Global Sensitivity Analysis of Randomized Trials with Missing Data

FDA Shortcourse

Daniel Scharfstein
Johns Hopkins University
dscharf@jhu.edu

May 8, 2017

Funding Acknowledgments

- FDA
- PCORI

Case Study: Quetiapine Bipolar Trial

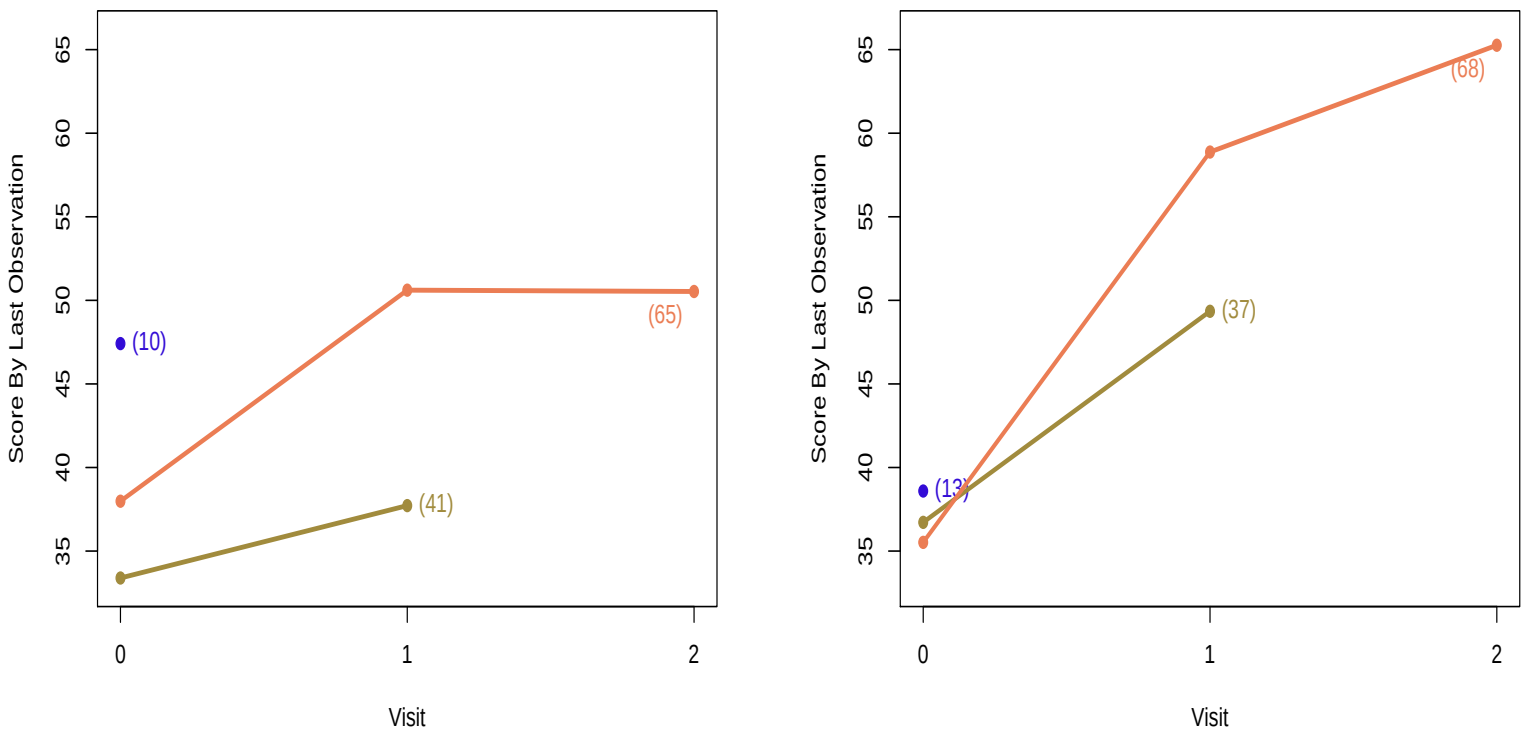
- Patients with bipolar disorder randomized equally to one of three treatment arms: placebo, Quetiapine 300 mg/day or Quetiapine 600 mg/day (Calabrese *et al.*, 2005).
- Randomization was stratified by type of bipolar disorder.
- Short-form version of the Quality of Life Enjoyment Satisfaction Questionnaire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

Quetiapine Bipolar Trial

- 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.
- 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.



Central Question

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?

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.
- For now, we will focus on monotone missing data.
- Later, we will address missingness prior to last visit on-study.

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 $R_k = 0$.

Notation: Quetiapine Bipolar Trial

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

Notation: Quetiapine Bipolar Trial

- 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 $Y_0 = y_0$ $Y_1 = y_1$.

Benchmark Assumption (Missing at Random)

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 $B_2(y_0, y_1)$.

Benchmark Assumption (Missing at Random)

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.

Benchmark Assumption (Missing at Random)

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

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

Missing Not at Random (MNAR)

The MAR assumption is not the only one that is (a) unverifiable and (b) allows identification of μ^*.

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

$$\begin{aligned} & f^*(Y_2|A_0(y_0), Y_1 = y_1) \\ & = f^*(Y_2|B_1(y_0), Y_1 = y_1) \text{ for all } y_0, y_1 \end{aligned} \tag{3}$$

and

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

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:

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

and

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

where

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

and

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

Missing Not at Random (MNAR)

Written in this way:

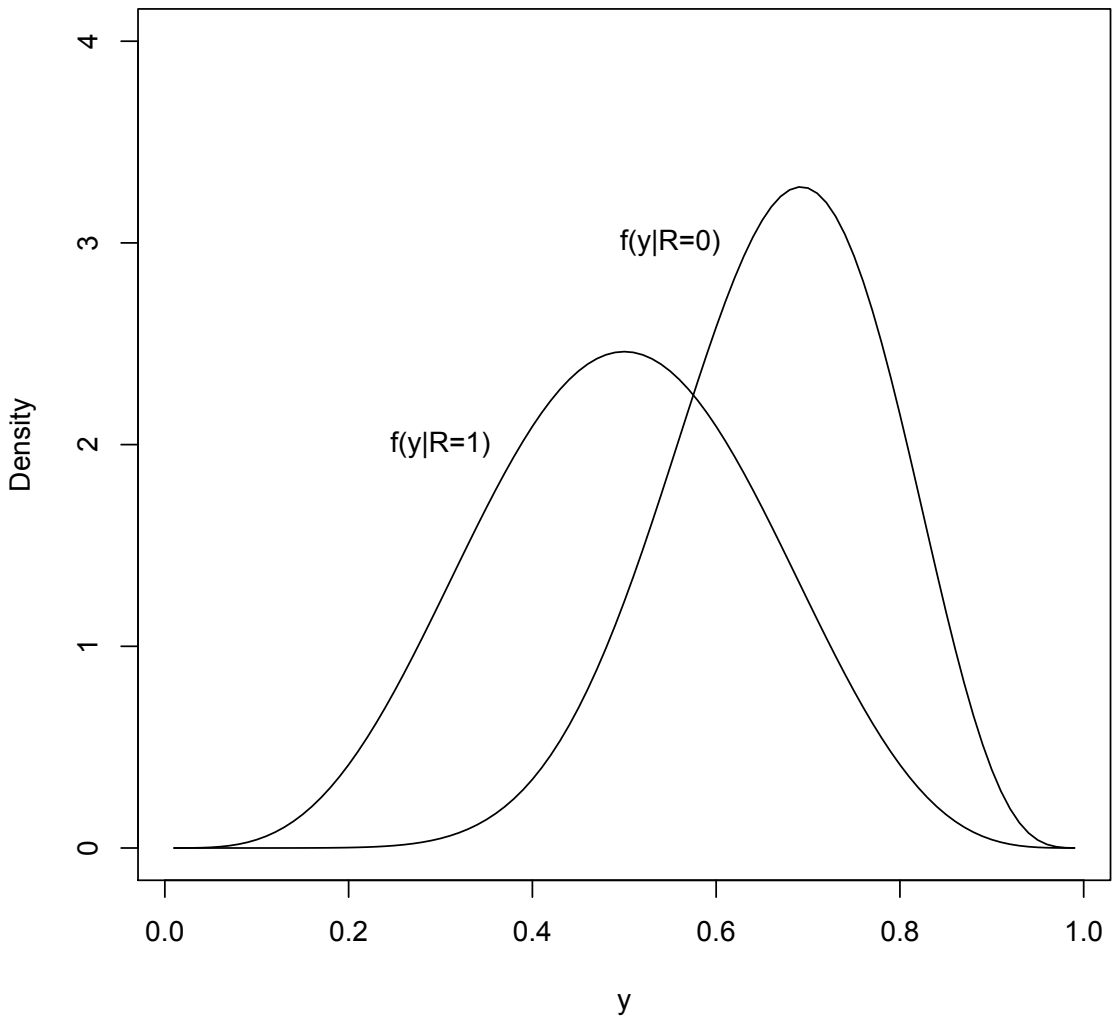
- The decision to discontinue the study before visit 1 is like the flip of a coin with probability depending on the value of the outcome at visit 0 *and (in a specified way)* the value of the outcome at visit 1.
- For those on-study at visit 1, the decision to discontinue the study before visit 2 is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

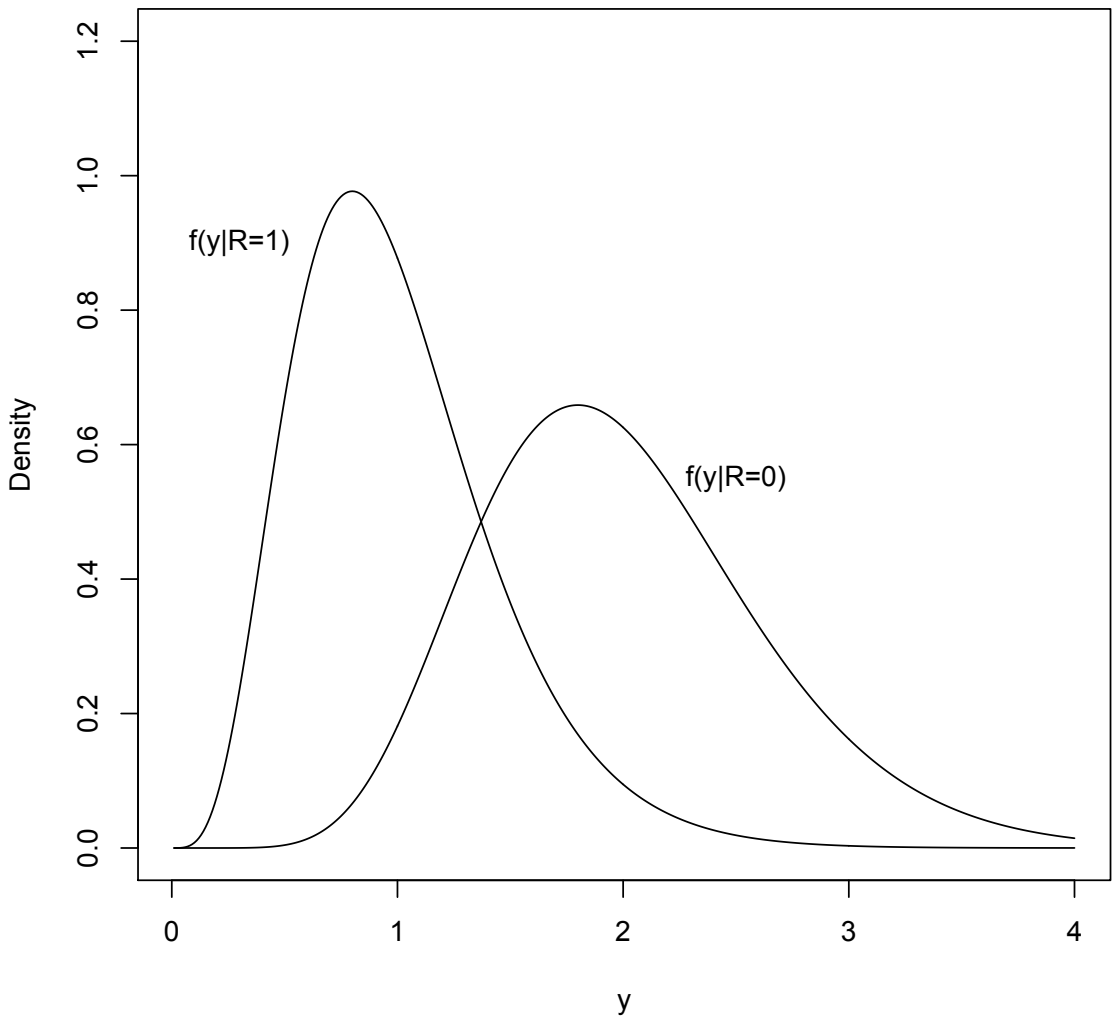
$$f(Y|R=0) \propto f(Y|R=1)\exp\{\alpha r(Y)\}$$

- 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

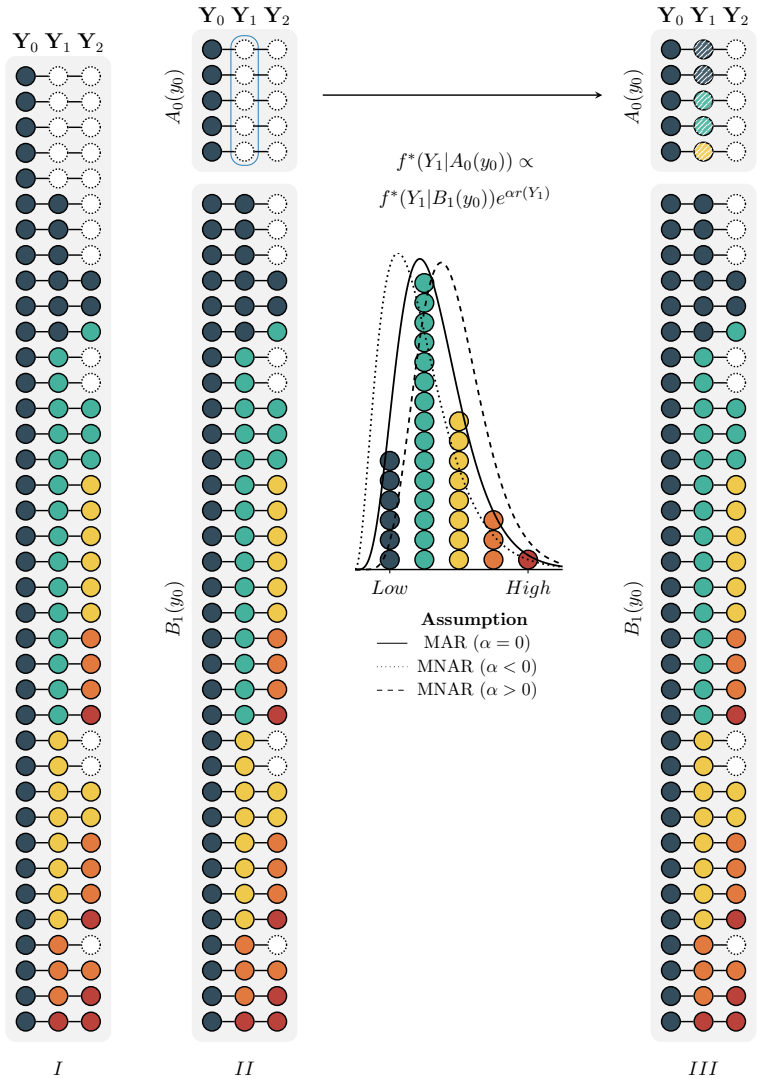
$$H_2^*(y_1,y_0) = P^*(R_2 = 0 | R_1 = 1, Y_1 = y_1, Y_0 = y_0)$$

and

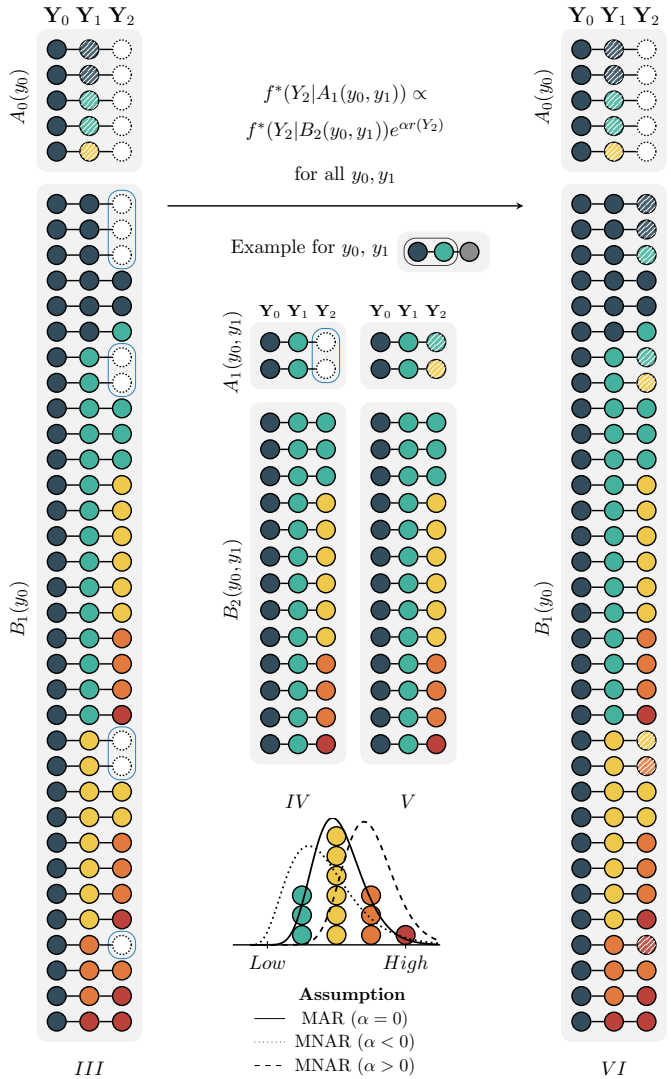
$$H_1^*(y_0) = P^*(R_1 = 0 | R_0 = 1, Y_0 = y_0)$$

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

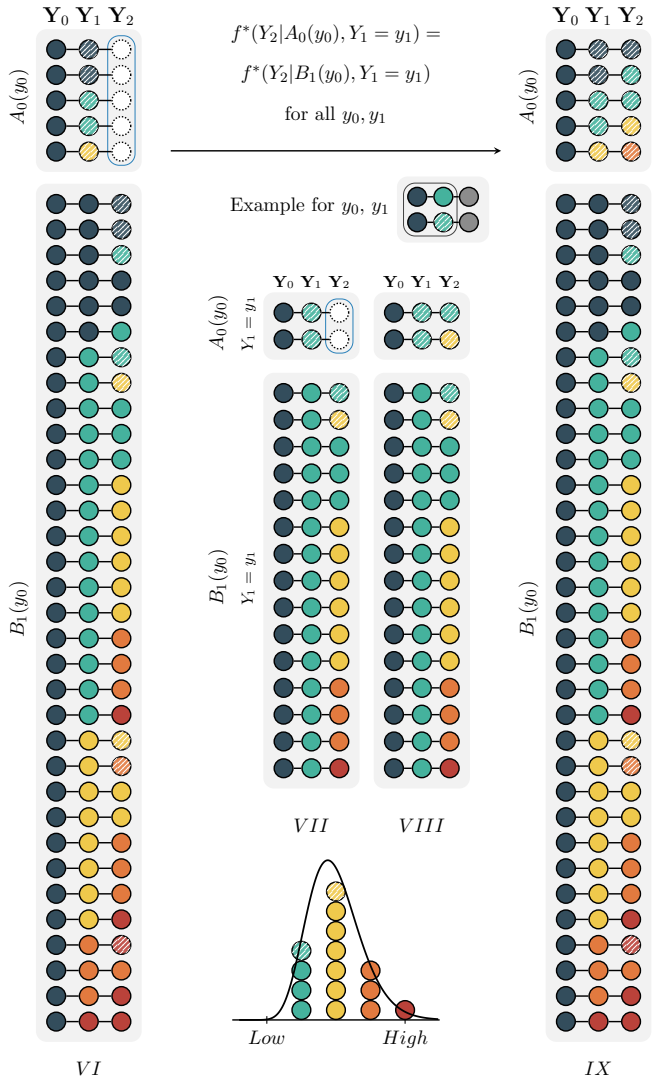
Global Sensitivity Analysis



Global Sensitivity Analysis



Global Sensitivity Analysis



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.

Inference - Type (ii) Assumptions

We make the estimation task slightly easier by assuming that

$$F_2^*(y_2|y_1, y_0) = F_2^*(y_2|y_1) \tag{7}$$

and

$$H_2^*(y_1, y_0) = H_2^*(y_1) \tag{8}$$

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 $\widehat{\sigma}_{F_2} = \text{argmin } CV_{F_2^*(\cdot|\cdot)}(\sigma_{F_2})$.

Inference - Cross-Validation

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.

In our software, we set $\sigma_{H_2} = \sigma_{H_1} = \sigma_H$ and minimize a single CV function.

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, *in theory*, 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

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

  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 $\tilde{\sigma}_\alpha^2 = \frac{1}{n}\sum_{i=1}^n \psi_{\widehat{P}}(O_i;\alpha)^2$.
- 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.

Inference - Efficient Influence Function/Gradient

Let

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

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

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

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

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

Inference - Efficient Influence Function/Gradient

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

where

Inference - Efficient Influence Function/Gradient

$$
\begin{aligned}
a_0^*(Y_0) &= E^*\left(\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \,\middle|\, Y_0\right) - \mu(P^*; \alpha) \\
b_1^*(Y_0, Y_1; \alpha) &= E^*\left(\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \,\middle|\, R_1 = 1, Y_1, Y_0\right]) - E^*\left(\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \,\middle|\, R_1 = 1, Y_0\right) \\
&\quad + E^*\left(\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \left[\frac{\exp\{\alpha r(Y_1)\}}{g_1^*(Y_0, Y_1; \alpha)}\right] \,\middle|\, R_1 = 1, Y_0\right) H_1^*(Y_0) \left\{1 - \frac{\exp\{\alpha r(Y_1)\}}{w_1^*(Y_0; \alpha)}\right\} \\
b_2^*(Y_1, Y_2; \alpha) &= E^*\left(\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \,\middle|\, R_2 = 1, Y_2, Y_1\right) - E^*\left(\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \,\middle|\, R_2 = 1, Y_1\right) \\
&\quad + E^*\left(\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \left[\frac{\exp\{\alpha r(Y_2)\}}{g_2^*(Y_1, Y_2; \alpha)}\right] \,\middle|\, R_2 = 1, Y_1\right) H_2^*(Y_1) \left\{1 - \frac{\exp\{\alpha r(Y_2)\}}{w_2^*(Y_1; \alpha)}\right\} \\
c_1^*(Y_0) &= E^*\left(\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \left[\frac{\exp\{\alpha r(Y_1)\}}{g_1^*(Y_0, Y_1; \alpha)}\right] \,\middle|\, Y_0\right) \\
&\quad - E^*\left(\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \left[\frac{1}{g_1^*(Y_0, Y_1; \alpha)}\right] \,\middle|\, Y_0\right) w_1^*(Y_0; \alpha) \\
c_2^*(Y_1) &= E^*\left(\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \left[\frac{\exp\{\alpha r(Y_2)\}}{g_2^*(Y_1, Y_2; \alpha)}\right] \,\middle|\, R_1 = 1, Y_1\right) \\
&\quad - E^*\left(\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \left[\frac{1}{g_2^*(Y_1, Y_2; \alpha)}\right] \,\middle|\, R_1 = 1, Y_1\right) w_2^*(Y_1; \alpha)
\end{aligned}
$$

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.
- In simulation studies, symmetric studentized-t bootstrap with jackknife standard errors performs best.

Bootstrap

- Non-parametric bootstrap is that it does not require a model for the distribution of the observed data.
- Since our analysis depends on correct specification and on estimation of such a model, it makes sense to use this model to bootstrap observed datasets.

Bootstrap

Using $\widehat{F}_0(y_0)$, $\widehat{F}_1(y_1|y_0;\widehat{\sigma}_F)$, $\widehat{F}_2(y_2|y_1;\widehat{\sigma}_F)$ and $\widehat{H}_1(y_0;\widehat{\sigma}_H)$and $\widehat{H}_2(y_1;\widehat{\sigma}_H)$, simulate observed data for an individual using the following procedure:

1. Draw Y_0 from $\widehat{F}_0(\cdot)$
2. Draw R_1 from $\widehat{H}_1(Y_0;\widehat{\lambda}_H)$. If $R_1 = 0$, stop.
3. Draw Y_1 from $\widehat{F}_1(\cdot \mid Y_0;\widehat{\lambda}_F)$.
4. Draw R_2 from $\widehat{H}_2(Y_1;\widehat{\lambda}_H)$. If $R_2 = 0$, stop.
5. Draw Y_2 from $\widehat{F}_2(\cdot \mid Y_1;\widehat{\lambda}_F)$.

We use this procedure to simulate observed datasets, each comprised of n individuals.

Goodness of Fit

- Simulate, using the parametric bootstrap procedure, a dataset with a large number, say N, individuals.
- Compare summary statistics of the simulated dataset to the comparable summary statistics on the actual dataset.

Comparison of Treatment Groups

To compare treatment groups, we draw inference about $\Delta^* = \mu^{(1)*} - \mu^{(2)*}$. Assuming treatment-specific sensitivity analysis parameters $\alpha^{(1)}$ and $\alpha^{(2)}$, we estimate Δ^* by

$$\tilde{\Delta}_{\alpha^{(1)},\alpha^{(2)}} = \tilde{\mu}^{(1)}_{\alpha^{(1)}} - \tilde{\mu}^{(2)}_{\alpha^{(2)}}$$

Using the fact that the treatment groups are independent, we can form a confidence interval for Δ^* using the same ideas discussed above.

Comparison of Treatment Groups

Specifically, we can draw B bootstrap samples within each treatment group and compute cutpoints based on the distribution:

$$\left\{ \frac{\tilde{\Delta}^{(b)}_{\alpha^{(1)},\alpha^{(2)}} - \tilde{\Delta}_{\alpha^{(1)},\alpha^{(2)}}}{\tilde{se}\left(\tilde{\Delta}^{(b)}_{\alpha^{(1)},\alpha^{(2)}}\right)} : b = 1, 2, \ldots, B \right\}, \tag{13}$$

where

$$\tilde{se}\left(\tilde{\Delta}^{(b)}_{\alpha^{(1)},\alpha^{(2)}}\right) = \sqrt{\left\{\tilde{se}\left(\tilde{\mu}^{(b,1)}_{\alpha}\right)\right\}^2 + \left\{\tilde{se}\left(\tilde{\mu}^{(b,2)}_{\alpha}\right)\right\}^2}$$

is an estimator of the standard error of $\tilde{\Delta}^{(b)}_{\alpha^{(1)},\alpha^{(2)}}$. Equal-tailed and symmetric confidence intervals can then be formed.

Plausibility of α

For each α, the estimated mean among non-completers can be back-calculated from the estimated overall mean ($\tilde{\mu}_\alpha$):

$$\frac{\tilde{\mu}_\alpha - \frac{1}{n}\sum_i R_{K,i} Y_{K,i}}{\frac{1}{n}\sum_i (1 - R_{K,i})} \tag{14}$$

As a plausibility check, this mean can be contrasted with the mean among completers. If the differences are scientifically implausible, then the choice of α can be ruled out.

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

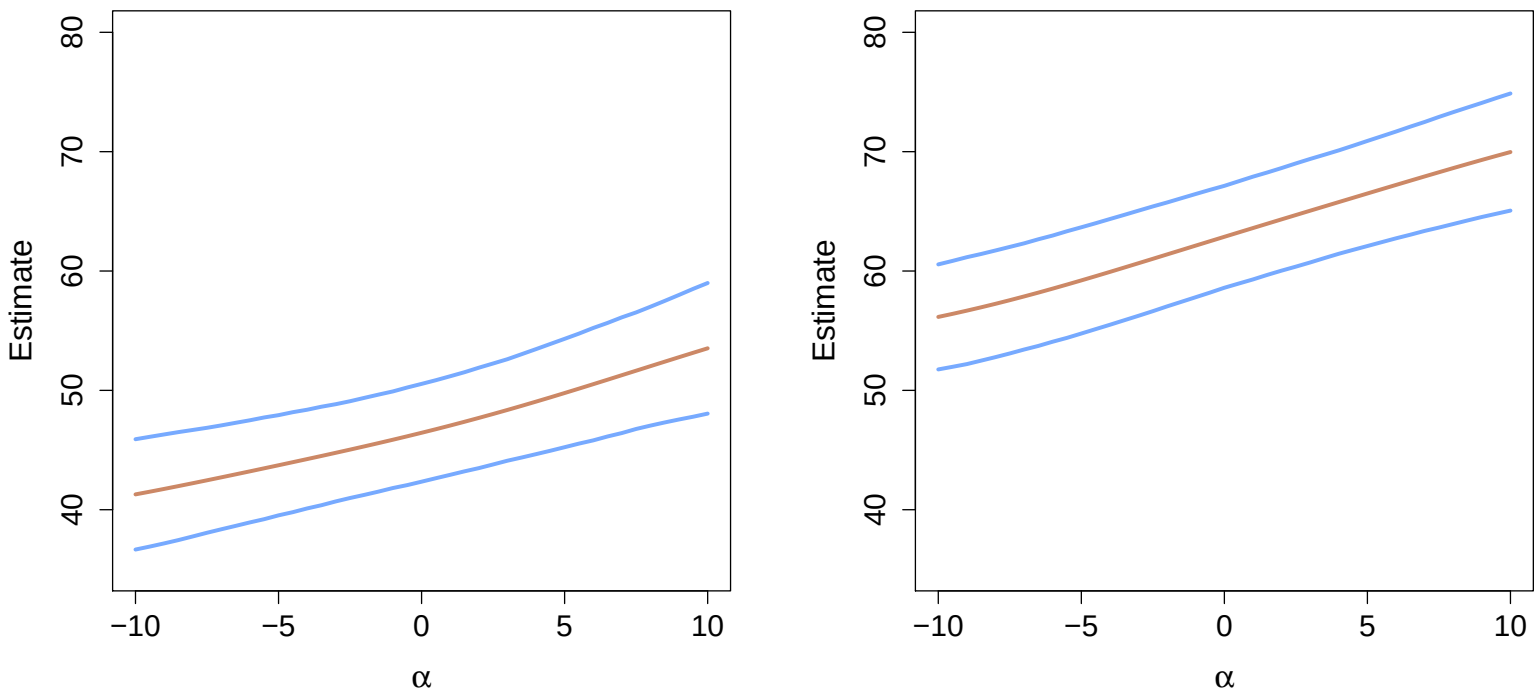
- Under MAR, the estimated values of μ^* are 46.45 (95% CI: 42.35,50.54) and 62.87 (95% CI: 58.60,67.14) for the placebo and 600 mg arms.
- The estimated difference between 600 mg and placebo is 16.42 (95% 10.34, 22.51)
- Statistically and clinically significant improvement in quality of life in favor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

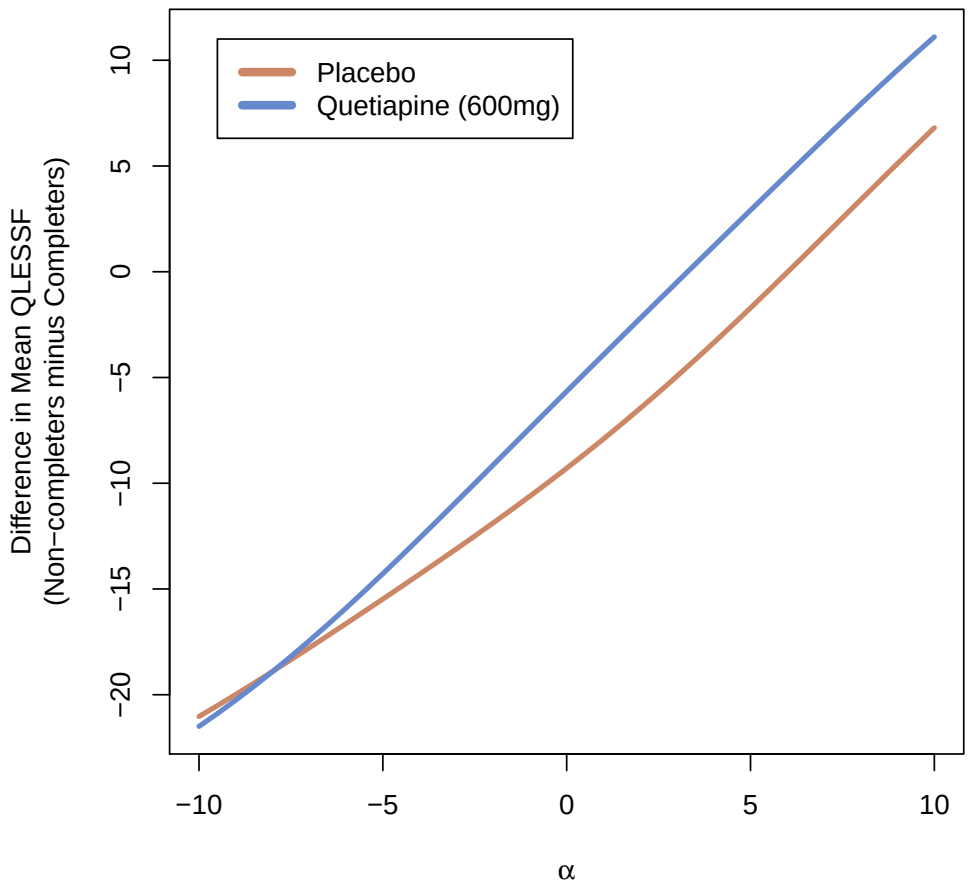
Quetiapine Bipolar Trial - Sensitivity Analysis

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



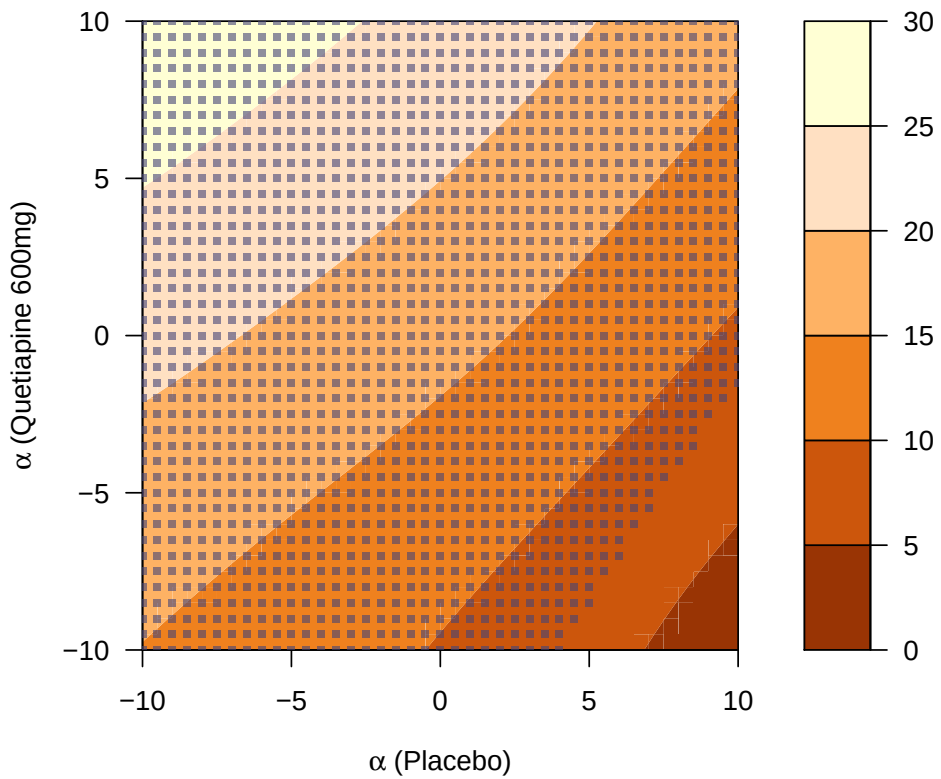
Quetiapine Bipolar Trial - Sensitivity Analysis

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.



Quetiapine Bipolar Trial - Sensitivity Analysis

- 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

		Placebo			Quetiapine		
α	Estimator	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%

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 $P \in \mathscr{M}$ can be represented by

- $F_0(y_0) := P(Y_0 \leq y_0)$;
- $F_{k+1}(y_{k+1} \mid \overline{y}_k) :=$
  $P\left(Y_{k+1} \leq y_{k+1} \mid R_{k+1} = 1, \overline{Y}_k = \overline{y}_k\right)$,
  $k = 0, 1, \ldots, K-1$;
- $H_{k+1}(\overline{y}_k) := P\left(R_{k+1} = 0 \mid R_k = 1, \overline{Y}_k = \overline{y}_k\right)$,
  $k = 0, 1, \ldots, K-1$.

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.

Missing at Random

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 $\overline{Y}_k = \overline{y}_k$.

Missing at Random

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:

$$f^*(\underline{Y}_k|A_k(\overline{y}_k)) = f^*(\underline{Y}_k|B_k(\overline{y}_k)) \text{ for all } \overline{y}_k \qquad (15)$$

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).*

Missing at Random

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:

$$f^*(\underline{Y}_{k+1}|A_k(\overline{y}_k), Y_{k+1} = y_{k+1}) = f^*(\underline{Y}_{k+1}|B_{k+1}(\overline{y}_k), Y_{k+1} = y_{k+1}) \text{ for all } \overline{y}_{k+1} \tag{16}$$

and

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

Equation (16) 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:

$$P^*(R_{k+1} = 0 | R_k = 1, \overline{Y}_K = \overline{y}_K) = P^*(R_{k+1} = 0 | R_k = 1, \overline{Y}_{k+1} = \overline{y}_{k+1}) \text{ for all } \overline{y}_K \quad (18)$$

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.

Missing Not at Random

Next, we generalize (17) and impose the following exponential tilting "linking" assumptions:

$$f^*(Y_{k+1}|A_k(\overline{y}_k)) \propto f^*(Y_{k+1}|B_{k+1}(\overline{y}_k)) \exp(\alpha r(Y_{k+1})) \text{ for all } \overline{y}_k \tag{19}$$

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 (16) and (19), where $r(\cdot)$ is considered fixed and α is a sensitivity analysis parameter that serves as the class index.
- (19) reduces to (17) when $\alpha = 0$. Thus, when $\alpha = 0$, the missing at random assumption is obtained.
- When $\alpha > 0$ (< 0), (19) 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

1. Assume
   - $F^*_{k+1}(y_{k+1} \mid \overline{y}_k) = F^*_{k+1}(y_{k+1} \mid y_k)$
   - $H^*_{k+1}(\bar{y}_k) = H^*_{k+1}(y_k)$
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 bootstrap alternatives to Wald-based confidence intervals.
5. Goodness-of-fit
6. Assess plausibility of α

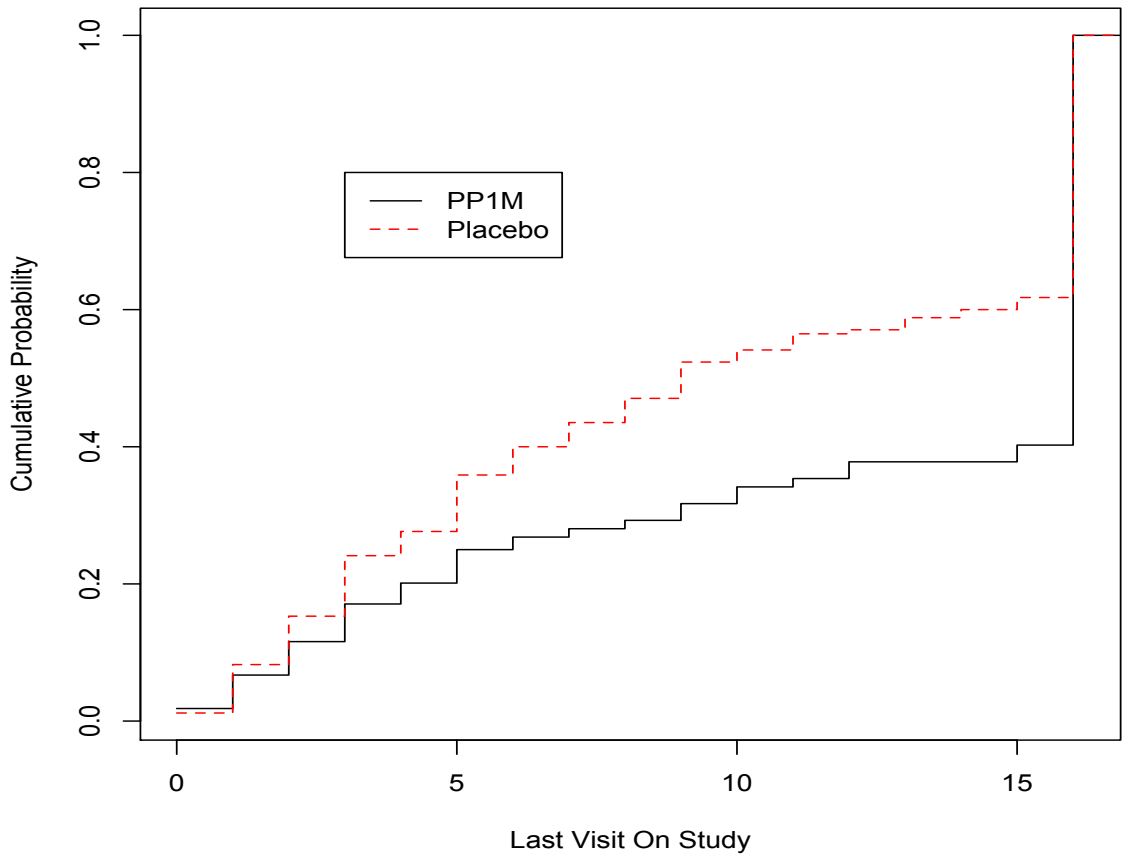
Case Study: SCA-3004

- 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.

Case Study: SCA-3004

- 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 ($p=0.0001$).

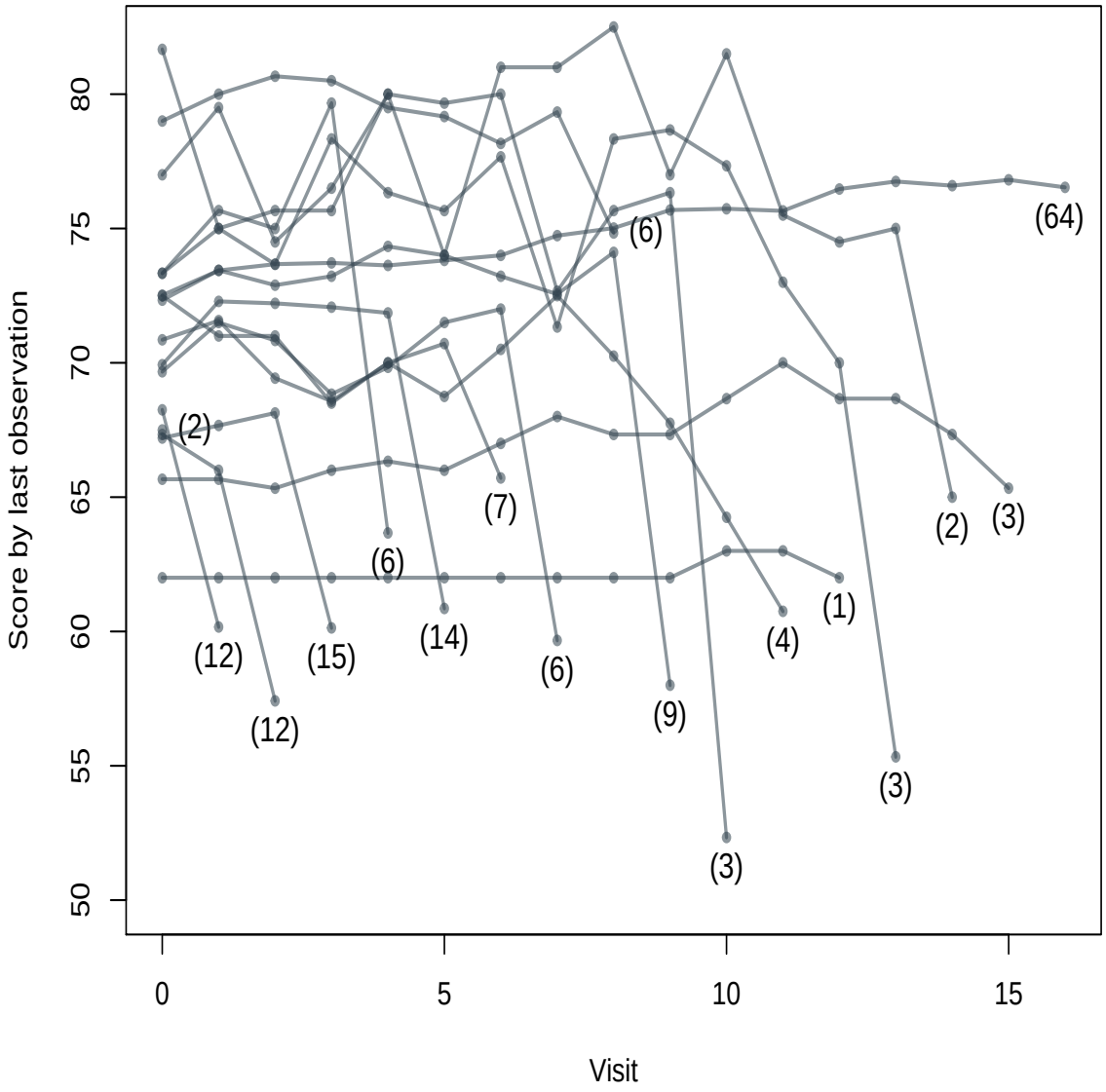
Case Study: SCA-3004



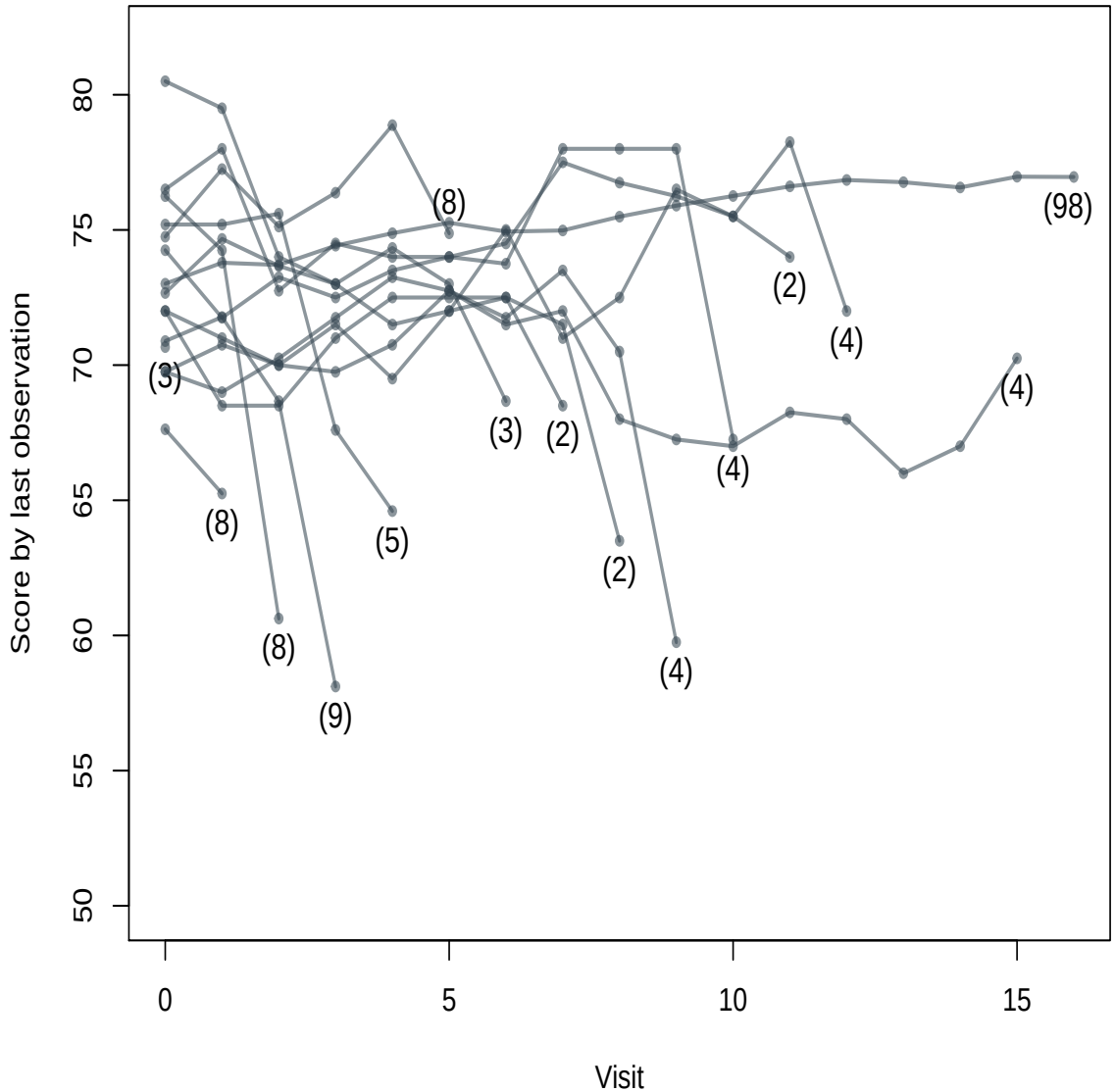
Case Study: SCA-3004

- 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).

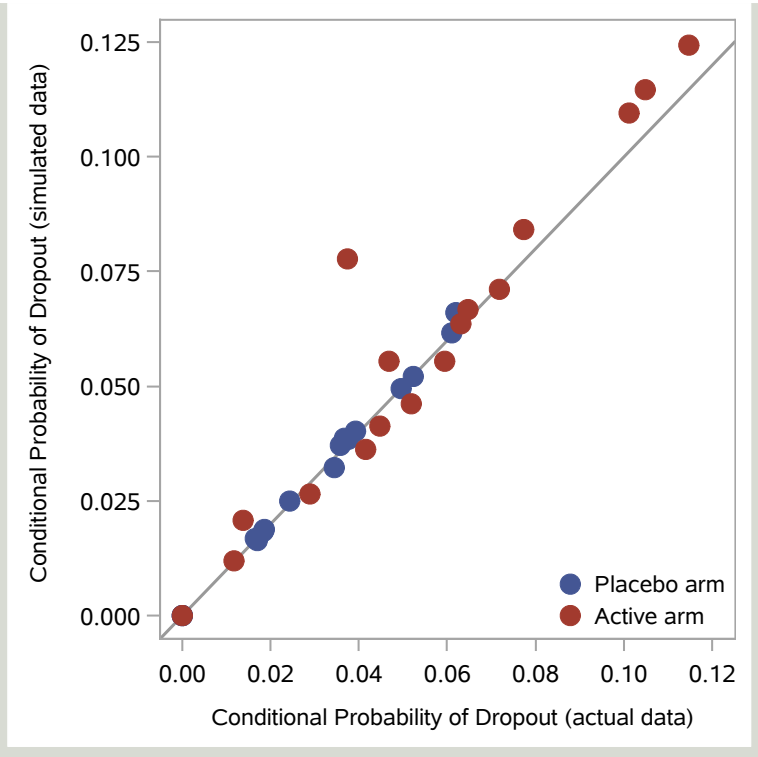
Case Study: SCA-3004 (PBO)

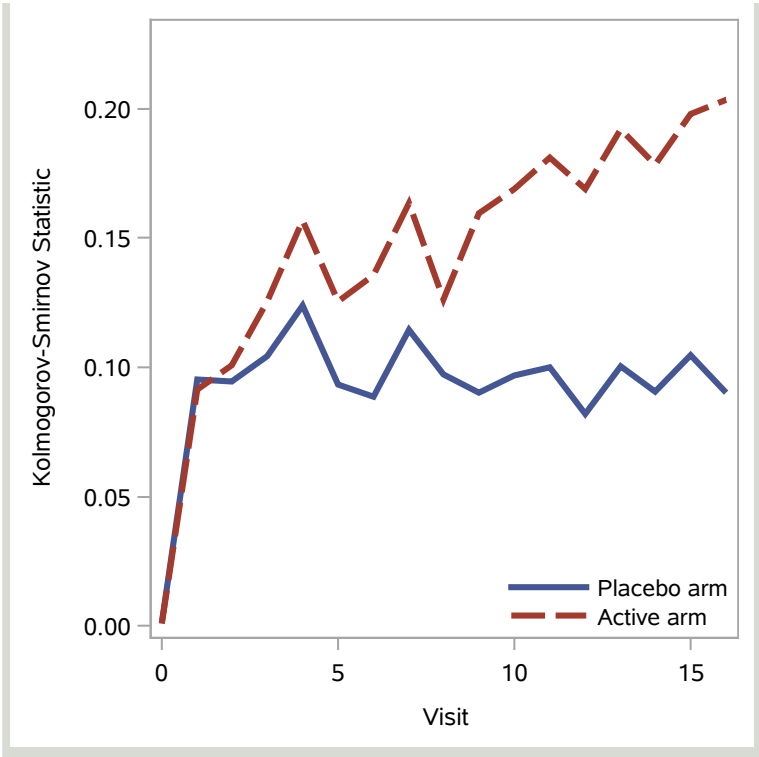


Case Study: SCA-3004 (PP1M)

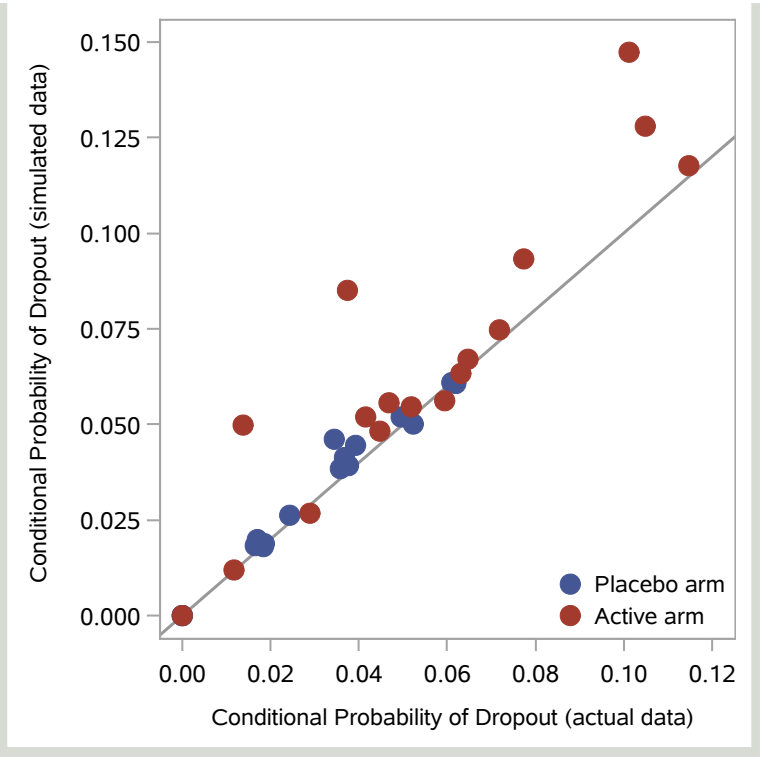


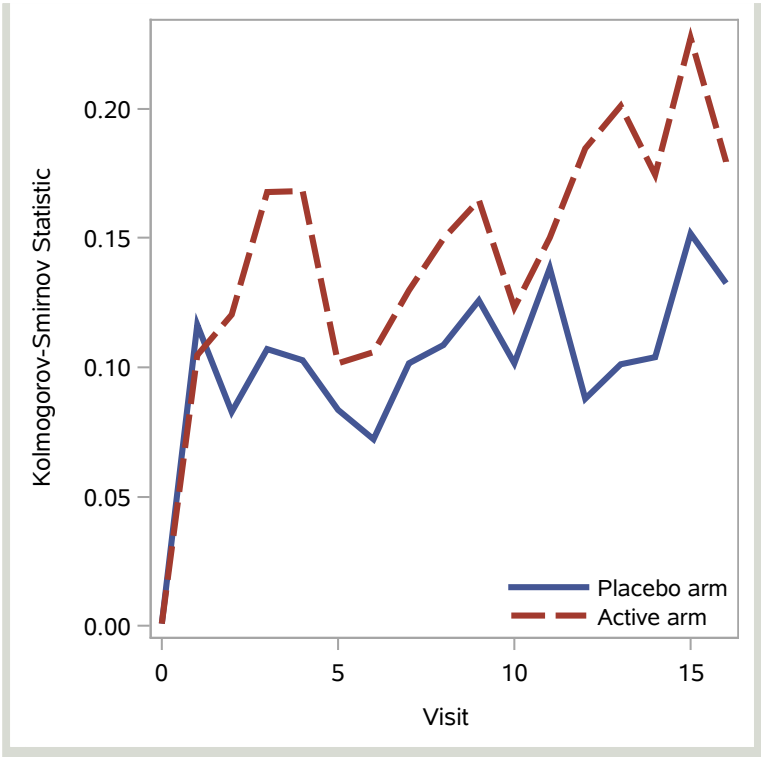
Case Study: SCA-3004



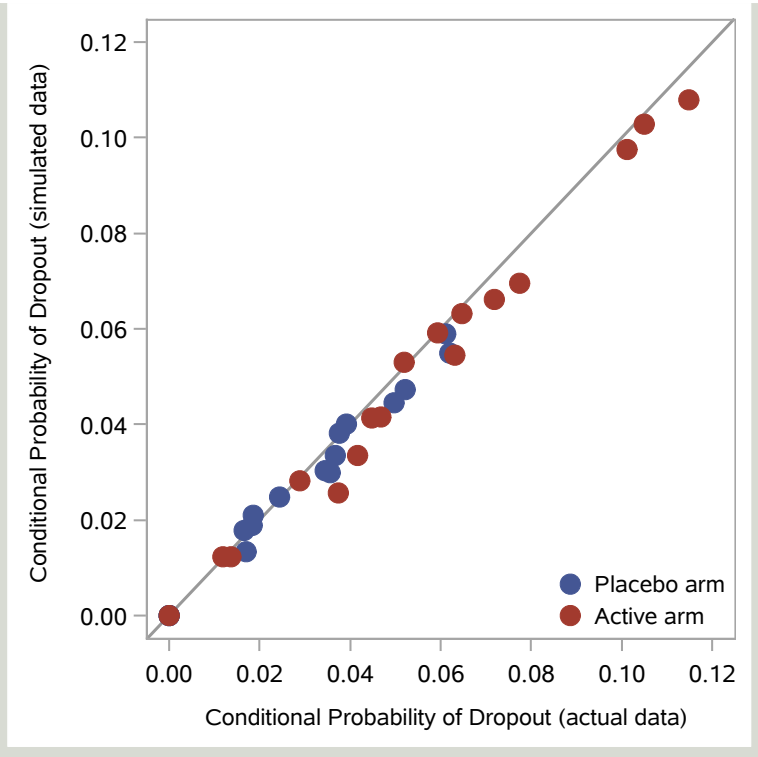


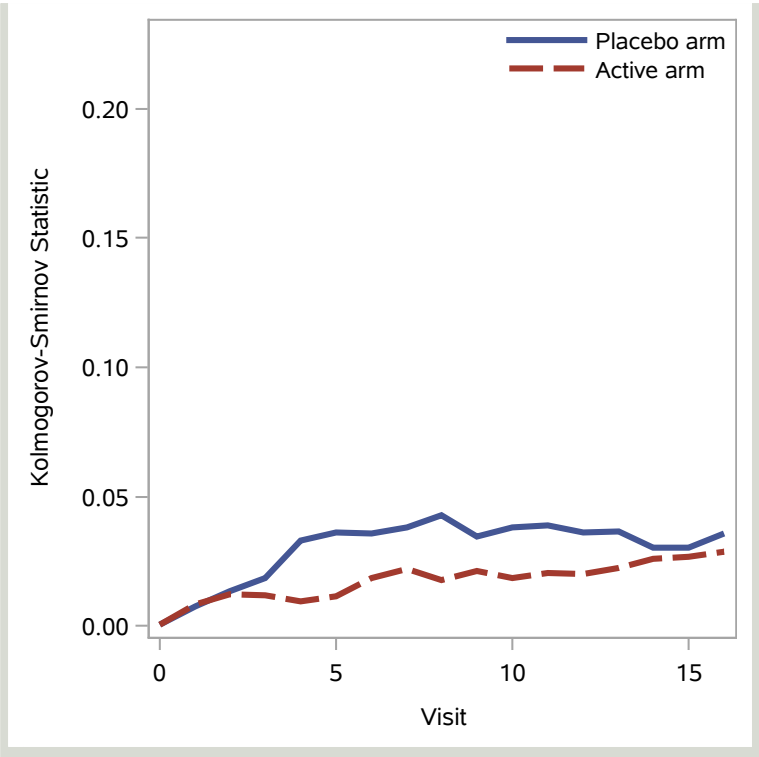
Case Study: SCA-3004





Case Study: SCA-3004

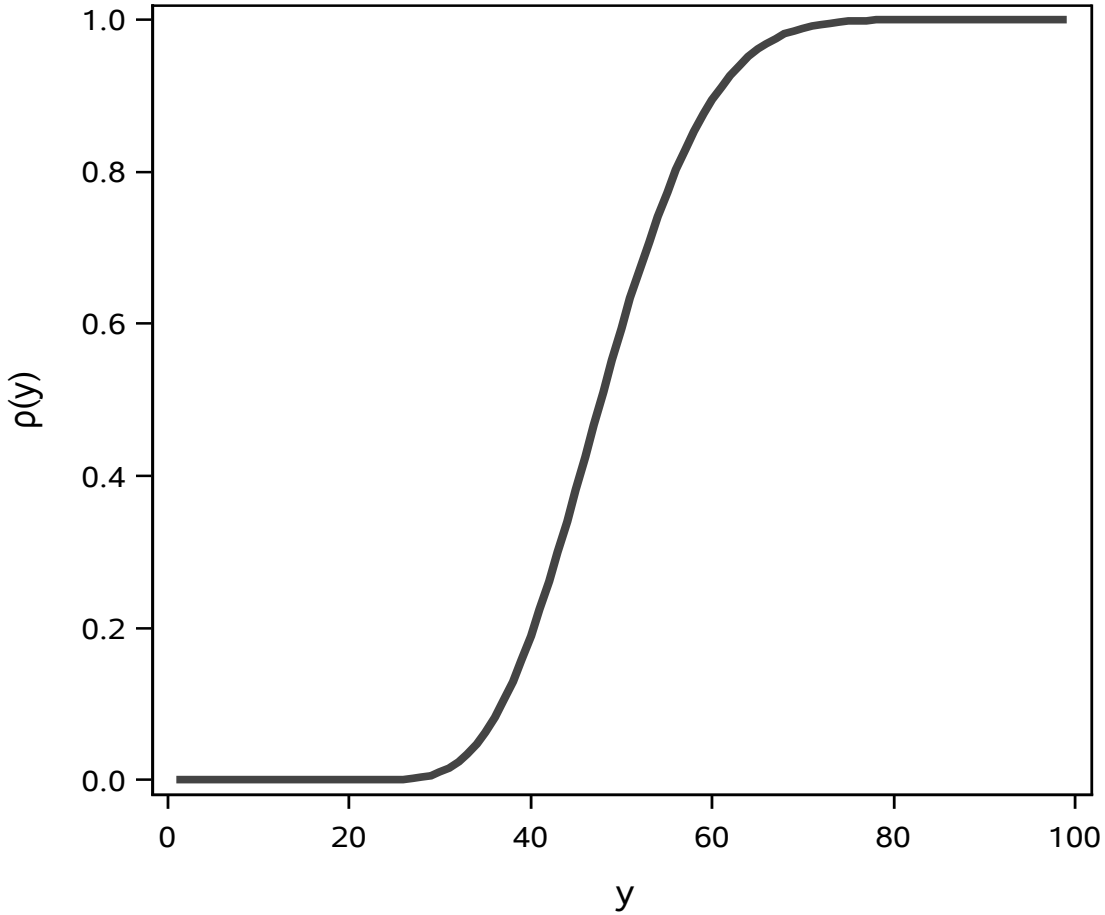




Case Study: SCA-3004

- 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).

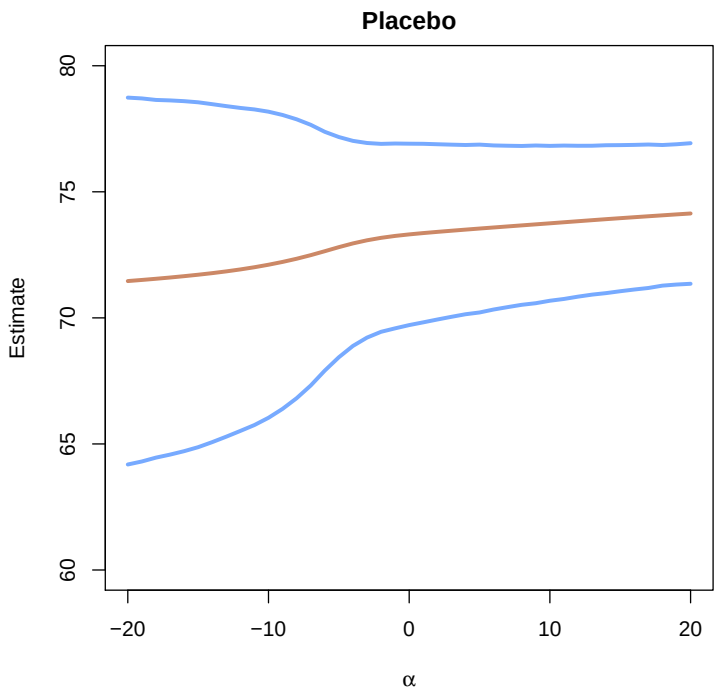
Case Study: SCA-3004

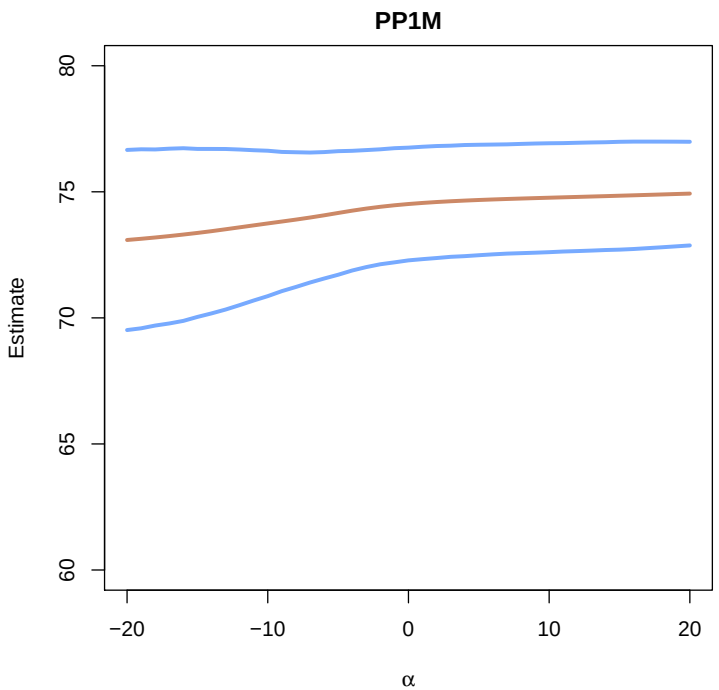


Case Study: SCA-3004

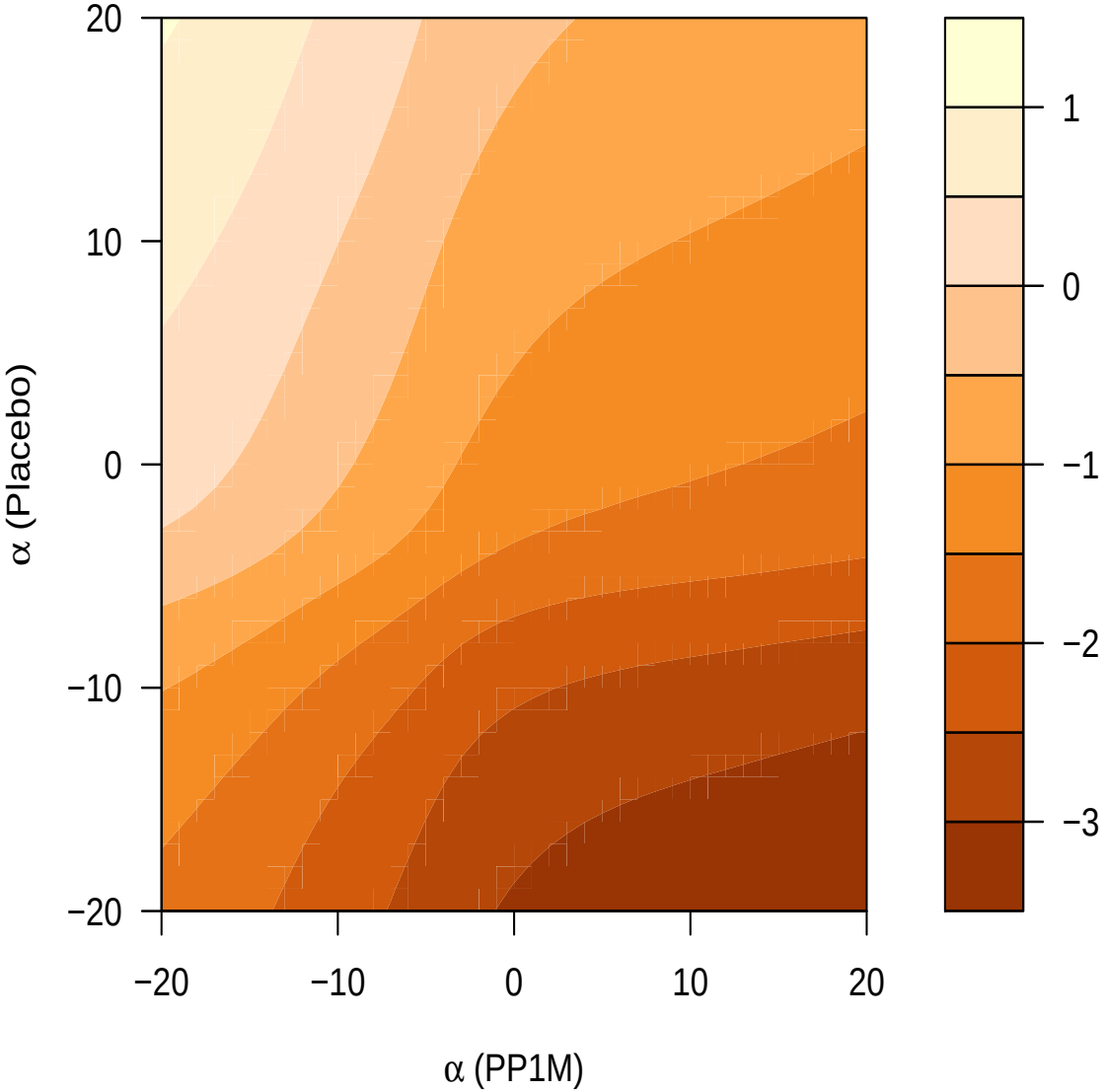
y_{k+1}	y^*_{k+1}	Log Odds Ratio
30	20	$\alpha \times 0.01$
40	30	$\alpha \times 0.18$
50	40	$\alpha \times 0.40$
60	50	$\alpha \times 0.30$
70	60	$\alpha \times 0.09$
80	700	$\alpha \times 0.01$

Case Study: SCA-3004

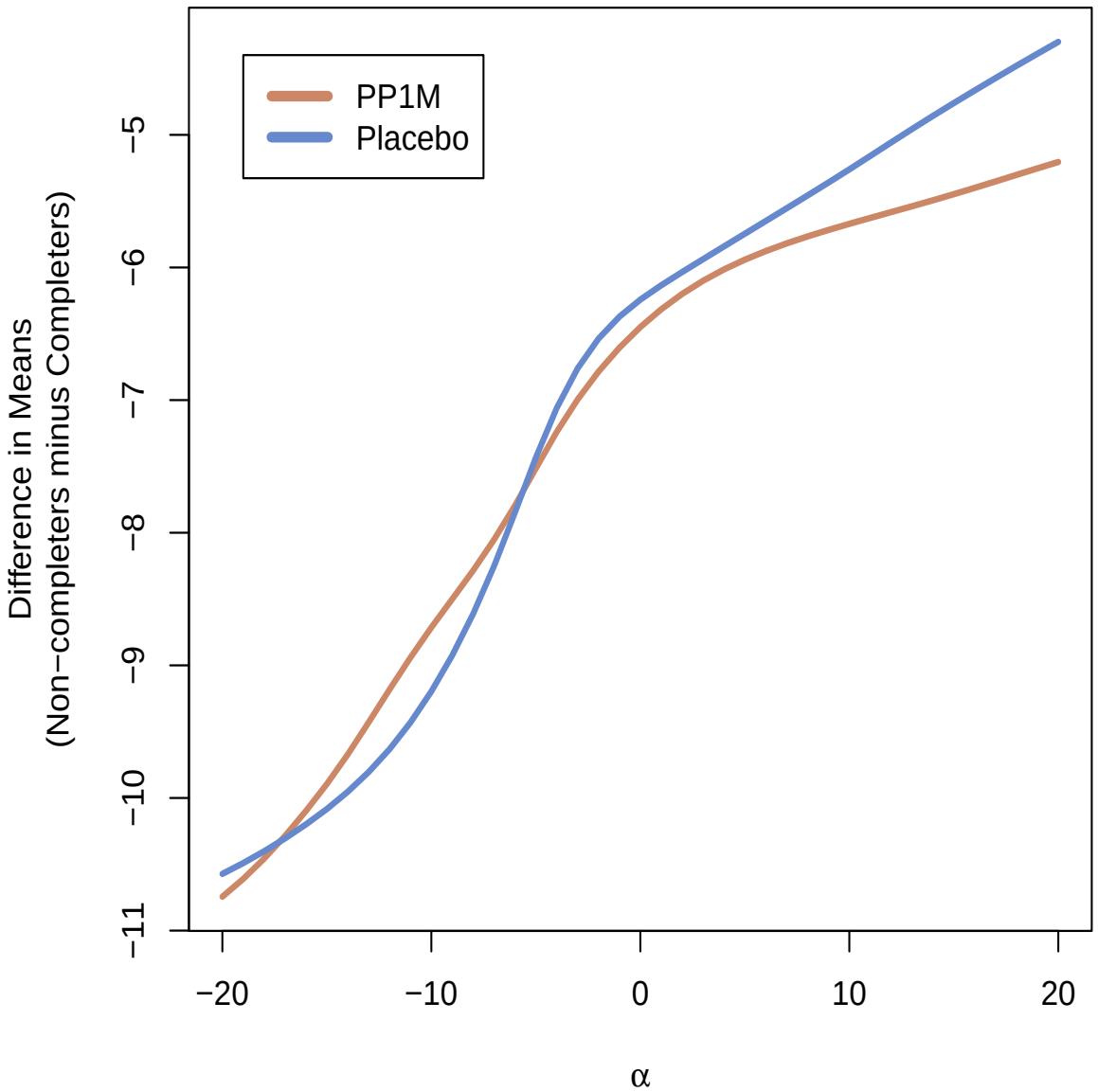




Case Study: SCA-3004



Case Study: SCA-3004



Simulation Study

		PBO			PP1M		
α	Estimator	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%
	Bootstap-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%
	Bootstap-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%
	Bootstap-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%
	Bootstap-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%
	Bootstap-JK-ET	93.8%	94.7%
	Bootstap-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%
	Bootstap-JK-ET	93.9%	94.8%
	Bootstap-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 $0 < k < C$,

$$M_k \perp Y_k \,|\, \overline{Y}_{k-1}, \underline{O}_k \tag{20}$$

- 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 $0 < k < C$,

$$f^*(Y_k|\underbrace{M_k = 1, \overline{Y}_{k-1}, \underline{O}_k}_{\text{Stratum A}}) = f^*(Y_k|\underbrace{M_k = 0, \overline{Y}_{k-1}, \underline{O}_k}_{\text{Stratum B}}) \quad (21)$$

Using Bayes' rule, (21) can be written as follows:

$$P^*(M_k = 1 \mid \overline{Y}_k, \underline{O}_k) = P^*(M_k = 1 \mid \overline{Y}_{k-1}, \underline{O}_k): \quad 0 < k < C. \quad (22)$$

Assumption

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

$$M_k \perp Y_k \mid \rho_k^*(\overline{Y}_{k-1}, \underline{O}_k): \quad 0 < k < C \tag{23}$$

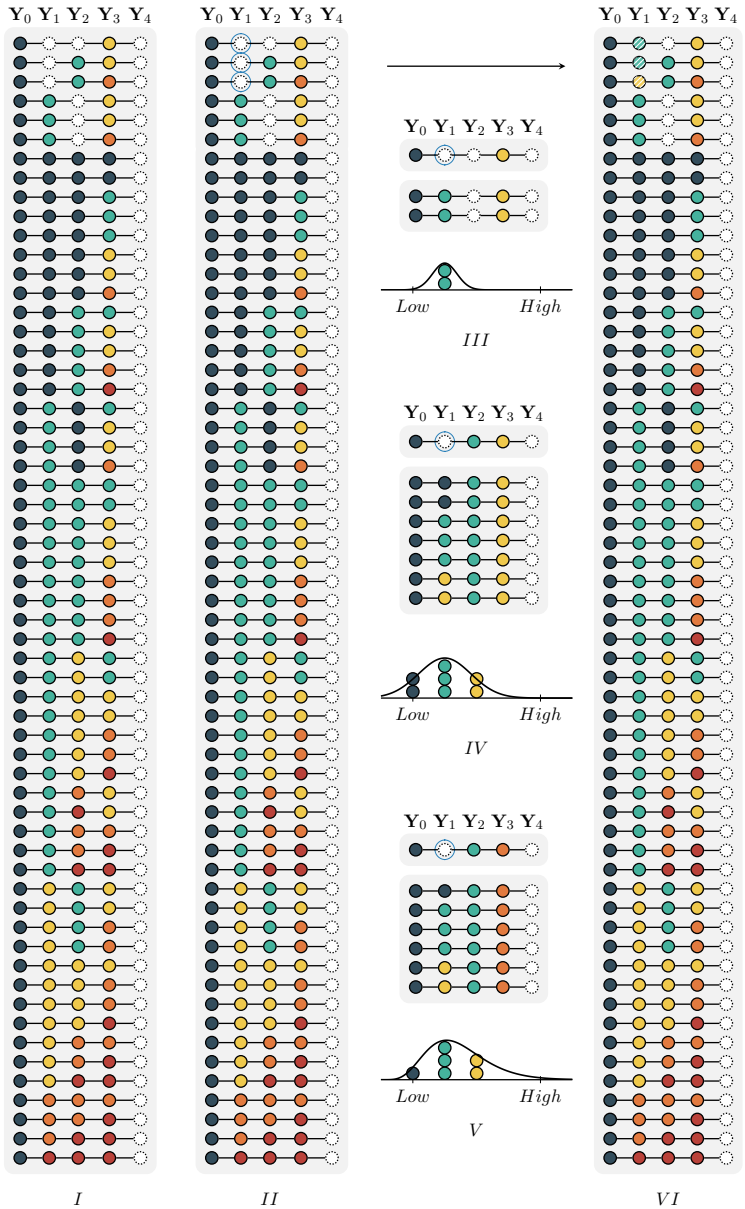
where

$$\rho_k^*(\overline{Y}_{k-1}, \underline{O}_k) = P^*(M_k = 1 \mid \overline{Y}_{k-1}, \underline{O}_k) \tag{24}$$

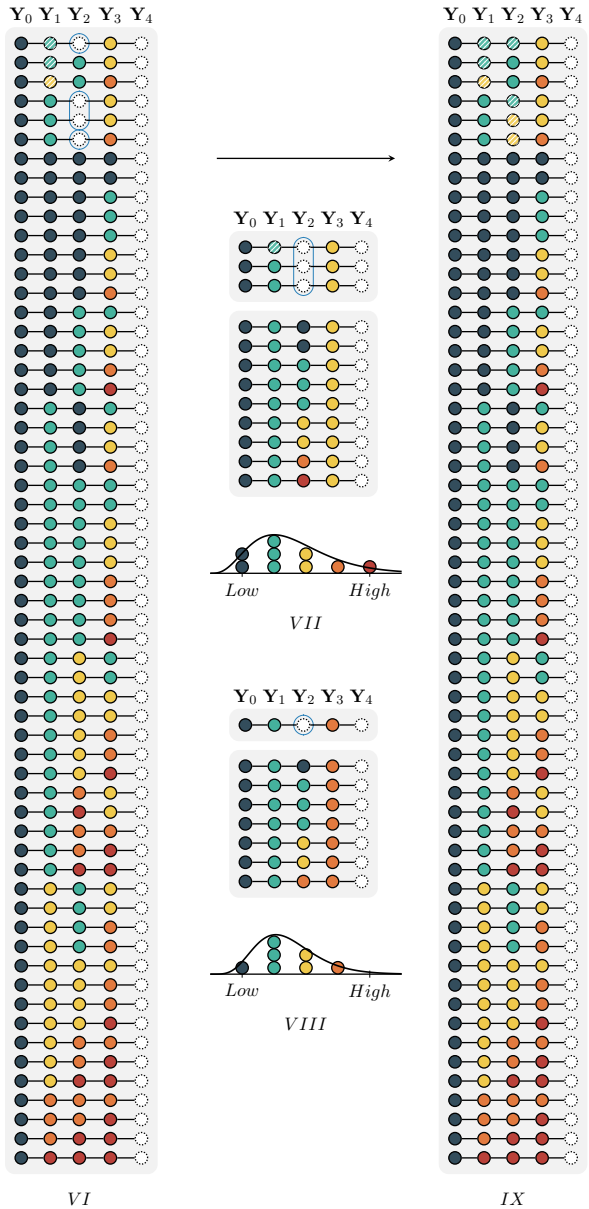
Assumption

Under assumption (20), the joint distribution of $(C, \overline{Y}_C)$ (i.e., the monotonized data) is identified by a recursive algorithm.

Illustration



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 (23) and match on estimates of $\rho_k^*(\overline{Y}_{k-1}, \underline{O}_k)$.

Smoothing Assumptions

$$\text{logit}\{\rho_k^*(\overline{Y}_{k-1}, \underline{O}_k)\} = w_k(\overline{Y}_{k-1}, \underline{O}_k; \nu_k^*); \quad k = 1, \ldots, K-1 \tag{25}$$

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

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

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 $\tilde{\sigma}^2_\alpha$ with Rubin's (1987) multiple imputation variance estimator, i.e.,

$$\tilde{\sigma}^2_\alpha = \frac{1}{M}\sum_{m=1}^{M}\tilde{\sigma}^2_{\alpha,m} + \left(1+\frac{1}{M}\right)\frac{1}{M-1}\sum_{m=1}^{M}(\tilde{\mu}_{\alpha,m}-\tilde{\mu}_\alpha)^2 \quad (26)$$

- In simulations, we have found success using (26) 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.

NP 001 and 002

- 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.

NP 001 and 002

- 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.

NP 001 and 002

- 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.

Missing Data Patterns

```
          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
```

Missing Data Patterns

```
          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
```

Central Question

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

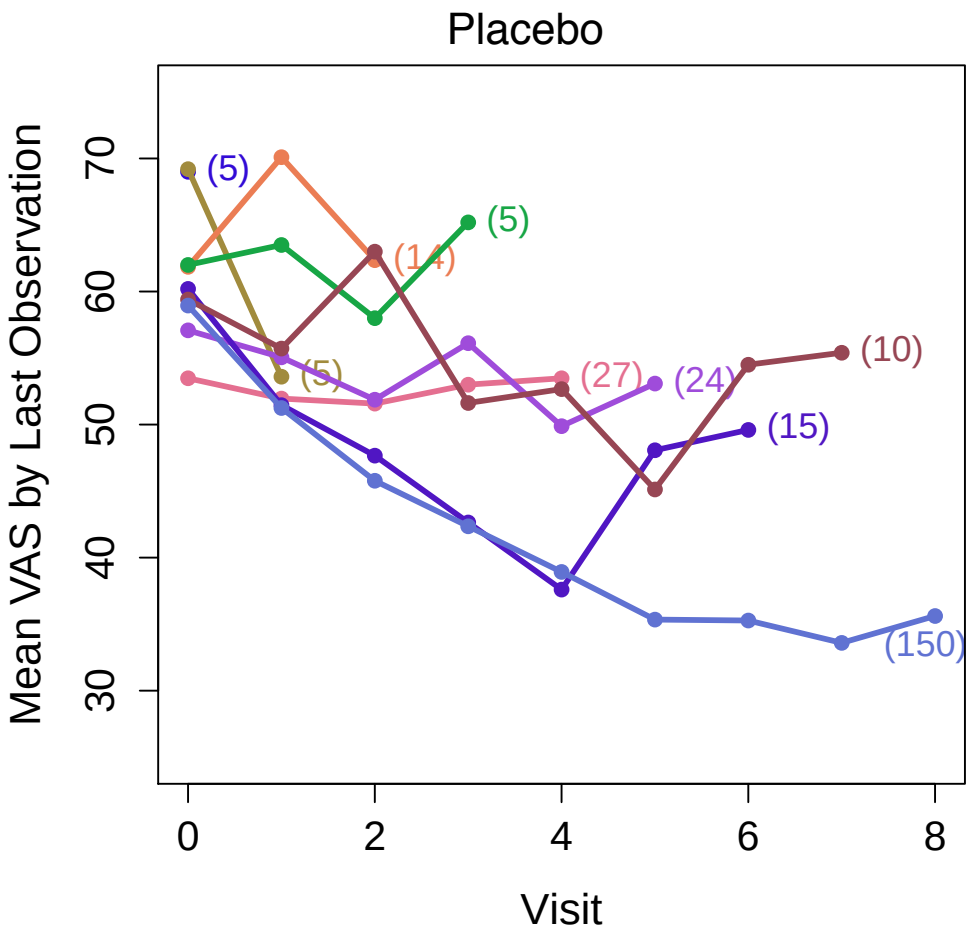
	Placebo				
	Number				
k	On-Study	Last Seen	Obs. Value	Mean	Std. Dev.
0	255	5	255	58.902	19.196
1	250	5	188	53.202	23.048
2	245	14	238	48.899	24.888
3	231	5	186	45.849	23.928
4	226	27	203	42.291	25.338
5	199	24	192	38.896	25.117
6	175	15	162	37.549	25.827
7	160	10	150	35.047	26.313
8	150	150	150	35.613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

Observed Data

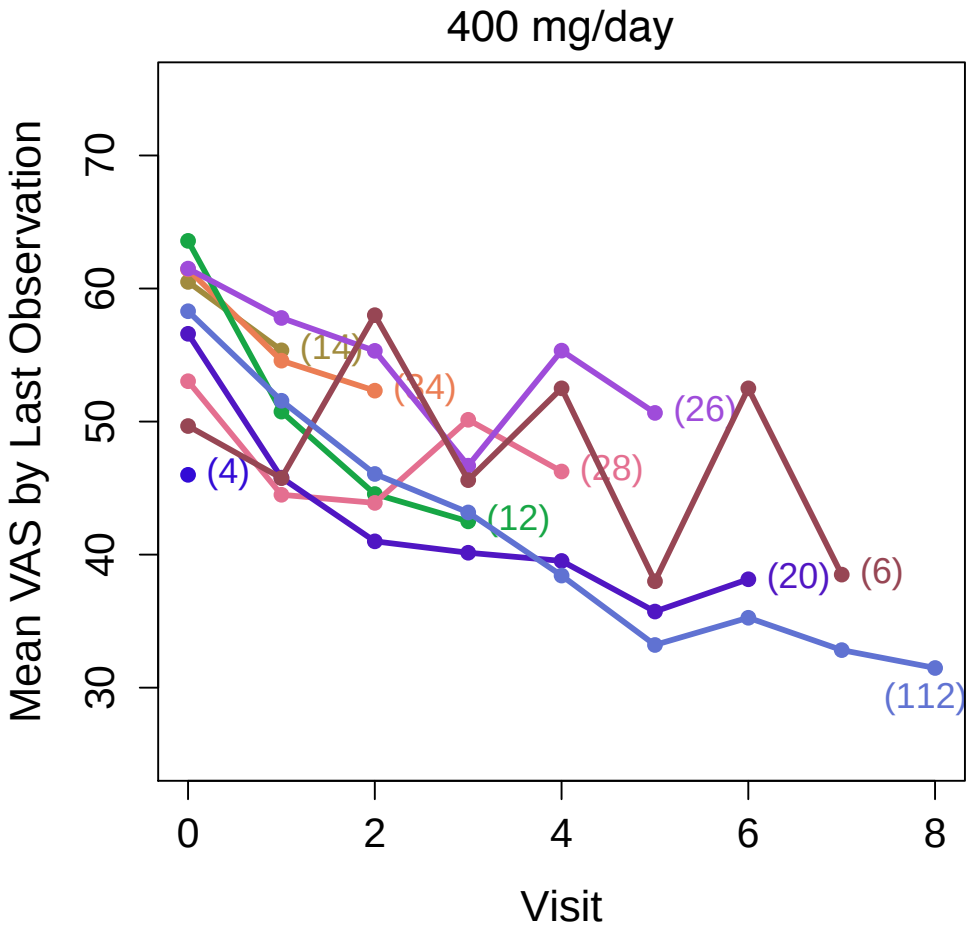
400 mg/day

k	Number On-Study	Number Last Seen	Number Obs. Value	Mean	Std. Dev.
0	256	4	256	58.305	19.958
1	252	14	192	51.297	22.605
2	238	34	223	47.466	25.268
3	204	12	162	44.228	22.956
4	192	28	174	41.879	23.851
5	164	26	159	36.528	24.101
6	138	20	133	36.211	24.334
7	118	6	109	33.138	21.842
8	112	112	112	31.482	22.149

Observed Data



Observed Data



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.

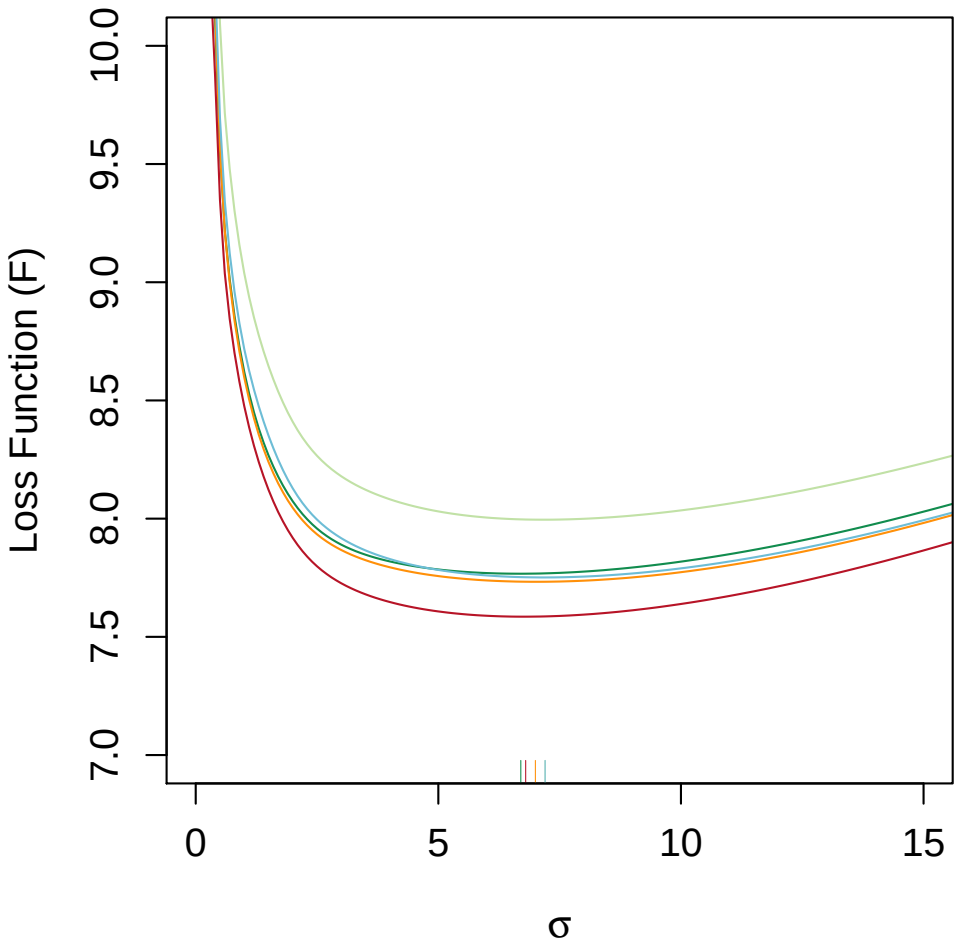
Observed Data

- 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.

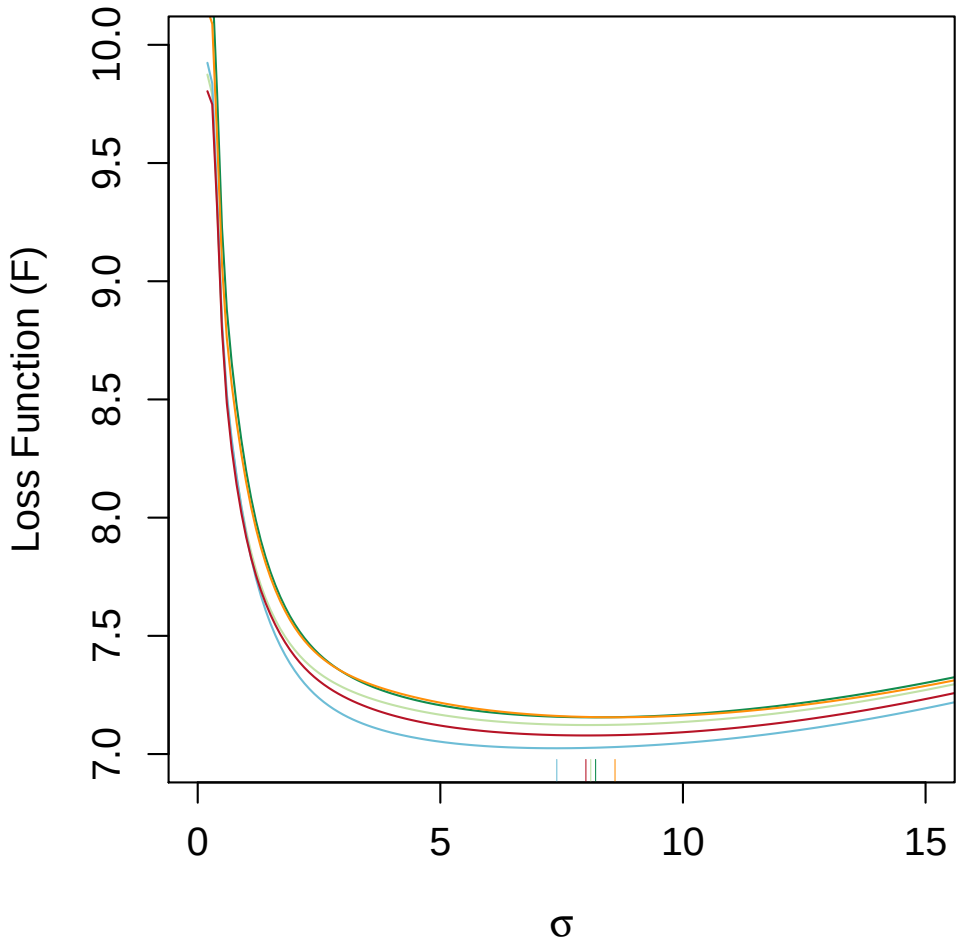
Observed Data

- 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 ($p = 0.18$).

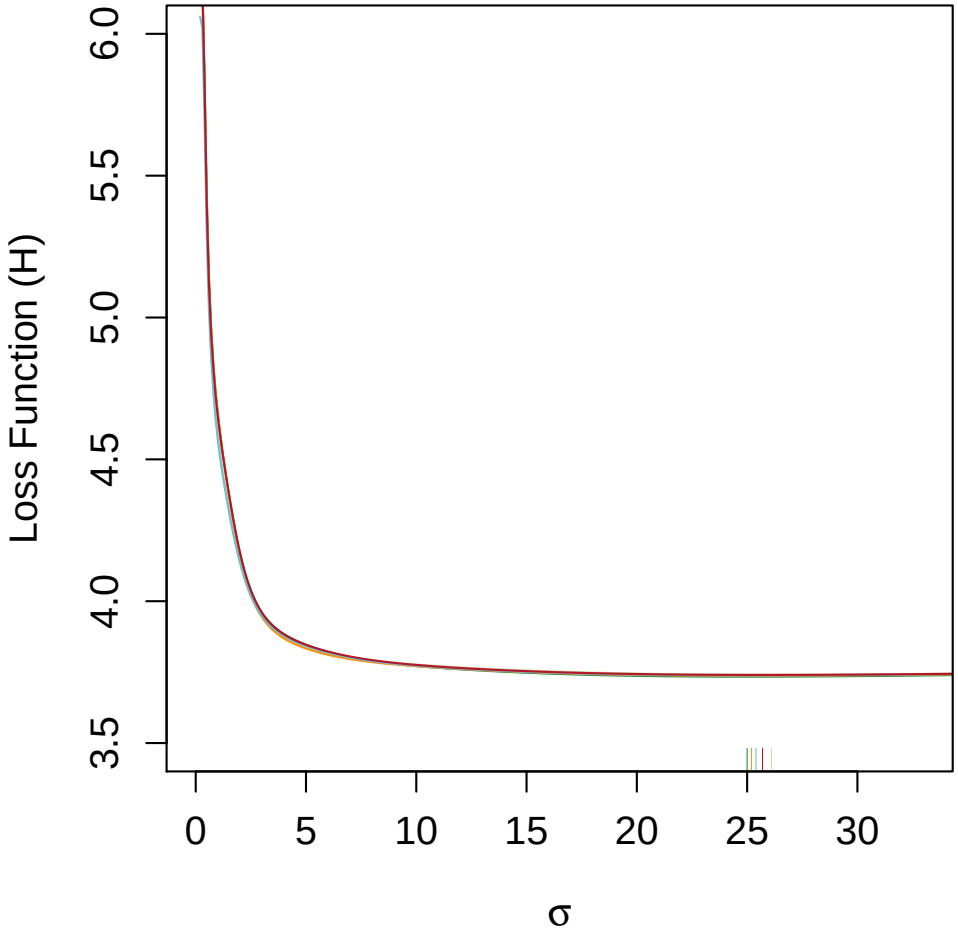
Estimation of Smoothing Parameters - Placebo



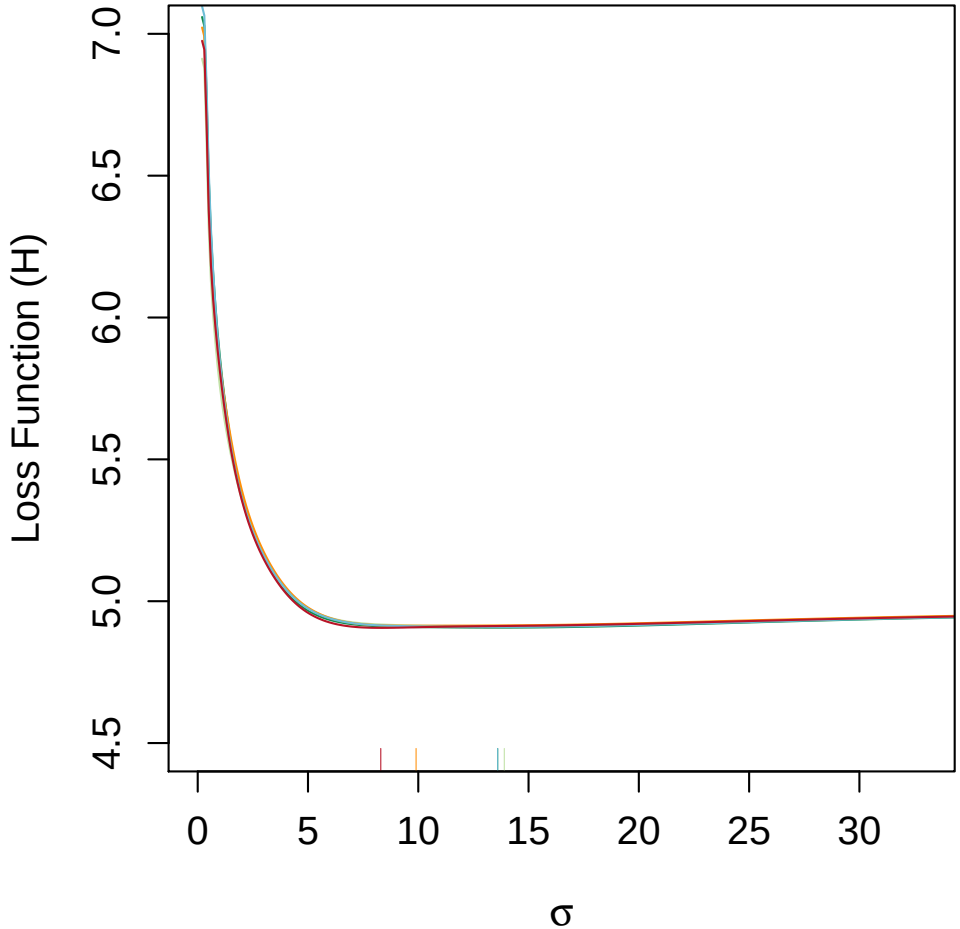
Estimation of Smoothing Parameters - 400 mg



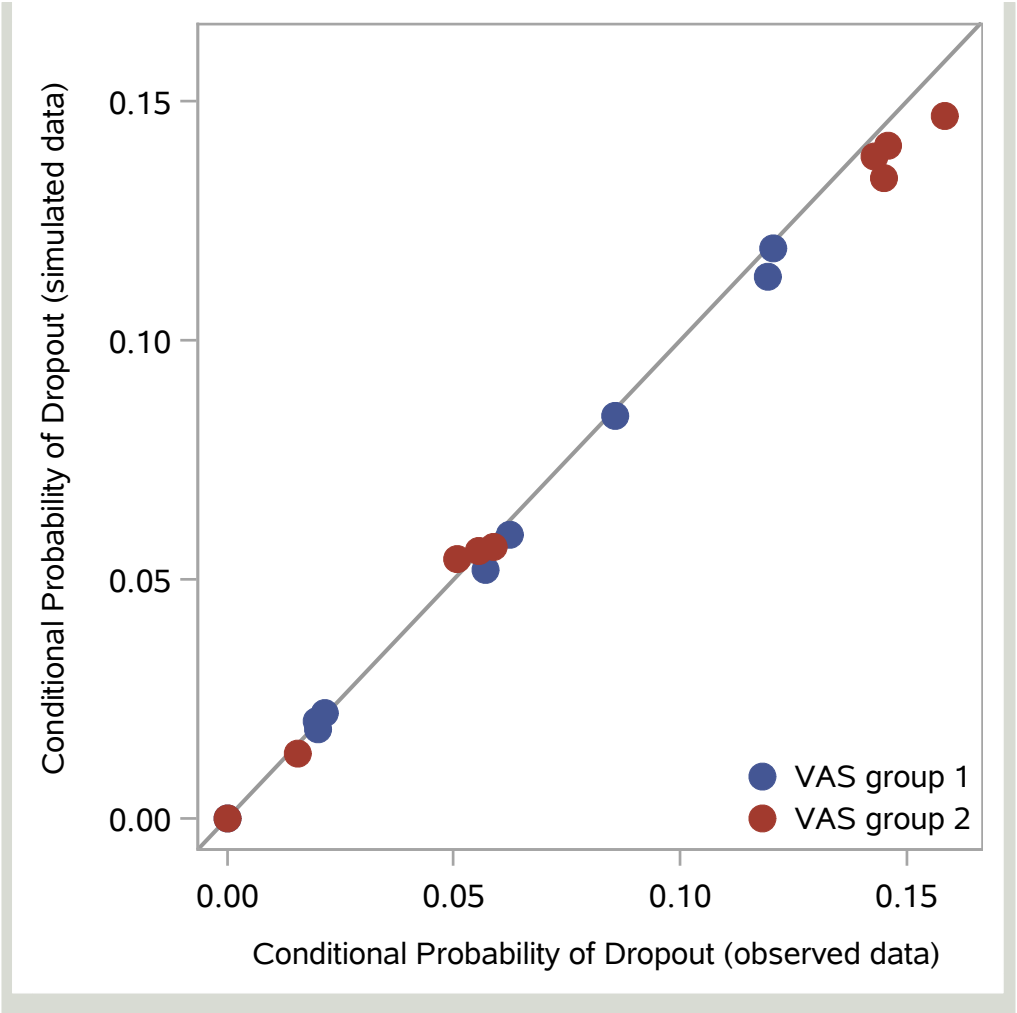
Estimation of Smoothing Parameters - Placebo



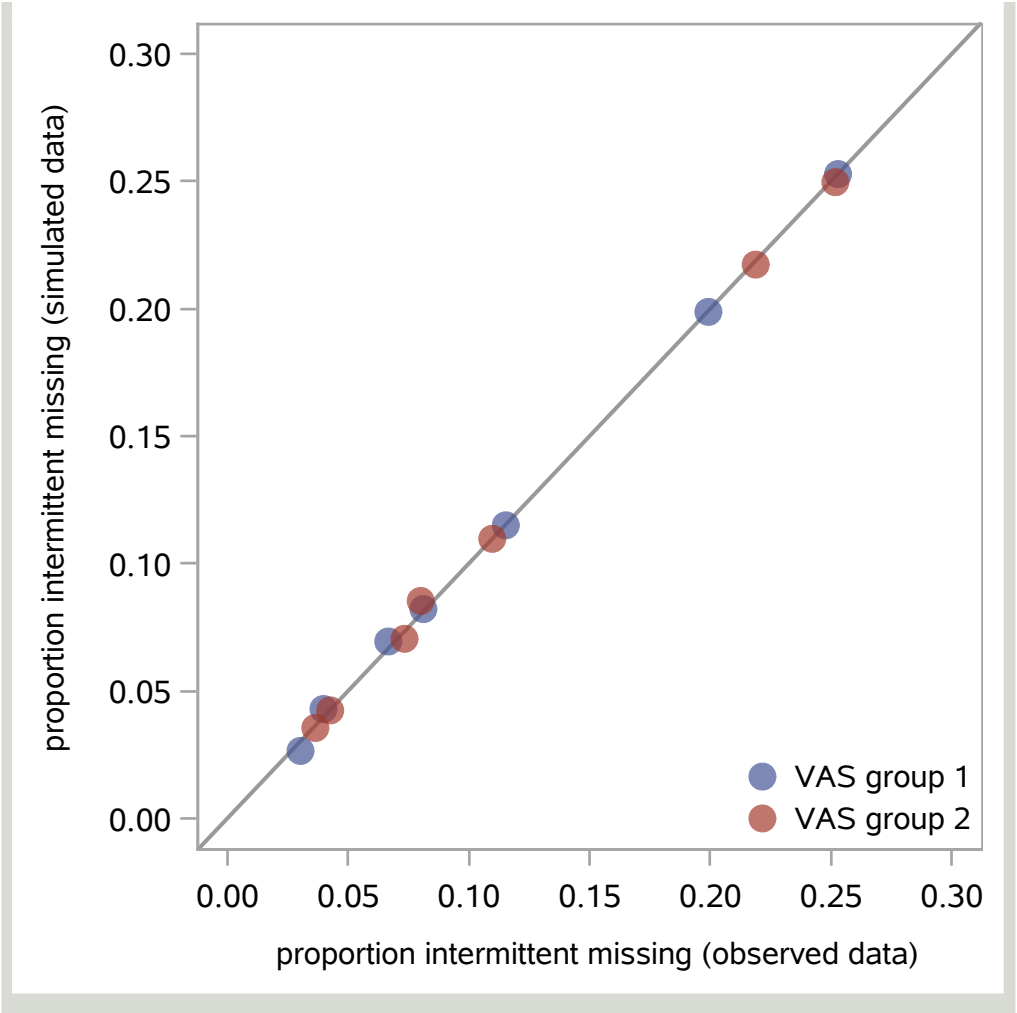
Estimation of Smoothing Parameters - 400 mg



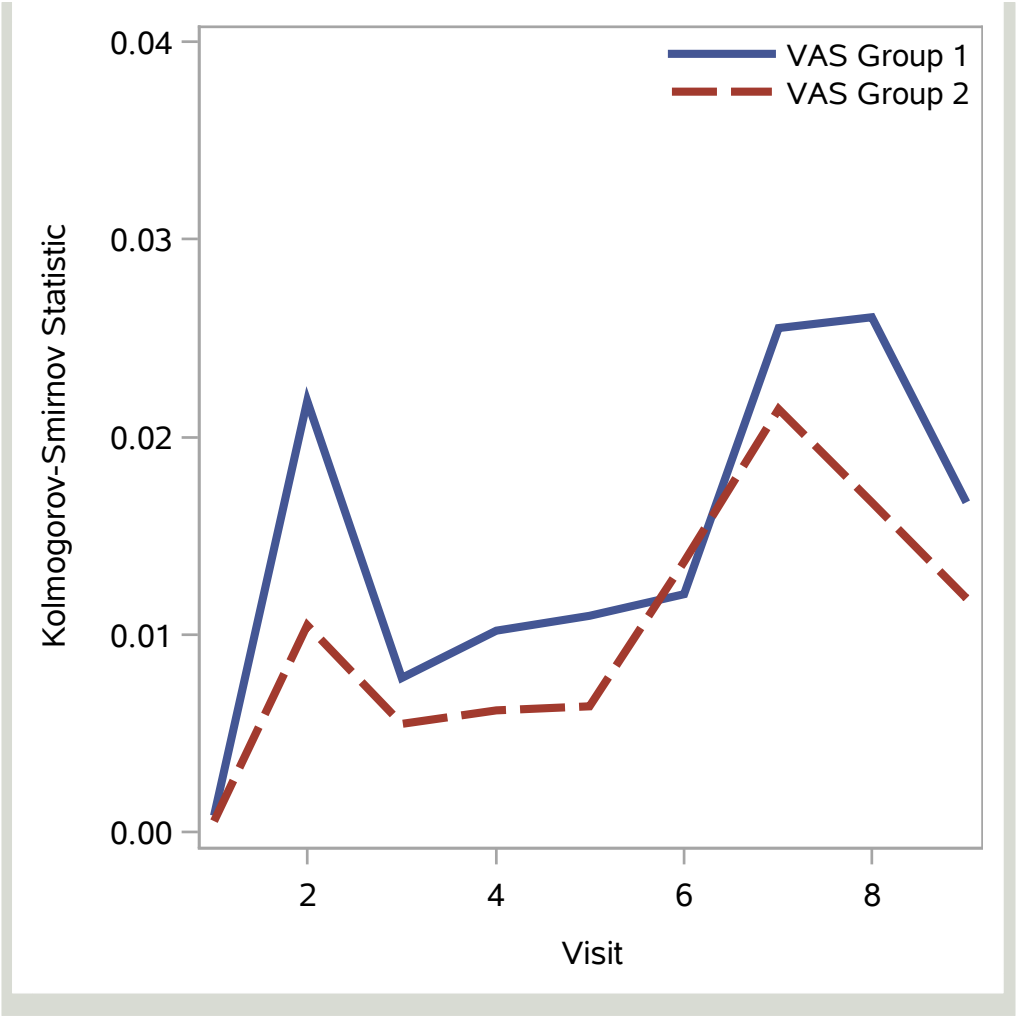
Goodness of Fit



Goodness of Fit



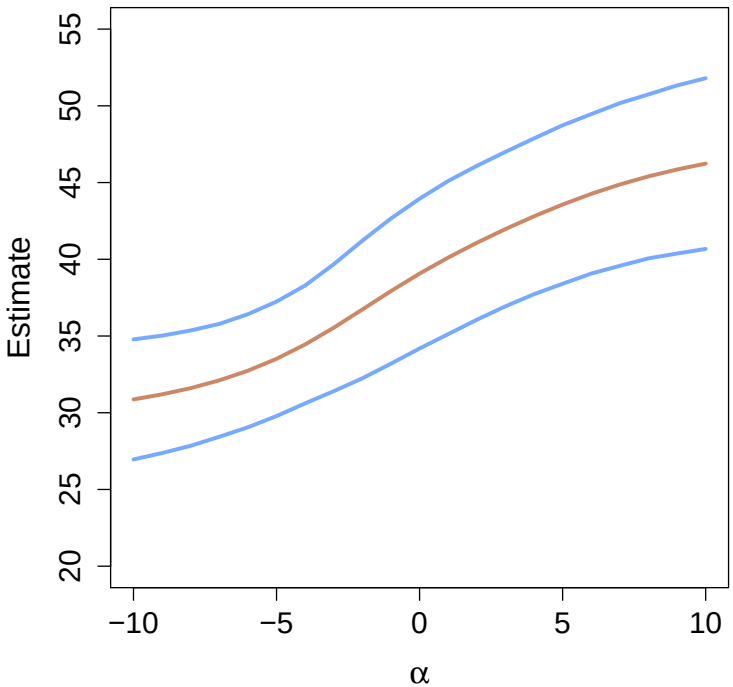
Goodness of Fit

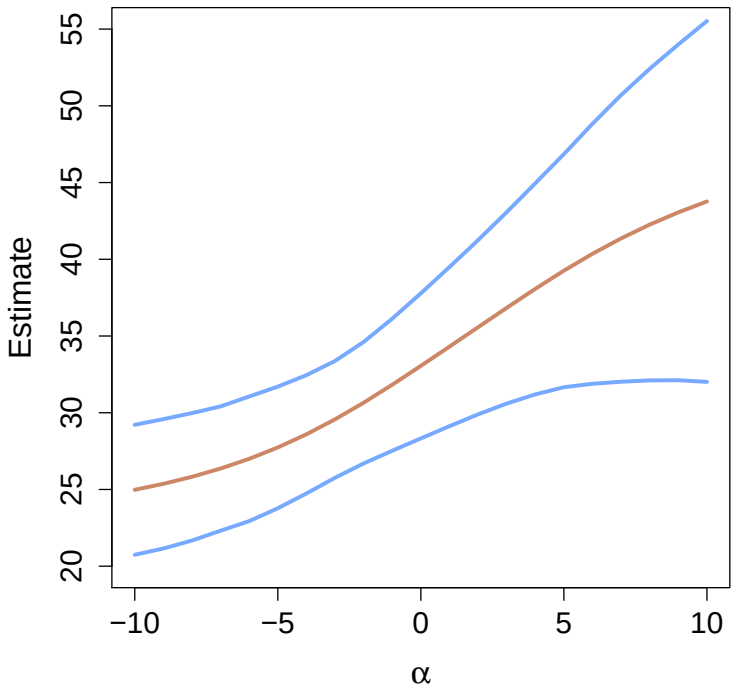


MAR Analysis

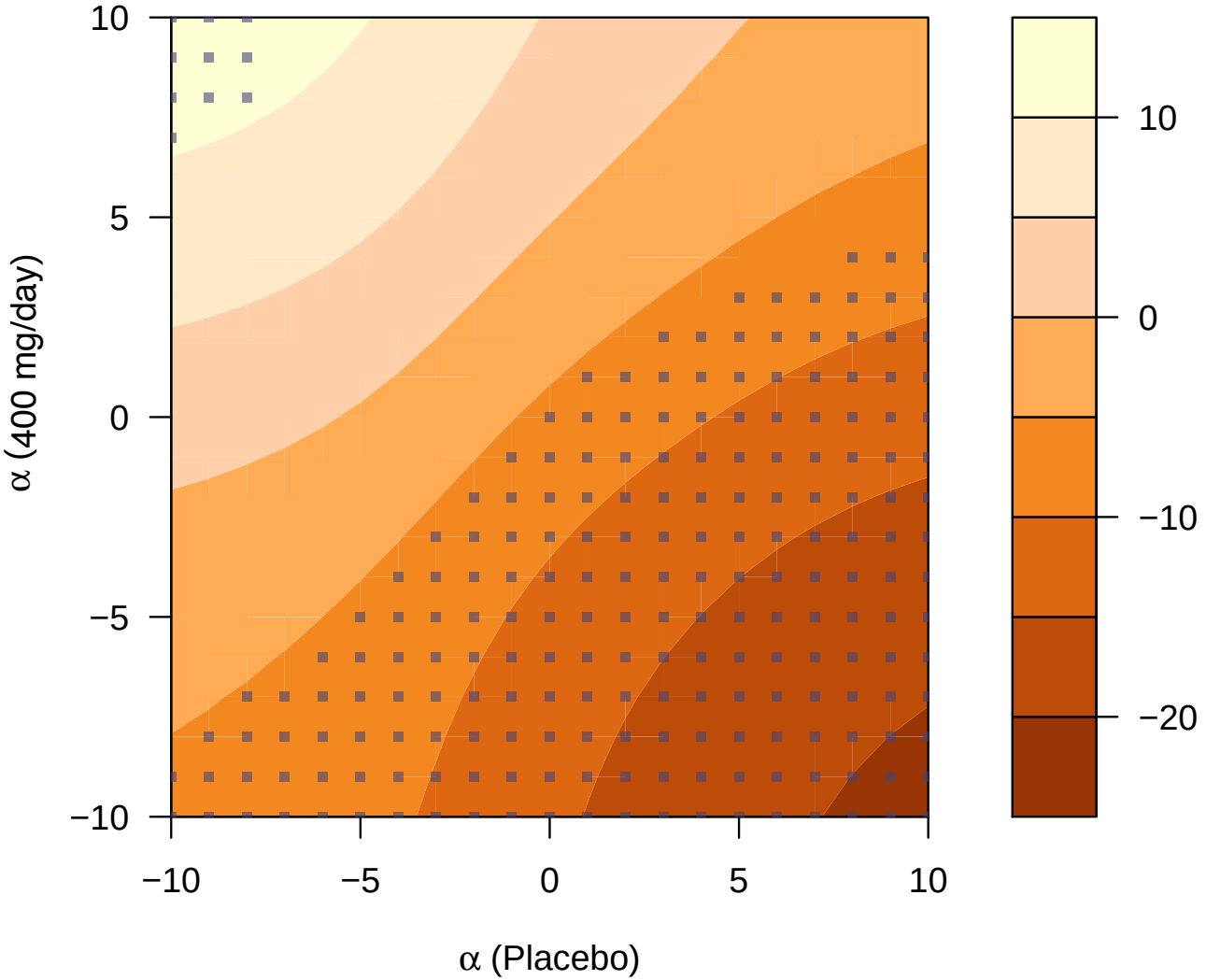
- The estimates of μ^* are 39.07 (95% CI: 34.19 to 43.95) and 33.06 (95% CI: 28.33 to 37.78) in the placebo and 400 mg/day arms, respectively.
- These estimates correct for the fact that individuals with higher VAS scores appear to be dropping out of the study.
- The correction is bigger for placebo 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

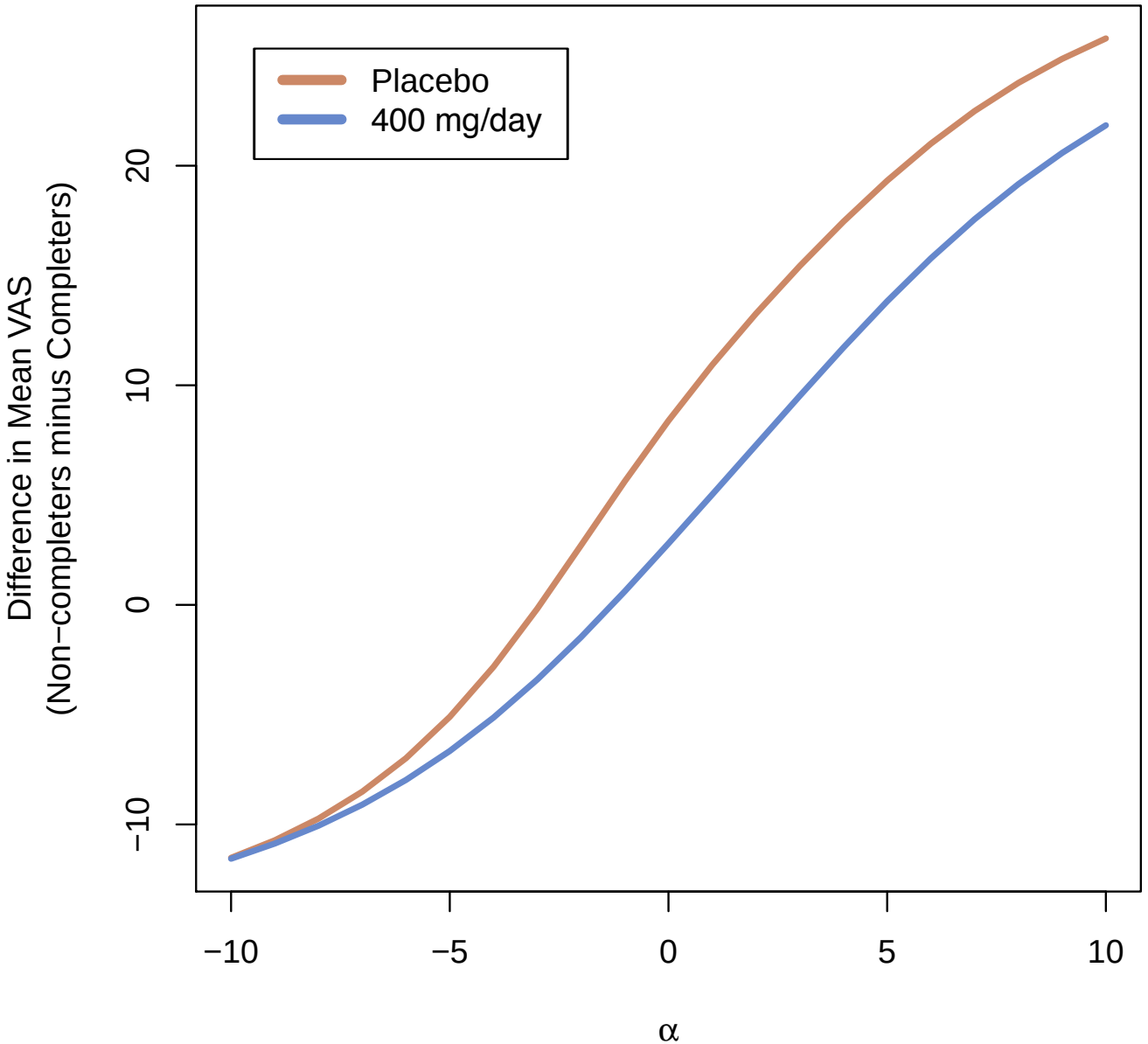




Sensitivity Analysis



Sensitivity Analysis



Simulation Study - Five Imputes

		PBO			400 mg/day		
α	Estimator	μ^*	Bias	MSE	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

samon 4.0 - the software

May 9, 2017

Two Software choices

1. R
   - samon library
   - functions with pass to C code
2. SAS
   - procedures and macros

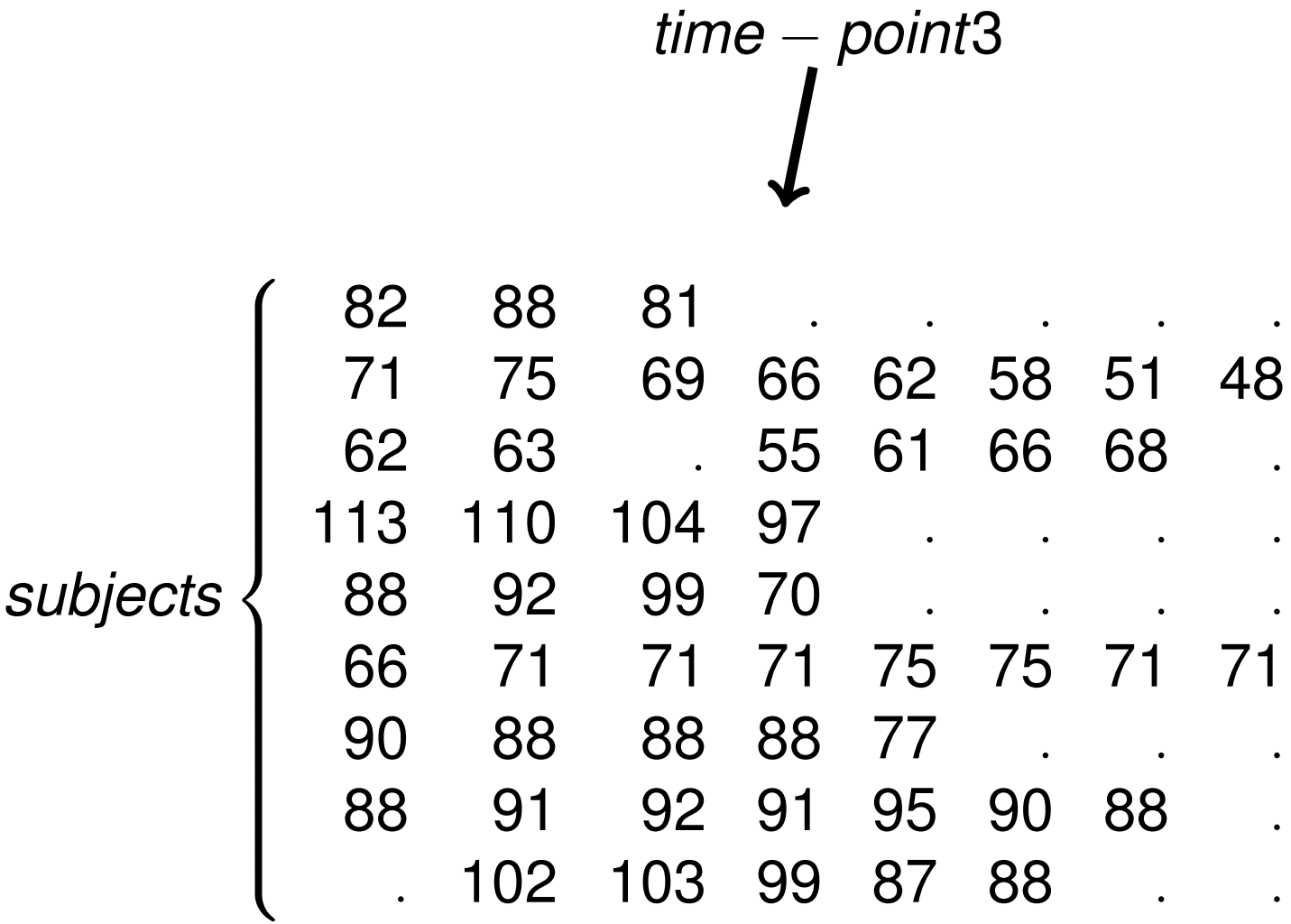
Background

- Randomized study with outcome measurements taken at fixed time-points
- Monotone missing data pattern
- Interest is in a comparison of treatment specific means at the last scheduled time-point
- Rows indicate individuals and columns indicate time-points
- Data at the first time-point (the baseline) is never missing

Missing Data

- Two patters of missing data are considered:
- Monotone missing data
- Intermittent missing data

Background


time – point3
subjects
82 88 81 . . . . .
71 75 69 66 62 58 51 48
62 63 . 55 61 66 68 .
113 110 104 97 . . . .
88 92 99 70 . . . .
66 71 71 71 75 75 71 71
90 88 88 88 77 . . .
88 91 92 91 95 90 88 .
. 102 103 99 87 88 . .

Case Study: Chronic Schizophrenia

- Patients scheduled for 5 post-baseline assessements at weeks 1,2,4,6, and 8.
- Primary efficiacy variable: PANSS score (positive and negative syndrome scale)
- Two treatment groups: placebo (treatment 1) with risperidone 6mg (treatment 2).
- Data are simulated from this trial but retain many of the original freatures.

PANSS Analysis (Monotone missing data)

- Descriptive stats for each treatment group and check the monotone missing condition.
- Decide on a sensitivity function.
- Use the SAMON procedure to estimate means.
- Compare results.

The PANSS Dataset

```
proc print data=PANSS1;
run;

Obs     V1     V2     V3     V4     V5     V6

  1     90     87     86     93     72     87
  2    112      .      .      .      .      .
  3     99     76     62     52     57     49
  4     86     78     91    113     89     68
  5     80     85      .      .      .      .
  6     72     64     78    113      .      .
  7     67      .      .      .      .      .
  8     96      .      .      .      .      .
  9     93     90      .      .      .      .
 10     78     70     53     85      .      .
 11     93     86     92     94      .      .
```

The samonDataCheck macro

- The samonDataCheck macro can be used to check data to ensure it is in samon canonical form.

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

samonDataCheck

```
%samonDataCheck(
    data        = data.panss1,
    vars        = v1 - v6,
    out         = panss1,
    stats       = stats1,
    mpattern    = mpattern1
    );
proc print data = stats1 label noobs;
run;
```

samonDataCheck

Time	N	Nim	Mean	Std. Dev.	Min	Max
1	88	0	91.443	18.013	56	132
2	80	0	87.188	19.488	44	153
3	70	0	85.171	17.771	53	125
4	45	0	83.556	19.484	52	120
5	30	0	83.933	21.571	52	144
6	23	0	78.261	19.495	47	111

samonDataCheck

```
proc print data = mpattern1 label noobs;
run;

Missing
pattern     Count     Percent

******         23     26.1364
*****_          7      7.9545
****__         15     17.0455
***___         25     28.4091
**____         10     11.3636
*_____          8      9.0909
```

samonDataCheck

```
%samonDataCheck(
      data        = data.panss2,
      vars        = v1 - v6,
      out         = panss2,
      stats       = stats2,
      mpattern    = mpattern2
      );

proc print data = stats2 label noobs;
run;
```

samonDataCheck

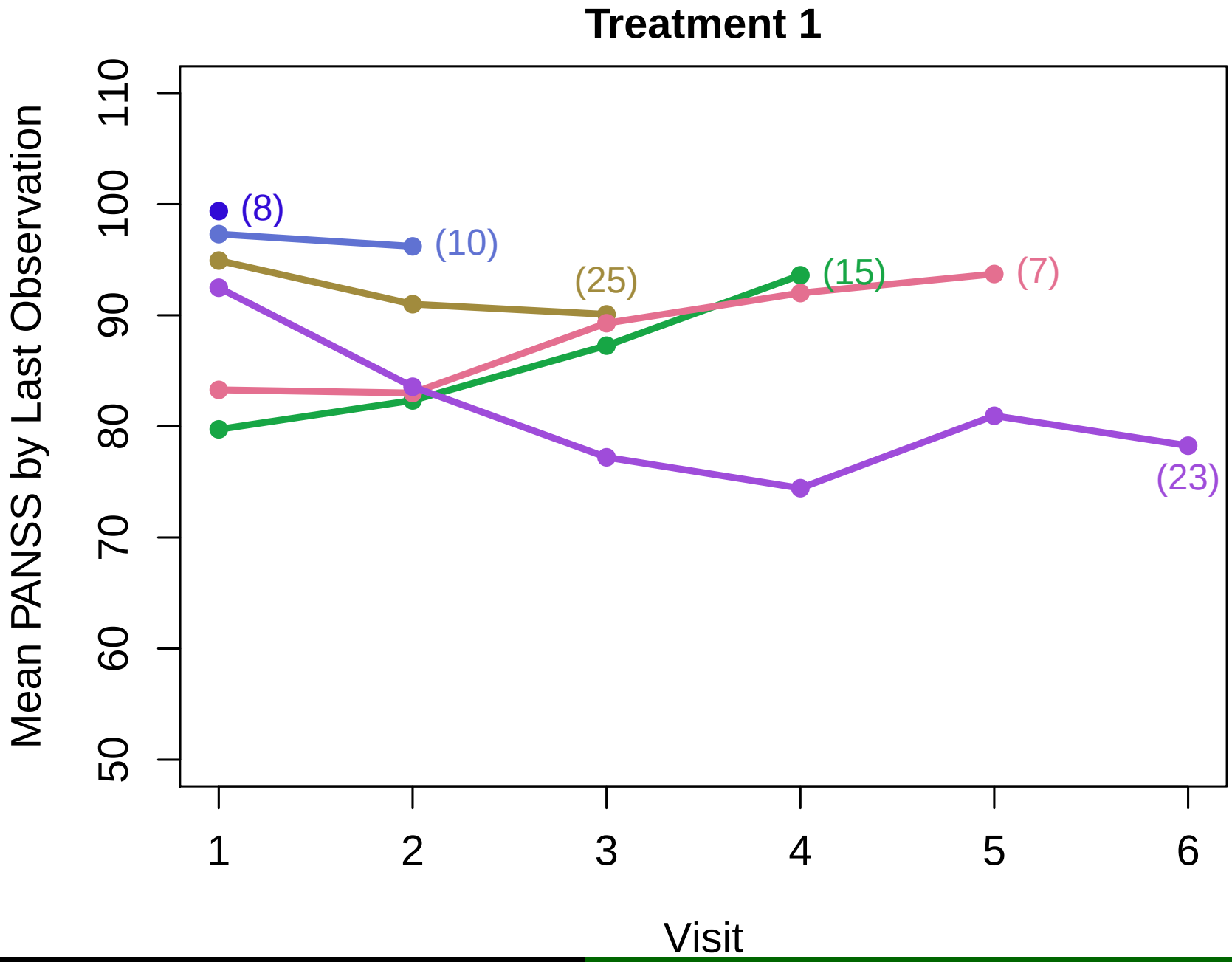
Time	N	Nim	Mean	Std. Dev.	Min	Max
1	86	0	89.849	18.891	54	135
2	81	0	77.914	17.174	47	120
3	77	0	75.649	18.541	42	119
4	68	0	74.721	18.264	38	118
5	53	0	70.547	21.182	38	107
6	51	0	68.628	20.434	37	114

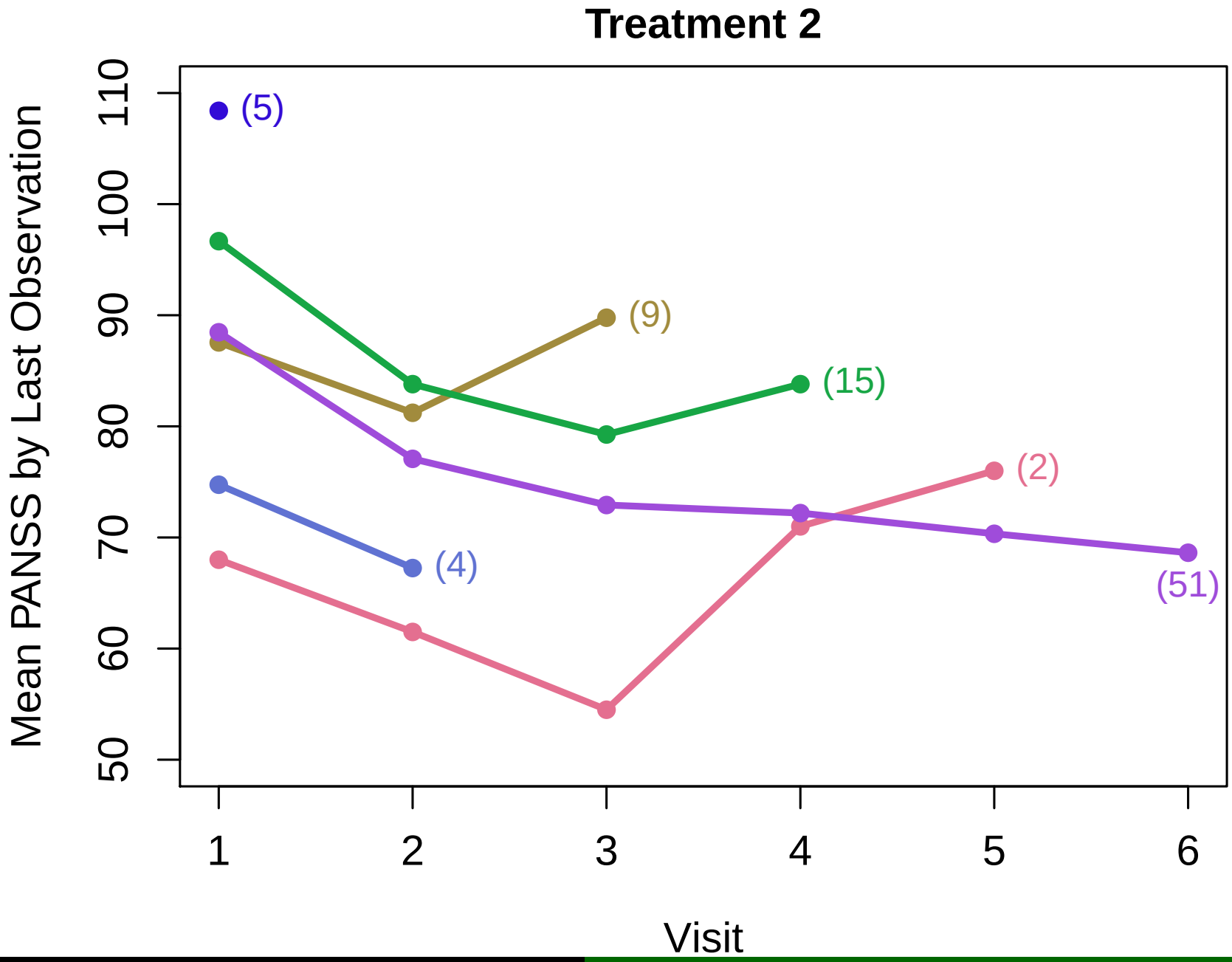
samonDataCheck

```
proc print data = mpattern2 label noobs;
run;

Missing
pattern     Count     Percent

******         51     59.3023
*****_          2      2.3256
****__         15     17.4419
***___          9     10.4651
**____          4      4.6512
*_____          5      5.8140
```


Treatment 1
Mean PANSS by Last Observation
50
60
70
80
90
100
110
1
2
3
4
5
6
Visit
(8)
(10)
(25)
(15)
(7)
(23)


Treatment 2
Mean PANSS by Last Observation
110
100
90
80
70
60
50
1
2
3
4
5
6
Visit
(5)
(9)
(15)
(2)
(4)
(51)

notation

- Consider each treatment group separately.
- Let K denote the number of time-points after baseline and n denote the number of individuals in a treatment group.
- $R_{t,j}$ indicates if individual j is on study at time-point t, $1 \leq j \leq n$ and $0 \leq t \leq K$. That is $R_{t,j} = 1$ if individual j is on study at time t and $R_{t,j} = 0$ otherwise.
- If $R_{t,j} = 1$ then $Y_{t,j}$ denotes individual j's outcome at time t.

notation

- We sometimes wish to refer to the condition of being on-study without specifying which individual is involved. In this case we drop the subscript j, so that, for example, $Prob(R_t = 1)$ refers to the probability of being on study at time-point t.
- In a similar fashion Y_t denotes an outcome value at time t.

Two natural questions

1. If an individual is on-study at time $t-1$, what is the probability of them staying on-study at time t? This probability may depend on the value Y_{t-1}. This leads to our first model:

   $$Prob[R_t = 1 \mid R_{t-1} = 1, Y_{t-1} = y] \sim smooth(y; \sigma)$$

   The smooth function of y depends on a single smoothing parameter σ.

Two natural questions

2. If individuals are on-study at time $t-1$ and remain on-study at time t, what is the distribution of their y values at time t? Again this distribution may depend on the value Y_{t-1}. This leads to our second model:

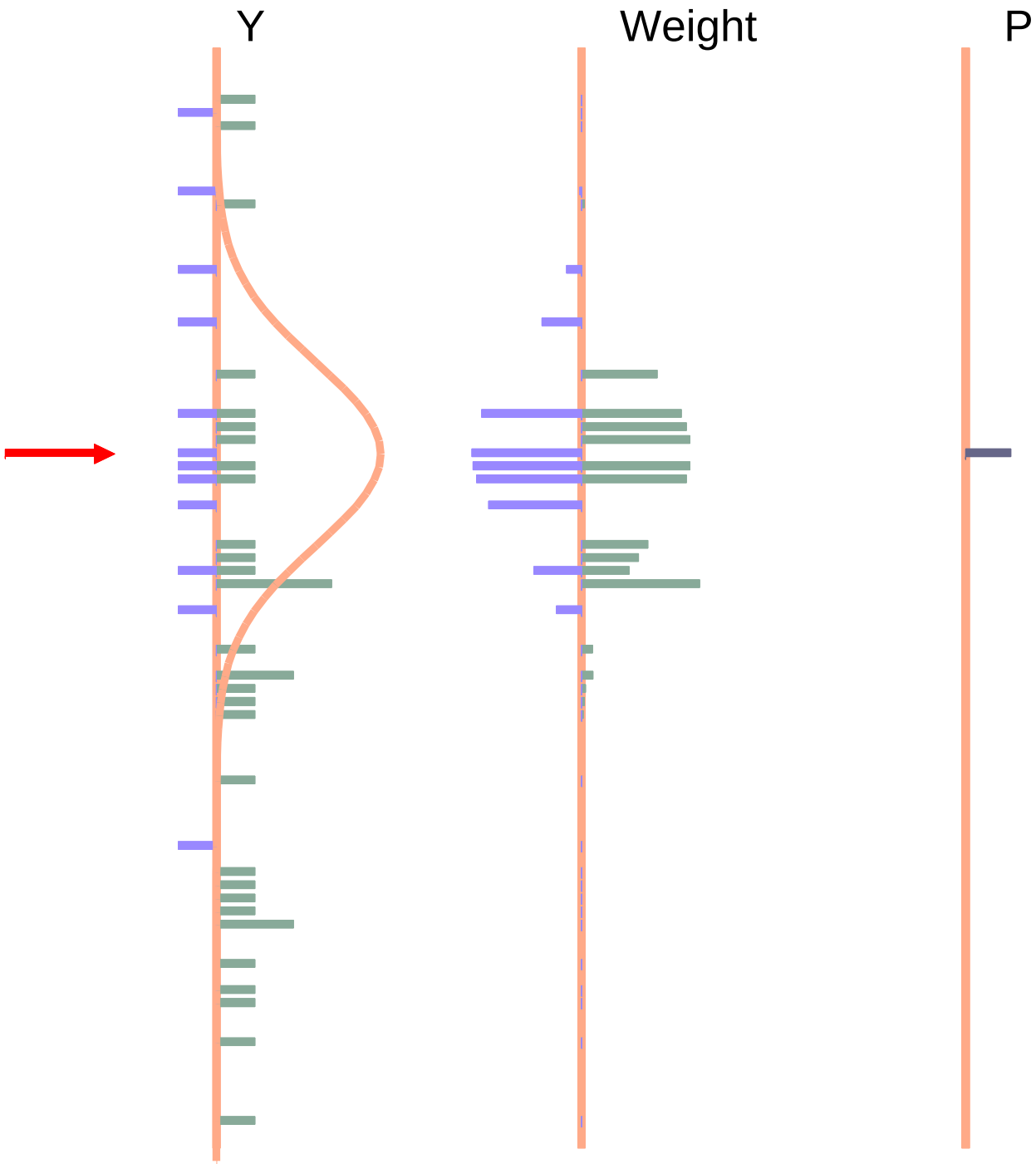
   $h[y' \mid R_t = 1, Y_{t-1} = y] \sim \mathit{smooth}(y; \sigma)$

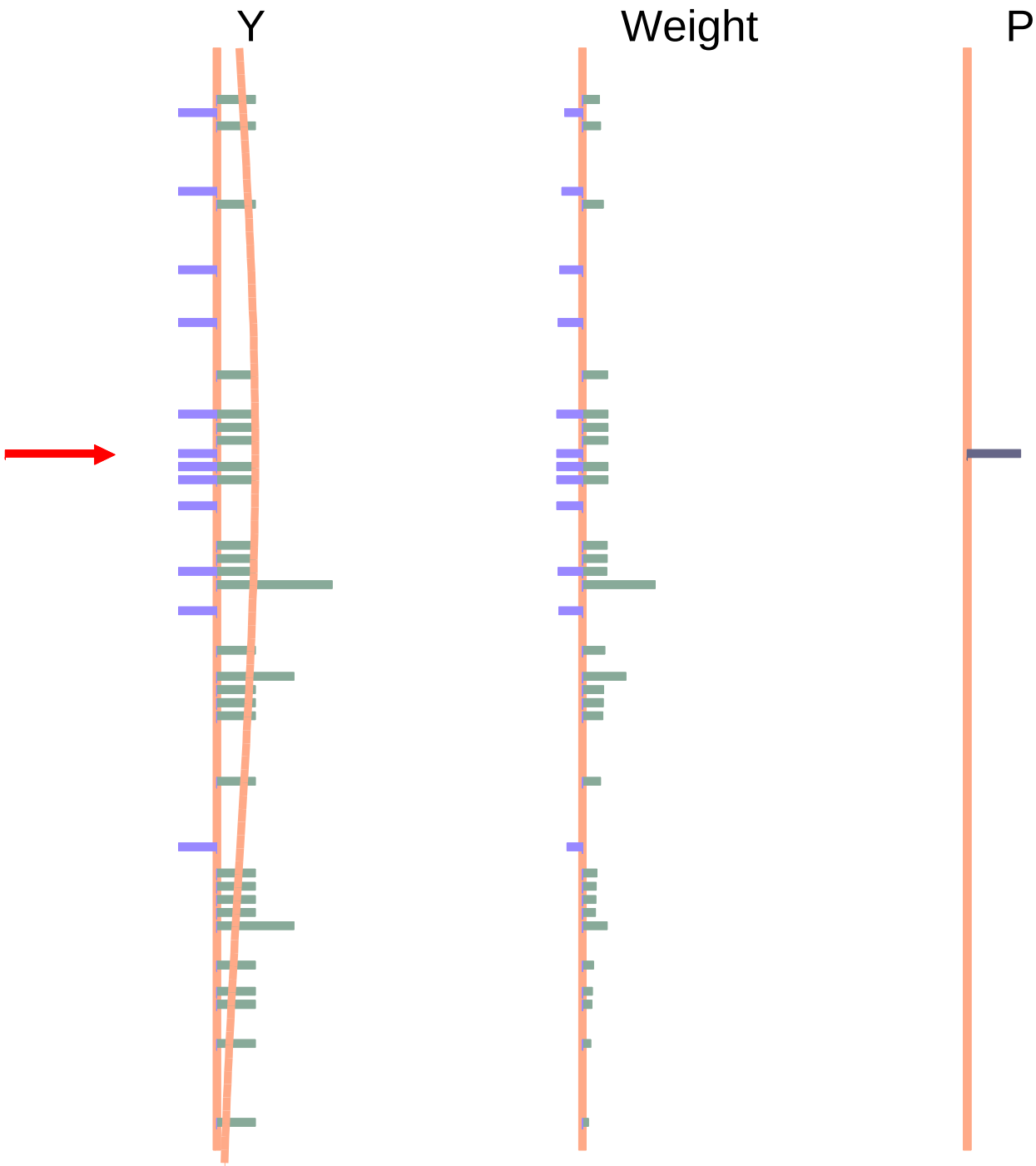
   The smooth function of y depends on a single smoothing parameter σ.

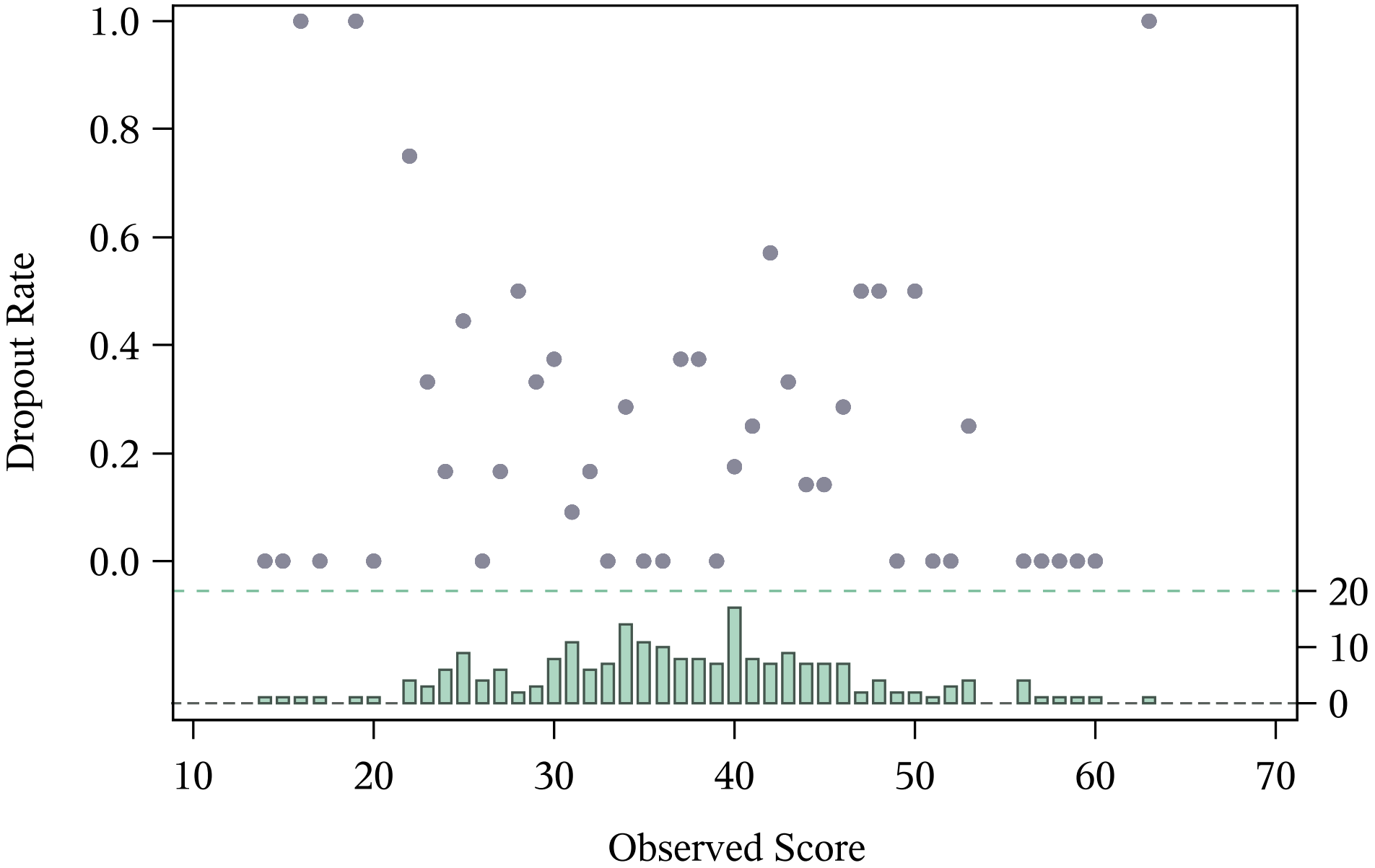
Estimating the smoothing parameters

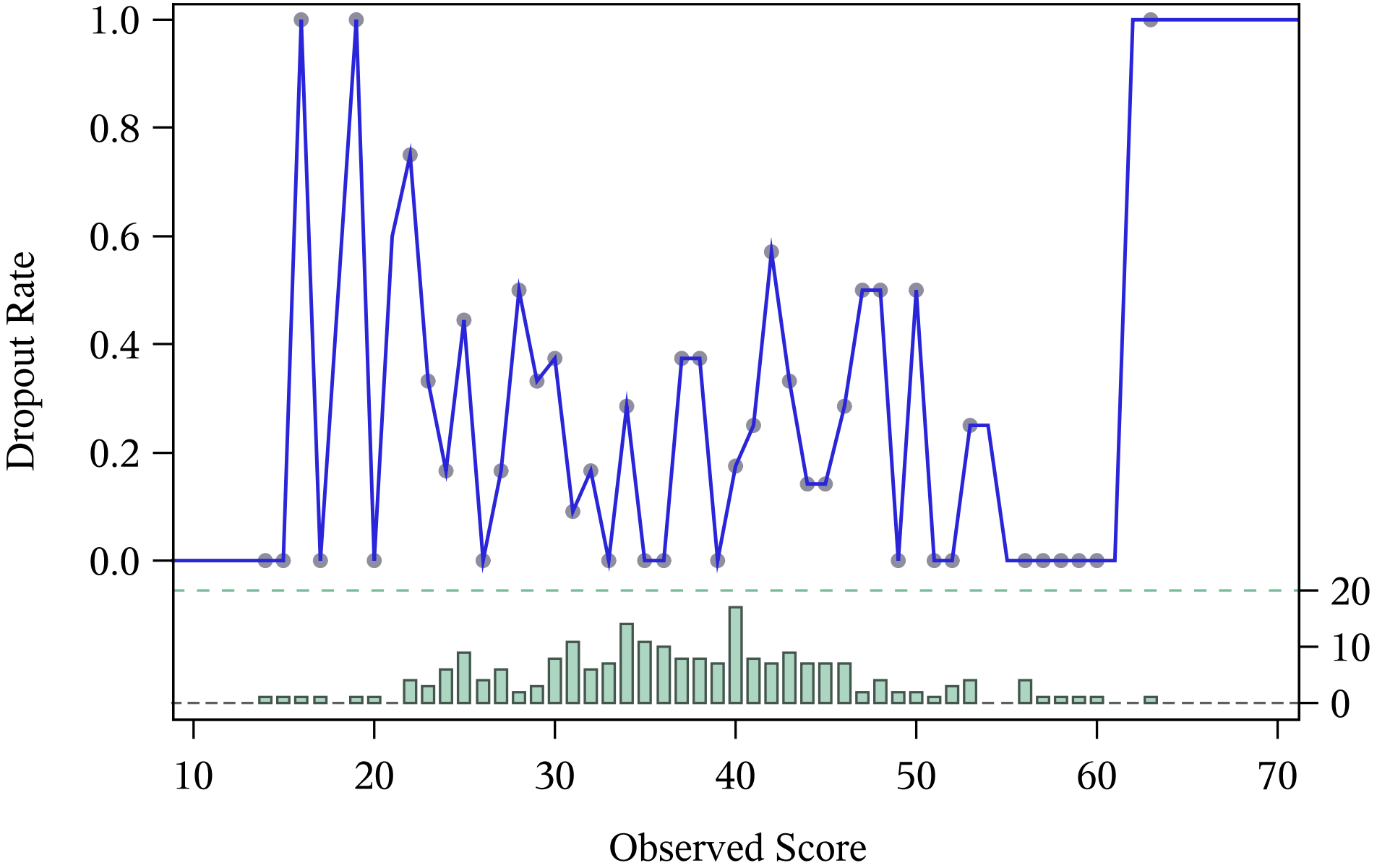
In order to estimate the two smoothing parameters we do the following:

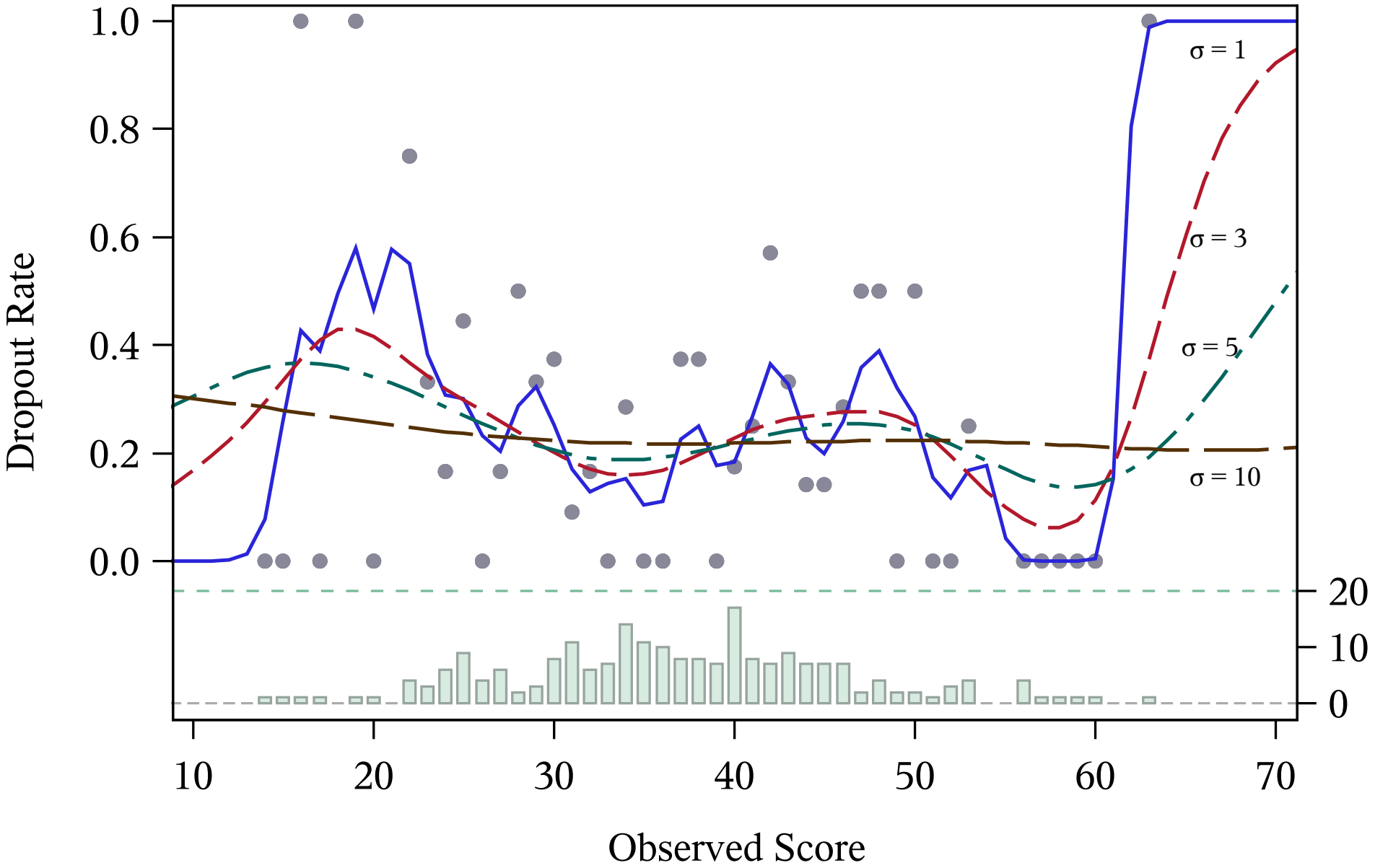
a. Partition the data into N_{part} pieces.
b. Set aside a partition of the data and use the remainder to "predict" a feature of the partition that has been set aside. How badly this "prediction" goes is given a numeric value – the loss associated with σ.
c. Repeat this dropping each partition in turn and computing the total loss.
d. Choose the smoothing parameter that minimizes the loss.

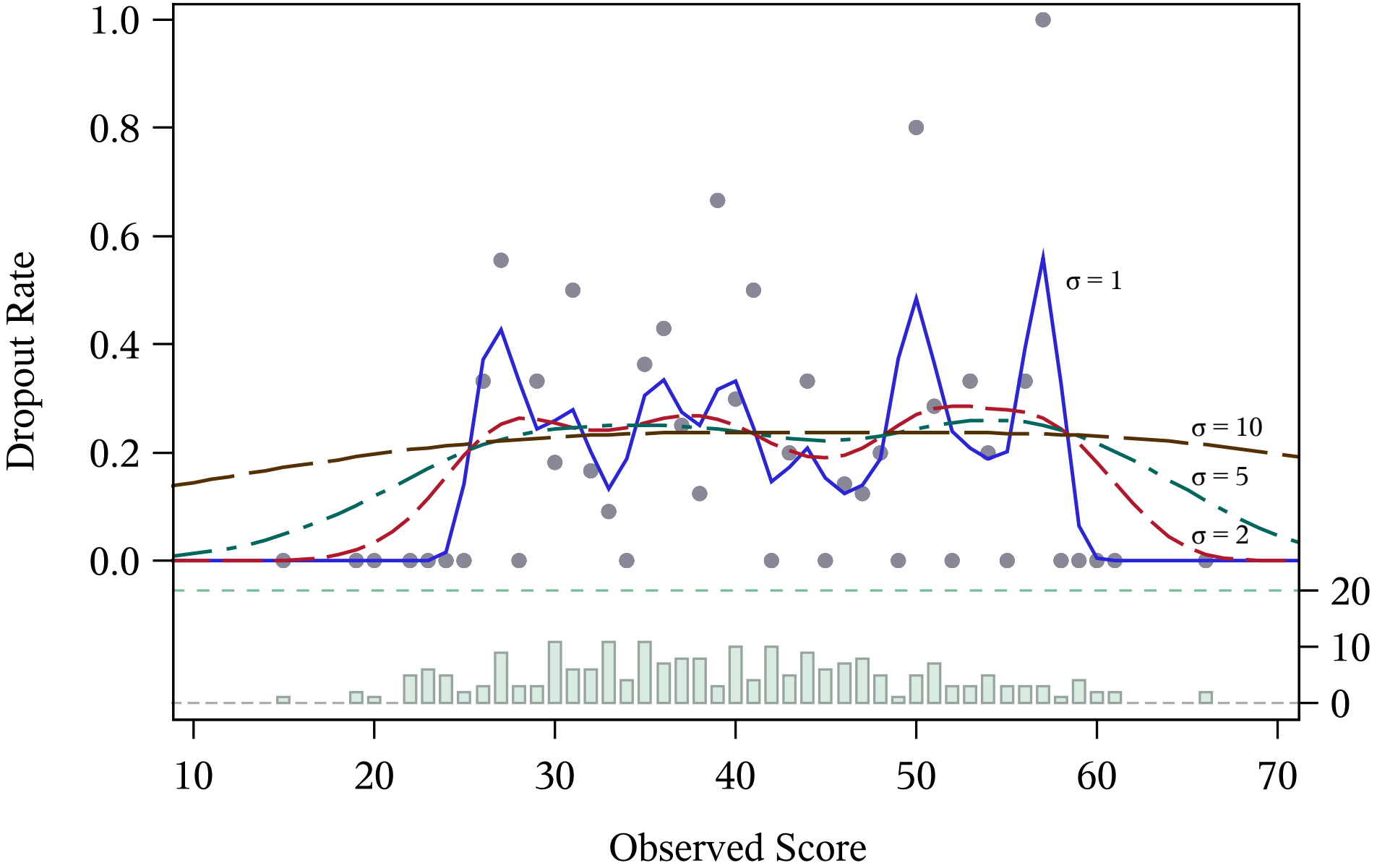

Y
Weight
P


Y
Weight
P


1.0
0.8
0.6
0.4
0.2
0.0
Dropout Rate
20
10
0
10
20
30
40
50
60
70
Observed Score


1.0
0.8
0.6
0.4
0.2
0.0
Dropout Rate
20
10
0
10
20
30
40
50
60
70
Observed Score


1.0
0.8
0.6
0.4
0.2
0.0
Dropout Rate
σ = 1
σ = 3
σ = 5
σ = 10
20
10
0
10
20
30
40
50
60
70
Observed Score


1.0
0.8
0.6
0.4
0.2
0.0
Dropout Rate
σ = 1
σ = 10
σ = 5
σ = 2
20
10
0
10
20
30
40
50
60
70
Observed Score

samonev

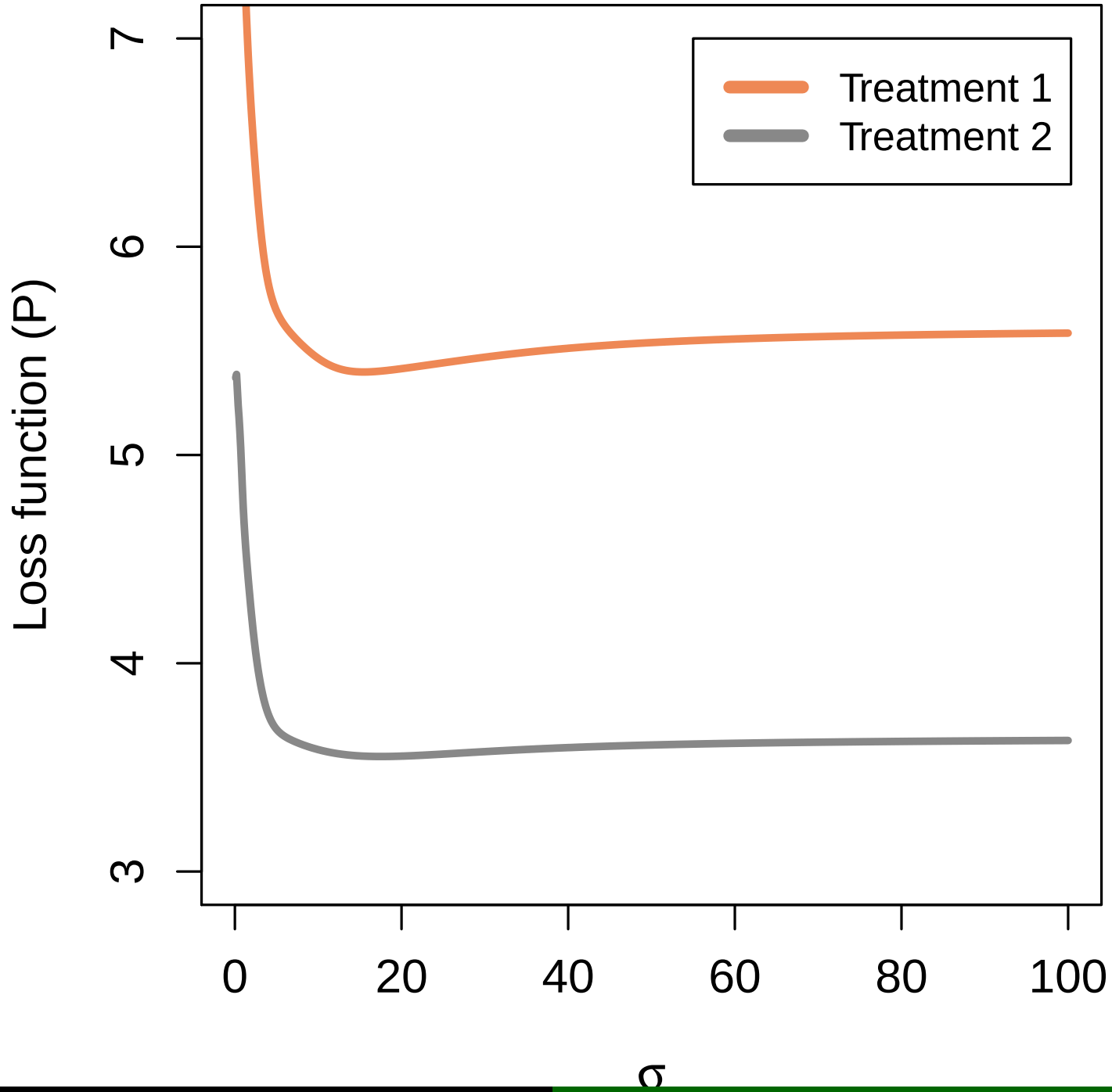
The SAMONEV procedure computes the loss function for a range of σ.

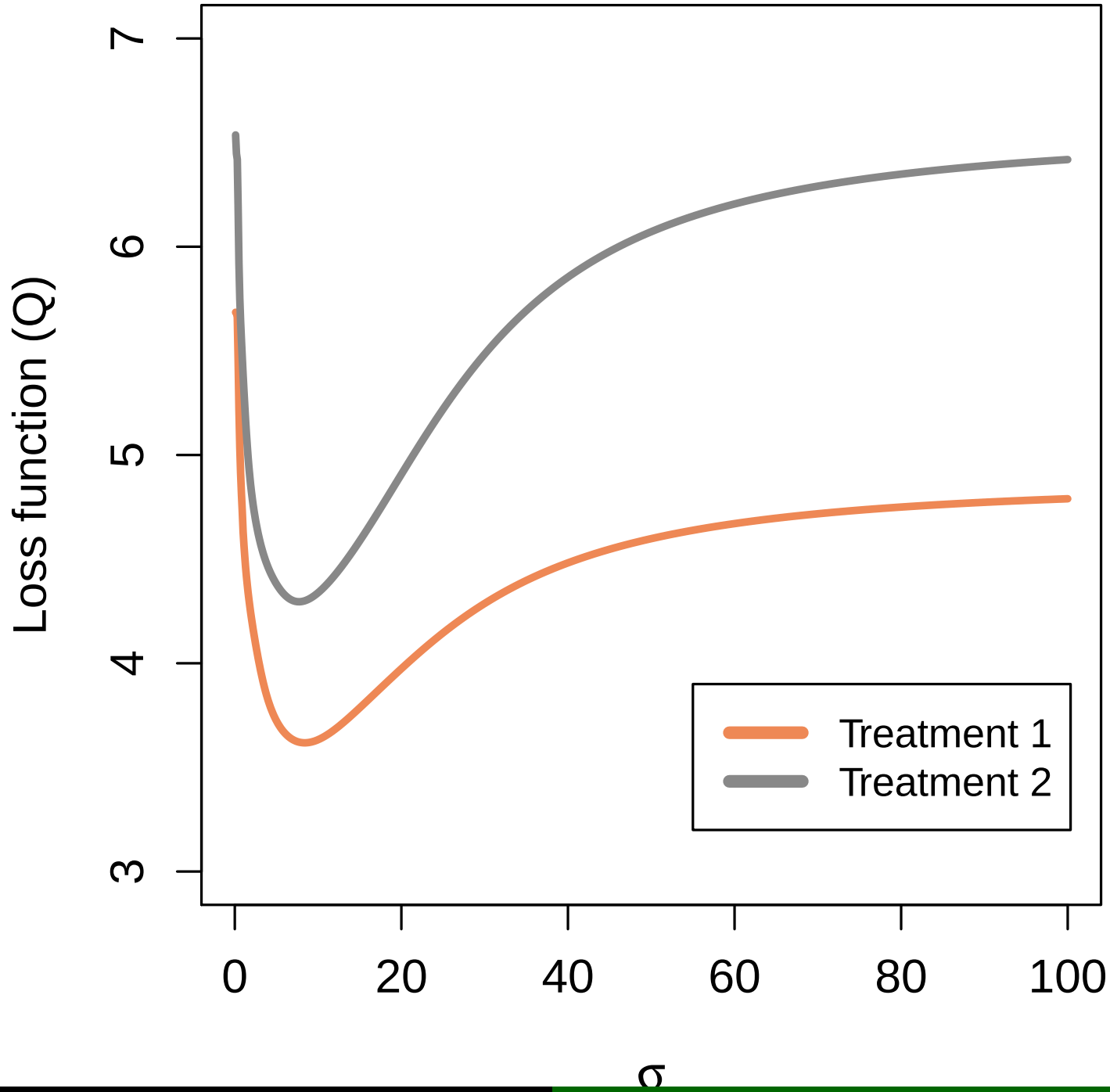
samonev	
data =	Input dataset
out =	Output dataset
npart =	Number of partitions
var varlist	list of variables in time order
sigma sigmalist	list of values

The samon library

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

    var v1 - v6;
    sigma 0.5 to 35 by 1;
run;
proc print data=ev1;
run;
```


Treatment 1
Treatment 2
Loss function (P)
7
6
5
4
3
0
20
40
60
80
100
σ


Loss function (Q)
7
6
5
4
3
0
20
40
60
80
100
σ
Treatment 1
Treatment 2

samon

The samon procedure can be used to find the optimal values of σ_H and σ_F. Aguments on the procedure statement include:

data	input dataset
out	output dataset
Npart	Number of partitions
HInit	initial value for σ_H
HHigh	upper bound for σ_H
FInit	initial value for σ_F
FHigh	upper bound for σ_F

```
* Finding optimal Sigma_H and Sigma_F.;
* ------------------------------------;
proc samon data   = panss1
           out    = samon1
           HOut   = Hout1
           FOut   = Fout1
           Npart  = 10
           Hinit  = 10.0
           Hhigh  = 50.0
           Finit  = 8.0
           Fhigh  = 50;
  var v1 - v6;
run;
```

Treatment 1

```
proc print data = Hout1 noobs;
run;
```

rc	Niter	Sigma	loss
2	4	15.4519	5.39857

```
proc print data = HOut1 noobs;
run;
```

rc	Niter	Sigma	loss
2	3	8.39927	3.61805

Treatment 2

```
proc print data = HOut1 noobs;
run;
```

rc	Niter	Sigma	loss
2	4	17.5248	3.55231

```
proc print data = FOut1 noobs;
run;
```

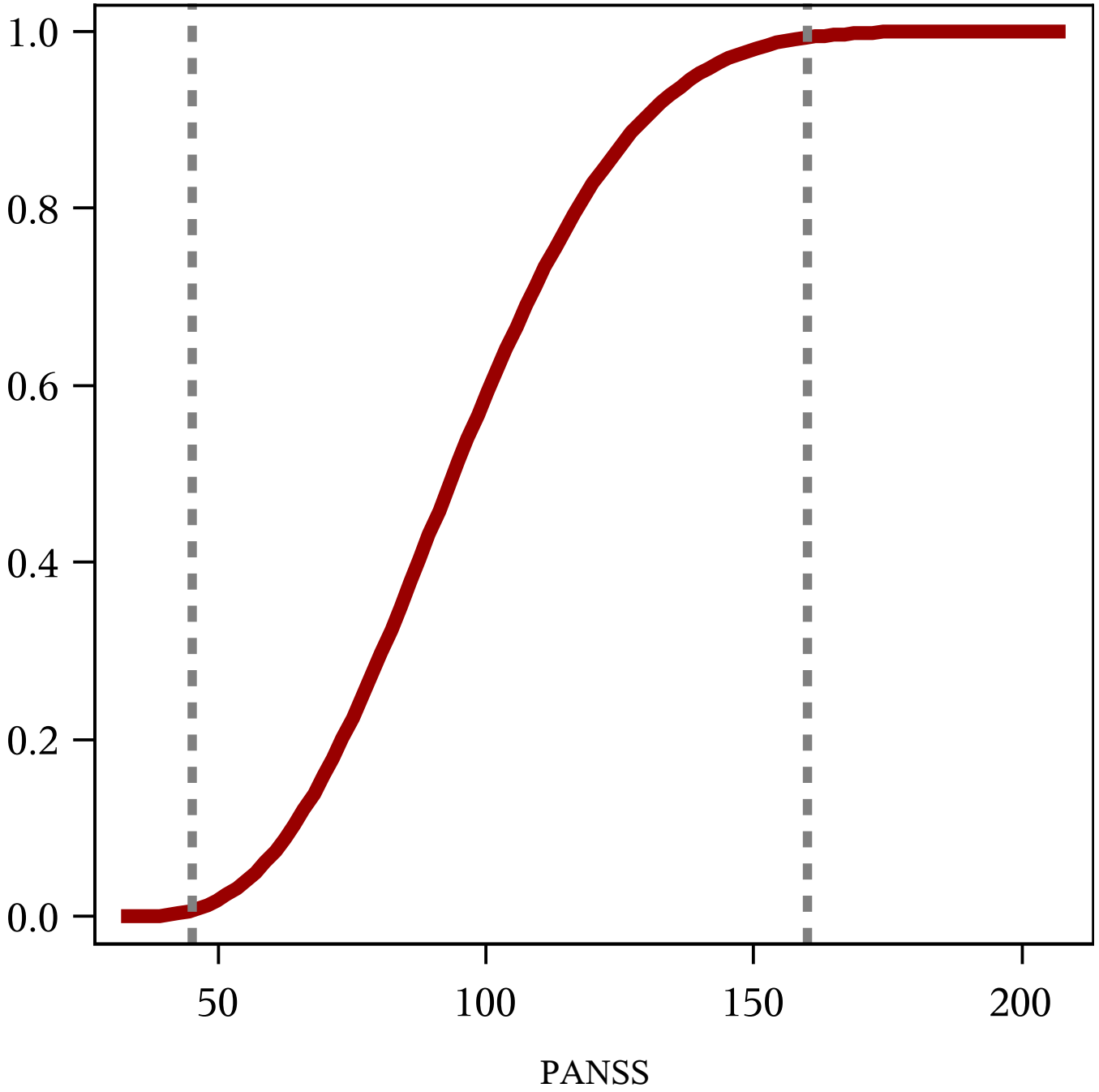
rc	Niter	Sigma	loss
2	2	7.70219	4.29539

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.
- $\alpha = 0$ is missing at random
- α quantifies the influence of Y_{t+1} on the decision to drop-out between t and $t + 1$.

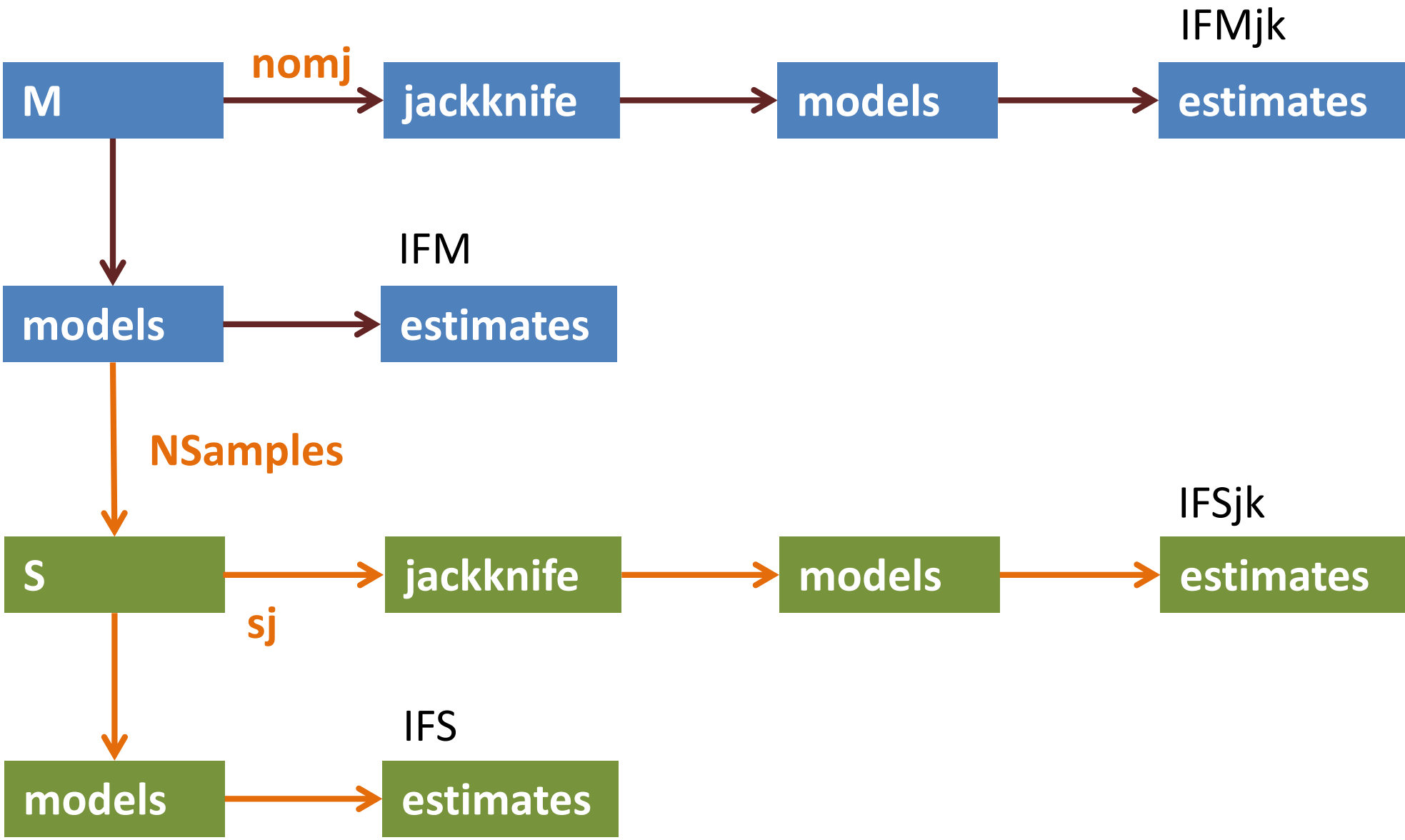
Sensitivity Analysis

- 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 PANSS data there are theoretical limits in that PANSS scores range between 30 and 210.
- Clinical practice gives a range of values over which a change in PANSS noticeable effect. This translates to parameters for the cumulative beta function ζ_1 and ζ_2.
- Another strategy might be to fit a beta distribution to the data (after suitible transformation) to determine ζ_1 and ζ_2.


1.0
0.8
0.6
0.4
0.2
0.0
50
100
150
200
PANSS

The SAMON procedure

samon	
data =	Input dataset
out =	Output dataset
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
sj	compute jackknifes for each bootstrap sample
nomj	suppress jackknife computation for the main dataset
var varlist	list of variables in time order
sensp senslist	list of sensitivity parameters

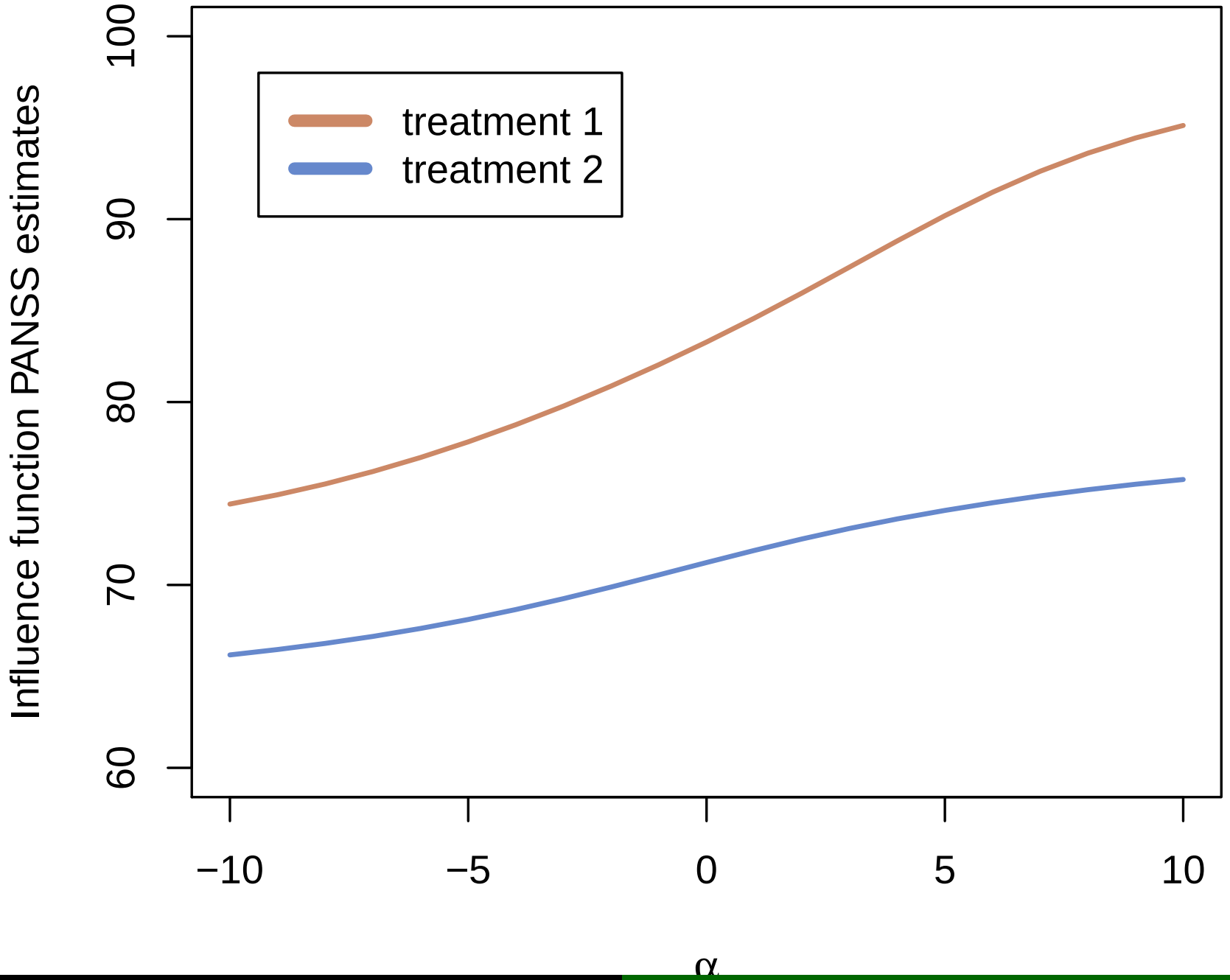

IFMjk
nomj
M
jackknife
models
estimates
IFM
models
estimates
NSamples
IFSjk
S
jackknife
models
estimates
sj
IFS
models
estimates

```
proc samon  data = panss1 out = samon1
    Npart = 10
    Hinit = 10.0    HHigh = 50.0
    Finit =  8.0    FHigh = 50.0

    lb    = 30.0    ub    = 210.0
    zeta1 = 4.0     zeta2 = 7.0
    nomj            nsamples = 0 ;

  var v1 - v6;
  sensp -10 to 10 by 1;
run;
proc print data = samon noobs;
  var alpha AEst AVar IFEst IFVar;
run;
```

alpha	AEst	AVar	IFEst	IFVar
-10	73.8065	0.003674	74.4212	11.3807
-9	74.2520	0.004203	74.9302	11.1160
-8	74.7698	0.004849	75.5221	10.8707
-7	75.3714	0.005630	76.2017	10.6534
-6	76.0690	0.006564	76.9697	10.4688
-5	76.8755	0.007670	77.8240	10.3181
-4	77.8043	0.008965	78.7618	10.2052
-3	78.8690	0.010454	79.7807	10.1429
-2	80.0814	0.012104	80.8767	10.1537
-1	81.4478	0.013809	82.0441	10.2633
0	82.9634	0.015361	83.2792	10.4883
1	84.6038	0.016482	84.5839	10.8161
2	86.3140	0.016987	85.9578	11.1820
3	88.0077	0.016943	87.3798	11.5014
4	89.5939	0.016586	88.8076	11.7346
5	91.0111	0.016092	90.1890	11.8620
6	92.2343	0.015535	91.4724	11.8725
7	93.2629	0.014954	92.6186	11.7935
8	94.1108	0.014382	93.6080	11.6770
9	94.7996	0.013853	94.4388	11.5673


Influence function PANSS estimates
100
90
80
70
60
treatment 1
treatment 2
−10
−5
0
5
10
α

- Use bootstrap with jackknife to compute confidence intervals for IF estimates.
- The NSamples arugment 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 bootstaps each with bootsrap estimates on 50 sensitivity parameters gives rise to 50 x 100 x 1000 = 5 million estimates.

```
proc samon data = panss1 out = samon1
    Npart    = 10

    Hinit    = 10.0    HHigh = 50.0
    FInit    =  8.0    FHigh = 50.0
    lb       = 30      ub    = 210
    zeta1    = 4.0     zeta2 = 7.0
    NSamples = 500     seed0 = 81881
    sj;
  var v1-v6;
  sensp -20 to 20 by 1;
run;
```

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 effect difference with confidence interval from a pair of samonSummary objects.
samonCrossSummary	Computes the effect difference with confidence interval for each pair of sensitivity paramaters 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	Difference in two trials
(	
IFM1 =	main results from samonSummary trt 1
sampIF1 =	sample results from samonSummary trt 1
IFM2 =	main results from samonSummary trt 2
sampIF2 =	sample results from samonSummary trt 2
out =	summary of difference
)	

samonCrossSummary macro

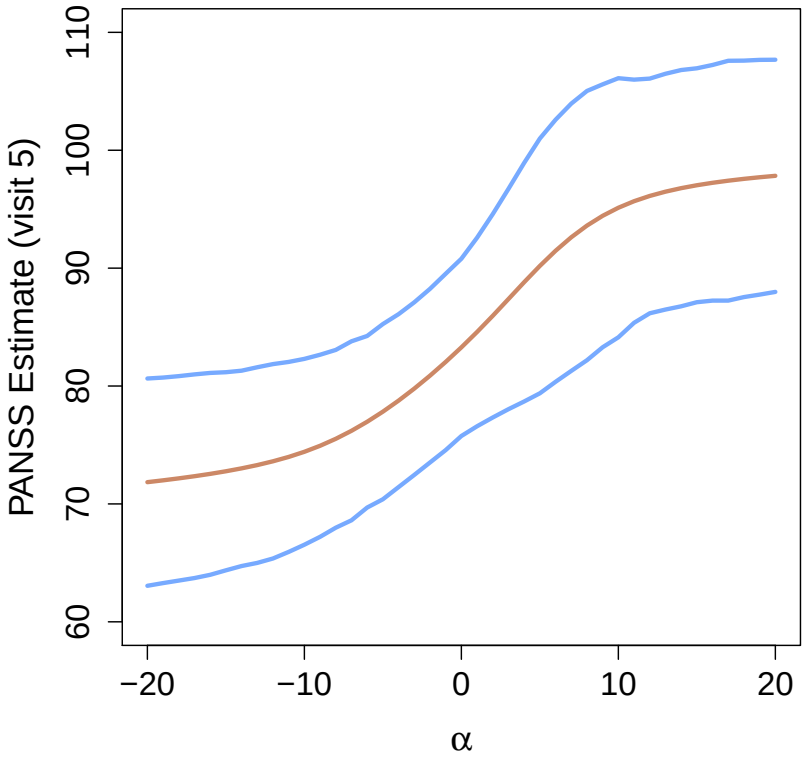
samonCrossSummary	Difference in two trials for all pairs of sensitivity parameter
(	
IFM1 =	main results from samonSummary trt 1
sampIF1 =	sample results from samonSummary trt 1
IFM2 =	main results from samonSummary trt 2
sampIF2 =	sample results from samonSummary trt 2
out =	summary of difference
)	

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

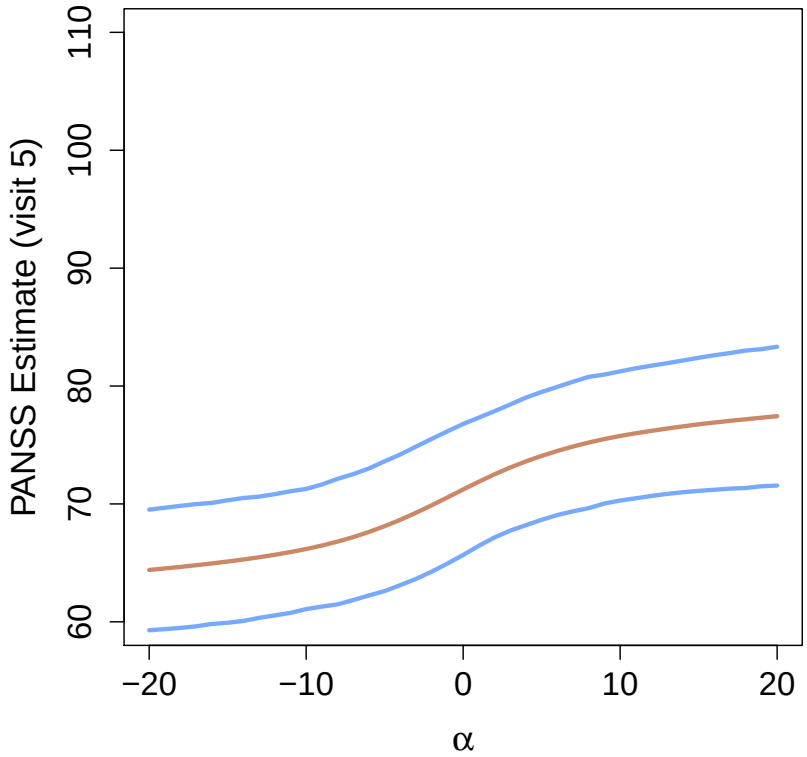
alpha	IFEst	IFVar	lb	ub
-10	74.4212	11.3807	66.4486	82.394
-9	74.9302	11.1160	67.0858	82.775
-8	75.5221	10.8707	68.0089	83.035
-7	76.2017	10.6534	68.7412	83.662
-6	76.9697	10.4688	69.5369	84.403
-5	77.8240	10.3181	70.4880	85.160
-4	78.7618	10.2052	71.6000	85.924
-3	79.7807	10.1429	72.5973	86.964
-2	80.8767	10.1537	73.7356	88.018
-1	82.0441	10.2633	74.9532	89.135
0	83.2792	10.4883	76.1091	90.449
1	84.5839	10.8161	77.3722	91.796
2	85.9578	11.1820	78.5012	93.414
3	87.3798	11.5014	79.2752	95.485
4	88.8076	11.7346	80.1181	97.497

Estimated PANSS score at visit 5

Placebo Arm

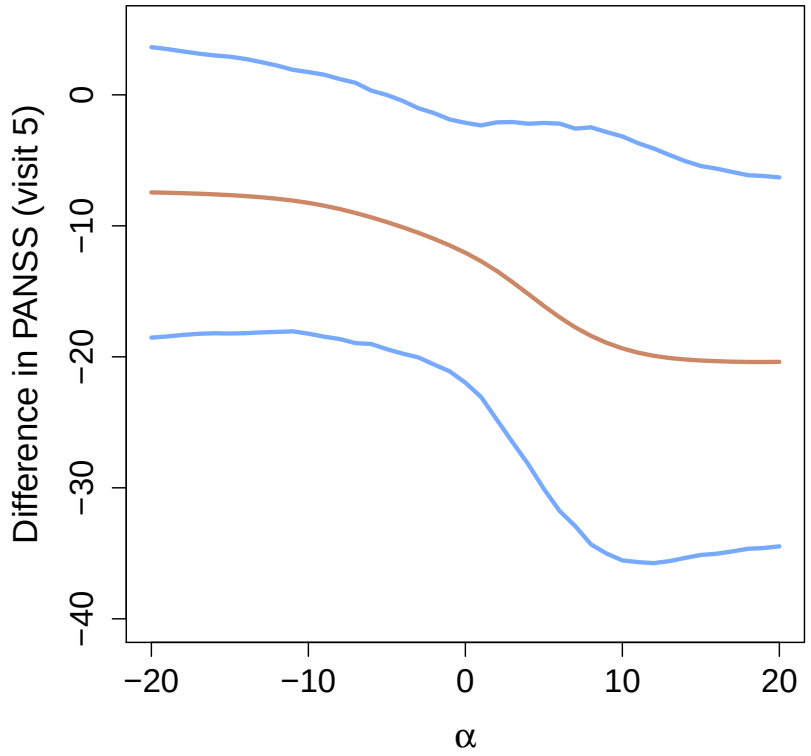


Active Arm



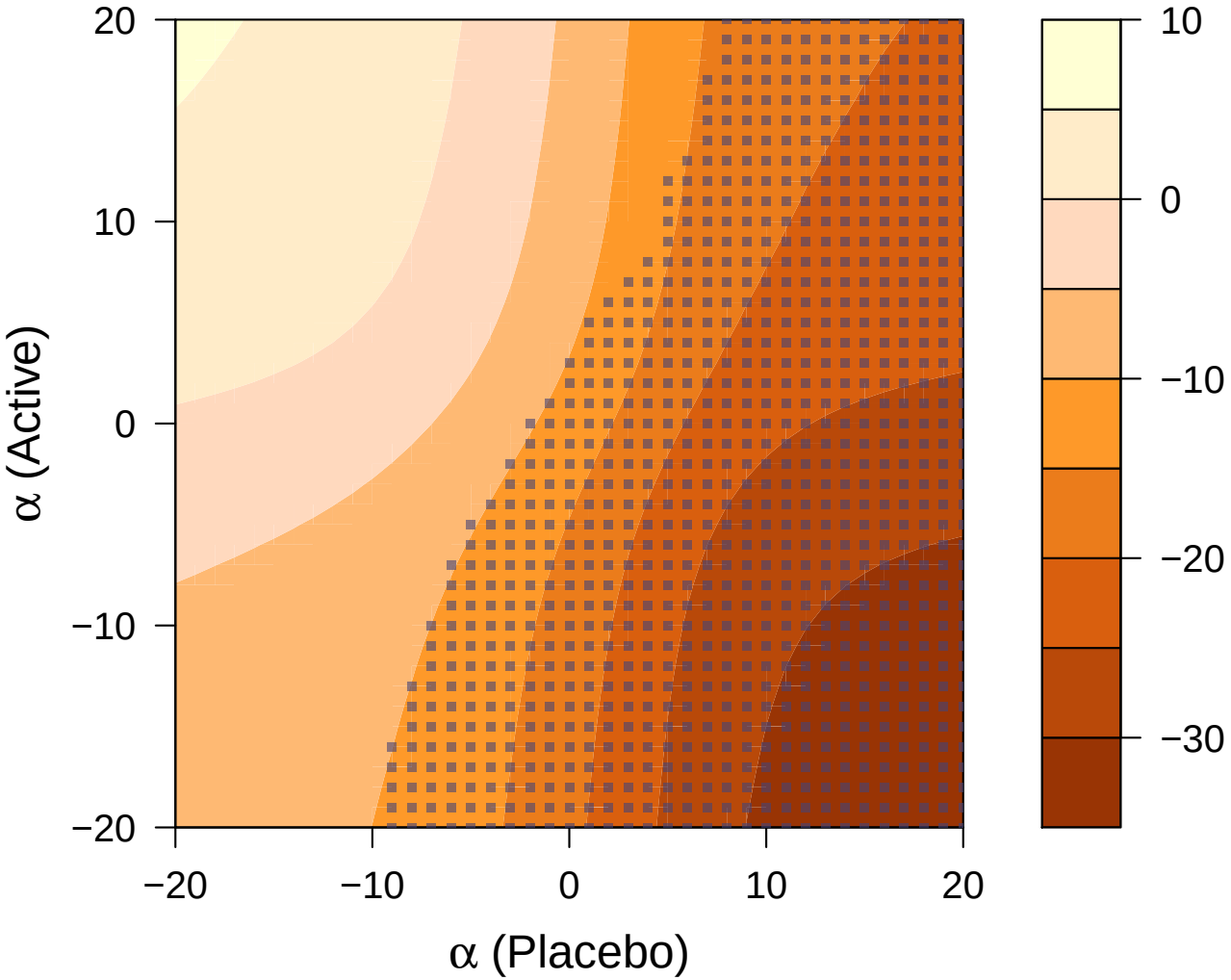
Estimated PANSS score at visit 5

Difference (active - placebo)



Another useful plot is a surface plot of the difference in the estimated mean value in the two treatment groups given as a function of the two alpha parameters. We use the samonCrossSummary function to compute the difference in estimates for each pair of alpha. The plotting is done with the filled.contour function.

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


20
10
0
−10
−20
α (Active)
−20
−10
0
10
20
α (Placebo)
10
0
−10
−20
−30

```
%samondatacheck(
  data     = data2.vas1,
  vars     = v1-v9,
  out      = chkv1,
  stats    = statsv1,
  mpattern = mpatternv1
);
%samondatacheck(
  data     = data2.vas2,
  vars     = v1-v9,
  out      = chkv2,
  stats    = statsv2,
  mpattern = mpatternv2
);
proc print data = statsv1 label noobs;
proc print data = statsv2 label noobs;
run;
```

```
proc samoneim
  data      = data2.vas1
  out       = evalv1
  npart     = 10;

  var v1-v9;
  sigma 0.5 to 35 by 0.5;
run;
proc samoneim
  data      = data2.vas2
  out       = evalv2
  npart     = 10;

  var v1-v9;
  sigma 0.5 to 35 by 0.5;
run;
```

```
proc samonIM
  data      = data2.vas1
  out       = samonv1

  npart     = 10
  Hinit     = 15     HHigh = 50
  Finit     =  8     FHigh = 50

  nimpute = 5

  lb = 0  ub = 102  zeta1 = 1  zeta2 = 1;

  var v1-v9;
  sigma 0.5 to 35 by 0.5;
run;
```

```
proc samonIM
   data      = data2.vas2
   out       = samonv2

   npart     = 10
   Hinit     = 15     HHigh = 50
   Finit     =  8     FHigh = 50

   nimpute = 5

   lb = 0  ub = 102  zeta1 = 1  zeta2 = 1;

   var v1-v9;
   sigma 0.5 to 35 by 0.5;
run;
```