Global Sensitivity Analysis of Randomized Trials with Missing Data
Harvard Shortcourse

Daniel Scharfstein
Johns Hopkins University
dscharf@jhu.edu

September 24, 2019
Funding Acknowledgments

- FDA
- PCORI
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.
"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."
In 2010, the National Research Council (NRC) issued a report 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.*
1998 International Conference of Harmonization (ICH) Guidance document (E9) entitled "Statistical Principles in Clinical Trials" states: "*it is important to evaluate the robustness of the results to various limitations of the data, assumptions, and analytic approaches to data analysis*"

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

European Medicines Agency 2009 draft "Guideline on Missing Data in Confirmatory Clinical Trials" states "*in all submissions with non-negligible amounts of missing data sensitivity analyses should be presented as support to the main 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.*
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
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.
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.
Restrict consideration to follow-up randomized study designs that prescribe that measurements of an outcome of interest are to be taken on each study participant at fixed time-points.

First part of course will focus on monotone missing data pattern. Second part will address how to handle intermittent missing data patterns.

Consider the case where interest is focused on a comparison of treatment arm means at the last scheduled visit in a counterfactual world without missingness.
Patients with bipolar disorder randomized equally to one of three treatment arms: placebo, Quetiapine 300 mg/day or Quetiapine 600 mg/day (Calabrese et al., 2005).

Randomization was stratified by type of bipolar disorder.

Short-form version of the Quality of Life Enjoyment Satisfaction Questionnaire (QLESSF, Endicott et al., 1993), was scheduled to be measured at baseline, week 4 and week 8.

Focus on the subset of 234 patients with bipolar 1 disorder who were randomized to either the placebo (n=116) or 600 mg/day (n=118) arms.
600 mg/day dose was titrated to achieve target by Day 8.

In each treatment group, a dose reduction of 100 mg was allowed to improve tolerability.

At discretion of the investigator, patients could be discontinued from study treatment and assessments at any time.

Patients were free to discontinue their participation in the study at any time.

Use of psychoactive drugs, except lorazepam and zolpidem tartrate during the first 3 weeks, was prohibited. Investigators were allowed to prescribe other medications for the safety and well-being of the participant.
Only 65 patients (56%) in placebo arm and 68 patients (58%) in the 600mg/day arm had a complete set of QLESSF scores.

Patients with complete data tend to have higher average QLESSF scores, suggesting that a complete-case analysis could be biased.
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?
Validity of assumptions will depend on what is *imagined* about treatments that patients receive off-study.

*Not* imagining the continuation of assigned treatment after occurrence of intolerable side effects or lack of efficacy.

Imagining that patients receive treatment as close to the assigned treatment as ethically possible.

The difference of the treatment-specific mean QLESSF outcomes at week 8 under this imaginary, yet plausible, treatment scenario is the target *estimand* of interest.
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.
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.
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.
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.
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.
Y₀, Y₁, Y₂: QLESSF scores scheduled to be collected at baseline, week 4 and week 8.

Let Rₖ be the indicator that Yₖ is observed.

We assume R₀ = 1 and that Rₖ = 0 implies Rₖ₊₁ = 0 (i.e., missingness is monotone).

Patient is on-study at visit k if Rₖ = 1

Patient discontinued prior to visit k if Rₖ = 0

Patient last seen at visit k − 1 if Rₖ₋₁ = 1 and Rₖ = 0.

Yₖ^{obs} equals to Yₖ if Rₖ = 1 and equals to nil if Rₖ = 0.
The observed data for an individual are

\[ O = (Y_0, R_1, Y_{1\text{obs}}, R_2, Y_{2\text{obs}}), \]

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

The superscript \( * \) will be used to denote the true value of the quantity to which it is appended.

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

- \( f(Y_0) \)
- \( P(R_1 = 1|Y_0) \)
- \( f(Y_1|R_1 = 1, Y_0) \)
- \( P(R_2 = 1|R_1 = 1, Y_1, Y_0) \)
- \( f(Y_2|R_2 = 1, Y_1, Y_0) \).
We assume that \( n \) independent and identically distributed copies of \( O \) are observed.

The goal is to use these data to draw inference about 
\[
\mu^* = E^*[Y_2].
\]

When necessary, we will use the subscript \( i \) to denote data for individual \( i \).
• $A_0(y_0)$: 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 $B_2(y_0, y_1)$. 
Mathematically, we can express these assumptions as follows:

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

and

\[ f^*(Y_2|A_1(y_0, y_1)) = f^*(Y_2|B_2(y_0, y_1)) \text{ for all } y_0, y_1 \quad (2) \]
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.
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, $\mu^*$ 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

- $F^*_2(y_2|y_1, y_0) = P^*(Y_2 \leq y_2|B_2(y_1, y_0))$
- $F^*_1(y_1|y_0) = P^*(Y_1 \leq y_1|B_1(y_0))$
- $F^*_0(y_0) = P^*(Y_0 \leq y_0)$. 

Benchmark Assumption (Missing at Random)
The MAR assumption is not the only one that is (a) unverifiable and (b) allows identification of the mean of $Y_2$. 
The first part of the MAR assumption (see (1) above) is

\[ f^*(Y_1, Y_2|A_0(y_0)) = f^*(Y_1, Y_2|B_1(y_0)) \] for all \( y_0 \)

It is equivalent to

\[ f^*(Y_2|A_0(y_0), Y_1 = y_1) = f^*(Y_2|B_1(y_0), Y_1 = y_1) \] for all \( y_0, y_1 \) \hspace{1cm} (3)

and

\[ f^*(Y_1|A_0(y_0)) = f^*(Y_1|B_1(y_0)) \] for all \( y_0 \) \hspace{1cm} (4)
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)
Generalizing (4) using Exponential Tilting

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

Generalizing (2) using Exponential Tilting

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

- \( r(y) \) is a specified increasing function; \( \alpha \) is a sensitivity analysis parameter.
- \( \alpha = 0 \) is MAR.
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 $\alpha$. 
Using Bayes’ rule, we can re-write (3), (5) and (6) as:

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

and

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

where

\[
l_1^*(y_0; \alpha) = \logit P^*(R_1 = 0 | R_0 = 1, Y_0 = y_0) - \log E^*(\exp\{\alpha r(Y_1)\}|B_1(y_0))
\]

and

\[
l_2^*(y_1, y_0; \alpha) = \logit P^*(R_2 = 0 | R_1 = 1, Y_0 = y_0, Y_1 = y_1) - \log E^*(\exp\{\alpha r(Y_2)\}|B_2(y_1, y_0))
\]
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.
Expontential 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 \( r(Y) = Y \),
  \( [Y|R = 0] \sim Bernoulli(\frac{p \exp(\alpha)}{p \exp(\alpha) + 1 - p}) \).
Gamma
For given $\alpha$, $\mu^*$ is identified. Specifically, $\mu^* = \mu(P^*; \alpha)$ equals

$$
\int_{y_0} \int_{y_1} \int_{y_2} \left\{ dF_2^*(y_2|y_1, y_0) \{ 1 - H^*_2(y_1, y_0) \} + \frac{dF_2^*(y_2|y_1, y_0) \exp\{\alpha r(y_2)\}}{\int_{y_2'} dF_2^*(y_2'|y_1, y_0) \exp\{\alpha r(y_2')\}} H^*_2(y_1, y_0) \right\} \times
$$

$$
\left\{ dF_1^*(y_1|y_0) \{ 1 - H^*_1(y_0) \} + \frac{dF_1^*(y_1|y_0) \exp\{\alpha r(y_1)\}}{\int_{y_1'} dF_1^*(y_1'|y_0) \exp\{\alpha r(y_1')\}} H^*_1(y_0) \right\} dF_0^*(y_0)
$$

where

$$
H^*_2(y_1, y_0) = P^*(R_2 = 0| R_1 = 1, Y_1 = y_1, Y_0 = y_0)
$$

and

$$
H^*_1(y_0) = P^*(R_1 = 0| R_0 = 1, Y_0 = y_0)
$$

$\mu^*$ is written as a function of the distribution of the observed data (depending on $\alpha$).
Global Sensitivity Analysis

\[ f^*(Y_1 | A_0(y_0)) \propto f^*(Y_1 | B_1(y_0)) e^{\alpha r(Y_1)} \]

Assumption
- MAR (\( \alpha = 0 \))
- MNAR (\( \alpha < 0 \))
- MNAR (\( \alpha > 0 \))
Global Sensitivity Analysis

For all $y_0, y_1$

Example for $y_0, y_1$

Assumption
- MAR ($\alpha = 0$)
- MNAR ($\alpha < 0$)
- MNAR ($\alpha > 0$)
Global Sensitivity Analysis

\[ f^*(Y_2|A_0(y_0), Y_1 = y_1) = f^*(Y_2|B_1(y_0), Y_1 = y_1) \]
for all \( y_0, y_1 \)

Example for \( y_0, y_1 \)

\[ Y_1 = y_1 \]
\[ A_0(y_0) \]

\[ B_1(y_0) \]

\[ V0, Y1, Y2 \]

\[ A_0(y_0) \]

\[ V0, Y1, Y2 \]

\[ B_1(y_0) \]

\[ V0, Y1, Y2 \]
For given $\alpha$, the above formula shows that $\mu^*$ 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 $\mu^*$ 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.
A non-parametric estimate of $F_2^*(y_2|y_1, y_0)$ would take the form:

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

- 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 plug-in estimator of $\mu^*$ that uses such non-parametric estimators will perform poorly.
We make the estimation task slightly easier by assuming that

\[ F_2^*(y_2 | y_1, y_0) = F_2^*(y_2 | y_1) \]  \hspace{1cm} (7)

and

\[ H_2^*(y_1, y_0) = H_2^*(y_1) \]  \hspace{1cm} (8)
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 $F_2^*(y_2|y_1)$

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

- This estimator will still perform poorly, although better than $\hat{F}_2(y_2|y_1, y_0)$.
- 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 $\sigma_{F_2}$ is a tuning parameter.

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

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

$$CV_{F_2^*(\cdot|\cdot)}(\sigma_{F_2}) = \frac{1}{J} \sum_{j=1}^{J} \left\{ \frac{1}{n_j} \sum_{i \in V_j} R_{2,i} \int \left\{ I(Y_{2,i} \leq y_2) - \hat{F}_2^{(j)}(y_2|Y_{1,i}; \sigma_{F_2}) \right\}^2 \hat{F}_2^o(y_2) \right\}$$

will be small, where $\hat{F}_2^o(y_2)$ is the empirical distribution of $Y_2$ among subjects on-study at visit 2.
For each individual \( i \) in the \( j \)th 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
\[
\{ I(Y_{2,i} \leq y_2) : d\hat{F}_2^0(y_2) > 0 \}
\]
and the corresponding collection of predicted values
\[
\{ \hat{F}_2^{(j)}(y_2 | Y_{1,i}; \sigma_{F_2}) : d\hat{F}_2^0(y_2) > 0 \}.
\]

The distances for each of these individuals are then summed and divided by the number of subjects in the \( j \)th validation set.

An average across the \( J \) validation/training sets is computed.

We can then estimate \( F_2^*(y_2 | y_1) \) by \( \hat{F}_2(y_2 | y_1; \hat{\sigma}_{F_2}) \), where \( \hat{\sigma}_{F_2} = \arg\min CV_{F_2^*}(\cdot | \cdot)(\sigma_{F_2}) \).
We use similar ideas to estimate

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

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

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 $\sigma_P$. 
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 $\mu^*$. The plug-in estimator of $\mu^*$ 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.
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

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

(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 $P^*$. 
Equation (9) plus some simple algebraic manipulation teaches us that

\[
\mu(\hat{P}; \alpha) - \mu(P^*; \alpha)
\]

Plug-in

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

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

\[
+ \text{Rem}(\hat{P}, P^*; \alpha) \tag{12}
\]

where \( \hat{P} \) is the estimated distribution of \( P^* \) discussed in the previous section.
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 multiplied by $\sqrt{n}$).

Provided $\hat{P}$ 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.
The corrected estimator is

\[
\tilde{\mu}_\alpha = \mu(\hat{P}; \alpha) + \frac{1}{n} \sum_{i=1}^{n} \psi(\hat{P}(O_i; \alpha))
\]

The practical implication is that \(\tilde{\mu}_\alpha\) converges in probability to \(\mu^*\) 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.
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 \( \sigma^2_{\alpha} = E^*\left[\psi_{P^*}(O; \alpha)^2\right] \).

The asymptotic variance can be estimated by
\[ \tilde{\sigma}^2_{\alpha} = \frac{1}{n} \sum_{i=1}^{n} \psi_{\hat{P}}(O_i; \alpha)^2. \]

A \((1 - \gamma)\)% Wald-based confidence interval for \( \mu^* \) can be constructed as \( \tilde{\mu}_\alpha \pm z_{1-\gamma/2} \tilde{\sigma}_\alpha / \sqrt{n} \), where \( z_q \) is the \( q \)th quantile of a standard normal random variable.
Let

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

\[ w_1^*(y_0; \alpha) = E^*[\exp\{\alpha r(Y_1)\} \mid R_1 = 1, Y_0 = y_0], \]
\[ w_2^*(y_1; \alpha) = E^*[\exp\{\alpha r(Y_2)\} \mid R_2 = 1, Y_1 = y_1], \]
\[ g_1^*(y_0, y_1; \alpha) = \{1 - H_1^*(y_0)\} w_1^*(y_0; \alpha) + \exp\{\alpha r(y_1)\} H_1^*(y_0). \]
\[ g_2^*(y_1, y_2; \alpha) = \{1 - H_2^*(y_1)\} w_2^*(y_1; \alpha) + \exp\{\alpha r(y_2)\} H_2^*(y_1). \]
\[ \psi_{p^*}(O; \alpha) := a^*_0(Y_0; \alpha) + R_1 b^*_1(Y_0, Y_1; \alpha) + R_2 b^*_2(Y_1, Y_2; \alpha) + \{1 - R_1 - H^*_1(Y_0)\} c^*_1(Y_0; \alpha) + R_1\{1 - R_2 - H^*_2(Y_1)\} c^*_2(Y_1; \alpha) \]

where
Inference - Efficient Influence Function/Gradient

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

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

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

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

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

\[
- E^* \left[ \frac{R_2 Y_2}{\pi^*(Y_0, Y_1, Y_2; \alpha)} \right] \frac{1}{g_2^*(Y_1, Y_2; \alpha)} R_1 = 1, Y_1 w_2^*(Y_1; \alpha)
\]
Wald-based confidence intervals don’t always have adequate coverage properties in finite samples.

In equal-tailed studentized bootstrap, the confidence interval takes the form $[\hat{\mu} - t_{0.975} \hat{se}(\hat{\mu})$, $\hat{\mu} - t_{0.025} \hat{se}(\hat{\mu})]$, where $t_q$ is the $q$th quantile of $\{\frac{\hat{\mu}(b) - \hat{\mu}}{\hat{se}(\hat{\mu}(b))} : b = 1, \ldots, B\}$.

In symmetric studentized bootstrap, the confidence interval takes the form $[\hat{\mu} - t^{*}_{0.95} \hat{se}(\hat{\mu})$, $\hat{\mu} + t^{*}_{0.95} \hat{se}(\hat{\mu})]$, where $t^{*}_{0.95}$ is selected so that 95% of the distribution of $\{\frac{\hat{\mu}(b) - \hat{\mu}}{\hat{se}(\hat{\mu}(b))} : b = 1, \ldots, B\}$ falls between $-t^{*}_{0.95}$ and $t^{*}_{0.95}$.

Useful to replace influence-function based standard error estimator with jackknife standard error estimator.
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%.
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.
Under MAR, the estimated values of $\mu^*$ 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.
Figure: Treatment-specific (left: placebo; right: 600 mg/day Quetiapine) estimates (along with 95% pointwise confidence intervals) of $\mu^*$ as a function of $\alpha$. 
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 $\alpha$. 

![Graph showing the difference in mean QLESSF between non-completers and completers for Placebo and Quetiapine (600mg) as a function of $\alpha$.]
Figure: Contour plot of the estimated differences between mean QLESSF at Visit 2 for Quetiapine vs. placebo for various treatment-specific combinations of the sensitivity analysis parameters.
Only when the sensitivity analysis are highly differential (e.g., $\alpha(\text{placebo}) = 8$ and $\alpha(\text{Quetiapine}) = -8$) are the differences no longer statistically significant.
Quetiapine Bipolar Trial - Sensitivity Analysis
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 $\alpha$, these true data generating mechanisms can be mapped to a true value of $\mu^*$.
- For each dataset, the sample size was to set to 116 and 118 in the placebo and Quetiapine arms, respectively.
## Simulation Study - Bias/MSE

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Estimator</th>
<th>Placebo</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu^*$</td>
<td>Bias</td>
<td>MSE</td>
</tr>
<tr>
<td>-10</td>
<td>Plug-in</td>
<td>40.85</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Corrected</td>
<td>0.43</td>
<td>4.56</td>
</tr>
<tr>
<td>-5</td>
<td>Plug-in</td>
<td>43.45</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>Corrected</td>
<td>0.27</td>
<td>4.26</td>
</tr>
<tr>
<td>-1</td>
<td>Plug-in</td>
<td>46.02</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>Corrected</td>
<td>0.18</td>
<td>4.22</td>
</tr>
<tr>
<td>0</td>
<td>Plug-in</td>
<td>46.73</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>Corrected</td>
<td>0.17</td>
<td>4.27</td>
</tr>
<tr>
<td>1</td>
<td>Plug-in</td>
<td>47.45</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>Corrected</td>
<td>0.16</td>
<td>4.36</td>
</tr>
<tr>
<td>5</td>
<td>Plug-in</td>
<td>50.48</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>Corrected</td>
<td>0.14</td>
<td>5.11</td>
</tr>
<tr>
<td>10</td>
<td>Plug-in</td>
<td>54.07</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>Corrected</td>
<td>0.04</td>
<td>6.30</td>
</tr>
</tbody>
</table>
## Simulation Study - Coverage

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Procedure</th>
<th>Placebo Coverage</th>
<th>Quetiapine Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10</td>
<td>Wald-IF</td>
<td>91.5%</td>
<td>90.5%</td>
</tr>
<tr>
<td></td>
<td>Wald-JK</td>
<td>95.0%</td>
<td>94.6%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-ET</td>
<td>94.3%</td>
<td>93.8%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-ET</td>
<td>94.4%</td>
<td>93.4%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-S</td>
<td>95.2%</td>
<td>94.6%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-S</td>
<td>95.0%</td>
<td>94.6%</td>
</tr>
<tr>
<td>-5</td>
<td>Wald-IF</td>
<td>93.5%</td>
<td>92.9%</td>
</tr>
<tr>
<td></td>
<td>Wald-JK</td>
<td>95.0%</td>
<td>94.8%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-ET</td>
<td>95.2%</td>
<td>94.6%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-ET</td>
<td>94.8%</td>
<td>94.6%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-S</td>
<td>95.4%</td>
<td>95.2%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-S</td>
<td>95.1%</td>
<td>95.2%</td>
</tr>
<tr>
<td>-1</td>
<td>Wald-IF</td>
<td>93.9%</td>
<td>94.2%</td>
</tr>
<tr>
<td></td>
<td>Wald-JK</td>
<td>94.9%</td>
<td>95.4%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-ET</td>
<td>95.1%</td>
<td>94.8%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-ET</td>
<td>95.1%</td>
<td>94.6%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-S</td>
<td>95.3%</td>
<td>96.4%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-S</td>
<td>95.1%</td>
<td>96.3%</td>
</tr>
<tr>
<td>0</td>
<td>Wald-IF</td>
<td>93.8%</td>
<td>94.0%</td>
</tr>
<tr>
<td></td>
<td>Wald-JK</td>
<td>95.0%</td>
<td>95.4%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-ET</td>
<td>94.6%</td>
<td>94.5%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-ET</td>
<td>94.6%</td>
<td>94.6%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-S</td>
<td>95.5%</td>
<td>96.6%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-S</td>
<td>95.2%</td>
<td>96.7%</td>
</tr>
</tbody>
</table>
## Simulation Study - Coverage

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Procedure</th>
<th>Placebo Coverage</th>
<th>Quetiapine Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wald-IF</td>
<td>93.3%</td>
<td>93.7%</td>
</tr>
<tr>
<td></td>
<td>Wald-JK</td>
<td>95.1%</td>
<td>95.5%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-ET</td>
<td>94.6%</td>
<td>94.6%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-ET</td>
<td>94.6%</td>
<td>94.6%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-S</td>
<td>95.5%</td>
<td>96.5%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-S</td>
<td>95.2%</td>
<td>96.5%</td>
</tr>
<tr>
<td>5</td>
<td>Wald-IF</td>
<td>90.8%</td>
<td>91.3%</td>
</tr>
<tr>
<td></td>
<td>Wald-JK</td>
<td>95.3%</td>
<td>95.7%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-ET</td>
<td>93.2%</td>
<td>91.6%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-ET</td>
<td>93.8%</td>
<td>93.0%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-S</td>
<td>95.5%</td>
<td>95.4%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-S</td>
<td>95.8%</td>
<td>96.4%</td>
</tr>
<tr>
<td>10</td>
<td>Wald-IF</td>
<td>85.4%</td>
<td>87.8%</td>
</tr>
<tr>
<td></td>
<td>Wald-JK</td>
<td>94.9%</td>
<td>94.5%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-ET</td>
<td>88.2%</td>
<td>87.0%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-ET</td>
<td>92.2%</td>
<td>89.7%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-S</td>
<td>94.6%</td>
<td>93.9%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-S</td>
<td>95.5%</td>
<td>95.1%</td>
</tr>
</tbody>
</table>
1. R
   * samon library
   * functions with pass to C code

2. SAS
   * procedures and macros
### The Quet1/Quet2 Dataset

The following code snippet prints the Quet1 dataset:

```sas
proc print data = quet1;
run;
```

<table>
<thead>
<tr>
<th>Obs</th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>36</td>
<td>56</td>
<td>57</td>
</tr>
<tr>
<td>2</td>
<td>32</td>
<td>32</td>
<td>31</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>42</td>
<td>27</td>
</tr>
<tr>
<td>4</td>
<td>48</td>
<td>52</td>
<td>54</td>
</tr>
<tr>
<td>5</td>
<td>27</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>6</td>
<td>38</td>
<td>43</td>
<td>.</td>
</tr>
<tr>
<td>7</td>
<td>35</td>
<td>44</td>
<td>39</td>
</tr>
<tr>
<td>8</td>
<td>24</td>
<td>26</td>
<td>28</td>
</tr>
<tr>
<td>9</td>
<td>25</td>
<td>30</td>
<td>.</td>
</tr>
<tr>
<td>10</td>
<td>25</td>
<td>28</td>
<td>.</td>
</tr>
<tr>
<td>11</td>
<td>40</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>12</td>
<td>14</td>
<td>16</td>
<td>.</td>
</tr>
<tr>
<td>13</td>
<td>34</td>
<td>37</td>
<td>.</td>
</tr>
<tr>
<td>14</td>
<td>27</td>
<td>25</td>
<td>.</td>
</tr>
<tr>
<td>15</td>
<td>26</td>
<td>25</td>
<td>.</td>
</tr>
<tr>
<td>16</td>
<td>38</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>17</td>
<td>44</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>18</td>
<td>35</td>
<td>31</td>
<td>.</td>
</tr>
<tr>
<td>19</td>
<td>25</td>
<td>45</td>
<td>49</td>
</tr>
<tr>
<td>20</td>
<td>45</td>
<td>53</td>
<td>56</td>
</tr>
<tr>
<td>21</td>
<td>32</td>
<td>38</td>
<td>48</td>
</tr>
<tr>
<td>22</td>
<td>46</td>
<td>45</td>
<td>.</td>
</tr>
<tr>
<td>23</td>
<td>30</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>24</td>
<td>43</td>
<td>56</td>
<td>43</td>
</tr>
<tr>
<td>25</td>
<td>42</td>
<td>44</td>
<td>41</td>
</tr>
<tr>
<td>26</td>
<td>34</td>
<td>45</td>
<td>47</td>
</tr>
<tr>
<td>27</td>
<td>34</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>
The `samonDataCheck` macro can be used to check the missing data pattern in the data.

```sas
%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);

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

```r
samonDataCheck( quet1 )
```

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

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

Missing Patterns:

<table>
<thead>
<tr>
<th>Pattern</th>
<th>N</th>
<th>proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>*__</td>
<td>10</td>
<td>0.0862</td>
</tr>
<tr>
<td>**_</td>
<td>41</td>
<td>0.3534</td>
</tr>
<tr>
<td>***</td>
<td>65</td>
<td>0.5603</td>
</tr>
</tbody>
</table>
Samon Data Check:
-----------------------------------------------
Number of time-points: 3
Number of subjects: 118
Minimum observed value: 15
Maximum observed value: 67
Average number of timepoints on study: 2.47
Total number of observed values: 291
Subjects observed at final timepoint: 68
Subjects observed at all timepoint: 68

Missing Patterns:

<table>
<thead>
<tr>
<th>Pattern</th>
<th>N</th>
<th>proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>*__</td>
<td>13</td>
<td>0.1102</td>
</tr>
<tr>
<td>**_</td>
<td>37</td>
<td>0.3136</td>
</tr>
<tr>
<td>***</td>
<td>68</td>
<td>0.5763</td>
</tr>
</tbody>
</table>
R: samoneval function

- The `samoneval` function can be used to compute the loss function for a range of $\sigma$.

- Takes four arguments:

  ```r
  samoneval(
    mat = ,  # input matrix to evaluate
    Npart = 10,  # number of partitions
    sigmaList = c(0,1),  # vector of sigma values
    type = "both"  # compute the loss function for $\sigma_H$, $\sigma_F$, or both $\sigma_H$ and $\sigma_F$
  )
  ``

- Returns a matrix containing $\sigma_H$ and its corresponding loss function value and $\sigma_F$ and its corresponding loss function value.
> library(samon, lib.loc="../samlib")

> Results1 <- samoneval(
+   mat = quet1,
+   Npart = 10,
+   sigmaList = seq(0.2, 30.0, by=0.1),
+   type = "both"
+ )
> Results2 <- samoneval(
+   mat = quet2, 
+   Npart = 10, 
+   sigmaList = seq(0.2,30.0,by=0.1), 
+   type = "both"
+ )

> ResultsH <- cbind(Results1[,1:2],
+   Results2[,1:2])

> HFPlot(ResultsH, "H.pdf", 4.2, 4.2,
+   "Loss function (H)", c(0,30),
+   c( 2.5, 4.5), c(15,4.5) )
The `samonev` procedure computes the loss function for a range of $\sigma$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>samonev</code></td>
<td></td>
</tr>
<tr>
<td><code>data =</code></td>
<td>Input dataset</td>
</tr>
<tr>
<td><code>out =</code></td>
<td>Output dataset</td>
</tr>
<tr>
<td><code>npart =</code></td>
<td>Number of partitions</td>
</tr>
<tr>
<td><code>var varlist</code></td>
<td>list of variables in time order</td>
</tr>
<tr>
<td><code>sigma sigmalist</code></td>
<td>list of sigmas</td>
</tr>
</tbody>
</table>
proc samonev
  data = quet1
  out = ev1
  Npart = 10;

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

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

samonResults <- samon(
    mat = quet1,
    Npart = 10,
    InitialSigmaH = 6.0,
    HighSigmaH = 50.0,
    InitialSigmaF = 4.0,
    HighSigmaF = 50.0
)

# print the output
print(samonResults)
* Finding optimal Sigma_h and Sigma_f.;
* ----------------------------------------;
proc samon data = quet1
   out = samon1
   Npart = 10
   Hinit = 6.0
   Hhigh = 50.0
   Finit = 4.0
   Fhigh = 50
   Hout = HM1
   Fout = FM1;
   var v1 - v3;
run;
### Treatment 1

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

rc   Niter   Sigma   loss
  2     3   6.6918   2.7468

proc print data = FM1 noobs;
run;

crc  Niter  Sigma  loss
  2   6   3.6771  1.9057
```
proc print data = HM2 noobs;
run;

rc   Niter   Sigma   loss
 2     3   5.6938   2.9607

proc print data = FM2 noobs;
run;

rc   Niter   Sigma   loss
 2     3   4.6704   2.1872
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 $\alpha$ provides the mechanism by which we measure the sensitivity of the results to informative drop-out.
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 $\zeta_1$ and $\zeta_2$ to be 1.
### samon procedure

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>data</td>
<td>Input dataset</td>
</tr>
<tr>
<td>IFout</td>
<td>Influence function estimates</td>
</tr>
<tr>
<td>npart</td>
<td>Number of partitions</td>
</tr>
<tr>
<td>Hinit</td>
<td>initial value for smoothing parameter ( \sigma_H )</td>
</tr>
<tr>
<td>Hhigh</td>
<td>Highest value for smoothing parameter ( \sigma_H )</td>
</tr>
<tr>
<td>Finit</td>
<td>initial value for smoothing parameter ( \sigma_F )</td>
</tr>
<tr>
<td>Fhigh</td>
<td>Highest value for smoothing parameter ( \sigma_F )</td>
</tr>
<tr>
<td>lb</td>
<td>lower bound of data</td>
</tr>
<tr>
<td>ub</td>
<td>upper bound of data</td>
</tr>
<tr>
<td>zeta1</td>
<td>parameter for cumulative beta distribution</td>
</tr>
<tr>
<td>zeta2</td>
<td>parameter for cumulative beta distribution</td>
</tr>
<tr>
<td>nsamples</td>
<td>Number of bootstrap samples</td>
</tr>
<tr>
<td>seed0</td>
<td>Seed to pass to random number generator</td>
</tr>
<tr>
<td>var varlist</td>
<td>list of variables in time order</td>
</tr>
<tr>
<td>sensp senslist</td>
<td>list of sensitivity parameters</td>
</tr>
</tbody>
</table>

*varlist, senslist, varlist, senslist* are lists of variables or sensitivity parameters, respectively.
proc samon data = quet1 IFout = IFM1
   Npart = 10
   Hinit = 6.0   HHigh = 50.0
   Finit = 4.0   FHigh = 50.0
   lb = 13      ub = 71
   zeta1 = 1.0  zeta2 = 1.0
   nomj nsamples = 0 ;

   var v1 - v3;
   sensp -10 to 10 by 1;
run;
proc print data = IFM1 noobs;
   var alpha AEst AVar IFEst IFVar;
run;
# Produce one-step influence function estimates
# ----------------------------------------------
library(samon, lib.loc="..:/samlib")

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

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

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

    AlphaList = -10:10,  # alphas
    lb = 13, ub = 71,
    zeta1 = 1.0, zeta2 = 1.0
)
<table>
<thead>
<tr>
<th>alpha</th>
<th>AEst</th>
<th>AVar</th>
<th>IFEst</th>
<th>IFVar</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10</td>
<td>36.6909</td>
<td>0.08754</td>
<td>36.9454</td>
<td>1.3374</td>
</tr>
<tr>
<td>-9</td>
<td>36.9556</td>
<td>0.09050</td>
<td>37.2108</td>
<td>1.3438</td>
</tr>
<tr>
<td>-8</td>
<td>37.2402</td>
<td>0.09354</td>
<td>37.4881</td>
<td>1.3497</td>
</tr>
<tr>
<td>-7</td>
<td>37.5435</td>
<td>0.09653</td>
<td>37.7740</td>
<td>1.3547</td>
</tr>
<tr>
<td>-6</td>
<td>37.8641</td>
<td>0.09934</td>
<td>38.0654</td>
<td>1.3585</td>
</tr>
<tr>
<td>-5</td>
<td>38.2007</td>
<td>0.10185</td>
<td>38.3611</td>
<td>1.3609</td>
</tr>
<tr>
<td>-4</td>
<td>38.5526</td>
<td>0.10395</td>
<td>38.6609</td>
<td>1.3625</td>
</tr>
<tr>
<td>-3</td>
<td>38.9189</td>
<td>0.10556</td>
<td>38.9660</td>
<td>1.3641</td>
</tr>
<tr>
<td>-2</td>
<td>39.2993</td>
<td>0.10660</td>
<td>39.2787</td>
<td>1.3668</td>
</tr>
<tr>
<td>-1</td>
<td>39.6935</td>
<td>0.10701</td>
<td>39.6020</td>
<td>1.3715</td>
</tr>
<tr>
<td>0</td>
<td>40.1010</td>
<td>0.10678</td>
<td>39.9386</td>
<td>1.3792</td>
</tr>
<tr>
<td>1</td>
<td>40.5210</td>
<td>0.10590</td>
<td>40.2911</td>
<td>1.3898</td>
</tr>
<tr>
<td>2</td>
<td>40.9517</td>
<td>0.10436</td>
<td>40.6609</td>
<td>1.4026</td>
</tr>
<tr>
<td>3</td>
<td>41.3907</td>
<td>0.10222</td>
<td>41.0484</td>
<td>1.4160</td>
</tr>
<tr>
<td>4</td>
<td>41.8343</td>
<td>0.09950</td>
<td>41.4525</td>
<td>1.4278</td>
</tr>
<tr>
<td>5</td>
<td>42.2785</td>
<td>0.09626</td>
<td>41.8710</td>
<td>1.4356</td>
</tr>
<tr>
<td>6</td>
<td>42.7186</td>
<td>0.09252</td>
<td>42.3006</td>
<td>1.4374</td>
</tr>
<tr>
<td>7</td>
<td>43.1501</td>
<td>0.08832</td>
<td>42.7372</td>
<td>1.4320</td>
</tr>
<tr>
<td>8</td>
<td>43.5690</td>
<td>0.08370</td>
<td>43.1760</td>
<td>1.4190</td>
</tr>
<tr>
<td>9</td>
<td>43.9724</td>
<td>0.07871</td>
<td>43.6127</td>
<td>1.3989</td>
</tr>
<tr>
<td>10</td>
<td>44.3586</td>
<td>0.07345</td>
<td>44.0435</td>
<td>1.3729</td>
</tr>
</tbody>
</table>
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 $50 \times 100 \times 1000 = 5$ million estimates.
proc samon data = quet1 out = samon1
   Npart = 10

   Hinit = 6.0      HHigh = 50.0
   FInit = 4.0      FHigh = 50.0
   lb    = 13       ub    = 71
   zeta1 = 1.0      zeta2 = 1.0
   NSamples = 500   seed0 = 81881
   sj;
   var v1-v3;
   sensp -10 to 10 by 1;
run;
%samonSummary(
    data = results1,
    out  = data.Summary1,
    sampout = data.sampSummary1
);
proc print data=data.Summary1;
    var alpha IFEst lb ub;
run;
<table>
<thead>
<tr>
<th>alpha</th>
<th>IFEst</th>
<th>lb</th>
<th>ub</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10</td>
<td>36.9454</td>
<td>34.2645</td>
<td>39.6263</td>
</tr>
<tr>
<td>-9</td>
<td>37.2108</td>
<td>34.5635</td>
<td>39.8580</td>
</tr>
<tr>
<td>-8</td>
<td>37.4881</td>
<td>34.9005</td>
<td>40.0757</td>
</tr>
<tr>
<td>-7</td>
<td>37.7740</td>
<td>35.2496</td>
<td>40.2984</td>
</tr>
<tr>
<td>-6</td>
<td>38.0654</td>
<td>35.5840</td>
<td>40.5468</td>
</tr>
<tr>
<td>-5</td>
<td>38.3611</td>
<td>35.9267</td>
<td>40.7955</td>
</tr>
<tr>
<td>-4</td>
<td>38.6609</td>
<td>36.2627</td>
<td>41.0590</td>
</tr>
<tr>
<td>-3</td>
<td>38.9660</td>
<td>36.6038</td>
<td>41.3281</td>
</tr>
<tr>
<td>-2</td>
<td>39.2787</td>
<td>36.9218</td>
<td>41.6356</td>
</tr>
<tr>
<td>-1</td>
<td>39.6020</td>
<td>37.2546</td>
<td>41.9493</td>
</tr>
<tr>
<td>0</td>
<td>39.9386</td>
<td>37.5628</td>
<td>42.3144</td>
</tr>
<tr>
<td>1</td>
<td>40.2911</td>
<td>37.8957</td>
<td>42.6866</td>
</tr>
<tr>
<td>2</td>
<td>40.6609</td>
<td>38.2211</td>
<td>43.1007</td>
</tr>
<tr>
<td>3</td>
<td>41.0484</td>
<td>38.5820</td>
<td>43.5148</td>
</tr>
<tr>
<td>4</td>
<td>41.4525</td>
<td>38.9005</td>
<td>44.0045</td>
</tr>
<tr>
<td>5</td>
<td>41.8710</td>
<td>39.2378</td>
<td>44.5043</td>
</tr>
<tr>
<td>6</td>
<td>42.3006</td>
<td>39.5695</td>
<td>45.0318</td>
</tr>
<tr>
<td>7</td>
<td>42.7372</td>
<td>39.9254</td>
<td>45.5490</td>
</tr>
<tr>
<td>8</td>
<td>43.1760</td>
<td>40.2880</td>
<td>46.0641</td>
</tr>
<tr>
<td>9</td>
<td>43.6127</td>
<td>40.5862</td>
<td>46.6392</td>
</tr>
<tr>
<td>10</td>
<td>44.0435</td>
<td>40.8705</td>
<td>47.2165</td>
</tr>
</tbody>
</table>
We use the `samonCrossSummary` function to compute the difference in estimates for each pair of alpha.
%samonCrossSummary(
    IFM1 = data.Summary1,
    sampIF1 = data.sampSummary1,
    IFM2 = data.Summary2,
    sampIF2 = data.sampSummary2,
    out = data.Cross
);
<table>
<thead>
<tr>
<th>macro</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>samonCombine</td>
<td>Combines results from multiple runs of proc samon</td>
</tr>
<tr>
<td>samonSummary</td>
<td>Summarizes samon results. Combines bootstrap and jackknife results to produce confidence intervals</td>
</tr>
<tr>
<td>samonDifferenceSummary</td>
<td>Computes treatment effect differences and confidence intervals from a pair of samonSummary objects.</td>
</tr>
<tr>
<td>samonCrossSummary</td>
<td>Computes treatment effect differences and confidence intervals for each pair of sensitivity parameters $\alpha$.</td>
</tr>
<tr>
<td>samonECompleterStatus</td>
<td>Computes the difference in the expected value of non-completers and completers</td>
</tr>
</tbody>
</table>
**samonCombine macro**

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

<table>
<thead>
<tr>
<th>samonSummary macro</th>
<th>computes summary of samon object</th>
</tr>
</thead>
<tbody>
<tr>
<td>(</td>
<td></td>
</tr>
<tr>
<td>data =</td>
<td>input dataset to summarize</td>
</tr>
<tr>
<td>out =</td>
<td>summary of main data</td>
</tr>
<tr>
<td>sampSummary =</td>
<td>summary of parametric bootstrap</td>
</tr>
<tr>
<td>)</td>
<td>samples</td>
</tr>
</tbody>
</table>
The `samonDifferenceSummary` macro is used to calculate treatment-specific differences. The macro takes the following parameters:

- `IFM1 =` which is the main results from `samonSummary` for trt 1
- `sampIF1 =` which is the sample results from `samonSummary` for trt 1
- `IFM2 =` which is the main results from `samonSummary` for trt 2
- `sampIF2 =` which is the sample results from `samonSummary` for trt 2
- `out =` which is the summary of the difference

The macro calculates the difference between the main results and the sample results for each treatment and outputs the summary of the difference.
**samonCrossSummary macro**

| 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 |

Treatment-specific differences for all pairs of sensitivity parameter
<table>
<thead>
<tr>
<th>function</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>samonCombine</td>
<td>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.</td>
</tr>
<tr>
<td>samonDiff</td>
<td>Takes two samonMat objects and produces a samonMat object for the difference in influence function estimates</td>
</tr>
<tr>
<td>samonBiasCorrection</td>
<td>Takes a samonMat object and produces corrected influence function estimates</td>
</tr>
<tr>
<td>samonXBiasCorrection</td>
<td>Takes two samonMat objects (one from each treatment groups) and for each pair of alphas produces the difference in influence function estimates.</td>
</tr>
</tbody>
</table>
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 $z = (z_1, z_2, \ldots, z_K)$,
  - $\bar{z}_k = (z_0, z_1, \ldots, z_k)$
  - $\underline{z}_k = (z_{k+1}, z_{k+2}, \ldots, z_K)$.
- For each individual, the data unit $O = (C, \overline{Y}_C)$ is drawn from some distribution $P^*$ contained in the non-parametric model $M$ of distributions.
- The observed data consist of $n$ independent draws $O_1, O_2, \ldots, O_n$ from $P^*$. 
By factorizing the distribution of $O$ in terms of chronologically ordered conditional distributions, any distribution $P \in \mathcal{M}$ can be represented by

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

The main objective is to draw inference about $\mu^* := E^*(Y_K)$, the true mean outcome at visit $K$ in a hypothetical world in which all patients are followed to that visit.
For every $\bar{y}_k$, define the following strata:

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

- $B_{k+1}(\bar{y}_k)$: patients on-study at visit $k + 1$ (i.e., $R_{k+1} = 1$) with $\bar{Y}_k = \bar{y}_k$. 
For all $\bar{y}_k$, the distribution of $Y_k$ for patients in stratum $A_k(\bar{y}_k)$ is the same as the distribution of $Y_k$ for patients in stratum $B_{k+1}(\bar{y}_k)$.

Mathematically, we can express these assumptions as follows:

$$f^*(Y_k | A_k(\bar{y}_k)) = f^*(Y_k | B_{k+1}(\bar{y}_k)) \text{ for all } \bar{y}_k \quad (13)$$
Using Bayes’ rule, we can re-write these expressions as:

\[ P^*\left( R_{k+1} = 0 \mid R_k = 1, \bar{Y}_K = \bar{y}_K \right) \]

\[ = P^*\left( R_{k+1} = 0 \mid R_k = 1, \bar{Y}_k = \bar{y}_k \right) \text{ for all } \bar{y}_K \]

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

*Among those on study at visit \( k \), the decision to discontinue the study before the next visit is like the flip of a coin with probability depending only on the observable history of outcomes through visit \( k \) (i.e., no outcomes after visit \( k \)).*
Under missing at random, $\mu^*$ is identified. That is, it can be expressed as a functional of the distribution of the observed data. Specifically, $\mu^* = \mu(P^*)$ is

$$\int_{y_0} \cdots \int_{y_K} y_K \left\{ \prod_{k=0}^{K-1} dF_{k+1}^*(y_{k+1}|\overline{y}_k) \right\} dF_0^*(y_0)$$
Equation (13) is equivalent to the following two assumptions:

\[
f^* \left( Y_{k+1} | A_k(\bar{y}_k), Y_{k+1} = y_{k+1} \right) \\
= f^* \left( Y_{k+1} | B_{k+1}(\bar{y}_k), Y_{k+1} = y_{k+1} \right) \text{ for all } \bar{y}_{k+1}
\]  

(14)

and

\[
f^* \left( Y_{k+1} | A_k(\bar{y}_k) \right) = f^* \left( Y_{k+1} | B_{k+1}(\bar{y}_k) \right) \text{ for all } \bar{y}_k
\]  

(15)

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

For all \( \bar{y}_{k+1} \), the distribution of \( Y_{k+1} \) for patients in stratum \( A_k(\bar{y}_k) \) with \( Y_{k+1} = y_{k+1} \) is the same as the distribution of \( Y_{k+1} \) for patients in stratum \( B_{k+1}(\bar{y}_k) \) with \( Y_{k+1} = y_{k+1} \).
Using Bayes’ rule, this assumption can be re-written as:

\[ P^*(R_{k+1} = 0 | R_k = 1, \bar{Y}_k = \bar{y}_k) \]

\[ = P^*(R_{k+1} = 0 | R_k = 1, \bar{Y}_{k+1} = \bar{y}_{k+1}) \text{ for all } \bar{y}_k \]

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

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

We will retain this assumption.
Next, we generalize (15) and impose the following exponential tilting "linking" assumptions:

\[ f^*(Y_{k+1}|A_k(\bar{y}_k)) \propto f^*(Y_{k+1}|B_{k+1}(\bar{y}_k)) \exp(\alpha r(Y_{k+1})) \text{ for all } \bar{y}_k \]  

(17)

where \( r(\cdot) \) is a specified function which we will assume to be an increasing function of its argument and \( \alpha \) is a sensitivity analysis parameter.
The missing not at random class of assumptions that we propose involves Equations (14) and (17), where \( r(\cdot) \) is considered fixed and \( \alpha \) 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 \( \bar{y}_k \), the distribution of \( Y_{k+1} \) for patients in stratum \( A_k(\bar{y}_k) \) is weighted more heavily (i.e., tilted) to higher (lower) values than the distribution of \( Y_{k+1} \) for patients in stratum \( B_{k+1}(\bar{y}_k) \).

The amount of “tilting” increases with magnitude of \( \alpha \).
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.
Randomized trial designed to evaluate the efficacy and safety of once-monthly, injectable paliperidone palmitate (PP1M) relative to placebo (PBO) in delaying the time to relapse in subjects with schizoaffective disorder.

Open-label phase consisting of a flexible-dose, lead-in period and a fixed-dose, stabilization period.

Stable subjects entered a 15-month relapse-prevention phase and were randomized to receive PP1M or placebo injections at baseline (Visit 0) and every 28 days (Visits 1-15).

Additional clinic visit (Visit 16) scheduled for 28 days after the last scheduled injection.

170 and 164 subjects were randomized to the PBO and PP1M arms.
Research question: Are functional outcomes better in patients with schizoaffective disorder better maintained if they continue on treatment or are withdrawn from treatment and given placebo instead?

An ideal study would follow all randomized subjects through Visit 16 while maintaining them on their randomized treatment and examine symptomatic and functional outcomes at that time point.

Since clinical relapse can have a major negative impact, the study design required that patients who had signs of relapse were discontinued from the study.

In addition, some patients discontinued due to adverse events, withdrew consent or were lost to follow-up.

38% and 60% of patients in the PBO and PP1M arms were followed through Visit 16 ($p=0.0001$).
Case Study: SCA-3004

![Cumulative Probability Graph](image)

- **PP1M**
- **Placebo**

Cumulative Probability vs. Last Visit On Study
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)

![Graph showing Score by last observation over visits from 0 to 15.](graph.png)
Case Study: SCA-3004

![Graph 1: Conditional Probability of Dropout (simulated data)]

- **Conditional Probability of Dropout (actual data)**
  - Placebo arm
  - Active arm

![Graph 2: Kolmogorov-Smirnov Statistic]

- **Visit**
  - Placebo arm
  - Active arm
Case Study: SCA-3004

- Left graph: Conditional Probability of Dropout (simulated data)
  - Placebo arm (blue circles)
  - Active arm (red circles)

- Right graph: Kolmogorov-Smirnov Statistic
  - Placebo arm (solid blue line)
  - Active arm (dashed red line)

Visit axis ranges from 0 to 15.
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

<table>
<thead>
<tr>
<th>$y_{k+1}$</th>
<th>$y_{k+1}^*$</th>
<th>Log Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>20</td>
<td>$\alpha \times 0.01$</td>
</tr>
<tr>
<td>40</td>
<td>30</td>
<td>$\alpha \times 0.18$</td>
</tr>
<tr>
<td>50</td>
<td>40</td>
<td>$\alpha \times 0.40$</td>
</tr>
<tr>
<td>60</td>
<td>50</td>
<td>$\alpha \times 0.30$</td>
</tr>
<tr>
<td>70</td>
<td>60</td>
<td>$\alpha \times 0.09$</td>
</tr>
<tr>
<td>80</td>
<td>700</td>
<td>$\alpha \times 0.01$</td>
</tr>
</tbody>
</table>
Case Study: SCA-3004

![Graph showing data for Placebo and PP1M with α values ranging from -20 to 20 and estimated values ranging from 60 to 80.](image)
Case Study: SCA-3004

![Graph showing the difference in means (Non-completers minus Completers) over α values for PP1M and Placebo.]
## Simulation Study

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Estimator</th>
<th>PBO</th>
<th>PP1M</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu(\hat{P})$</td>
<td>$\mu^*$</td>
<td>$\mu^*$</td>
</tr>
<tr>
<td></td>
<td>$\hat{\mu}$</td>
<td>Bias</td>
<td>MSE</td>
</tr>
<tr>
<td>-10</td>
<td></td>
<td>72.89</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.50</td>
<td>1.58</td>
</tr>
<tr>
<td>-5</td>
<td></td>
<td>73.38</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.31</td>
<td>1.32</td>
</tr>
<tr>
<td>-1</td>
<td></td>
<td>73.74</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.19</td>
<td>1.18</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>73.80</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.18</td>
<td>1.17</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>73.84</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.17</td>
<td>1.15</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>74.00</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.13</td>
<td>1.11</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>74.15</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.10</td>
<td>1.08</td>
</tr>
</tbody>
</table>
## Simulation Study

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Procedure</th>
<th>PBO Coverage</th>
<th>PP1M Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10</td>
<td>Normal-IF</td>
<td>86.1%</td>
<td>88.6%</td>
</tr>
<tr>
<td></td>
<td>Normal-JK</td>
<td>92.1%</td>
<td>92.6%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-ET</td>
<td>90.2%</td>
<td>91.9%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-ET</td>
<td>92.4%</td>
<td>93.7%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-S</td>
<td>92.3%</td>
<td>92.7%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-S</td>
<td>93.9%</td>
<td>94.3%</td>
</tr>
<tr>
<td>-5</td>
<td>Normal-IF</td>
<td>89.0%</td>
<td>91.7%</td>
</tr>
<tr>
<td></td>
<td>Normal-JK</td>
<td>94.1%</td>
<td>94.2%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-ET</td>
<td>91.7%</td>
<td>92.6%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-ET</td>
<td>93.6%</td>
<td>94.9%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-S</td>
<td>94.1%</td>
<td>94.2%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-S</td>
<td>95.1%</td>
<td>95.1%</td>
</tr>
<tr>
<td>-1</td>
<td>Normal-IF</td>
<td>90.8%</td>
<td>93.4%</td>
</tr>
<tr>
<td></td>
<td>Normal-JK</td>
<td>94.9%</td>
<td>94.8%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-ET</td>
<td>91.0%</td>
<td>94.0%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-ET</td>
<td>92.8%</td>
<td>94.9%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-S</td>
<td>94.4%</td>
<td>94.7%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-S</td>
<td>95.0%</td>
<td>95.3%</td>
</tr>
<tr>
<td>0</td>
<td>Normal-IF</td>
<td>90.7%</td>
<td>93.5%</td>
</tr>
<tr>
<td></td>
<td>Normal-JK</td>
<td>95.0%</td>
<td>94.9%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-ET</td>
<td>92.8%</td>
<td>93.9%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-ET</td>
<td>94.3%</td>
<td>95.0%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-S</td>
<td>95.3%</td>
<td>94.7%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-S</td>
<td>96.0%</td>
<td>95.1%</td>
</tr>
<tr>
<td>α</td>
<td>Procedure</td>
<td>PBO Coverage</td>
<td>PP1M Coverage</td>
</tr>
<tr>
<td>----</td>
<td>-----------------</td>
<td>--------------</td>
<td>---------------</td>
</tr>
<tr>
<td>1</td>
<td>Normal-IF</td>
<td>90.9%</td>
<td>93.5%</td>
</tr>
<tr>
<td></td>
<td>Normal-JK</td>
<td>94.9%</td>
<td>94.8%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-ET</td>
<td>92.8%</td>
<td>93.5%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-ET</td>
<td>94.2%</td>
<td>95.0%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-S</td>
<td>95.3%</td>
<td>94.6%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-S</td>
<td>96.0%</td>
<td>95.2%</td>
</tr>
<tr>
<td>5</td>
<td>Normal-IF</td>
<td>91.5%</td>
<td>93.7%</td>
</tr>
<tr>
<td></td>
<td>Normal-JK</td>
<td>94.6%</td>
<td>95.1%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-ET</td>
<td>92.6%</td>
<td>93.8%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-ET</td>
<td>93.8%</td>
<td>94.7%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-S</td>
<td>94.9%</td>
<td>95.1%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-S</td>
<td>96.0%</td>
<td>95.5%</td>
</tr>
<tr>
<td>10</td>
<td>Normal-IF</td>
<td>92.1%</td>
<td>93.4%</td>
</tr>
<tr>
<td></td>
<td>Normal-JK</td>
<td>94.8%</td>
<td>95.0%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-ET</td>
<td>92.9%</td>
<td>93.8%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-ET</td>
<td>93.9%</td>
<td>94.8%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF-S</td>
<td>94.7%</td>
<td>95.0%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-JK-S</td>
<td>95.6%</td>
<td>95.4%</td>
</tr>
</tbody>
</table>
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.
\( M_k \): 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 \( \bar{O}_K \).
\( O_0 = Y_0 \) and \( C \) can be computed from \( \bar{O}_K \) as \( \max\{k : M_k = 0\} \).
For $0 < k < C$, 

$$M_k \perp Y_k \mid \overline{Y}_{k-1}, O_k$$  \hspace{1cm} (18) 

- While on-study, the probability of providing outcome data at time $k$ can depend on previous outcomes (observed or not) and observed data after time $k$.
- Imagine a stratum of individuals who share the same history of outcomes prior to time $k$ and same observed data after time $k$.
- Imagine splitting the stratum into two sets: those who provide outcome data at time $k$ (stratum B) and those who do not (stratum A).
- The distribution of the outcome at time $k$ is the same for these two strata.
For $0 < k < C$, 

$$f^*(Y_k \mid M_k = 1, \overline{Y}_{k-1}, O_k) = f^*(Y_k \mid M_k = 0, \overline{Y}_{k-1}, O_k) \quad (19)$$

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

$$P^*(M_k = 1 \mid \overline{Y}_k, O_k) = P^*(M_k = 1 \mid \overline{Y}_{k-1}, O_k) : \quad 0 < k < C. \quad (20)$$
In our imputation algorithm, we will use the following fact:

\[ M_k \perp Y_k \mid \rho_k^*(\overline{Y}_{k-1}, O_k) : 0 < k < C \quad (21) \]

where

\[ \rho_k^*(\overline{Y}_{k-1}, O_k) = P^*(M_k = 1 \mid \overline{Y}_{k-1}, O_k) \quad (22) \]
Under assumption (18), the joint distribution of \((C, \overline{Y}_C)\) (i.e., the monotonized data) is identified by a recursive algorithm.
Imputation

- The number of individuals contributing to the histograms that form the basis of the imputation strategy may be quite small.
- Rather than matching on the past outcomes and future observed data, we plan to use (21) and match on estimates of $\rho_k^*(\bar{Y}_{k-1}, O_k)$. 
where $w_k(\bar{Y}_{k-1}, O_k; \nu_k)$ is a specified function of its arguments and $\nu_k$ is a finite-dimensional parameter with true value $\nu_k^*$. 

Simultaneous Estimation/Imputation

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

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

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

where

$$d_k(\overline{Y}_{k-1}, O_{k}; \nu^*_k) = \frac{\partial w_k(\overline{Y}_{k-1}, O_{k}; \nu_k)}{\partial \nu_k}$$
For each individual $i$ with $R_{k,i} = 1$, compute

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

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

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

Set $k = k + 1$. If $k = K$ then stop. Otherwise, return to Step 2.
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 corrected estimator of $\mu^*$ based on the $m$th imputed dataset.
When $M > 1$, we can replace $\tilde{\sigma}_\alpha^2$ with Rubin’s (1987) multiple imputation variance estimator, i.e.,

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

- In simulations, we have found success using (24) coupled with symmetric bootstrap to form confidence intervals.
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 $\hat{\nu}_k$’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 $\hat{\nu}_k$’s from Step 1 to create a non-monotone dataset $\mathcal{D}^{(b)}$. 
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 \(U < \hat{\rho}_k(Y_{k-1,i}, O_{k,i})\), 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.
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 activities and quality of life.
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).
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 period.
• The primary efficacy variable was the pain score measured on a 100-mm Visual Analog Scale (VAS), where higher levels of VAS indicate worse pain.

• VAS scores were scheduled on day 1 of the baseline phase, every two weeks during titration, and then monthly during the maintenance phase.

• Treatment effects were based on the difference in the mean VAS scores at the final scheduled follow-up visit.

• Adverse events and use of rescue medications was also scheduled to be monitored throughout the double-blind phase.

• The trials were not designed to follow patients after they discontinued their assigned therapy.
In NP 001, 531 subjects were randomized to one of four study arms: placebo ($n = 137$), 100 mg/day ($n = 129$), 200 mg/day ($n = 132$), and 400 mg/day ($n = 133$).

In NP 002, 370 subjects were randomized to one of three study arms: placebo ($n = 123$), 200 mg/day ($n = 118$), and 400 mg/day ($n = 129$).

Seven subjects in NP 001 and six subjects NP 002 did not have at least one follow-up visit and were not considered part of the intent-to-treat (ITT) population.
In our analysis, we merge the data from the two studies. We focus our analysis on a comparison of the placebo versus 400 mg/day arms. One individual from the 400 mg/day arm was excluded because of undue influence on the analysis. The sample sizes are 255 and 256 in the placebo and 400 mg/day arms, respectively.
### Placebo

#### Monotone:

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>*_______</td>
<td>5</td>
<td>0.0196</td>
</tr>
<tr>
<td>**_______</td>
<td>5</td>
<td>0.0196</td>
</tr>
<tr>
<td>***_______</td>
<td>10</td>
<td>0.0392</td>
</tr>
<tr>
<td>****_______</td>
<td>3</td>
<td>0.0118</td>
</tr>
<tr>
<td>*****_______</td>
<td>19</td>
<td>0.0745</td>
</tr>
<tr>
<td>******_______</td>
<td>12</td>
<td>0.0471</td>
</tr>
<tr>
<td>*******_______</td>
<td>12</td>
<td>0.0471</td>
</tr>
<tr>
<td>********_______</td>
<td>5</td>
<td>0.0196</td>
</tr>
<tr>
<td>*********_______</td>
<td>81</td>
<td>0.3176</td>
</tr>
</tbody>
</table>

#### Intermittent:

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>_</em>_*****</td>
<td>14</td>
<td>0.0549</td>
</tr>
<tr>
<td><em>_</em>******</td>
<td>13</td>
<td>0.0510</td>
</tr>
<tr>
<td><strong><strong>_</strong></strong></td>
<td>7</td>
<td>0.0275</td>
</tr>
<tr>
<td><em><strong>_</strong></em>**</td>
<td>6</td>
<td>0.0235</td>
</tr>
<tr>
<td>*****<strong>_</strong></td>
<td>5</td>
<td>0.0196</td>
</tr>
<tr>
<td>Other</td>
<td>47</td>
<td>0.1843</td>
</tr>
</tbody>
</table>
### Monotone

<table>
<thead>
<tr>
<th>Pattern</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>*________</td>
<td>4</td>
<td>0.0156</td>
</tr>
<tr>
<td>**________</td>
<td>14</td>
<td>0.0547</td>
</tr>
<tr>
<td>***_______</td>
<td>19</td>
<td>0.0742</td>
</tr>
<tr>
<td>****_____</td>
<td>7</td>
<td>0.0273</td>
</tr>
<tr>
<td>*****____</td>
<td>19</td>
<td>0.0742</td>
</tr>
<tr>
<td>*****___</td>
<td>10</td>
<td>0.0391</td>
</tr>
<tr>
<td>********__</td>
<td>9</td>
<td>0.0352</td>
</tr>
<tr>
<td>********_</td>
<td>2</td>
<td>0.0078</td>
</tr>
<tr>
<td>*********</td>
<td>67</td>
<td>0.2617</td>
</tr>
</tbody>
</table>

### Intermittent

<table>
<thead>
<tr>
<th>Pattern</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>_</em>______</td>
<td>15</td>
<td>0.0586</td>
</tr>
<tr>
<td><em>_</em>_*****</td>
<td>9</td>
<td>0.0352</td>
</tr>
<tr>
<td><em>_</em>******</td>
<td>8</td>
<td>0.0312</td>
</tr>
<tr>
<td><em><strong>_</strong></em>**</td>
<td>7</td>
<td>0.0273</td>
</tr>
<tr>
<td><em>_</em>_***__</td>
<td>5</td>
<td>0.0195</td>
</tr>
<tr>
<td>Other</td>
<td>56</td>
<td>0.2188</td>
</tr>
</tbody>
</table>
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?
### Placebo

<table>
<thead>
<tr>
<th>$k$</th>
<th>On-Study</th>
<th>Last Seen</th>
<th>Obs. Value</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>255</td>
<td>5</td>
<td>255</td>
<td>58.902</td>
<td>19.196</td>
</tr>
<tr>
<td>1</td>
<td>250</td>
<td>5</td>
<td>188</td>
<td>53.202</td>
<td>23.048</td>
</tr>
<tr>
<td>2</td>
<td>245</td>
<td>14</td>
<td>238</td>
<td>48.899</td>
<td>24.888</td>
</tr>
<tr>
<td>3</td>
<td>231</td>
<td>5</td>
<td>186</td>
<td>45.849</td>
<td>23.928</td>
</tr>
<tr>
<td>4</td>
<td>226</td>
<td>27</td>
<td>203</td>
<td>42.291</td>
<td>25.338</td>
</tr>
<tr>
<td>5</td>
<td>199</td>
<td>24</td>
<td>192</td>
<td>38.896</td>
<td>25.117</td>
</tr>
<tr>
<td>6</td>
<td>175</td>
<td>15</td>
<td>162</td>
<td>37.549</td>
<td>25.827</td>
</tr>
<tr>
<td>7</td>
<td>160</td>
<td>10</td>
<td>150</td>
<td>35.047</td>
<td>26.313</td>
</tr>
<tr>
<td>8</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>35.613</td>
<td>26.446</td>
</tr>
</tbody>
</table>
### Observed Data

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

Mean VAS by Last Observation

Visit

400 mg/day

Mean VAS by Last Observation

Visit
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.
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 on 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.
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 - 400 mg

Loss Function (F) vs. $\sigma$
Estimation of Smoothing Parameters - Placebo

![Graph showing the relationship between smoothing parameters and loss function. The x-axis represents \( \sigma \) and the y-axis represents the loss function (H). The graph demonstrates a decrease in the loss function as \( \sigma \) increases.](image-url)
Goodness of Fit

The diagram illustrates the conditional probability of dropout for two groups, VAS group 1 and VAS group 2. The x-axis represents the conditional probability of dropout (observed data), while the y-axis shows the conditional probability of dropout (simulated data). The scatter plot indicates a reasonable fit for both groups, with points closely following the line of identity.
Goodness of Fit

The image shows a scatter plot with two groups labeled VAS group 1 and VAS group 2. The x-axis represents the proportion intermittent missing (observed data), while the y-axis represents the proportion intermittent missing (simulated data). The data points appear to align closely with the line of best fit, indicating a good fit between the observed and simulated data.
Goodness of Fit

![Graph showing Goodness of Fit for VAS Group 1 and VAS Group 2. The x-axis represents visits (2, 4, 6, 8), and the y-axis represents the Kolmogorov-Smirnov Statistic. The graph shows the statistical comparison between the two groups over the visits.](image-url)
The estimates of $\mu^*$ 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 Graph]

The graph illustrates the sensitivity analysis for α (400 mg/day) versus α (Placebo). The plot uses a color scale to represent different values, ranging from -20 to 10.
Sensitivity Analysis

Placebo

400 mg/day

Difference in Mean VAS (Non-completers minus Completers)

Placebo
400 mg/day

\(\alpha\)
<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Estimator</th>
<th>PBO</th>
<th>400 mg/day</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu(\hat{P})$</td>
<td>$\hat{\mu}$</td>
<td>$\mu^\ast$</td>
</tr>
<tr>
<td>-10</td>
<td>31.69</td>
<td>0.53</td>
<td>4.13</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>3.76</td>
<td>2.58</td>
</tr>
<tr>
<td>-5</td>
<td>33.60</td>
<td>0.03</td>
<td>4.03</td>
</tr>
<tr>
<td></td>
<td>-0.02</td>
<td>3.93</td>
<td>3.75</td>
</tr>
<tr>
<td>-1</td>
<td>37.10</td>
<td>-0.70</td>
<td>4.93</td>
</tr>
<tr>
<td></td>
<td>-0.32</td>
<td>4.41</td>
<td>3.93</td>
</tr>
<tr>
<td>0</td>
<td>38.12</td>
<td>-0.82</td>
<td>5.23</td>
</tr>
<tr>
<td></td>
<td>-0.36</td>
<td>4.53</td>
<td>3.77</td>
</tr>
<tr>
<td>1</td>
<td>39.10</td>
<td>-0.89</td>
<td>5.47</td>
</tr>
<tr>
<td></td>
<td>-0.37</td>
<td>4.65</td>
<td>3.92</td>
</tr>
<tr>
<td>5</td>
<td>42.75</td>
<td>-1.06</td>
<td>6.43</td>
</tr>
<tr>
<td></td>
<td>-0.49</td>
<td>5.35</td>
<td>6.26</td>
</tr>
<tr>
<td>10</td>
<td>45.59</td>
<td>-1.43</td>
<td>8.10</td>
</tr>
<tr>
<td></td>
<td>-0.70</td>
<td>6.50</td>
<td>8.11</td>
</tr>
</tbody>
</table>
## Simulation Study - Five Imputes

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Procedure</th>
<th>PBO Coverage</th>
<th>400 mg/day Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10</td>
<td>Normal-IF(Rubin)</td>
<td>94.6%</td>
<td>93.6%</td>
</tr>
<tr>
<td></td>
<td>Normal-BootstrapSE</td>
<td>94.4%</td>
<td>93.8%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-Percentile</td>
<td>92.6%</td>
<td>86.6%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF(Rubin)-ET</td>
<td>93.8%</td>
<td>94.5%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF(Rubin)-S</td>
<td>95.4%</td>
<td>95.5%</td>
</tr>
<tr>
<td>-5</td>
<td>Normal-IF(Rubin)</td>
<td>93.9%</td>
<td>94.5%</td>
</tr>
<tr>
<td></td>
<td>Normal-BootstrapSE</td>
<td>94.3%</td>
<td>94.6%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-Percentile</td>
<td>93.8%</td>
<td>93.3%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF(Rubin)-ET</td>
<td>94.3%</td>
<td>95.8%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF(Rubin)-S</td>
<td>94.7%</td>
<td>95.2%</td>
</tr>
<tr>
<td>-1</td>
<td>Normal-IF(Rubin)</td>
<td>92.8%</td>
<td>94.8%</td>
</tr>
<tr>
<td></td>
<td>Normal-BootstrapSE</td>
<td>92.8%</td>
<td>94.7%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-Percentile</td>
<td>92.0%</td>
<td>93.4%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF(Rubin)-ET</td>
<td>94.0%</td>
<td>96.0%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF(Rubin)-S</td>
<td>94.8%</td>
<td>95.2%</td>
</tr>
<tr>
<td>0</td>
<td>Normal-IF(Rubin)</td>
<td>92.7%</td>
<td>95.1%</td>
</tr>
<tr>
<td></td>
<td>Normal-BootstrapSE</td>
<td>92.9%</td>
<td>95.0%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-Percentile</td>
<td>91.3%</td>
<td>92.5%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF(Rubin)-ET</td>
<td>93.8%</td>
<td>95.9%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF(Rubin)-S</td>
<td>94.9%</td>
<td>96.1%</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Procedure</td>
<td>PBO Coverage</td>
<td>400 mg/day Coverage</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------</td>
<td>-------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>1</td>
<td>Normal-IF(Rubin)</td>
<td>93.1%</td>
<td>94.6%</td>
</tr>
<tr>
<td></td>
<td>Normal-BootstrapSE</td>
<td>92.9%</td>
<td>94.4%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-Percentile</td>
<td>90.7%</td>
<td>91.8%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF(Rubin)-ET</td>
<td>94.3%</td>
<td>95.7%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF(Rubin)-S</td>
<td>95.0%</td>
<td>96.2%</td>
</tr>
<tr>
<td>5</td>
<td>Normal-IF(Rubin)</td>
<td>93.3%</td>
<td>93.2%</td>
</tr>
<tr>
<td></td>
<td>Normal-BootstrapSE</td>
<td>93.4%</td>
<td>93.2%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-Percentile</td>
<td>89.8%</td>
<td>84.5%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF(Rubin)-ET</td>
<td>94.3%</td>
<td>93.3%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF(Rubin)-S</td>
<td>95.2%</td>
<td>96.0%</td>
</tr>
<tr>
<td>10</td>
<td>Normal-IF(Rubin)</td>
<td>94.0%</td>
<td>88.6%</td>
</tr>
<tr>
<td></td>
<td>Normal-BootstrapSE</td>
<td>93.5%</td>
<td>88.4%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-Percentile</td>
<td>86.8%</td>
<td>70.6%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF(Rubin)-ET</td>
<td>93.6%</td>
<td>89.7%</td>
</tr>
<tr>
<td></td>
<td>Bootstrap-IF(Rubin)-S</td>
<td>96.4%</td>
<td>94.8%</td>
</tr>
</tbody>
</table>
Honey-do List

- Develop data adaptive technique for handling outliers
- Incorporate auxiliary covariates
No substitute for better trial design and procedures to minimize missing data.

Global sensitivity analysis should be a mandatory component of trial reporting.

Visit us at www.missingdatamatters.org or email me at dscharf@jhu.edu
Inference in Randomized Trials with Death and Missingness
Harvard Shortcourse

Daniel Scharfstein
Johns Hopkins University
dscharf@jhu.edu

September 24, 2019
Anamorelin is a drug developed for the treatment of cancer cachexia and anorexia.

HT-ANAM 302 was a randomized, double-blind, placebo-controlled Phase III study designed to evaluate the efficacy of anamorelin in patients with advanced non-small cell lung cancer.

Lean body mass (LBM) was scheduled to be measured at baseline ($Y_0$), 6 weeks ($Y_1$) and 12 weeks ($Y_2$).

Primary functional endpoint: $Z = \frac{(Y_2 + Y_1)}{2} - Y_0$
## Death and missingness

<table>
<thead>
<tr>
<th>Category</th>
<th>Placebo (n = 157)</th>
<th>Anamorelin (n = 322)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Died Prior to Wk 12</td>
<td>24 (15.3%)</td>
<td>54 (16.8%)</td>
</tr>
<tr>
<td>Survivors with complete data</td>
<td>93 (59.2%)</td>
<td>185 (57.5%)</td>
</tr>
<tr>
<td>Survivors missing only Wk 6</td>
<td>3 (1.9%)</td>
<td>17 (5.3%)</td>
</tr>
<tr>
<td>Survivors missing only Wk 12</td>
<td>17 (10.8%)</td>
<td>31 (9.6%)</td>
</tr>
<tr>
<td>Survivors missing both Wks 6, 12</td>
<td>20 (12.7%)</td>
<td>35 (10.9%)</td>
</tr>
</tbody>
</table>
Central Question

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

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

1. Evaluate treatment effect on functional outcome conditional on survival
   - Conditioning on post-baseline factor

2. Joint modeling survival and functional outcomes
   - Allows extrapolation of outcomes after death

3. Principal stratification
   - Applies to a subset of patients who are not identifiable at baseline

4. Composite endpoint combining survival and functional outcomes
   - May be hard to separate effect on function.
NO PERFECT SOLUTIONS

Not a fan of Approaches 1 and 2.
Goal

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

- $T = 0, 1$: treatment assignment
- $X$: vector baseline covariates
- $Y_0$: baseline functional measure at $t_0$
- $Y_1, \ldots, Y_K$: functional outcomes at $t_1, \ldots, t_K$
- $L$: survival time
- $A_k = I(L > t_k)$: survival status at $t_k$
- $Z = g(Y_0, \ldots, Y_K)$: primary functional endpoint
  - e.g. $K = 2$, $Z = (Y_2 + Y_1)/2 - Y_0$
  - only defined when $A_K = 1$
Finite-valued random variable $U$ which assigns a score to each patient such that

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

Only the ordering of $U$ is important, not the actual score assignments.
Treatment effect ($\theta$) is measured by the probability that the outcome for an individual with $T = 0$ is less than the outcome of an individual with $T = 1$ minus the probability that the outcome for an individual with $T = 0$ is greater than the outcome of an individual with $T = 1$

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

Can also compare the treatment-specific quantiles of $U$. 
In the absence of missing data,

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

where \( n_0 = \sum_i (1 - T_i) \) and \( n_1 = \sum_i T_i \).
$R_k$: missing data indicator (defined when $A_k = 1$)

$S = (R_1, \ldots, R_K)$ (defined when $A_K = 1$)

- $Y_{obs}^{(s)} = \{ Y_k : R_k = 1, k \geq 1, S = s \}$
- $Y_{mis}^{(s)} = \{ Y_k : R_k = 0, k \geq 1, S = s \}$

$Z$ is unobserved when $S \neq 1$.

To estimate $\theta$, need to impute $Z$ or equivalently $Y_{mis}^{(s)}$ for $s \neq 1$. 
Missing Data Assumptions

\[ f(Y_{mis}^{(s)}|A_K = 1, Y_{obs}^{(s)}, Y_0, X, T, S = s) \]
\[ \propto \exp(\beta_T Z) f(Y_{mis}^{(s)}|A_K = 1, Y_{obs}^{(s)}, Y_0, X, T, S = 1) \]

Reference Distribution

for all \( s \neq 1 \),

- \( \beta_T \) is a treatment-specific sensitivity parameter.
- \( \beta_T = 0 \) (i.e., benchmark assumption) reduces to the complete case missing value (CCMV) restrictions applied to the missing data patterns for patients alive at \( t_K \).
- CCMV is different than missing at random (MAR) assumption.
\[ K = 2, \ Z = (Y_1 + Y_2)/2 - Y_0. \]
\[ \beta_T' = 2\beta_T \]

\[
f(Y_2|A_2 = 1, Y_1, Y_0, X, T, S = (1, 0)) \\
\propto \exp(\beta_T' Y_2) f(Y_2|A_2 = 1, Y_1, Y_0, X, T, S = 1)
\]

Reference Distribution

For subjects alive at \( t_2 \), who are observed at time \( t_1 \), who share the same functional measure at \( t_1 \) and who share the same baseline factors, the distribution of \( Y_2 \) for those whose functional measure at \( t_2 \) is missing is, when \( \beta_T' > 0 \) (< 0), more heavily weighted toward higher (lower) values of \( Y_2 \) than those whose functional measure at \( t_2 \) is observed.
\[ f(Y_1|A_2 = 1, Y_2, Y_0, X, T, S = (0, 1)) \]
\[ \propto \exp(\beta'_T Y_1) f(Y_1|A_2 = 1, Y_2, Y_0, X, T, S = 1) \]

Reference Distribution

For subjects alive at \( t_2 \), who are observed at time \( t_2 \), who share the same functional measure at \( t_2 \) and who share the same baseline factors, the distribution of \( Y_1 \) for those whose functional measure at \( t_1 \) is missing is, when \( \beta'_T > 0 \) (\(< 0\)), more heavily weighted toward higher (lower) values of \( Y_1 \) than those whose functional measure at \( t_1 \) is observed.
\[ f(Y_1, Y_2|A_2 = 1, Y_0, X, T, S = (0, 0)) \]

\[ \propto \exp (\beta_T' (Y_1 + Y_2)) f(Y_1, Y_2|A_2 = 1, Y_0, X, T, S = 1) \]

For subjects alive at \( t_2 \) and who share the same baseline factors, the joint distribution of \( Y_1 \) and \( Y_2 \) for those whose functional measures at \( t_1 \) and \( t_2 \) are missing is, when \( \beta_T' > 0 \) \((< 0)\), more heavily weighted toward higher (lower) values of \( Y_1 \) and \( Y_2 \) than those whose measures are fully observed.
Ignore conditioning on $Y_0$ and $X$ and suppose $f(Y_1, Y_2|A_2 = 1, T, S = 1)$ is multivariate normal with mean $(\mu_{T,1}, \mu_{T,2})$ and variance-covariance matrix

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

- $f(Y_2|A_2 = 1, Y_1, T, S = (1, 0))$ is normal with mean $\mu_{T,2} + \beta'_T (1 - \rho^2_T) \sigma^2_{T,2} + \rho_T \frac{\sigma_{T,2}}{\sigma_{T,1}} (Y_1 - \mu_{T,1})$ and variance $(1 - \rho^2_T) \sigma^2_{T,2}$
- $f(Y_1|A_2 = 1, Y_2, T, S = (0, 1))$ is normal with mean $\mu_{T,1} + \beta'_T (1 - \rho^2_T) \sigma^2_{T,1} + \rho_T \frac{\sigma_{T,1}}{\sigma_{T,2}} (Y_2 - \mu_{T,2})$ and variance $(1 - \rho^2_T) \sigma^2_{T,1}$
\( f(Y_1, Y_2|A_2 = 1, T, S = (0, 0)) \) is multivariate normal with mean 
\( (\mu_{T,1} + \beta_T \sigma_{T,1}^2 + \beta_T \rho_T \sigma_{T,1} \sigma_{T,2}, \mu_{T,2} + \beta_T \sigma_{T,2}^2 + \beta_T \rho_T \sigma_{T,1} \sigma_{T,2}) \) and variance-covariance matrix \( \Sigma_T \).

- If \( \rho_T > 0 \), then the means increase linearly in \( \beta'_T \)
- \( \beta'_T \) has no impact on the variances and covariances.
- \( \beta'_T > 0 \) (\( \beta'_T < 0 \)) implies that the non-identified distributions have more (less) mass at higher values than their reference distributions.
Example: Exponential tilting

\begin{align*}
\beta &= -1 \\
\beta &= 0 \\
\beta &= 1
\end{align*}
Modeling

Need to specify of a model for

\[ f(\bar{Y}_K | A_K = 1, Y_0, X, T, S = 1) \]

- To respect bounds, define

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

- \( Y_k^\dagger = \phi(Y_k) \) and \( \bar{Y}_K^\dagger = (Y_1^\dagger, \ldots, Y_K^\dagger) \).

- One-to-one mapping between

\[ h(\bar{Y}_K^\dagger | A_K = 1, Y_0, X, T, S = 1) \]

and

\[ f(\bar{Y}_K | A_K = 1, Y_0, X, T, S = 1) \].
Modeling

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

- Posit a model for each component of the product.
Modeling

\[ h(Y_k^\dagger | A_K = 1, \bar{Y}_{k-1}^\dagger, Y_0, X, T = t, S = 1) = h_{k,t}(Y_k^\dagger - \mu_{k,t}(\bar{Y}_{k-1}^\dagger, Y_0, X; \alpha_{k,t})) \]

- \( \mu_{k,t}(\bar{Y}_{k-1}^\dagger, Y_0, X; \alpha_{k,t}) \) is a specified function
- \( \alpha_{k,t} \) is an unknown parameter vector
- \( h_{k,t} \) is an unspecified time/treatment-specific density function.
The parameter vectors $\alpha_{k,t}$ can be estimated by minimizing the least squares objective function

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

The density function $h_{k,t}$ can be estimated by kernel density estimation based on the residuals

$$\{ Y_{k,i}^{\dagger} - \mu_{k,t}(\overline{Y}_{k-1}^{\dagger}, Y_0, X; \hat{\alpha}_{k,t}) : T_i = t, A_{K,i} = 1, R_{1,i} = \ldots, R_{K,i} = 1, i = 1, \ldots, n \}$$

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

$$\prod_{k=1}^{K} \hat{h}_{k,t}(Y_k^{\dagger} - \mu_{k,t}(\overline{Y}_{k-1}^{\dagger}, Y_0, X; \hat{\alpha}_{k,t})) \left| \frac{d\phi(Y_k)}{dY_k} \right|. $$
For each individual $i$ alive at $t_K$ and who is in a stratum $s \neq 1$, impute the missing functional outcomes by drawing (using rejection sampling techniques) from the density that is proportional to

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

- Draw $M$ copies of the missing functional outcomes to create $M$ complete datasets.
- For each complete dataset $m$, estimate $\theta$ by $\hat{\theta}_m$.
- Overall estimator of $\theta$ is $\tilde{\theta} = \frac{1}{M} \sum_{m=1}^{M} \hat{\theta}_m$.
- Confidence intervals can be constructed by non-parametric bootstrap.
Baseline covariates: ECOG performance status, age, gender, BMI, weight loss in prior 6 months

LBM is bounded between 24 and 140

10 imputed datasets

Under benchmark assumptions,

\( \hat{\theta} = 0.30 \) (95% CI: 0.16 to 0.37, \( p < 0.0001 \))

Placebo: Median -0.98 kg (95% CI: -1.27 kg to -0.28 kg).

Anamorelin: Median 0.69 kg (95% CI: 0.43 kg to 0.93 kg).
HT-ANAM 302 Study

Cumulative Probability

L: Time to Death (days)

Z: Average change in LBM (kg)

Placebo

Anamorelin
\( \beta_1 (\text{Anamorelin}) = -0.5 \)

\( \beta_1 (\text{Anamorelin}) = 0.5 \)
Method presumes that death and the functional outcome can be ordered in a scientifically meaningful way.

Use mixed methods to confirm that ordering is consistent with the health preferences of patient population.

Ranking scheme is similar to ‘untied worst-rank score analysis’ for missing data of Lachin (1999).

The “worst-rank score analysis” ranks all the patients who died ($A_K = 0$) the same and is also commonly used.

CCMV is a strong benchmark assumption.

Assumed survival time is always known, need to extend methods to handle censoring.

idem software is available on CRAN.