Global Sensitivity Analysis of Randomized Trials with Missing Data
ASA Webinar

Daniel Scharfstein
Johns Hopkins University
dscharf@jhu.edu

September 20, 2016

Funding Acknowledgments

- FDA
- PCORI

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*"

- 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.
- Focus on monotone missing data pattern
- Consider the case where interest is focused on a comparison of treatment arm means at the last scheduled visit.

## 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.
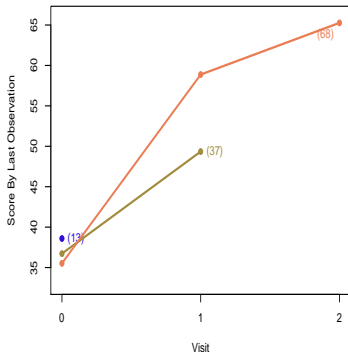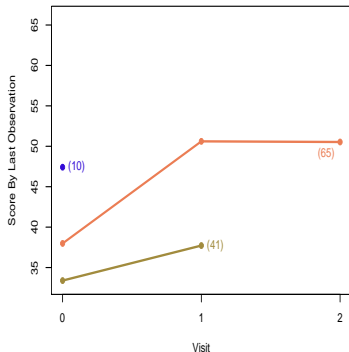
What is the difference in the mean QLESSF score at week 8 between Quetiapine 600 mg/day and placebo in the counterfactual world in which all patients were followed to that week?

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.
- 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 $R_k = 0$.

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 $Y_0 = y_0$.
- $B_0(y_0)$: 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 $Y_1 = y_1$.
- $B_1(y_0, y_1)$: 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 $B_0(y_0)$.
- 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 $B_1(y_0, y_1)$.

Mathematically, we can express these assumptions as follows:

$$f^*(Y_1, Y_2 | A_0(y_0)) = f^*(Y_1, Y_2 | B_0(y_0)) \text{ for all } y_0 \qquad (1)$$

and

$$f^*(Y_2 | A_1(y_0, y_1)) = f^*(Y_2 | B_1(y_0, y_1)) \text{ for all } y_0, y_1 \qquad (2)$$

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

- $F_2^*(y_2|y_1, y_0) = P^*[Y_2 \le y_2 | B_1(y_1, y_0)]$
- $F_1^*(y_1|y_0) = P^*[Y_1 \le y_1 | B_0(y_0)]$
- $F_0^*(y_0) = P^*[Y_0 \le y_0]$.

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

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

$$f^*(Y_1, Y_2|A_0(y_0)) = f^*(Y_1, Y_2|B_0(y_0)) \text{ for all } y_0$$

It is equivalent to

$$f^*(Y_2|A_0(y_0), Y_1 = y_1)$$
$$= f^*(Y_2|B_0(y_0), Y_1 = y_1) \text{ for all } y_0, y_1 \qquad (3)$$

and

$$f^*(Y_1|A_0(y_0)) = f^*(Y_1|B_0(y_0)) \text{ for all } y_0 \qquad (4)$$

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 $B_0(y_0)$ 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

$$f^*(Y_1|A_0(y_0))$$
$$\propto f^*(Y_1|B_0(y_0)) \exp\{\alpha r(Y_1)\} \text{ for all } y_0 \qquad (5)$$

Generalizing (2) Using Exponential Tilting

$$f^*(Y_2|A_1(y_0, y_1))$$
$$\propto f^*(Y_2|B_1(y_0, y_1)) \exp\{\alpha r(Y_2)\} \text{ for all } y_0, y_1 \qquad (6)$$

- $r(y)$ 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 $B_0(y_0)$.
- 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 $B_1(y_0, y_1)$.

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

$$l_1^*(y_0; \alpha) = \text{logit } P^*[R_1 = 0 | R_0 = 1, Y_0 = y_0] -$$
$$\log E^*[\exp\{\alpha r(Y_1)\} | B_0(y_0)]$$

and

$$l_2^*(y_1, y_0; \alpha) = \text{logit } P^*[R_2 = 0 | R_1 = 1, Y_0 = y_0, Y_1 = y_1] -$$
$$\log E^*[\exp\{\alpha r(Y_2)\} | B_1(y_1, y_0)]$$
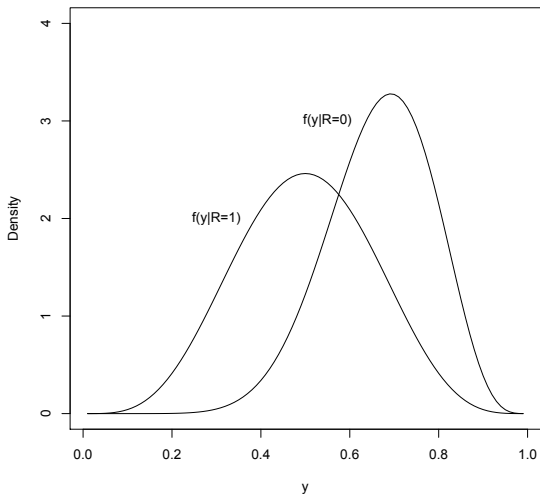
Missing Not at Random (MNAR)

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.
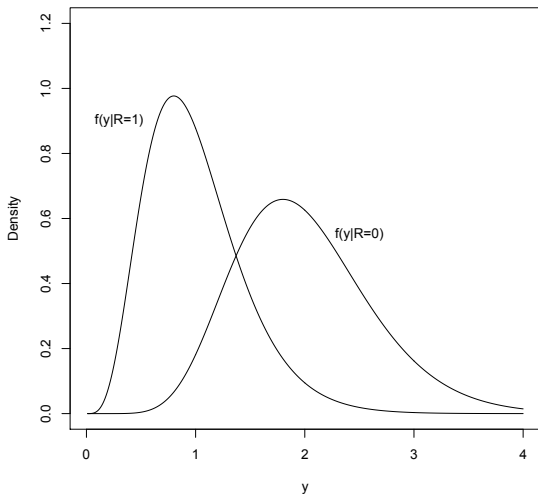
Exponential Tilting Explained

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

- 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

$$\int_{y_0} \int_{y_1} \int_{y_2} y_2 \left\{ dF_2^*(y_2|y_1,y_0)\{1 - H_2^*(y_1,y_0)\} + \frac{dF_2^*(y_2|y_1,y_0)\exp\{\alpha r(y_2)\}}{\int_{y_2'} dF_2^*(y_2'|y_1,y_0)\exp\{\alpha r(y_2')\}} H_2^*(y_1,y_0) \right\} \times$$
$$\left\{ dF_1^*(y_1|y_0)\{1 - H_1^*(y_0)\} + \frac{dF_1^*(y_1|y_0)\exp\{\alpha r(y_1)\}}{\int_{y_1'} dF_1^*(y_1'|y_0)\exp\{\alpha r(y_1')\}} H_1^*(y_0) \right\} dF_0^*(y_0)$$
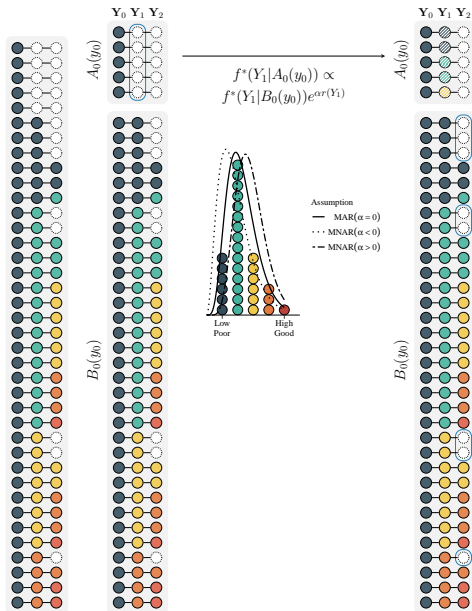
where

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

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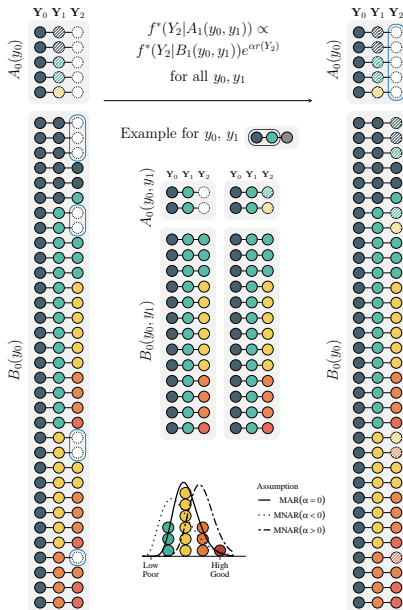
$$H_1^*(y_0) = P^*[R_1 = 0 | R_0 = 1, Y_0 = y_0]$$

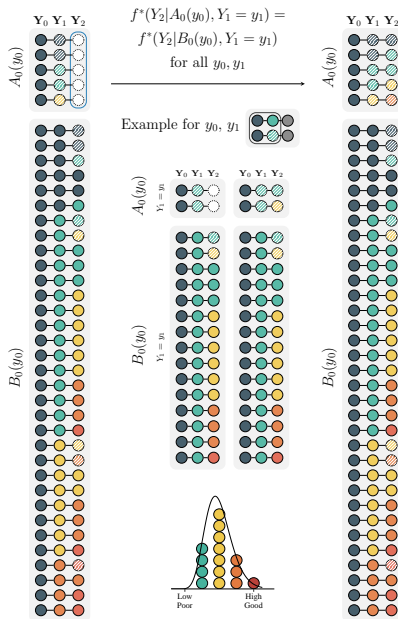- μ^* is written as a function of the distribution of the observed data (depending on α).

$$f^*(Y_1 | A_0(y_0)) \propto f^*(Y_1 | B_0(y_0)) e^{\alpha r(Y_1)}$$

$$f^*(Y_2|A_0(y_0), Y_1 = y_1) =$$
$$f^*(Y_2|B_0(y_0), Y_1 = y_1)$$
$$\text{for all } y_0, y_1$$

Example for y_0, y_1

Inference

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

- $F_2^*(y_2|y_1, y_0) = P^*[Y_2 \leq y_2|B_1(y_1, y_0)]$
- $F_1^*(y_1|y_0) = P^*[Y_1 \leq y_1|B_0(y_0)]$
- $H_2^*(y_1, y_0) = P^*[R_2 = 0|R_1 = 1, Y_1 = y_1, Y_0 = y_0]$
- $H_1^*(y_0) = P^*[R_1 = 0|R_0 = 1, Y_0 = y_0]$.

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

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

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 divided 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. 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

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

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

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

$$\qquad (11)$$

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

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 probability to zero even when it is multipled by $\sqrt{n}$).
- Provided $\widehat{P}$ converges to P^* at a reasonably fast rate, term (19) 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

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

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

Inference - Efficient Influence Function/Gradient

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

$$b_1^*(Y_0, Y_1; \alpha) = E^*\left[\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)}\,\middle|\, R_1 = 1, Y_1, Y_0\right] - E^*\left[\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)}\,\middle|\, R_1 = 1, Y_0\right]$$
$$+ E^*\left[\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)}\left[\frac{\exp\{\alpha r(Y_1)\}}{g_1^*(Y_0, Y_1; \alpha)}\right]\,\middle|\, R_1 = 1, Y_0\right] H_1^*(Y_0)\left\{1 - \frac{\exp\{\alpha r(Y_1)\}}{w_1^*(Y_0; \alpha)}\right\}$$

$$b_2^*(Y_1, Y_2; \alpha) = E^*\left[\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)}\,\middle|\, R_2 = 1, Y_2, Y_1\right] - E^*\left[\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)}\,\middle|\, R_2 = 1, Y_1\right]$$
$$+ E^*\left[\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)}\left[\frac{\exp\{\alpha r(Y_2)\}}{g_2^*(Y_1, Y_2; \alpha)}\right]\,\middle|\, R_2 = 1, Y_1\right] H_2^*(Y_1)\left\{1 - \frac{\exp\{\alpha r(Y_2)\}}{w_2^*(Y_1; \alpha)}\right\}$$

$$c_1^*(Y_0) = E^*\left[\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)}\left[\frac{\exp\{\alpha r(Y_1)\}}{g_1^*(Y_0, Y_1; \alpha)}\right]\,\middle|\, Y_0\right]$$
$$- E^*\left[\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)}\left[\frac{1}{g_1^*(Y_0, Y_1; \alpha)}\right]\,\middle|\, Y_0\right] w_1^*(Y_0; \alpha)$$

$$c_2^*(Y_1) = E^*\left[\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)}\left[\frac{\exp\{\alpha r(Y_2)\}}{g_2^*(Y_1, Y_2; \alpha)}\right]\,\middle|\, R_1 = 1, Y_1\right]$$
$$- E^*\left[\frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)}\left[\frac{1}{g_2^*(Y_1, Y_2; \alpha)}\right]\,\middle|\, R_1 = 1, Y_1\right] w_2^*(Y_1; \alpha)$$

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.
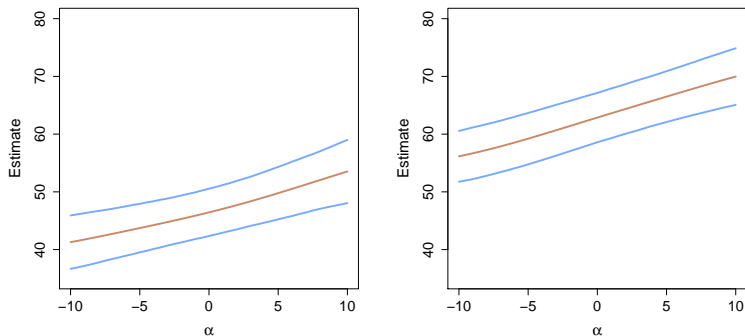
Quetiapine Bipolar Trial - MAR

- 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.

- 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.
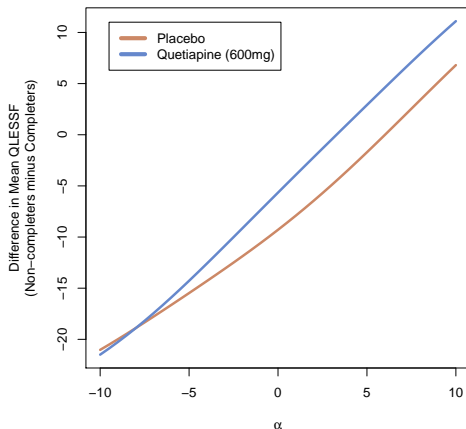
Quetiapine Bipolar Trial - Sensitivity Analysis

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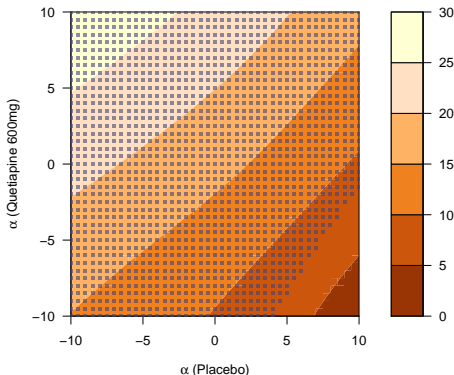
Quetiapine Bipolar Trial - Sensitivity Analysis

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.

- Only when the sensitivity analysis are highly differential (e.g., α(placebo) $= 8$ and α(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

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%

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%
	Bootstrap-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%
	Bootstrap-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%
	Bootstrap-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 $\mathcal{M}$ 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

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

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.

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 $\overline{Y}_k = \overline{y}_k$.
- $B_k(\overline{y}_k)$: 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 $B_k(\overline{y}_k)$

Mathematically, we can express these assumptions as follows:

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

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).

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:

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

and

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

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 $B_k(\overline{y}_k)$ with $Y_{k+1} = y_{k+1}$.

Missing Not at Random

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

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

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.

Missing Not at Random

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

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

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 $B_k(\overline{y}_k)$.

  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.

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).
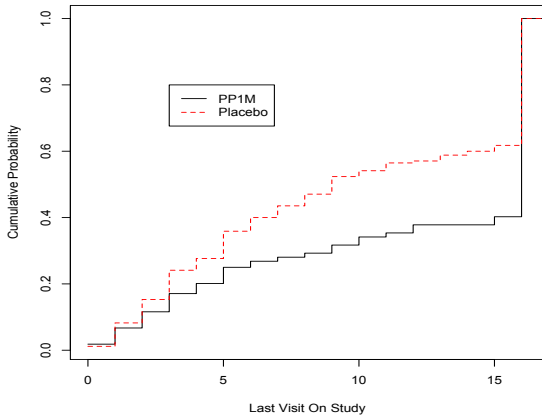
Case Study: SCA-3004

- 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).
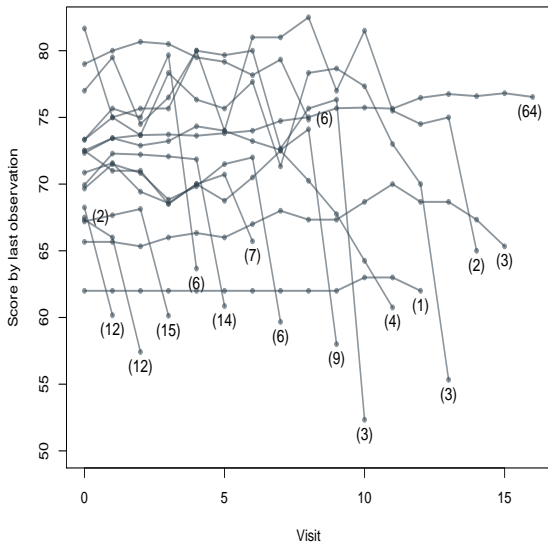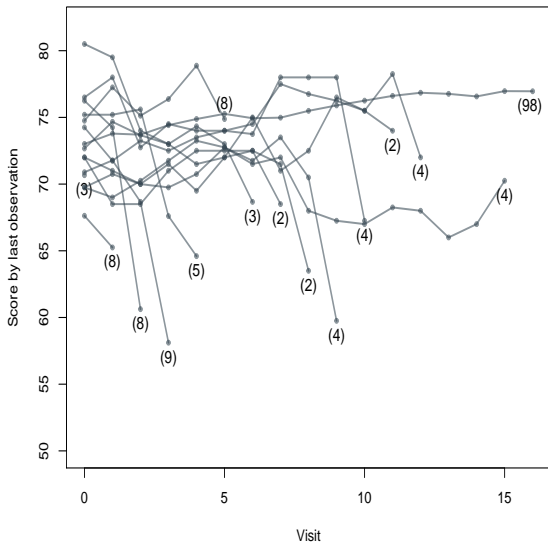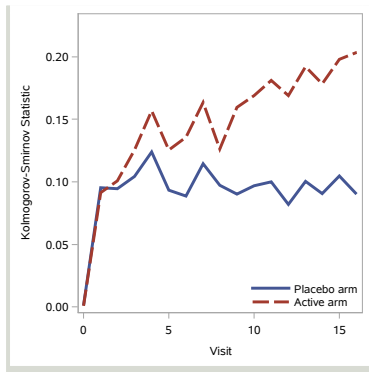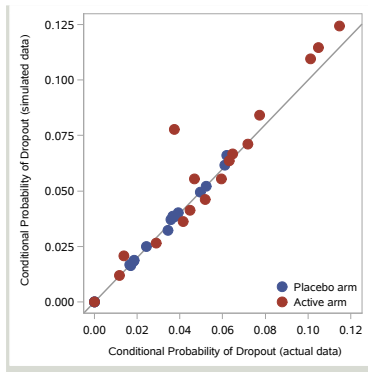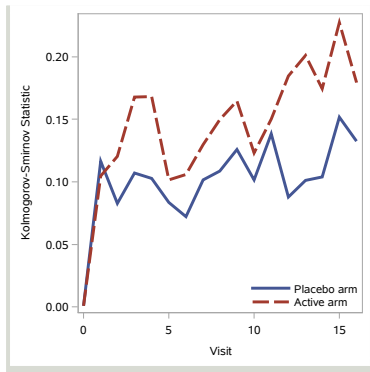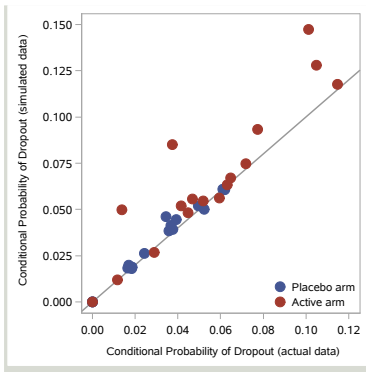
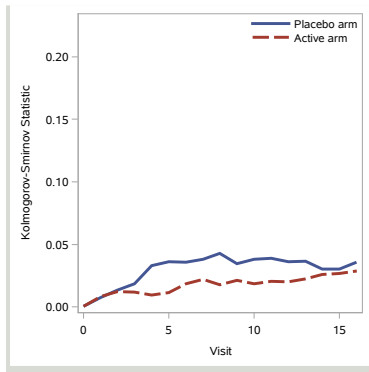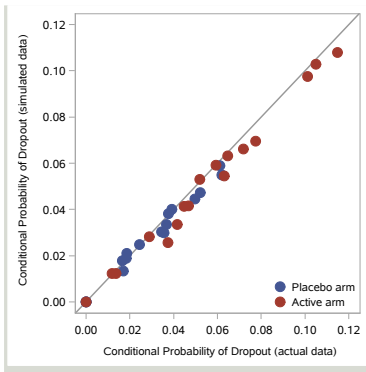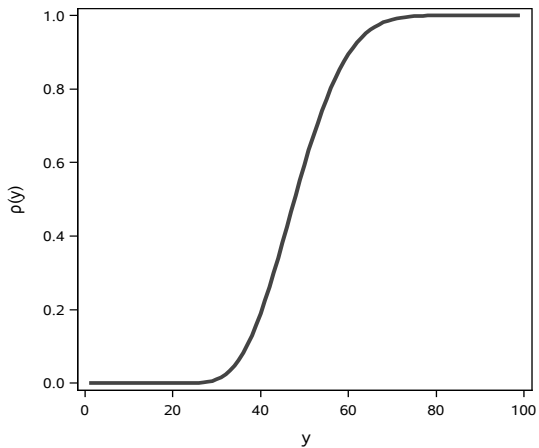# Case Study: SCA-3004

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	$\alpha \times 0.01$
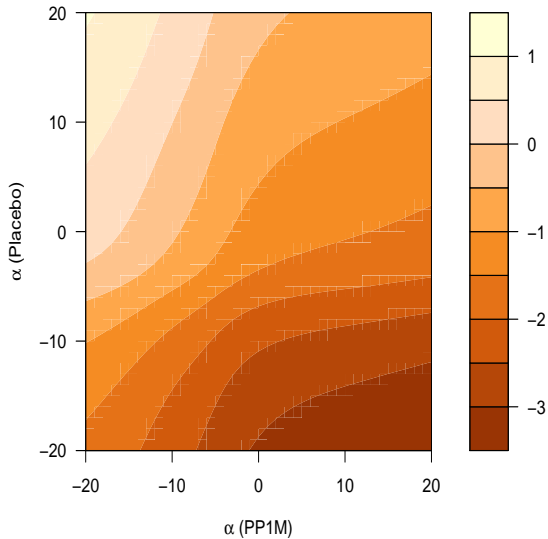
Case Study: SCA-3004
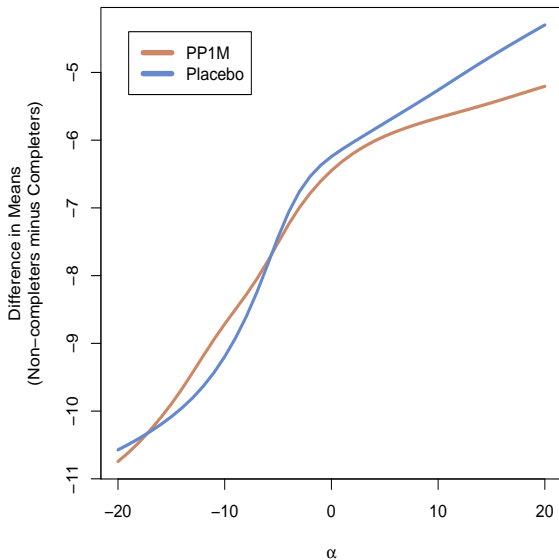
Simulation Study

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%
	Bootstap-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%
	Bootstap-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%
	Bootstap-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%
	Bootstap-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%
	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%

Issues

- Non-monotone missing data; partial imputation
- Outliers

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

Software

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 arm means at the last scheduled time-point
- Outcomes are coded as positive integers
- Missing values are coded as -1
- Rows indicate individuals and columns indicate time-points
- Data at the first time-point (the baseline) is never missing

Background

time — *point*3

$$\text{subjects} \begin{cases} 82 & 88 & 81 & -1 & -1 & -1 & -1 & -1 \\ 71 & 75 & 69 & 66 & 62 & 58 & 51 & 48 \\ 62 & 63 & -1 & 55 & 61 & 66 & 68 & -1 \\ 72 & 63 & -1 & -1 & 62 & 44 & 55 & -1 \\ 83 & 62 & 74 & 67 & -1 & -1 & -1 & -1 \\ 88 & 92 & 99 & 70 & -1 & -1 & -1 & -1 \\ 66 & 71 & 71 & 71 & 75 & 75 & 71 & 71 \\ 90 & 88 & 88 & 88 & 77 & -1 & -1 & -1 \\ 88 & 91 & 92 & 91 & 95 & 90 & 88 & -1 \\ -1 & 52 & 33 & 99 & 87 & 88 & -1 & -1 \end{cases}$$

Case Study: Quetiapine Fumarate

- Among Bipolar I subjects
- Subjects scheduled for 3 assessments at days 1, 29 and 57.
- Primary efficacy variable: QLESSF score (Quality of Life Enjoyment and Satisfaction – Short Form)
- Two treatment groups considered here: placebo (treatment 1) and quetiapine fumarate 600mg (treatment 2).

QLESSF Analysis

- Descriptive analysis for each treatment group.
- Check the monotone missing condition.
- Decide on a sensitivity function.
- In SAS use the `samon` procedure, or, in R use the `samon` function to estimate means.
- Examine the results.

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    -1
      7    35    44    39
      8    24    26    28
      9    25    30    -1
     10    25    28    -1
     11    40    45    45
     12    14    16    -1
```

SAS: `samonDataCheck` macro

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

%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 )
```

samonDataCheck

```
Samon Data Check:
--------------------------------
Number of time-points:        3
Number of subjects:         116
Minimum value:               14
Maximum value:               63


No Samon problems found

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 value:                15
Maximum value:                67

No Samon problems found

Missing Patterns:
               N     proportion
*__ :         13       0.1102
**_ :         37       0.3136
*** :         68       0.5763
```
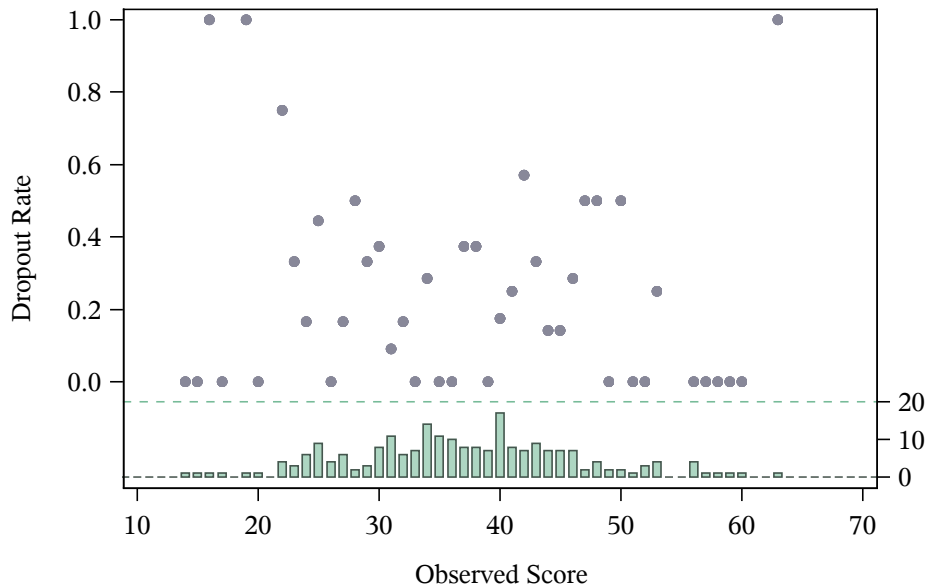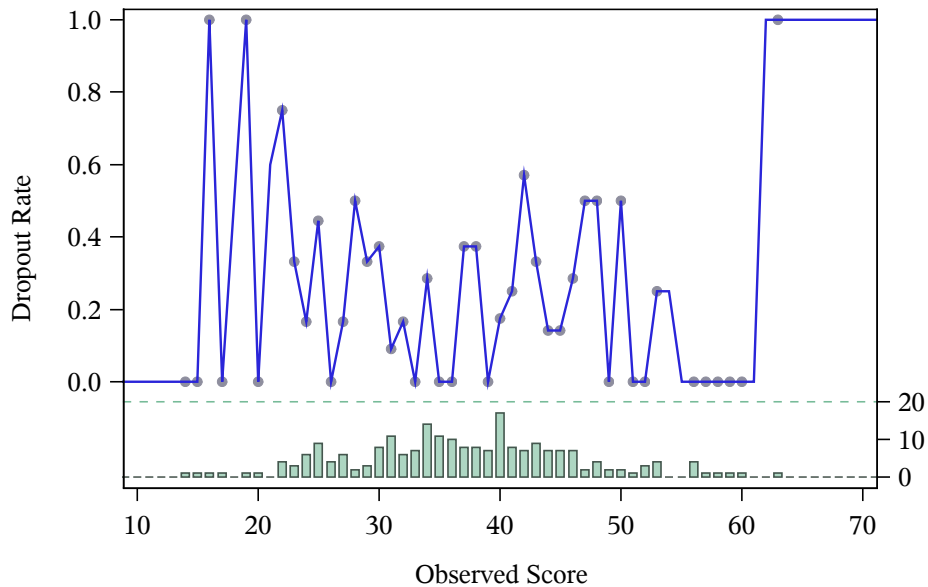
Two natural questions

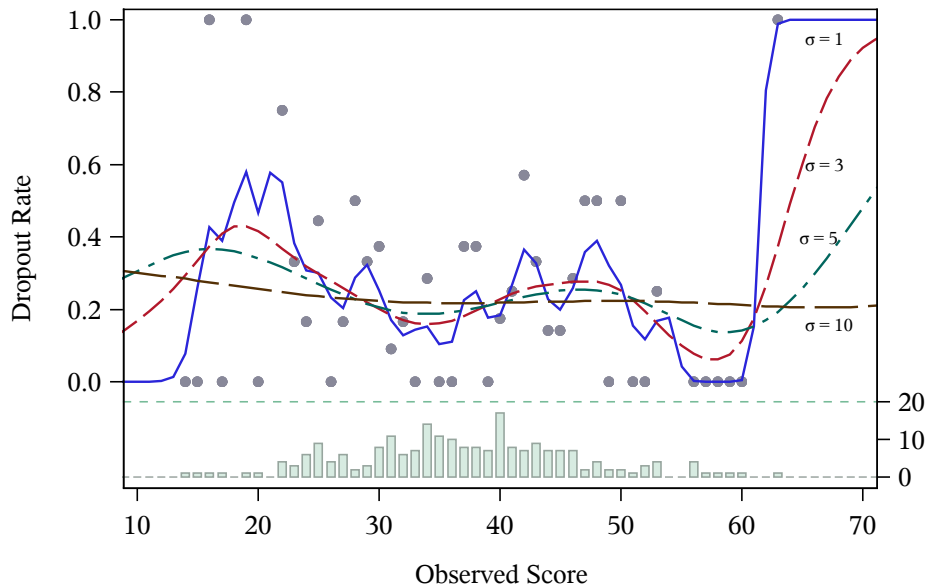1. The probability of dropout or of staying on study

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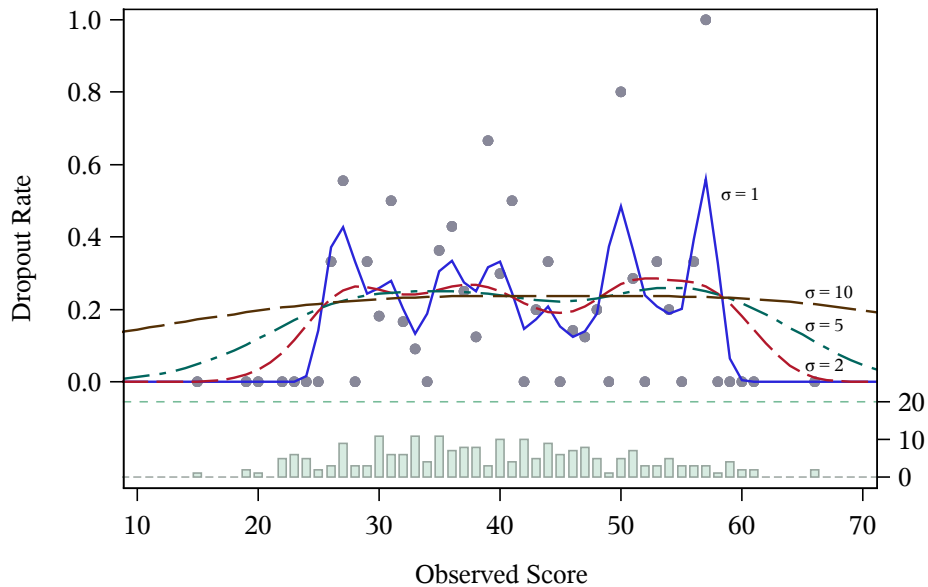
   $Prob[R_t = 1 \mid R_{t-1} = 1, Y_{t-1} = y] \sim smooth(y; \sigma)$

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

Two natural questions

2. The distribution of the outcome

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

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

   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.

The loss function

We choose an optimal σ_P by minimizing the loss function L_P, which up to a constant is

$$L_P(\sigma) = \sum_{t=2}^{N_t} \sum_{j=1}^{n} R_{t-1,j} \left(R_{t,j} - P_{-j}(R_{t,j} = 1 \mid Y_{t-1,j}) \right)^2$$

where P_{-j} is the propability distribution derived from the data after removing the partition to which j belongs.

R: samoneval function

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

```
samoneval(
  mat = MAT,
  Npart = 10,
  LowSigma = 0.1,
  HighSigma = 100,
  IncrementSigma = 0.1
  type = "both"
)
```

- Returns a matrix containing σ_P and its corresponding loss function value and σ_Q and its corresponding loss function value.

```
> library(samon, lib.loc="Rver2.0/samlib")
>
> Results1 <- samoneval(
+     mat              = quet1,
+     Npart            = 10,
+
+     IncrementSigma   = 0.1,
+     LowSigma         = 0.2,
+     HighSigma        = 30.0,
+
+     type             = "both"
+                        )
```

R: samoneval function

```
> Results2 <- samoneval(
+      mat              = quet2,
+      Npart            = 10,
+
+      IncrementSigma   = 0.1,
+      LowSigma         = 0.2,
+      HighSigma        = 30.0,
+
+      type             = "both"
+                         )
> ResultsP <- cbind(Results1[,1:2],
+                   Results2[,1:2])
> PQPlot(ResultsP, "P.pdf", 4.2, 4.2,
+        "Loss function (P)",  c(0,30),
+        c( 2.5, 4.5), c(15,4.5) )
```

The samonev procedure computes the loss function for a range of σ.
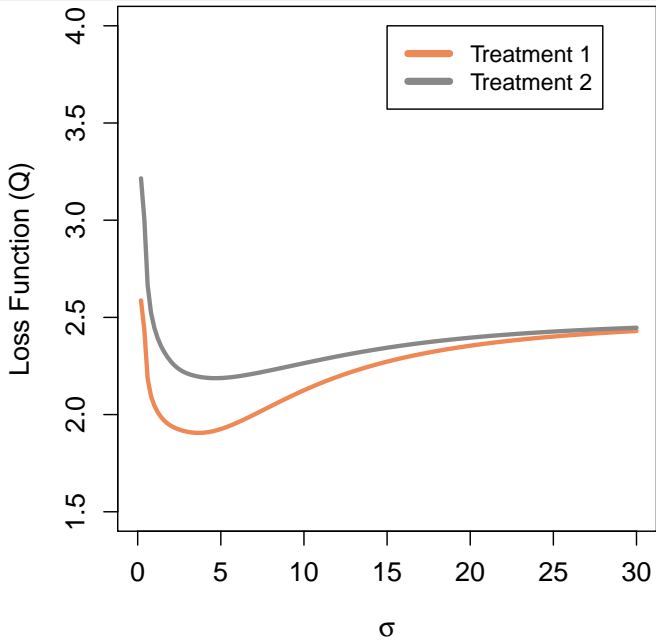
samonev	
data =	Input dataset
out =	Output dataset
npart =	Number of partitions
lowSigma =	low value of sigma
highSigma =	high value of sigma
incSigma =	increment sigma
var varlist	list of variables in time order

samonev procedure

```
proc samonev
   data     = quet1
   out      = ev1
   Npart    = 10
   incSigma = 0.1
   lowSigma = 0.2
   highSigma = 30;

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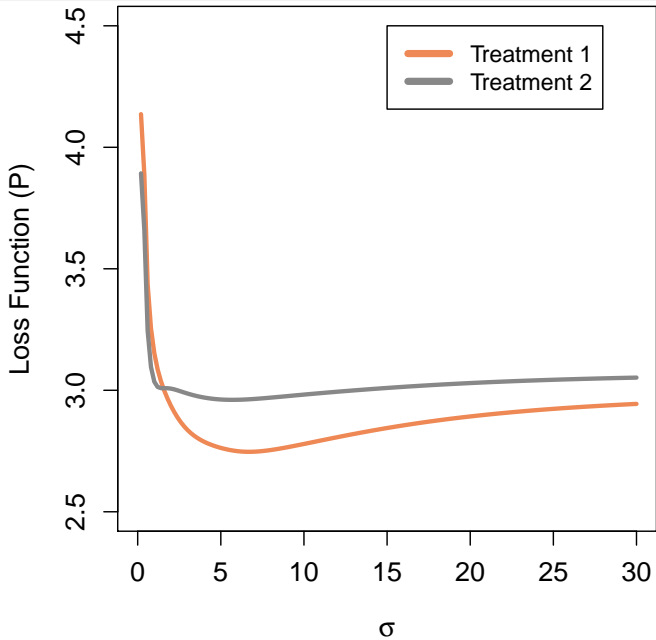
The samon function and the samon procedure can be used to find optimal values of σ_P and σ_Q. Like many optimization techniques providing good initial estimates can improve the efficiency of convergence of the optimization. Within samon we also provide an upper bound for σ_P and σ_Q. Should the algorithm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

R: samon function

```
# Example1.R
# Finding optimal Sigma_p and Sigma_q.
# ----------------------------------------
library(samon, lib.loc="Rver2.0/samlib")

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

    InitialSigmaP  = 6.0,
    HighSigmaP     = 50.0,

    InitialSigmaQ  = 4.0,
    HighSigmaQ     = 50.0

    )

# print the output
print(samonResults)
```

```
* Finding optimal Sigma_p and Sigma_q.;
* ----------------------------------------;
proc samon data    = quet1
           out     = samon1
           Npart   = 10
           pinit   = 6.0
           phigh   = 50.0
           qinit   = 4.0
           qhigh   = 50;
  var v1 - v3;
run;

%split( data = samon1, tag = 1 );
```

```
proc print data = PM1 noobs;
run;

  rc    Niter    Sigma      loss
   2        3    6.6918    2.7468


proc print data = QM1 noobs;
run;

  rc    Niter    Sigma      loss
   2        6    3.6771    1.9057
```

```
proc print data = PM2 noobs;
run;

  rc    Niter    Sigma      loss
  2       3      5.6938    2.9607


proc print data = QM2 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.
- $\alpha = 0$ is missing at random
- α quantifies the influence of Y_{t+1} on the decision to drop-out between t and $t + 1$.

- 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.
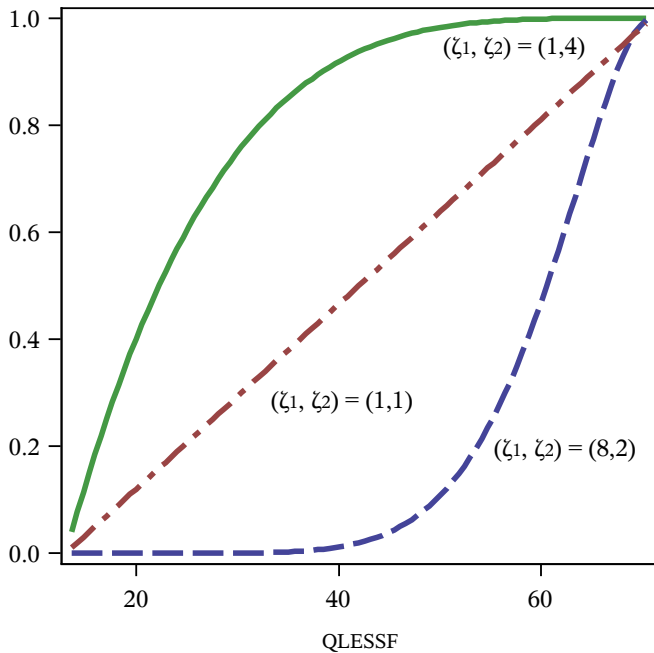
samon procedure

samon	
data =	Input dataset
out =	Output dataset
npart =	Number of partitions
pinit =	initial value for smoothing parameter sigma p
phigh =	Highest value for smoothing parameter sigma p
qinit =	initial value for smoothing parameter sigma q
qhigh =	Highest value for smoothing parameter sigma q
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
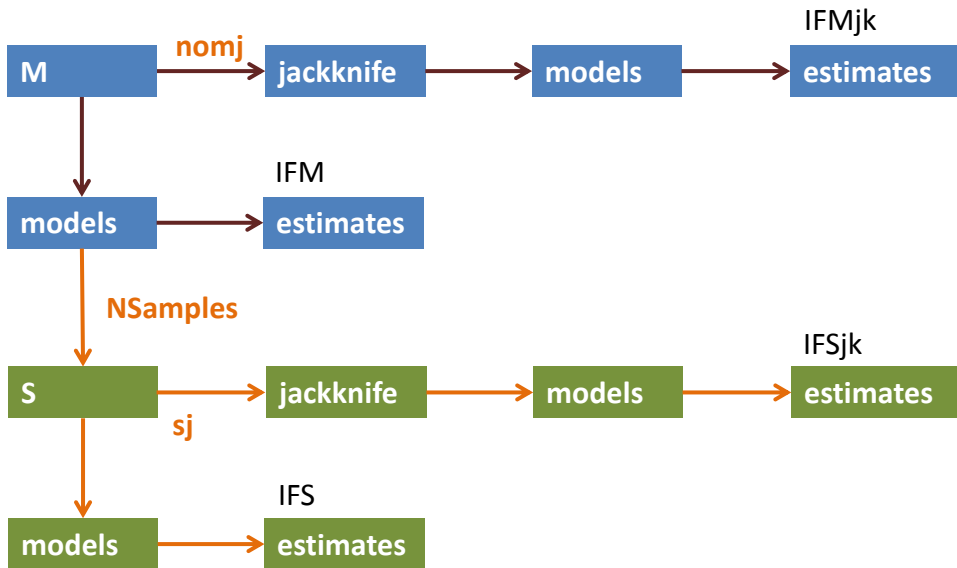
```
proc samon  data = quet1 out = samon1
   Npart = 10
   Pinit =  6.0    PHigh = 50.0
   Qinit =  4.0    QHigh = 50.0

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

  var v1 - v3;
  sensp -10 to 10 by 1;
run;
%split( data = samon1, tag = 1 );
proc print data = IFM1 noobs;
  var alpha AEst AVar IFEst IFVar;
run;
```

```
# Produce one-step influence function estimates
# ------------------------------------------------
library(samon, lib.loc="Rver2.0/samlib")

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

    # initial value and upper bound for sigmaP
    InitialSigmaP      = 6.0,
    HighSigmaP         = 50.0,

    # initial value and upper bound for sigmaQ
    InitialSigmaQ      = 4.0,
    HighSigmaQ         = 50.0,

    LowAlphaList       = -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
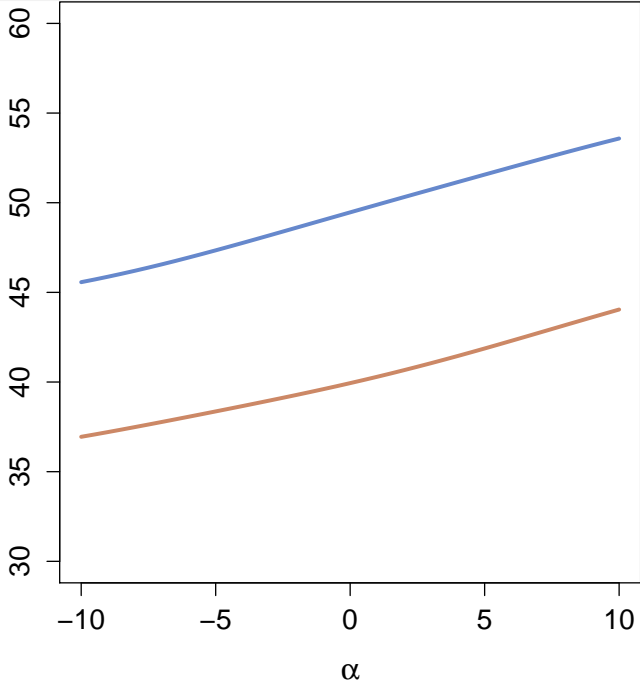
Confidence intervals

- 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

   Pinit   =  6.0    PHigh = 50.0
   QInit   =  4.0    QHigh = 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    = results.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

Estimated QLESSF score at visit 3
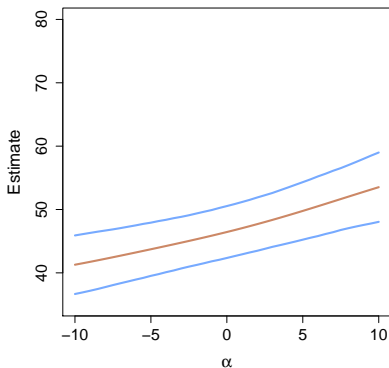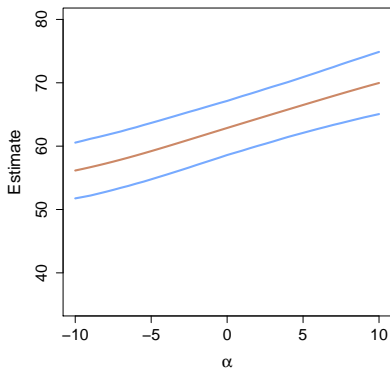
Estimated QLESSF score at visit 3

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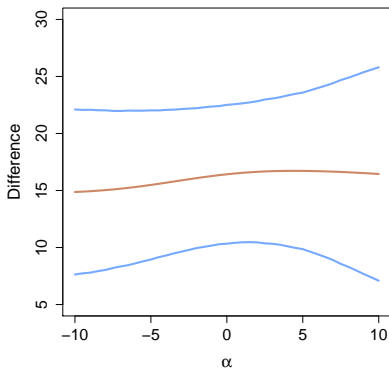
Another useful plot is a surface plot of the difference in the estimated mean value in the two treatment groups given as a function of the two alpha parameters. We use the `samonCrossSummary` function to compute the difference in estimates for each pair of alpha. The plotting is done with the `filled.contour`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

```
%samonCrossSummary(
    IFM1      = data.Summary1,
    sampIF1   = data.sampSummary1,
    IFM2      = data.Summary2,
    sampIF2   = data.sampSummary2,
    out       = data.Cross
    );
```
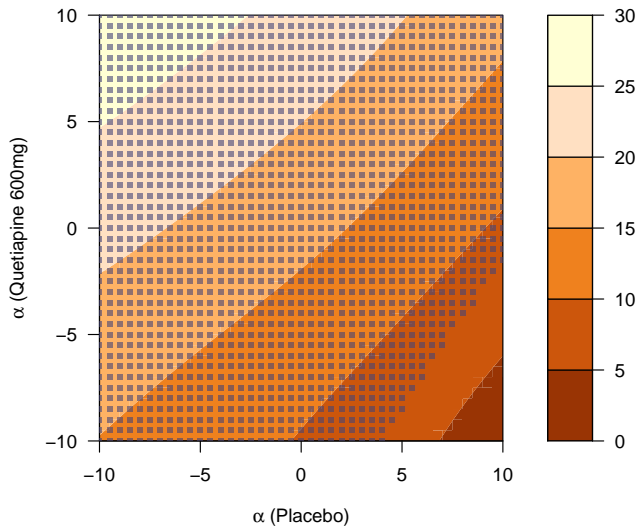
macro	description
samonCombine	Combines results from multiple runs of proc samon
samonSummary	Summarizes samon results. Combines bootstrap and jackknife results to produce confidence intervals
samonDifferenceSummary	Computes treatment effect differences and confidence intervals from a pair of samonSummary 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.