol{Y}_{mis}^{(s)}|A_K = 1, \boldsymbol{Y}_{obs}^{(s)} = \boldsymbol{Y}_{obs,i}, Y_0 = Y_{0,i}, \boldsymbol{X} = \boldsymbol{X}_i, T = T_i, \boldsymbol{S} = \boldsymbol{1})$ using MCMC sampling techniques to create M complete datasets.

Wang *et al.* (2017) suggested the Metropolis Hastings algorithm for the MCMC sampling. To improve the Markov Chain convergence, **idem** implements a rejection sampling approach with $\xi \widehat{f}(\boldsymbol{Y}_{mis}^{(s)}|A_K = 1, \boldsymbol{Y}_{obs}^{(s)} = \boldsymbol{Y}_{obs,i}, Y_0 = Y_{0,i}, \boldsymbol{X} = \boldsymbol{X}_i, T = T_i, \boldsymbol{S} = \boldsymbol{1})$ being the proposal distribution, where ξ is a constant that is large enough such that $\exp(\Delta_T Z) < \xi$ for all Z. Such a constant exists in the settings we consider where the functional outcome is bounded by its biological boundaries. Rejection sampling candidates are then drawn from $\widehat{f}(\boldsymbol{Y}_{mis}^{(s)}|A_K = 1, \boldsymbol{Y}_{obs}^{(s)} = \boldsymbol{Y}_{obs,i}, Y_0 = Y_{0,i}, \boldsymbol{X} = \boldsymbol{X}_i, T = T_i, \boldsymbol{S} = \boldsymbol{1})$ in **idem** via **rstan** (Carpenter *et al.* 2017) by Adaptive Hamiltonian Monte Carlo.

For each complete dataset m, we estimate θ by $\widehat{\theta}_m$. The overall estimator of θ is then $\tilde{\theta} = \frac{1}{M}\sum_{m=1}^{M} \widehat{\theta}_m$. Confidence intervals can be constructed by applying the non-parametric bootstrap procedure.

Similarly, computations are applied to generate overall estimates of and confidence intervals for θ_{surv} and θ_{SACE}.

3. The idem package

3.1. Installation and overall scheme

The **idem** package is available from CRAN at <http://CRAN.R-project.org/package=idem>.
To install and load **idem**, type the following in R:

```
R> install.packages("idem")
R> require(idem)
```

The major steps of conducting an analysis using **idem** include data preparation, imputation
model fitting, missing data imputation, and treatment effect estimation and hypothesis test-
ing. Intermediate results are organized and passed between steps as **idem**-specific classes.
Figure 1 presents the overall scheme and the major functions in **idem**.
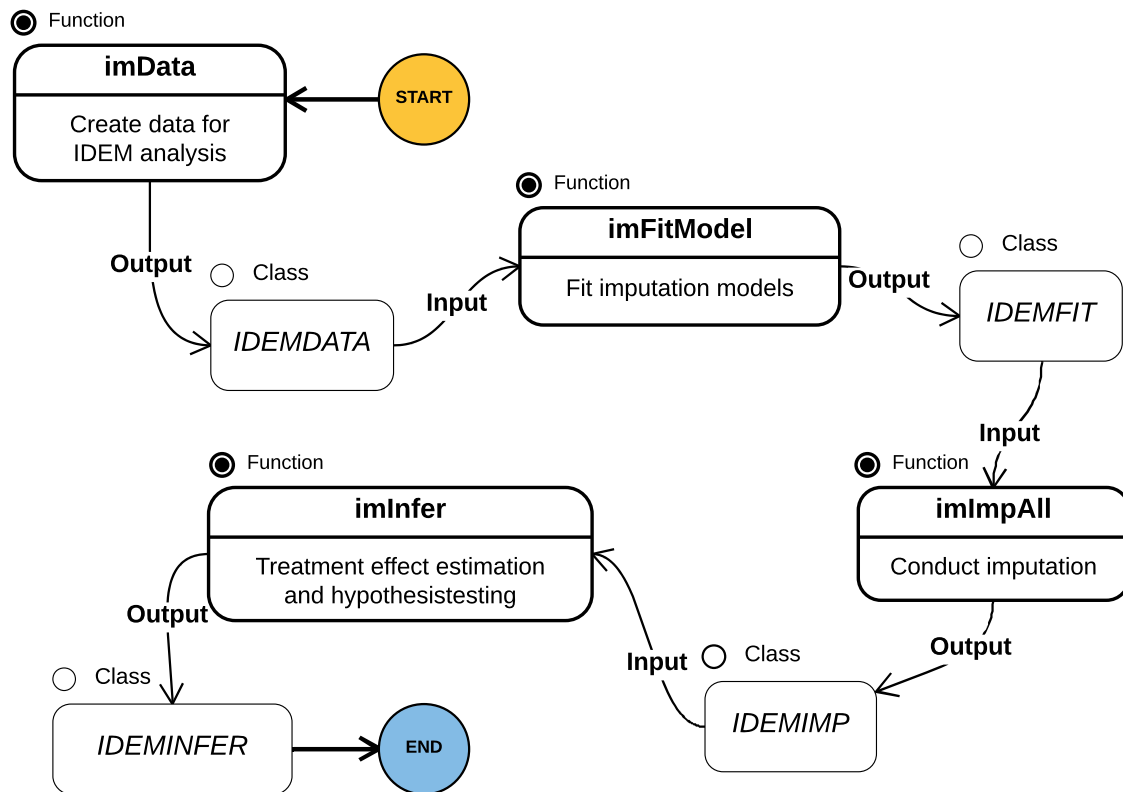
Figure 1: Overall scheme of the **idem** package.

3.2. Data preparation

Data format

The **idem** package requires the dataset to be formatted as follows: each row represents a
subject and includes the treatment assignment, baseline covariates, baseline outcome (if ap-
plicable), post-randomization functional outcomes and survival time. It is assumed that there

is no censoring of the survival time prior to time l_K. For patients who were censored after time l_K, their survival time can be entered as any arbitrary number that is longer than l_K.

The **idem** package provides the dataset `abc` from the ABC trial as an example dataset with a single baseline covariate, `Age`. Note: baseline cognition was not measured in the ABC trial.

```
R> head(abc)


    AGE TRT SURV Y2 Y1
1 59.63   1  999 NA NA
2 66.89   0  999 52 49
3 59.70   1    1 NA NA
4 81.41   0   72 NA NA
5 66.52   1  999 45 51
6 40.27   0   65 NA NA
```

Create analysis data object

As the first step, the function `imData` combines the original dataset and analysis specification parameters to create a class `IDEMDATA` object for the **idem** analysis. The parameters include variable names in the dataset, functional outcome specification, functional endpoint specification, duration of the study, etc.. Details can be found in the help document of `imData`.

When there are mis-specifications in the parameters, error and inconsistency messages will be returned by `imData`. Otherwise, the return value is class `IDEMDATA` and contains the original dataset and the specification parameters.

```
R> err.data <- imData(abc, trt = "TRT", outcome = c("Y1","Y2"),
+   y0 = NULL, endfml = "Y2", bounds = c(10,20), duration = 365)
R> err.data


Model specification is invalid. Please check the following:
    No survival time specified
    Upper bound is smaller than some observed outcomes


R> im.abc <- imData(abc, trt = "TRT", surv = "SURV",
+   outcome = c("Y1","Y2"), unitTime = "days",
+   trt.label = c("UC+SBT", "SAT+SBT"),
+   cov = c("AGE"), endfml = "Y2", duration = 365, bounds = c(0,100))
R> im.abc


There are 187 observations of 5 variables in the data.
Detailed specifications are as follows:
    Treatment: TRT
    Survival time: SURV
    Study duration: 365
    Outcomes (ordered chronically): Y1 Y2
```

```
Endpoint (in R formula): Y2
Treatment labels: UC+SBT SAT+SBT
Covariates: AGE
Biological boundary of the outcomes: 0 100
```

Data visualization

The class `IDEMDATA` result from `imData` provides S3 methods `summary` and `plot` for data visualization.

The missing data patterns among survivors will be generated as a data frame by its `summary` function:

```
R> summary(im.abc)

                       Y1       Y2  Control Intervention
Deaths on study                    58 (62%)     38 (41%)
S=1              Observed Observed 18 (19%)     32 (34%)
S=2              Observed  Missing  8 (9%)       8 (9%)
S=3               Missing Observed  1 (1%)       0 (0%)
S=4               Missing  Missing  9 (10%)     15 (16%)
Total                                    94           93
```

Spaghetti plots of the functional outcomes for survivors (Figure 2), missing data pattern heatmaps (Figure 3) and Kaplan-Meier survival curves (Figure 4) can be generated by the S3 `plot` method using options `survivor`, `missing` and `KM`, respectively.

```
R> plot(im.abc, opt = "survivor")
R> plot(im.abc, opt = "missing", cols = c("blue", "gray"))
R> plot(im.abc, opt = "KM")
```
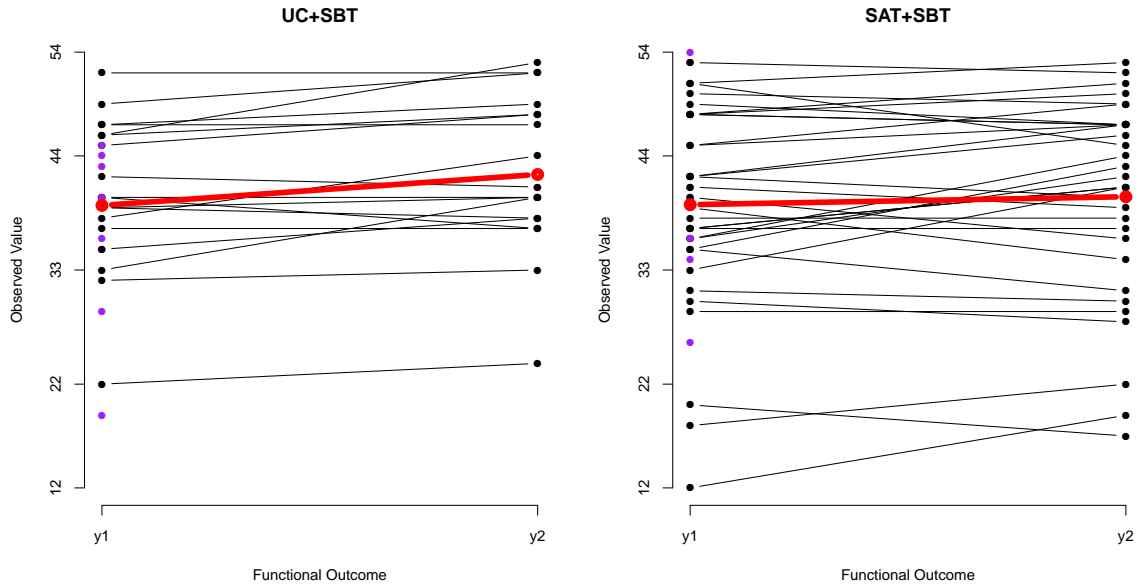
Figure 2: Spaghetti plot of the functional outcome among survivors in the ABC trial. The purple dots represent patients with missing functional outcomes. The red line represents the mean of the observed functional outcomes as a function of time.
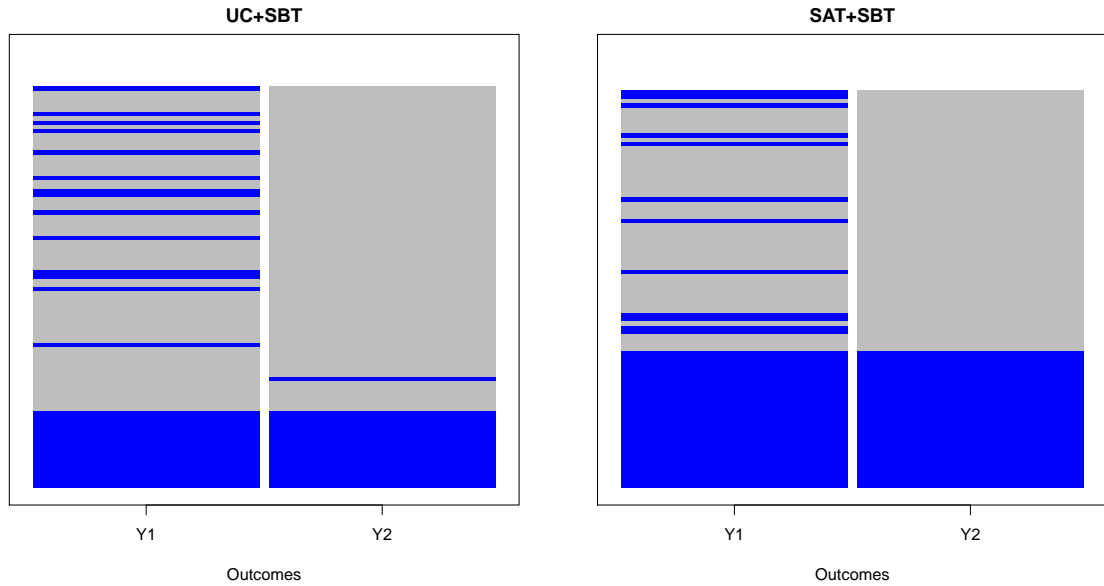
Figure 3: Missing data pattern heatmap for survivors in the ABC trial. The blue and gray cells represent observed and missing functional outcomes, respectively.
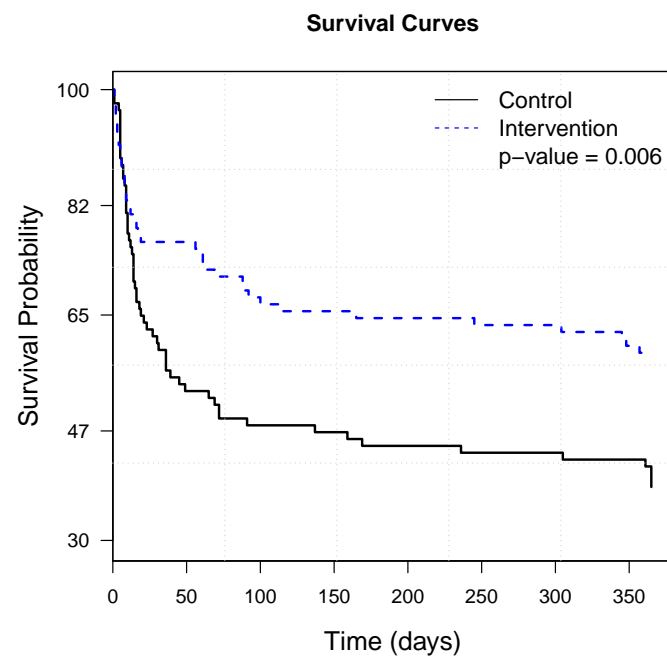
Figure 4: Kaplan-Meier survival curves for patients in the ABC trial. The p-value from the log-rank test is displayed.

In addition, through the S3 `summary` method, the `IDEMDATA` class returns the row indices that correspond to the subjects who were alive at the end of the study but had missing functional outcomes, i.e., the subjects that need missing data imputation.

```
R> summary(im.abc, opt = "missid")

[1]    1  15  25  27  47  50  57  61  63  67  70  73  79  80  83  86
[17]  87  88  89  95 106 112 122 127 132 133 142 155 158 161 162 167
[33] 169 171 172 174 178 180 183 185 187
```

3.3. Missing data imputation

Fit imputation models

For the missing data imputation, the function `imFitModel` needs to be called first to fit the imputation model(s) (3) among survivors with $S = 1$, i.e., the patients who were alive at the end of the study without missing functional outcomes. The return value of the `imFitModel` function has class `IDEMFIT` and contains `lm` results for all the imputation models.

```
R> rst.fit <- imFitModel(im.abc)
R> rst.fit

-- Treatment UC+SBT
---- Y1 ~ AGE

Call:
lm(formula = as.formula(cur.f), data = cur.data)

Residuals:
    Min      1Q   Median      3Q      Max
-0.83313 -0.11755 -0.01075  0.22708  0.48152

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.255046   0.415477  -0.614    0.548
AGE         -0.002226   0.006473  -0.344    0.735

Residual standard error: 0.3271 on 16 degrees of freedom
Multiple R-squared:  0.007339,       Adjusted R-squared:  -0.0547
F-statistic: 0.1183 on 1 and 16 DF,  p-value: 0.7354

...

-- Treatment SAT+SBT

...
```

```
---- Y2 ~ Y1+AGE

Call:
lm(formula = as.formula(cur.f), data = cur.data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.34779 -0.10677 -0.01963  0.15173  0.27374

...
```

The S3 `plot` of the `IDEMFIT` class generates the goodness of fit diagnostic plots (Figure 5). If the normality assumption of the distribution of the residuals does not seem to hold, imputation of the missing data using kernel density estimation of the residuals should be considered (see Section 2.2 for more details).

```
R> plot(rst.fit, mfrow=c(2,4))
```
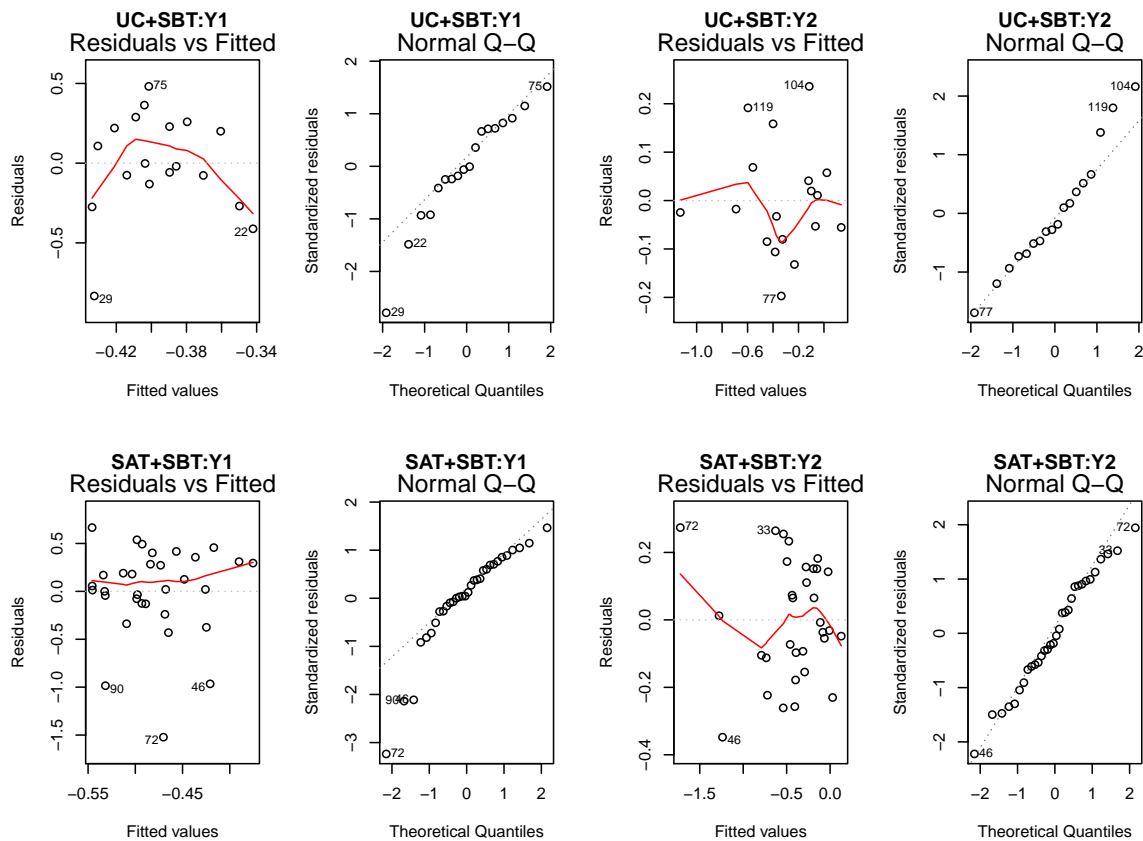

Figure 5: Goodness of fit diagnostic plots.

MCMC convergence checking

Before conducting the imputation for the entire dataset, it is recommended that the MCMC sampling convergence be checked. The **idem** package provides the function `imImpSingle` that implements the MCMC sampling under the benchmark assumption (i.e., with $\Delta_T = 0$) for an individual subject. The convergence of the MCMC chains can then be checked by a trace plot of the results (Figure 6). If the mixing of the Markov chains are not satisfactory, users should refer to the **rstan** documents for options (e.g., `adapt_delta`) that can improve the convergence.

```
R> rst.mixing <- imImpSingle(abc[1,], rst.fit, chains = 4,
+    normal = F, iter = 2000, warmup = 1000)
R> plot(rst.mixing)
```
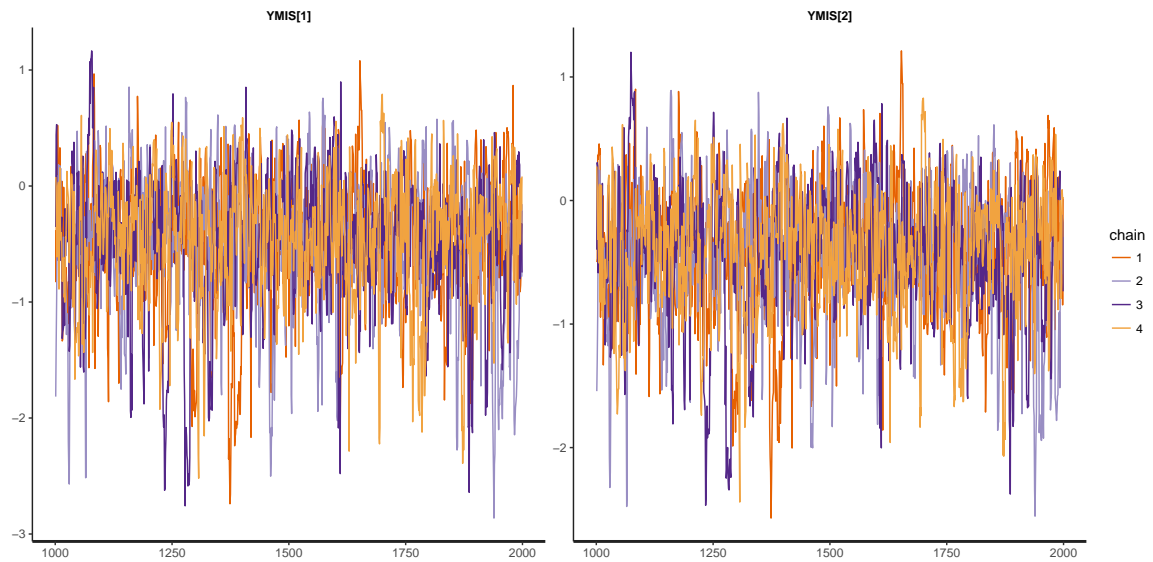

Figure 6: Trace plot of the imputed missing functional outcomes of an individual subject.

Conduct imputation

The function `imImpAll` imputes missing outcomes for all survivors with missing functional outcomes to generate complete datasets. The following code shows how to use `imImpAll` to get $M = 5$ (`n.imp = 5`) imputed complete datasets for sensitivity parameters $\Delta_T = -0.2, -0.15, \ldots, 0.2$; in this example, the residuals are not assumed to follow a Normal distribution (`normal = F`).

```
R> rst.imp <- imImpAll(rst.fit, deltas = seq(-0.2, 0.2, 0.05),
+    n.imp = 5, normal = F, chains = 4, iter = 2000, warmup = 1000)
R> rst.imp

A total of 5 complete datasets were imputed. Normality assumption
was NOT made for the imputation model residual distribution.
```

```
The sensitivity parameters considered were
[1] -0.20 -0.15 -0.10 -0.05  0.00  0.05  0.10  0.15  0.20

The last 5 records in the complete dataset
are given below as an example:

          ID DELTA IMP   AGE TRT SURV Y1       Y2 ORGY1 ORGY2     ENDP
1943 187   0.2   1 66.12   1  999 26 34.68728    26    NA 34.68728
1944 187   0.2   2 66.12   1  999 26 28.41199    26    NA 28.41199
1945 187   0.2   3 66.12   1  999 26 32.02637    26    NA 32.02637
1946 187   0.2   4 66.12   1  999 26 36.16493    26    NA 36.16493
1947 187   0.2   5 66.12   1  999 26 26.95370    26    NA 26.95370
```

The returned value from function `imImpAll` is class `IDEMIMP`. Its S3 `plot` method provides options to generate treatment-specific densities of the imputed functional outcomes. Figure 7 presents the treatment-specific densities of the imputed Y_2 (the functional endpoint) for the ABC trial.

```
R> plot(rst.imp, opt = "imputed", deltas = c(-0.2,0,0.2),
+    xlim = c(0,100), endp = TRUE)
```
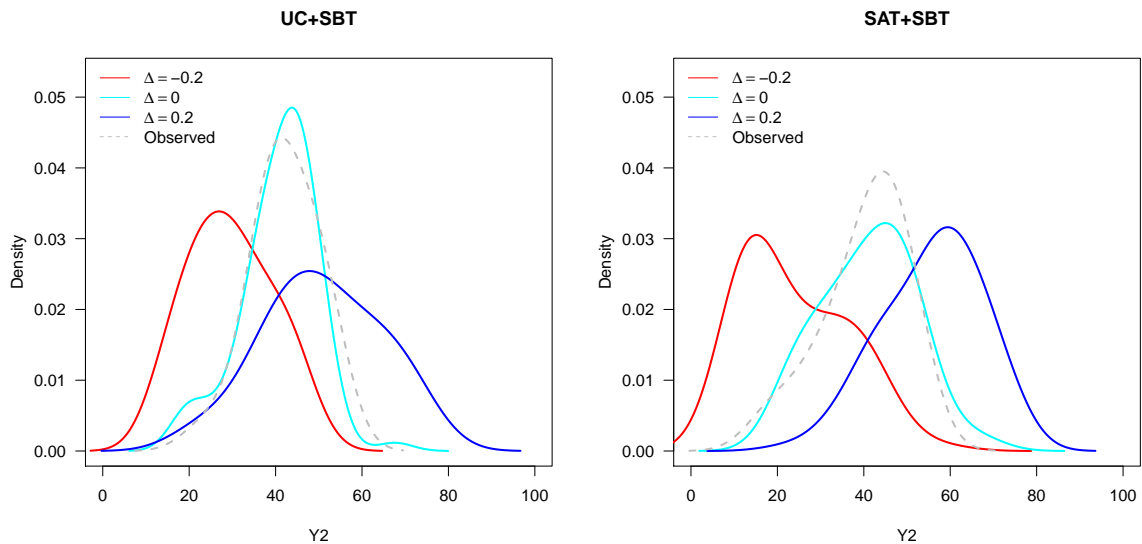

Figure 7: Treatment-specific densities of the imputed Y_2 for different choices of the sensitivity parameters Δ_T.

The other option provided in the `plot` method of the `IDEMIMP` class is `composite`. The `composite` option generates the treatment-specific cumulative distribution function of the composite endpoint, where the values of the composite endpoint are labeled according to the survival time and functional endpoint among survivors (Figure 8).

```
R> plot(rst.imp, opt = "composite", delta = 0)
```
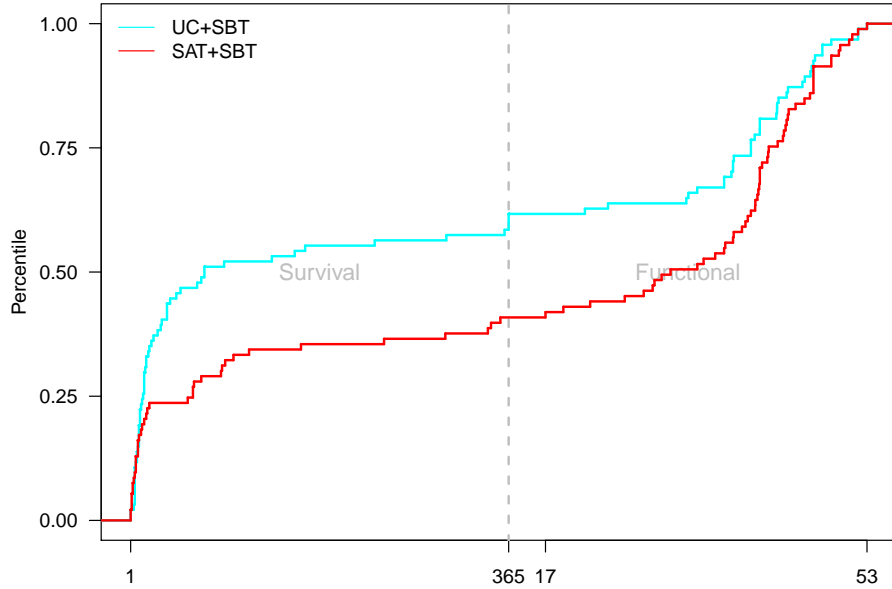

Figure 8: Cumulative distribution function of the composite endpoint for each treatment group based on the multiple imputation algorithm with the benchmark assumptions ($\Delta_T = 0$).

3.4. Treatment effect estimation and hypothesis testing

Composite endpoint approach

Given a class IDEMIMP object that contains complete datasets with imputed outcomes, **idem** uses function imInfer to estimate the treatment effect and quantiles of the composite endpoint distribution. Note that the results of quantiles of the composite endpoint may be a survival time or a value of the functional outcome, which are reported in columns QuantY and QuantSurv, respectively.

```
R> rst.est <- imInfer(rst.imp, n.boot = 0,
+    effect.quantiles = c(0.25,0.5,0.75))
R> print(rst.est, delta0=0)

The sensitivity parameters considered were
[1] -0.20 -0.15 -0.10 -0.05  0.00  0.05  0.10  0.15  0.20

The estimated treatment effect theta under different
sensitivity parameters are:
```

```
     Delta0 Delta1   Theta
5         0  -0.20 -0.1266
14        0  -0.15 -0.1379
23        0  -0.10 -0.1460
32        0  -0.05 -0.1688
41        0   0.00 -0.1992
50        0   0.05 -0.2161
59        0   0.10 -0.2378
68        0   0.15 -0.2537
77        0   0.20 -0.2647
```

The estimated treatment effect quantiles under different
sensitivity parameters are:

```
     Delta TRT    Q QuantY QuantSurv
123      0   0 0.25     NA        14
128      0   0 0.50     NA        72
133      0   0 0.75     38        NA
138      0   1 0.25     NA        61
143      0   1 0.50     30        NA
148      0   1 0.75     44        NA
```

When choosing the number of bootstrap samples to be bigger than 0, the function `imInfer` performs non-parametric boostrap to conduct hypothesis testing for the treatment effect including evaluating the uncertainties of the estimated quantiles from the composite endpoint distribution. For bootstrap analysis, the function `imInfer` supports parallel computation by specifying `ncore > 1`. For the other imputation parameters (e.g. normality assumption, number of MCMC chains, etc.), the function `imInfer` takes the same settings contained in the `IDEMIMP` class object.

Two-sided p-values for testing the null hypothesis of $\theta = 0$, the standard deviation of the bootstraps for the estimated θ, and confidence intervals for quantiles of the composite endpoint are obtained by summarizing the results from the bootstrap analysis. Note that the 2.5% and 97.5% credible intervals are reported in columns `Q2.5` and `Q97.5`, respectively. The columns `Q2.5_Surv` and `Q97.5_Surv` are indicators for `Q2.5` and `Q97.5`, respectively, of being a survival time.

```
R> rst.final <- imInfer(rst.imp, n.boot = 100, n.cores = 5)
R> print(rst.final, delta0 = 0)
```

```
The sensitivity parameters considered were
[1] -0.20 -0.15 -0.10 -0.05  0.00  0.05  0.10  0.15  0.20
```

```
Treatment effect (theta) under different
sensitivity parameters are:
```

```
     Delta0 Delta1   Theta      SD   PValue
5         0  -0.20 -0.1225 0.09129 0.179680
14        0  -0.15 -0.1270 0.09163 0.165669
23        0  -0.10 -0.1438 0.09116 0.114760
32        0  -0.05 -0.1544 0.09091 0.089364
41        0   0.00 -0.1905 0.09008 0.034468
50        0   0.05 -0.2071 0.08874 0.019584
59        0   0.10 -0.2395 0.08771 0.006316
68        0   0.15 -0.2534 0.08669 0.003462
77        0   0.20 -0.2573 0.08591 0.002742


Treatment effect (quantiles) under different
sensitivity parameters are:

   Delta TRT   Q QuantY QuantSurv Q2.5  Q97.5 Q2.5_Surv Q97.5_Surv
43     0   0 0.5     NA        72   31 365.00         1          1
48     0   1 0.5     29        NA  348  37.73         1          0



The hypothesis testing and confidence intervals are
based on 100 bootstrap samples.
```

A contour plot of two-sided p-values for the null hypothesis of $\theta = 0$ as a function of the multiple imputation sensitivity parameters, Δ_T, can be generated by the S3 `plot` method of the `imInfer` function result. Alternatively, the contour plot of the estimated treatment effect $\widehat{\theta}$ can be generated by specifying the option to be `effect`. Figure 9 presents these two types of plots.

```
R> plot(rst.final, nlevels = 30, con.v = 0.05, main = 'P-Value')
R> plot(rst.final, opt = "effect", nlevels = 30,
+    con.v = c(-0.1, -0.2), main = expression(theta))
```

Survivors only approach

The default summary of the `IDEMINFER` class, returned by the `imInfer` function, generates the *survivors only* analysis results. As a cautious note, the print out emphasizes that the *survivors only* analysis is only valid when the treatment has no effect on survival.

```
R> rst.survonly <- summary(rst.final)
R> rst.survonly

The imputation sensitivity parameters considered were
[1] -0.20 -0.15 -0.10 -0.05  0.00  0.05  0.10  0.15  0.20


The estimated survivors only treatment effects are
```
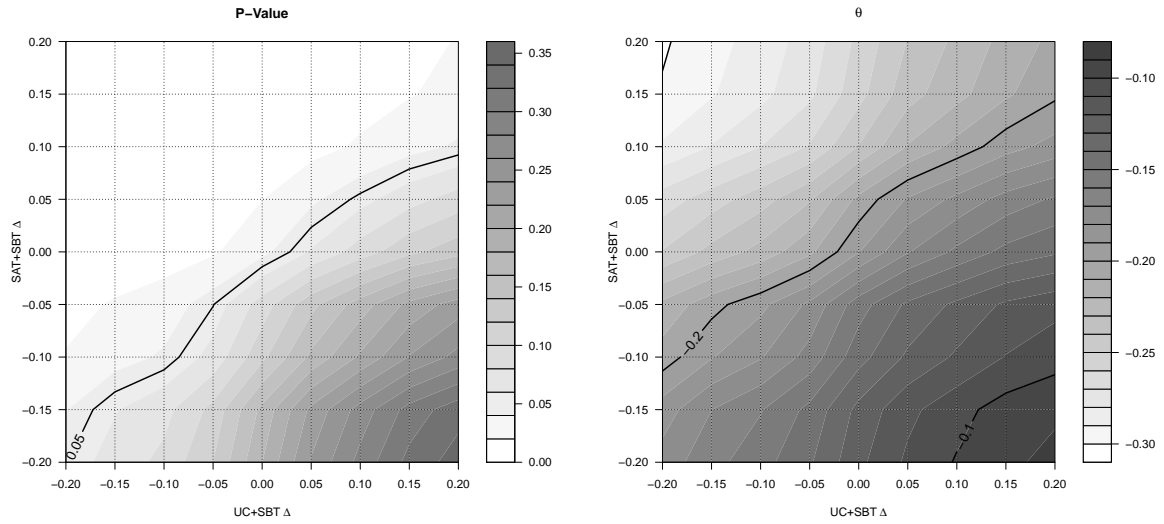
Figure 9: The contour plots of the two-sided p-values obtained by testing the null hypothesis of $\theta = 0$ and the estimated treatment effect $\widehat{\theta}$ as functions of treatment-specific sensitivity analysis parameters.

```
    Delta0 Delta1    Effect       LB       UB    PValue
1    -0.20  -0.20  -5.38365 -14.4889   3.7216 2.465e-01
2    -0.20  -0.15  -2.39526 -11.2445   6.4539 5.957e-01
3    -0.20  -0.10   2.22139  -5.9230  10.3658 5.929e-01
4    -0.20  -0.05   4.94428  -2.7069  12.5955 2.053e-01
5    -0.20   0.00  10.92172   4.3017  17.5418 1.222e-03
...
80    0.20   0.15   3.71120  -5.3493  12.7717 4.221e-01
81    0.20   0.20   5.08226  -4.2587  14.4233 2.862e-01


PLEASE BE CAUTIOUS that survivors only analysis is only valid
when the treatment has no impact on survival.
```

Similar as for the composite endpoint approach, contour plots of p-values and the estimated treatment effect on the functional outcomes for *survivors only* analysis can be generated by the plot function of the summary results (Figure 10).

```
R> plot(rst.survonly, nlevels = 30, con.v = 0.05,
+    main = 'Survivors Only: P-Value')
R> plot(rst.final, opt = "effect", nlevels = 30, con.v = c(-15, 0, 15),
+    main = expression(theta[surv]))
```

Figure 10: Survivors only analysis results.

SACE approach
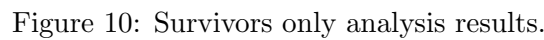
The summary function of the `IDEMINFER` class will generate the SACE analysis results when the option (`opt`) is specified as `SACE`. The sensitivity parameters Δ_{SACE} (7) are passed to the `summary` function by its argument `sace.delta`. The default values of `sace.delta` are provided based on the standard deviation of the bootstraps for the estimated treatment effect on the functional outcomes for survivors.

```
R> rst.sace <- summary(rst.final, opt = "SACE",
+    sace.deltas = seq(-2, 0, by = 0.5))
R> rst.sace

The imputation sensitivity parameters considered were
[1] -0.20 -0.15 -0.10 -0.05  0.00  0.05  0.10  0.15  0.20

The SACE sensitivity parameters considered were
[1]  0.0 -0.5 -1.0 -1.5 -2.0

The estimated SACE are
    Delta0 Delta1    Effect SACE_Delta        LB       UB    PValue
1    -0.20  -0.20  -5.38365        0.0 -14.48891  3.72160 2.465e-01
2    -0.20  -0.15  -2.39526        0.0 -11.24447  6.45394 5.957e-01
3    -0.20  -0.10   2.22139        0.0  -5.92299 10.36577 5.929e-01
4    -0.20  -0.05   4.94428        0.0  -2.70695 12.59550 2.053e-01
5    -0.20   0.00  10.92172        0.0   4.30167 17.54177 1.222e-03
...
```

```
404    0.20   0.15   5.71120        -2.0  -3.34927  14.77167 2.167e-01
405    0.20   0.20   7.08226        -2.0  -2.25873  16.42325 1.373e-01
```

The **idem** package provides two different types plots for visualizing the SACE analysis results. With `by.sace = FALSE`, the `plot` function generates the contour plots of p-values and θ_{SACE} for given Δ_{SACE}. With `by.sace = TRUE`, the `plot` function of the summary results displays the estimates of and confidence intervals for θ_{SACE} for given imputation sensitivity parameters Δ_0 and Δ_1. Figure 4 presents the different types of plots.

```
R> plot(rst.sace, by.sace = FALSE, sace.delta = -1,
+    main = "SACE: P-Value")
R> plot(rst.sace, by.sace = TRUE, delta0 = 0, delta1 = 0,
+    main = expression(theta[SACE]))
```
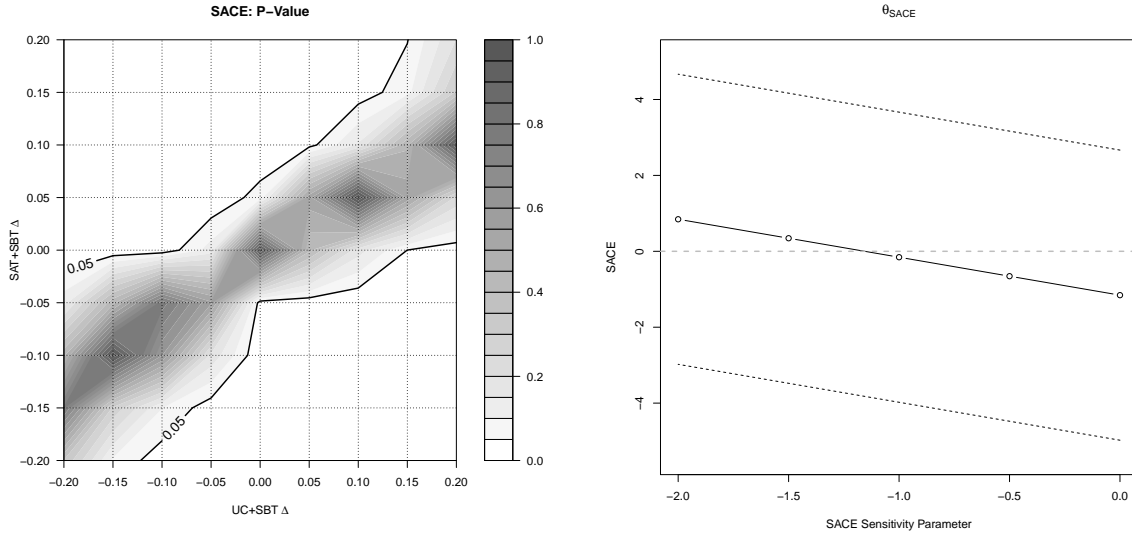

Figure 11: SACE analysis results. The left panel is the contour plot of the p-values (obtained by testing the null hypothesis of $\theta_{SACE} = 0$) as a function of Δ_0 and Δ_1 for $\Delta_{SACE} = -1.5$. The right panel presents $\widehat{\theta}_{SACE}$ with its 95% confidence intervals as a function of Δ_{SACE} for $\Delta_0 = \Delta_1 = 0$.

4. The idem GUI

The **idem** GUI is web-based and developed in R using the Shiny (RStudio, Inc 2013) web application framework. The GUI can be accessed within R using the function `imShiny`, which calls the `runApp` function in the R package **shiny** (Chang, Cheng, Allaire, Xie, and McPherson 2016).

```
R> imShiny()
```

The **idem** GUI provides a series of tab panels that sequentially walk the user through the analysis which include `About`, `Upload Data`, `Model Specification`, `Data Exploration`, `Model Fitting`, `Configuration`, `Imputation` and `Report`. The details of each tab panel are given as follows:

About Panel:

The `About` panel serves as an introduction page for the software. The sections on this panel present the background information for **idem** and the purpose of the software. It also explains the basic steps to use the software.

Upload Data Panel:

The `Upload Data` panel provides an interface for users to upload the data to be analyzed. The sections and items within each section on this panel include:

- `Upload Data`

Choose File	Clicking the Browse... button will load local data files in `csv` or `plain text` format.
Separator	Field separating character.
Quote	Quoting character.
NA String	String for `NA` values.
Other	There are two additional options: the *Header* Checkbox indicates if the first line of the file are the names of the columns, the *Show Data* Checkbox indicates whether to present the uploaded data in the `Review Data` section on this panel.

- `Try An Example`
  Clicking the Try it button will load the example `abc` dataset.

- `Review Data`
  Presents the uploaded dataset in a table view.

Model Specification Panel:

The `Model Specification` panel is designed to specify the `idem-parameters`. This panel is only available after a dataset has been successfully uploaded. Items on this panel include:

Define Variables	The columns `Treatment`, `Time to death`, `Outcome`, `Baseline outcome`, `Baseline covariates` correspond to the `idem-parameters` `trt`, `surv`, `outcome`, `y0` and `cov`, respectively. The user selects the appropriate variables from the uploaded dataset that define `Treatment`, `Time to death`, `Outcome`, `Baseline outcome`, and `Baseline covariates`.
Functional Endpoint	Specify `enfml` in `idem-parameters`. This is an R expression indicating the user-specified final functional outcome of interest.
Study Duration	Specify `duration`, l_K, in `idem-parameters`. This is the length of the study.
Boundary	Specify `bounds` in `idem-parameters`. These create a numeric vector of lower and upper bounds for the functional outcomes
Unit Time	A drop-down list that specifies `unitTime` in `idem-parameters`. This is the unit of time measurement for survival.
Ranking Rules	Reserved for advanced users.

After the parameters are specified, click the [Validate Model] button which calls the **idem** function `imChkPars` to check if there are any errors or inconsistencies in the specifications.

Data Exploration Panel:

The `Data Exploration` panel provides summary tables and figures for the users to visualize the uploaded dataset including the missing data patterns survival status and functional outcomes among survivors. The items on this panel include:

Missing Table	Missingness frequency table generated by `imMisTable` in **idem**.
Missing Heatmap	Missingness heatmap plot generated by `imPlotMisPattern` in **idem**.
Survival	Kaplan-Meier survival curve generated by `imPlotSurv` in **idem**.
Survivors	Spaghetti plot of the observed functional outcomes for survivors generated by `imPlotCompleters` in **idem**.

Model Fitting Panel:

The `Model Fitting` panel provides R output and diagnostic plots for each component in the factorized joint distribution of the functional outcomes among survivors with no missing data (Equation 3). The diagnostic plots include the `Residuals vs. Fitted` plot and the `Normal Q-Q` plot.

Configuration Panel:

The `Configuration` panel sets the parameters for the multiple imputation and MCMC sampling. The sections and items within each section on this panel include:

- General Imputation Settings

    Imputed Datasets   Number of complete datasets to be generated.

    Bootstrap Samples   Number of boostrap samples for bootstrap analysis.

    Cores   Number of cores for parallel bootstap analysis .

    Random Seed   Random seed for multiple imputation.

- MCMC Parameters

    Iterations   STAN parameter specifying how many iterations including burn-in for posterior sampling.

    Number of burn-in   STAN parameter specifying how many burn-in for posterior sampling.

    Number of thinning   STAN parameter specifying the period for saving posterior samples.

    Number of Chains   STAN parameter specifying the number of MCMC chains for sampling.

    Acceptance Rate   STAN parameters that affect the MCMC convergence.

    Initial Step-size   STAN parameters that affect the MCMC convergence.

- Sensitivity Parameters And Additional Quantile Output

    Percentiles   Percentiles of the composite endpoint to be analyzed and reported.

    Sensitivity   Choices of sensitivity parameters Δ_T.
    Parameters

- Check Convergence
    Clicking the $\boxed{\text{Check Convergence}}$ button will randomly select a subject with at least one missing functional outcome, draw samples of the missing functional outcome(s) by MCMC sampling and present the trace plots of the Markov chains. The trace plots serve as a diagnostic tool for evaluating the mixing of the Markov chains in the imputation.

Imputation Panel:

The Imputation Panel conducts the imputation and bootstap analysis, presents the results and provides a link to download the imputed data. The sections and items within each section on this panel include:

- Get Imputed Data
    After clicking the $\boxed{\text{Get Imputed Data}}$ button, a progress bar will show up during the imputation. Once the imputation is finished, the following results are presented in this section:

Imputed Data	Contains three panels. The *Imputed Dataset* panel provides a table view of the complete dataset. The *Imputed Outcome* and the *Imputed Endpoint* panel provide the density plots of the imputed functional outcomes and the functional endpoint, respectively, that are generated by the **idem** function `imPlotImputed`.
Analysis Results	Presents the tables of the estimated θ and quantiles of the composite endpoint for all values of the sensitivity analysis parameters. It also presents the cumulative distribution function of the composite endpoint under the benchmark assumption that is generated by the `imPlotComposite` function in **idem**.
Download	Select the `Download` button to download the complete datasets as a delimited text file.

- Hypothesis Testing By Bootstrap

  Clicking the button `Hypothesis Testing by Bootstrap` will conduct the bootstrap analysis. The results are further presented in three panels.

  The `Ranks` panel presents the table of $\widehat{\theta}$'s, the corresponding standard deviation of the bootstraps and p-values for all sensitivity analysis scenarios.

  The `Quantiles` panel presents the table of requested quantiles and the corresponding lower and upper bounds for all sensitivity analysis scenarios.

  The `Contour Plot` panel presents the contour plot of the p-values obtained by testing the null hypothesis of $\theta = 0$ as a function of the treatment-specific sensitivity analysis. The plot is generated by the **idem** function `imPlotContour`.

Report Panel:

The `Report` panel provides a `Download` button for downloading the analysis results as a report. The available document formats for the report include PDF, HTML and Word.

5. Demonstration of idem GUI

In this section, we demonstrate the **idem** GUI using the ABC trial data. The imputation incorporates patient age (AGE) as the baseline covariate. There is no Y_0 and we set $B_L = 0$ and $B_U = 100$. The variable TRT is 0 and 1 for the UC+SBT and the SAT+SBT arm, respectively. We specify the following models for $\mu_{k,t}(\overline{Y}^{\dagger}_{k-1}, X; \boldsymbol{\alpha}_{k,t})$:

$$\mu_{1,t}(X, \boldsymbol{\alpha}_{1,t}) = \alpha_{1,t,1} + \alpha_{1,t,2} AGE$$
$$\mu_{2,t}(\overline{Y}^{\dagger}_1, X; \boldsymbol{\alpha}_{2,t}) = \alpha_{2,t,1} + \alpha_{2,t,2} AGE + \alpha_{2,t,3} Y_1^{\dagger}.$$

The entire analysis can be performed using the following steps:

Step 1. Upload the ABC data file to **idem** from the `Upload Data` panel (Figure 12). One can also load the data from **idem** by clicking the `Try it` button.

Step 2. Specify the `idem-parameters` on the `Model Specification` panel (Figure 13). Set the `TRT` column to be the `Treatment`, `SURV` column to be `Time to Death`, `Y1` and `Y2` columns to be `Outcome` and `AGE` column to be `Baseline covariates`. Specify the functional endpoint Z as `Y2` and study duration to be 365 days. Set the boundaries of the cognition score (i.e., functional outcomes) to be $(0, 100)$.

Click the `Validate Model` button to validate the model specification settings and may proceed to the next step if the result is *Model specification is valid* (Figure 14).

Step 3. In the `Data Exploration` panel, review the missing data pattern table (Figure 15), the missing data pattern heatmap, the Kaplan-Meier survival curves (Figure 16) and the spaghetti plot of the functional outcome among survivors. The results show that there is a statistically significant difference between Kaplan-Meier survival functions for the two treatment arms (p-value = 0.006).

Step 4. The model fitting results are presented on the `Model Fitting` panel (Figures 17-18). The residuals vs. fitted plot and the normal Q-Q plot of the model fitting results indicate that the normality assumption for the residuals may not hold.

Step 5. Move to the `Configuration` panel to specify imputation and bootstrap analysis parameters. Because of the concern about the normality assumption based on the model fitting results, specify the `Normality assumption` to be `No`. Specify 100 bootstrap samples for the bootstrap analysis and specify the sensitivity parameters to be $-0.2, 0, 0.2$. Choose 4 cores for bootstap parallel analysis (Figure 19).

This panel provides a `Check Convergence` button to randomly select an individual with missing functional outcomes, conduct the imputation under the benchmark assumption for the individual and present the traceplot of the MCMC samples (Figure 20). If there appears to be an issue with the convergence, the user should consider running a longer Markov chain and adjusting the target Metropolis acceptance rate or initial step-size. In our example, it can be seen that the MCMC chains are mixed well.

Step 6. On the `Imputation` panel, click the `Get Imputed Data` button to conduct the imputation and the `Hypothesis Testing by Bootstrap` button to conduct the bootstrap analysis and draw inference. Selected results for the example are presented on Figures 21 and 22.

The results on the `Ranks` panel under hypothesis testing suggest that the SAT+SBT group is favored over the control group under all the sensitivity analysis scenarios we consider, i.e., $\theta > 0$, Figure 22. Under the benchmark assumptions, $\theta = 0.18$ (SD 0.08, p-value = 0.02). The statistical test for θ results in a statistically significant finding for scenarios when $\Delta_1 = 0$ or 0.2 for the SAT+SBT arm except when $\Delta_1 = 0$ and $\Delta_0 = 0.2$ for the UC+SBT arm. When $\Delta_1 = -0.2$, the test for θ is significant when $\Delta_0 = -0.2$.

For the UC+SBT group, we estimate that 50% of the subjects will survive past 72 days (95% CI: survive past 34 to 364 days). In the SAT+SBT group, we estimate that 50% of subjects will survive to 12 months with cognitive scores of 29 or greater (95% CI: cognitive score of 17 to 38 or greater). These results are reported on the `Quantiles` panel under hypothesis testing (not shown).

Based on the primary and sensitivity analysis results, we conclude that there is relatively robust evidence that a difference exists between the control and the intervention arms in the composite endpoint of survival and cognitive performance which favors the intervention arm.

Step 7. After conducting the analysis, choose to download a report as a `PDF`, `HTML`, or `Word` document from the `Report` panel, (Figure 23). The report contains sections for `Data Summary`, `Analysis Summary`, `Missingness Summary`, `Imputation Results` and `Bootstrap Results`. Figure 24 shows the content page of report for the example.

Figure 12: Upload data.

Figure 13: Model specification.

Figure 14: Model specification.
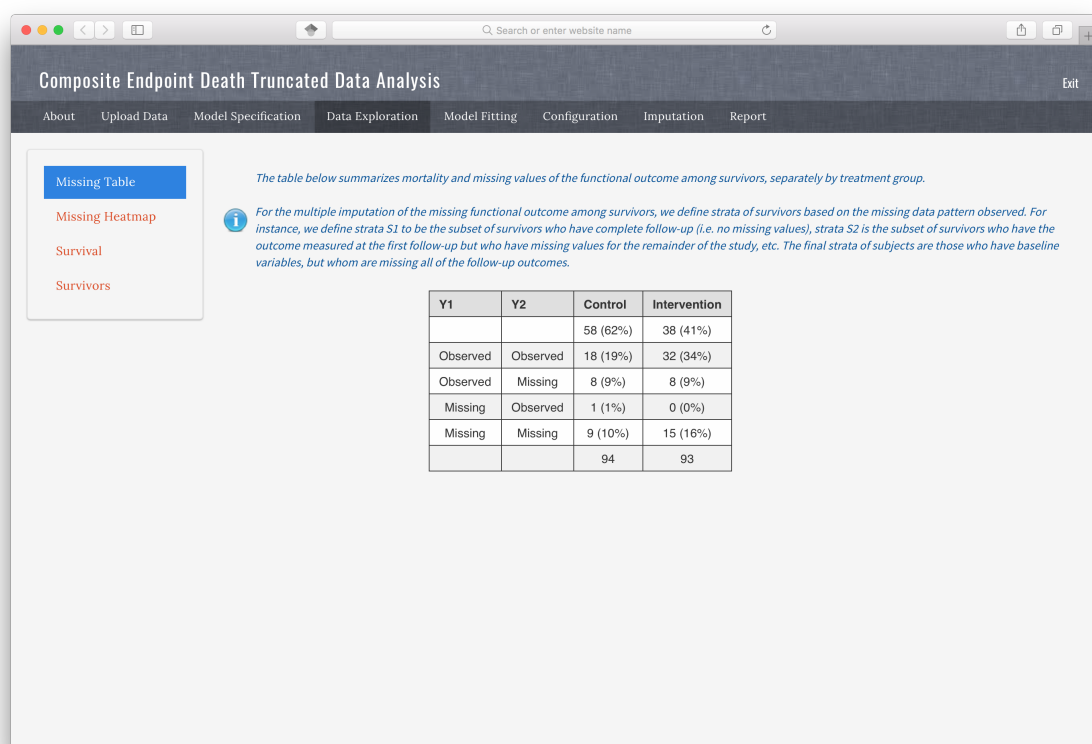
Figure 15: Data exploration: Missigness frequency table.

Figure 16: Data exploration: Survival curves.
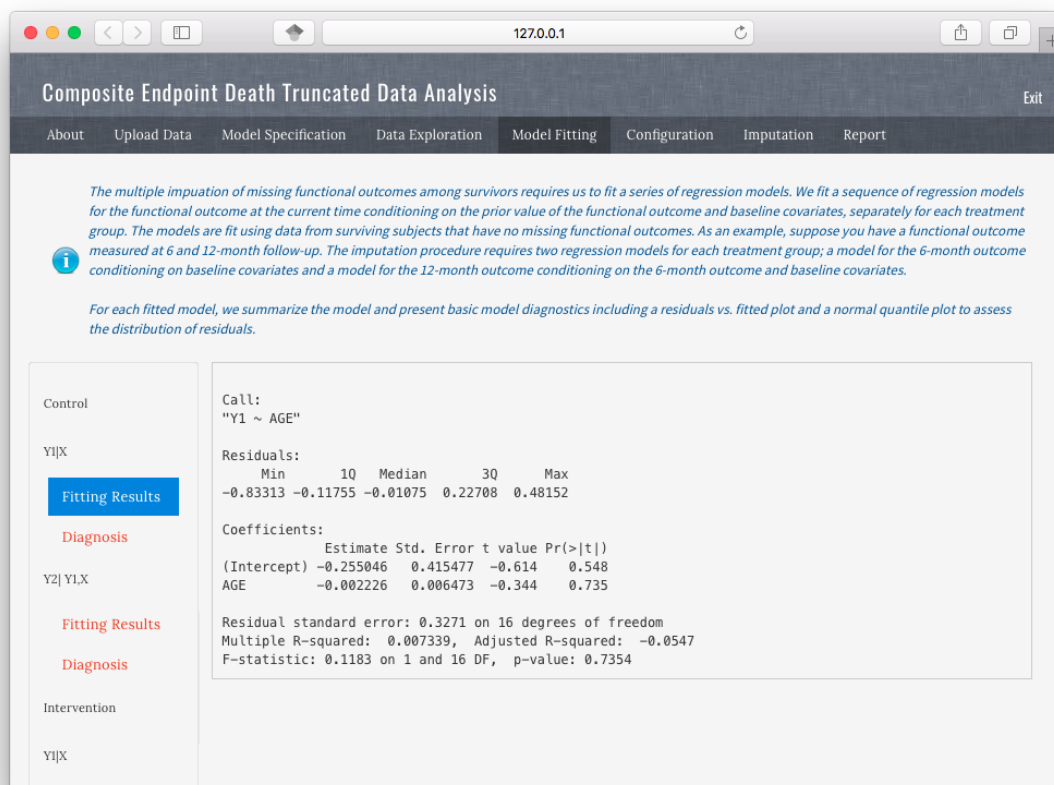
Figure 17: Model fitting: Raw R output.

Figure 18: Model fitting: Model fitting diagnostic plots.

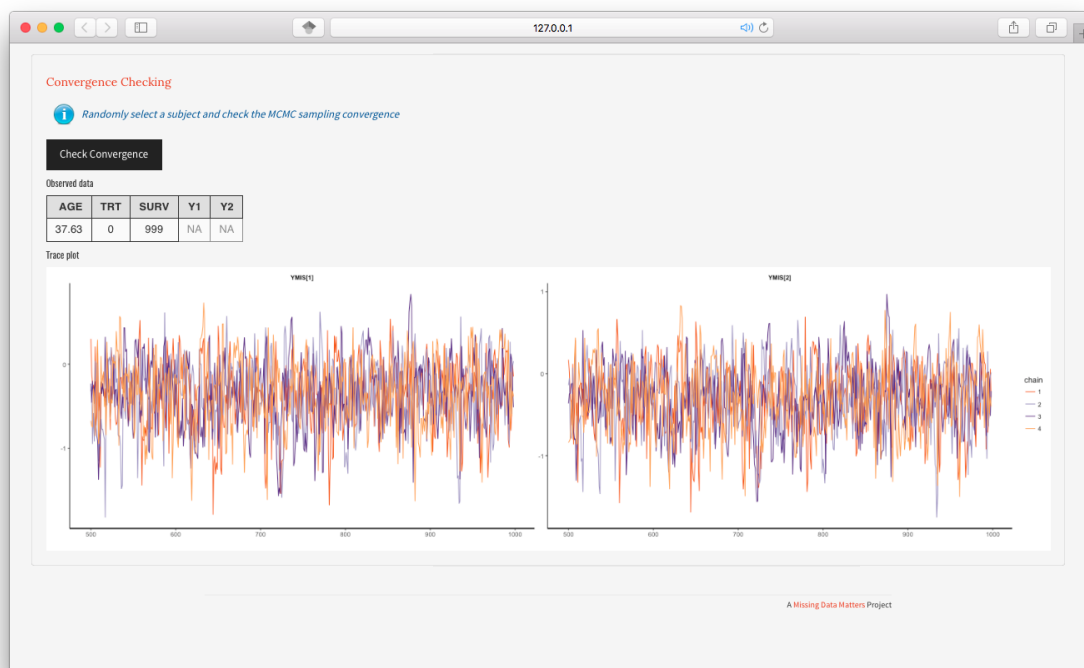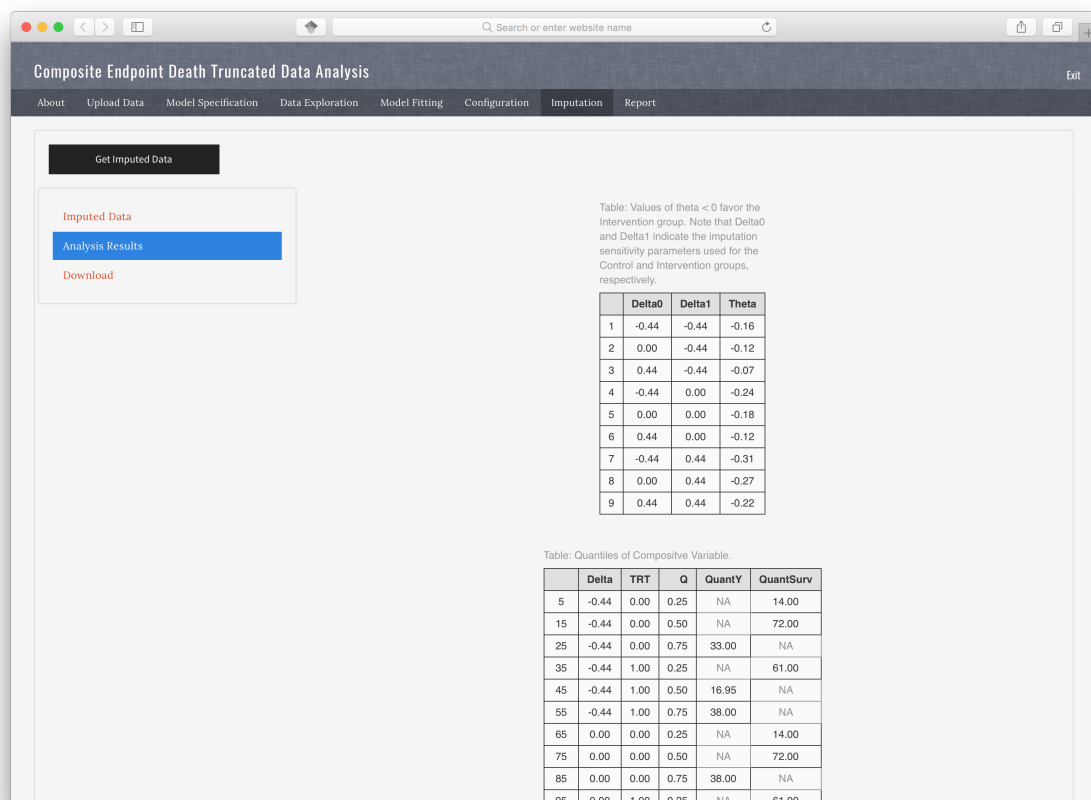Figure 19: Configuration: Parameter specification

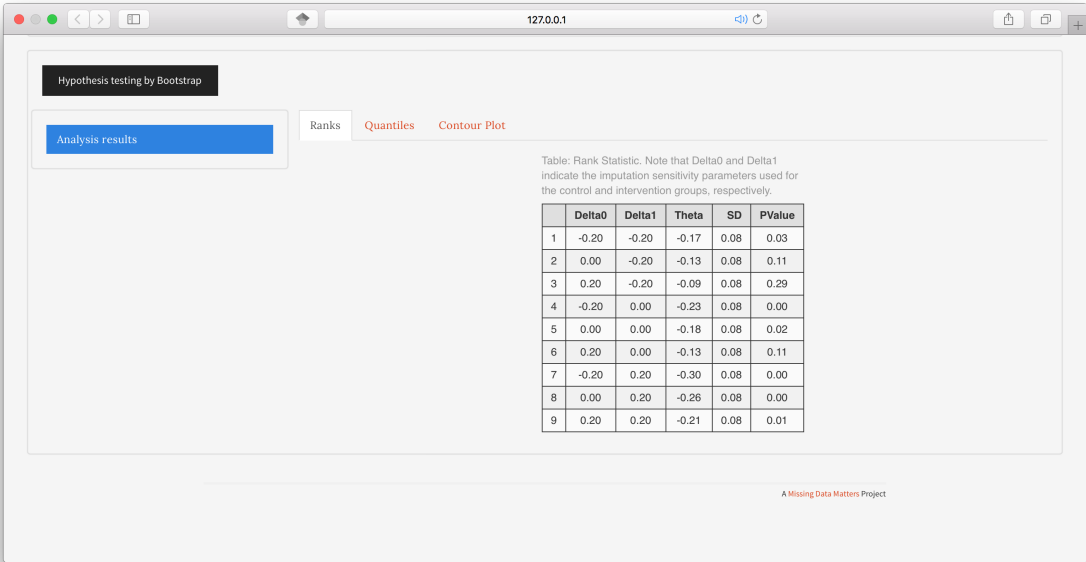Figure 20: Configuration: Convergence.

Figure 21: Imputation results
.

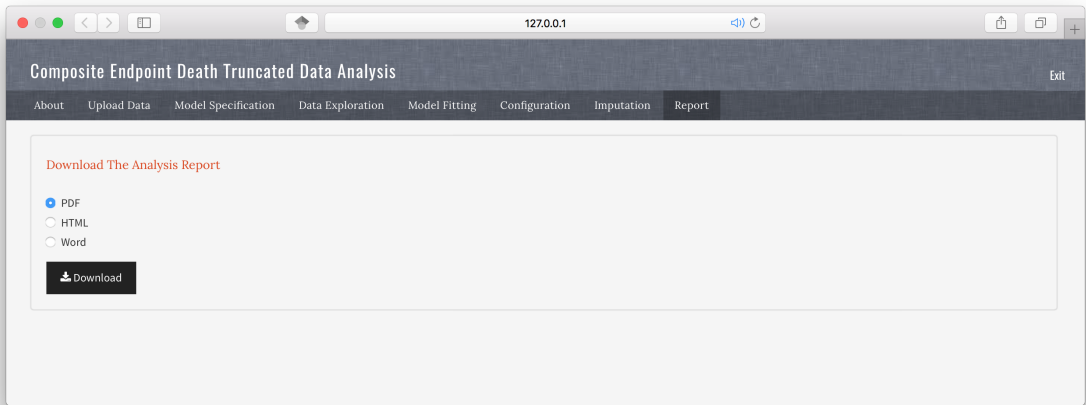Figure 22: Hypothesis analysis results
.

Figure 23: Report panel.

Composite Endpoint Analysis Report

Missing Data Matters

2017-04-22

Contents

Figure 24: Content page of a downloaded report.

6. Conclusion

Missing data and data "truncated due to death" occur frequently in randomized clinical trials. Wang *et al.* (2017) proposed an approach that was based on the composite of mortality and the functional outcomes among survivors that accounts for both intermittent missing data and data "truncated due to death". Their proposal applied the complete case missing value constraints for missing data imputation and suggested a global sensitivity analysis framework to further assess the robustness of the findings.

In this paper, we introduce the R package **idem** that implements the proposed method in Wang *et al.* (2017). The **idem** package provides functions for users to visualize the missing data patterns, the observed functional outcomes among survivors and the survival curves for all randomized patients. The imputation functions in **idem** implement the imputation using the Adaptive Hamiltonian Monte Carlo algorithm provided by **rstan**. The **idem** also provides functions for conducting bootstrap analysis and drawing inference. In addition, the **idem** package also provides functions to evaluate the *survivors only* treatment effect and *survivor average causal effect* on the functional outcomes based on the same missing data imputation strategy proposed by Wang *et al.* (2017).

A unique feature of **idem** is that it provides a Shiny-based graphical user interface for users to apply functions in **idem** in an interactive and user-friendly manner. With the GUI feature, **idem** can be used by not only statisticians but also analysts that are not familiar with the R environment.

Acknowledgments

References

Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A (2017). "Stan: A Probabilistic Programming Language." *Journal of Statistical Software*, **76**(1).

Chang W, Cheng J, Allaire J, Xie Y, McPherson J (2016). *shiny: Web Application Framework for R*. R Package Version 0.13.2.

Chiba Y, VanderWeele TJ (2011). "A Simple Method for Principal Strata Effects When the Outcome Has Been Truncated Due to Death." *American journal of epidemiology*, **173**(7), 745–751.

Diehr P, Patrick DL, Spertus J, Kiefe CI, Donell M, Fihn SD (2001). "Transforming Self-Rated Health And The SF-36 Scales To Include Death To Improve Interpretability." *Medical Care*, **39**, 670–680.

Dupont C, Shepherd B (2014). *sensitivityPStrat: Principal Stratification Sensitivity Analysis Functions*. R package version 1.0-6.

Frangakis CE, Rubin DB (2002). "Principal Stratification In Causal Inference." *Biometrics*, **58**(1), 21–29.

Girard TD, Kress JP, Fuchs BD, Thomason JWW, Schweickert WD, Pun BT, Taichman DB, Dunn JG, Pohlman AS, Kinniry PA, Jackson JC, Canonico AE, Light RW, Shintani AK, Thompson JL, Gordon SM, Hall JB, Dittus RS, Bernard GR, Ely EW (2008). "Efficacy and Safety of A Paired Sedation and Ventilator Weaning Protocol For Mechanically Ventilated Patients In Intensive Care (Awakening And Breathing Controlled Trial): A Randomised Controlled Trial." *Lancet*, **371**, 126–134.

Hayden D, Pauler DK, Schoenfeld D (2005). "An Esitmator for Treatment Comparisons Among Survivors in Randomized Trials." *Biometrics*, **61**(1), 305–310.

Joshua Chen Y, Gould AL, Nessly ML (2005). "Treatment Comparisons for A Partially Categorical Outcome Applied to A Biomarker With Assay Limit." *Statistics in Medicine*, **24**(2), 211–228.

Lachin JM (1999). "Worst-Rank Score Analysis with Informatively Missing Observations in Clinical Trials." *Controlled Clinical Trials*, **20**(5), 408–422.

Little RJ (1993). "Pattern-Mixture Models for Multivariate Incomplete Data." *Journal of the American Statistical Association*, **88**(421), 125–134.

Rizopoulos D (2010). "**JM**: An R Package for the joint modelling of longitudinal and time-to-event data." *Journal of Statistical Software*, **35**(9), 1–33.

Rizopoulos D (2016). "The R Package **JMbayes** for Fitting Joint Models for Longitudinal and Time-to-Event Data Using MCMC." *Journal of Statistical Software, Articles*, **72**(7), 1–46. ISSN 1548-7660.

RStudio, Inc (2013). *Easy Web Applications in R.*

Wang C, Scharfstein DO, Colantuoni E, Girard TD, Yan Y (2017). "Inference in Randomized Trials with Death and Missingness." *Biometrics*, **73**(2), 431–440.

Affiliation:

Chenguang Wang
Division of Biostatistics and Bioinformatics
Sidney Kimmel Comprehensive Cancer Center
Johns Hopkins University
550 N. Broadway Suite 1103
Baltimore MD, 21205
Email: cwang68@jhmi.edu

Elizabeth Colantuoni
Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
615 North Wolfe Street
Baltimore MD, 21205
Email: ejohnso2@jhmi.edu

Andrew Leroux
Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
615 North Wolfe Street
Baltimore MD, 21205
Email: aleroux2@jhu.edu

Daniel O. Scharfstein
Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
615 North Wolfe Street
Baltimore MD, 21205
Email: dscharf68@gmail.com