Final Report: Sensitivity Analysis Tools for Randomized Trials with Missing Data

Daniel Scharfstein, Johns Hopkins University

April 29, 2017

Contents

1 Executive Summary

Missing outcome data are a widespread problem in randomized trials, including those used as the basis of regulatory approval of drugs and devices. Inference about treatment effects relies on unverifiable assumptions about the nature of the mechanism that generates the missing data, leading to concerns about the validity and robustness of trial results. A 2010 FDA-sponsored National Research Council Report (NRC) recommends that "examining sensitivity to the assumptions about the missing data mechanism should be a mandatory component of reporting." [2] Chapter 5 of the NRC report outlined a framework for conducting global sensitivity analysis.

In this three year contract, we (1) developed methods for global sensitivity analysis of clinical trials with monotone and non-monotone missing data, (2) implemented the methods in free, open source R and SAS software, (3) demonstrated the methods and software using real clinical trial data, and (4) disseminated the methods and software through presentations, short courses, webinars and manuscripts.

2 Project Objectives

The objectives of this project were to:

1. create unified and coherent methods for global sensitivity analysis of clinical trials with monotone and non-monotone missing data,
2. develop free, open source and reproducible software in SAS and R to implement the methods,
3. demonstrate the methods and software using real clinical trial data, and
4. disseminate the methods and software through presentations, short courses, webinars and manuscripts.

3 Background on Global Sensitivity Analysis

Chapter 5 of the NRC report lays out a general framework for global sensitivity analysis. In this framework, inference about treatment effects requires two types of assumptions: (i) untestable assumptions about the distribution of outcomes among those with missing data and (ii) testable assumptions that serve to increase the efficiency of estimation. Type (i) assumptions are required to "identify" parameters of interest: identification means that one can mathematically express parameters of interest (e.g., treatment arm-specific means, treatment effects) in terms of the distribution of the observed data. In other words, if one were given the distribution of the observed data and given a type (i) assumption, then one could compute the value of the parameter of interest. In the absence of identification, one cannot learn the value of the parameter of interest based only on knowledge of the distribution of the observed data. Identification implies that the parameters of interest can, *in theory*, be estimated if the sample size is large enough.

There are an infinite number of ways of positing type (i) assumptions. It is impossible to consider all such assumptions. A reasonable way of positing these assumptions is to

(a) stratify individuals with missing outcomes based on some features, and

(b) separately for each stratum, hypothesize a connection (or link) between the distribution of the missing outcomes with the distribution of these outcomes for patients who share the same features and for whom the distribution is identified.

The connection that is posited in (b) is a type (i) assumption. The problem with this approach is that the stratum of people who share the same features will typically be very small. As a result, it is necessary to draw strength across strata by "smoothing." Smoothing is required because, *in practice*, we are not working with large enough sample sizes. Without smoothing, the data analysis will not be informative because the uncertainty (i.e., standard errors) of the parameters of interest will be too large to be of substantive use. Thus, it is necessary to impose type (ii) smoothing assumptions. Type (ii) assumptions are testable (i.e., place restrictions on the distribution of the observed data) and should be scrutinized via model checking.

The global sensitivity framework proceeds by parameterizing (i.e., indexing) the connections (i.e., type (i) assumptions) in (b) above 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.

4 Deliverables

The deliverables of the project were to create:

1. a project website
2. methods, SAS/R modules, user documentation, case studies, short courses and manuscripts for *monotone* missing data
3. methods, SAS/R modules, user documentation, case studies, short courses and manuscripts for *non-monotone* missing data
4. methods, SAS/R modules, user documentation, case studies, short courses and manuscripts based *user-feedback*.

4.1 Project Website

We created a project website `www.missingdatamatters.org`. The website serves as a repository for software and presentation materials. To date, 320 individuals have registered to download software.

4.2 Monotone Missing Data

4.2.1 Methods

Notation: Let $k = 0, 1, \ldots, K$ refer in chronological order to the scheduled assessment times, with $k = 0$ corresponding to baseline. Let Y_k denote the outcome scheduled to be measured at assessment k. Define R_k to be the indicator that an individual is on-study at assessment k. We assume that all individuals are present at baseline. Furthermore, we assume that individuals do not contribute any further data once they have missed a visit. Let $C = \max\{k : R_k = 1\}$ and note that $C = K$ implies that the individual must have completed the study. For any given vector $z = (z_1, z_2, \ldots, z_K)$, we define $\overline{z}_k = (z_0, z_1, \ldots, z_k)$ and $\underline{z}_k = (z_{k+1}, z_{k+2}, \ldots, z_K)$. For each individual, $O = (C, \overline{Y}_C)$ is drawn from some distribution P^* contained in the non-parametric model $\mathscr{M}$ of distributions. The observed data consist of n independent draws $O_1, O_2, \ldots, O_n$ from P^*. The superscript $*$ will be used to denote the true value of the quantity to which it is appended.

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

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

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

Assumptions: Assumptions are required to draw inference about μ^* based on the available data. We consider a class of assumptions whereby an individual's decision to drop out in the interval between visits k and $k+1$ is not only influenced by past observable outcomes but by the outcome at visit $k+1$. Towards this end, we adopt the following two assumptions introduced in [3]: For $k = 0, 1, \ldots, K-2$,

$$P^*\left(Y_K \leq y \mid R_{k+1} = 0, R_k = 1, \overline{Y}_{k+1} = \overline{y}_{k+1}\right) = P^*\left(Y_K \leq y \mid R_{k+1} = 1, \overline{Y}_{k+1} = \overline{y}_{k+1}\right). \tag{1}$$

This says that in the cohort of patients who (1) are on-study at assessment k, (2) share the same outcome history through that visit and (3) have the same outcome at assessment $k+1$, the distribution of Y_K is the same for those last seen at assessment k and those still on-study at $k+1$.

For $k = 0, 1, \ldots, K-1$,

$$dG^*_{k+1}(y_{k+1} \mid \overline{y}_k) \propto \exp\{\rho_{k+1}(\overline{y}_k, y_{k+1})\} dF^*_{k+1}(y_{k+1} \mid \overline{y}_k), \tag{2}$$

where $G^*_{k+1}(y_{k+1} \mid \overline{y}_k) := P^*\left(Y_{k+1} \leq y_{k+1} \mid R_{k+1} = 0, R_k = 1, \overline{Y}_k = \overline{y}_k\right)$ and $\rho_{k+1}(\overline{y}_k, y_{k+1})$ is a known, pre-specified function of $\overline{y}_k$ and y_{k+1}.

Conditional on any given history $\overline{y}_k$, this assumption relates the distribution of Y_{k+1} for those patients who drop out between assessments k and $k+1$ to those patients who are on study at $k+1$. The special case whereby ρ_{k+1} is constant in y_{k+1} for all k implies that, conditional on the history $\overline{y}_k$, individuals who drop out between assessments k and $k+1$ have the same distribution of Y_{k+1} as those on-study at $k+1$. If instead ρ_{k+1} is an increasing (decreasing) function of y_{k+1} for some k, then individuals who drop-out between assessments k and $k+1$ tend to have higher (lower) values of Y_{k+1} than those who are on-study at $k+1$.

For specified ρ_{k+1}, Assumptions (1) and (2) are type(i) assumptions; they place no restriction on the distribution of the observed data. As such, ρ_{k+1} is not an empirically verifiable function.

Identifiability of target parameter: Under Assumptions 1 and 2 with given ρ_{k+1}, the parameter μ^* is identifiable. To establish identifiability, it suffices to demonstrate that μ^* can be expressed as a functional of the distribution of the observed data. The functional $\mu(P^*)$ can be equivalently expressed as

$$\int_{y_0} \cdots \int_{y_K} y_K \prod_{k=0}^{K-1} \Bigg\{ dF^*_{k+1}(y_{k+1} \mid \overline{y}_k) \left\{1 - H^*_{k+1}(\overline{y}_k)\right\} + \frac{\exp\{\rho_{k+1}(\overline{y}_k, y_{k+1})\} dF^*_{k+1}(y_{k+1} \mid \overline{y}_k)}{\int \exp\{\rho_{k+1}(\overline{y}_k, u)\} dF^*_{k+1}(u \mid \overline{y}_k)} H^*_{k+1}(\overline{y}_k) \Bigg\} dF^*_0(y_0) \ . \quad (3)$$

Statistical inference: Given a fixed function ρ_{k+1}, [3] proposed to estimate μ^* via the plug-in principle. Specifically, they specify type(ii) smoothing assumptions in the form of parametric models for both F^*_{k+1} and H^*_{k+1}, estimate parameters in these models by maximum likelihood, estimate F^*_0 nonparametrically using the empirical distribution function, and finally, estimate (3) by Monte Carlo integration using repeated draws from the resulting estimates of F^*_{k+1}, H^*_{k+1} and F^*_0. Since (3) is a smooth functional of F^*_0 and of the finite-dimensional parameters of the models for F^*_{k+1} and H^*_{k+1}, the resulting estimator of μ^* is $n^{1/2}$-consistent and, suitably normalized, tends in distribution to a mean-zero Gaussian random variable.

While simple to describe and easy to implement, this approach has a major drawback: the inferences it generates will be sensitive to correct specification of the parametric models imposed on F^*_{k+1} and H^*_{k+1}. Since the fit of these models is empirically verifiable, the plausibility of the models imposed can be scrutinized in any given application. In several instances, we have found it difficult to find models providing an adequate fit to the observed data. This is a serious problem since model misspecification will generally lead to inconsistent inference, which can translate into inappropriate and misleading scientific conclusions. To provide greater robustness, we instead adopted a more flexible modeling approach.

Instead, we assumed that P^* is contained in the submodel $\mathscr{M}_0 \subset \mathscr{M}$ of distributions that exhibit a first-order Markovian structure in the sense that $F_{k+1}(y_{k+1} \mid \overline{y}_k) = F_{k+1}(y_{k+1} \mid y_k)$ and $H_{k+1}(\bar{y}_k) = H_{k+1}(y_k)$. We can then estimate F^*_{k+1} and H^*_{k+1} by Nadaraya-Watson kernel estimators and select the associated tuning parameters by J-fold cross validation.

The tuning parameters are generally chosen to achieve an optimal finite-sample bias-variance trade-off for the quantity requiring smoothing - here, conditional distribution and probability mass functions. However, this trade-off may be problematic, since the resulting plug-in estimator $\mu(\widehat{P})$ may suffer from excessive and asymptotically nonnegligible bias due to inadequate tuning. This may prevent the plug-in estimator from having regular asymptotic behavior. In particular, 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. Therefore, the plug-in estimator must be regularized in order to serve as an appropriate basis for drawing statistical inference.

To address this problem, we employ a one-step bias correction procedure. This procedure involves adding a bias correction term to the plug-in estimator. The bias correction term is the average of the estimated "influence function", which measures the impact of "infinitesimal" contamination of P^* on (3). The resulting corrected estimator can be shown to have second-order asymptotic bias that ensures regular asymptotic behavior.

To characterize the uncertainty of our estimation procedure, we utilize bootstrapping techniques.

4.2.2 SAS/R Modules and User Documentation

A software package called `SAMON` was created. R Version 1.0 was posted on 12/12/2014 and SAS Version 1.0 was posted on 3/6/2015. R and SAS versions 2.0 were posted on 8/18/2015. R and SAS versions 3.0 were posted on 10/29/2016. R and SAS versions 4.0 are scheduled to be posted by 5/31/2017. Each software release includes user documentation.

4.2.3 Case Studies

Three case studies were developed that illustrate the methods developed to handle studies with monotone missing data. These include:

1. A randomized, placebo controlled trial comparing the effectiveness of four fixed doses of risperidone and one dose of haloperidol in schizophrenic patients. The primary outcome was patient function as measured by the total Positive and Negative Syndrome Scale score.

2. A randomized trial designed to evaluate the efficacy and safety of once-monthly, injectable paliperidone palmitate, as monotherapy or as an adjunct to pre-study mood stabilizers or antidepressants, relative to placebo in delaying the time to relapse in patients with schizoaffective disorder. The primary outcome was patient function as measured by the Personal and Social Performance scale.

3. A randomized trial designed to evaluate the efficacy of different doses of Quetiapine on treating patients with bipolar disorder. A key outcome was quality of life as measured by the Quality of Life Enjoyment Satisfaction Questionnaire.

4.2.4 Presentations

Over the course of the contract, we gave six short courses:

1. Scharfstein: Global Sensitivity Analysis of Randomized Trials with Missing Data: Recent Advances, Deming Conference, 12/2014.

2. Scharfstein, McDermott, Wang: Analysis of Randomized Trials with Missing Data, Johns Hopkins University, 1/2015.
3. Scharfstein: Global Sensitivity Analysis of Randomized Trials with Missing Data, Society for Clinical Trials, 5/2015.
4. Scharfstein, McDermott, Wang: Analysis of Randomized Trials with Missing Data, FDA, 11/2015.
5. Scharfstein, McDermott, Wang: Analysis of Randomized Trials with Missing Data, Johns Hopkins University, 6/2016.
6. Scharfstein: Analysis of Randomized Trials with Missing Data, University of Washington, 7/2016.

eight oral presentations:

1. McDermott: Global Sensitivity Analysis of Repeated Measures Studies with Informative Dropout: A Semi- Parametric Approach, Joint Statistical Meetings of American Statistical Association, 8/2104.
2. Scharfstein: Global Sensitivity Analysis of Repeated Measures Studies with Informative Dropout: A Semi- Parametric Approach, Joint Statistical Meetings of American Statistical Association, University of Rochester, 9/2014.
3. Scharfstein: Global Sensitivity Analysis of Randomized Trials with Missing Data: A Frequentist Perspective. FDA - Center for Tobacco Products, 11/2015.
4. Scharfstein: Missing Data and Sensitivity Analyses in Randomized Trials, GlaxoSmithKline, 11/2015.
5. Scharfstein: Global Sensitivity Analysis of Randomized Trials with Missing Data: From the Software Development Trenches, National Institute of Statistical Sciences. 11/2015.
6. Scharfstein: Inference in Randomized Trials with Death and Missingness, Brown University, 4/2016.
7. Scharfstein: Analysis of Randomized Trials with Missing Data, Novartis, 12/2016.
8. Scharfstein: Global Sensitivity Analysis of Randomized Trials with Missing Data, Evidera, 3/2017.

two webinars:

1. Scharfstein: Analysis of Randomized Trials with Missing Data, American Statistical Association, 5/2016.
2. Scharfstein: Analysis of Randomized Trials with Missing Data, American Statistical Association, 9/2016.

one poster:

1. Scharfstein: Global Sensitivity Analysis of Randomized Trials with Missing Data, FDA ORSI Symposium, 4/2015.

one on-line lecture:

1. Scharfstein, Li: Analysis of Prospective Studies with Missing Data, Johns Hopkins University, 7/2016.

4.2.5 Manuscripts

A technical manuscript (see Appendix A) describing the statistical methods described in Section 4.2.1 has been accepted for publication in *Biometrics*. The citation for the manuscript is:

- Scharfstein DO, McDermott A, Diaz I, Carone M, Lunardon N and Turkoz I (2017): "Global Sensitivity Analysis of Repeated Measures Studies with Informative Drop-out: A Semi-Parametric Perspective". *Biometrics*. To Appear. [5]

A second manuscript (see Appendix B) that provides a more accessible explication of the statistical methods described in Section 4.2.1 is under review at *Statistical Methods in Medical Research*. The citation for the manuscript is:

- Scharfstein DO and McDermott A (2017): "Global Sensitivity Analysis of Clinical Trials with Missing Patient Reported Outcomes". *Statistical Methods in Medical Research*. Under Review.

4.2.6 Discussion

In the original application, we proposed to specify type (ii) smoothing assumptions in the form of parametric models for both F^*_{k+1} and H^*_{k+1} (see Section 4.2.1). After the contract was awarded, we discovered, in a couple of case studies, that we were unable to posit parametric models that provided adequate fits to the observed data. This led us to develop methods that rely on more flexible models involving Markovian assumptions and non-parametric smoothing. This required major methodological development since the standard plug-in estimator was no longer guaranteed to have adequate large sample properties.

Using our new methods, we found, in realistic simulation studies, that standard Wald-type confidence intervals did not provide adequate coverage. This led us to explore resampling-based techniques for constructing confidence intervals. Ultimately, we discovered that confidence intervals constructed using a combination of jackknife standard errors coupled with symmetric parametric bootstrap provided reasonable coverage.

We have found that our new procedure can be sensitive to outliers. That is, there can be observations in a given dataset that can have excessive influence on the results. To date, we have not yet found a data-adaptive solution to this problem. We plan to explore this issue in the future.

4.3 Non-monotone Missing Data

In the original application, we planned to develop separate sensitivity analysis procedures for studies with non-monotone missing data. These procedures were to be based on parametric models for the distribution of the observed data. Given the problem discussed above and the fact that, in many studies, missingness prior to last visit on study is a second order issue, we decided to adopt the following imputation strategy.

4.3.1 Methods

Notation: Let M_k denote the indicator that Y_k is unobserved at time k. We assume that $M_0 = 0$ and $M_C = 0$. By construction, $M_k = 1$ if $R_k = 0$. Let $O_k = (M_k, Y_k : M_k = 0)$. The observed data for an individual are $\overline{O}_K$. With this notation, $O_0 = Y_0$ and C can be computed from $\overline{O}_K$ as $\max\{k : M_k = 0\}$.

Identifiability Asssumption: To handle missing data prior to last visit on study, we adapt an untestable identifiability assumption from Robins (1997). Specifically, we assume that, for $0 < k < C$, M_k is independent of Y_k given $\overline{Y}_{k-1}$ and $\underline{O}_k$. In words, this assumption says that, 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. Alternatively, imagine a stratum of individuals who share the same history of outcomes prior to time k and same observed data after time k. Now, imagine splitting the stratum into two sets: those who provide outcome data at time k (stratum B) and those who do not (stratum A). This assumption says that the distribution of the outcome at time k is the same for these two strata. Mathematically, we write this assumption as follows:

$$dF^*(Y_k | \underbrace{M_k = 1, \overline{Y}_{k-1}, \underline{O}_k}_{\text{Stratum A}}) = dF^*(Y_k | \underbrace{M_k = 0, \overline{Y}_{k-1}, \underline{O}_k}_{\text{Stratum B}}) : \quad 0 < k < C. \tag{4}$$

Using Bayes' rule, (4) can be written as follows:

$$P^*(M_k = 1 \mid \overline{Y}_k, \underline{O}_k) = P^*(M_k = 1 \mid \overline{Y}_{k-1}, \underline{O}_k) : \quad 0 < k < C. \tag{5}$$

Letting $\rho^*_k(\overline{Y}_{k-1}, \underline{O}_k) = P^*(M_k = 1 \mid \overline{Y}_{k-1}, \underline{O}_k)$, it can be shown that

$$M_k \perp Y_k \mid \rho^*_k(\overline{Y}_{k-1}, \underline{O}_k) : \quad 0 < k < C \tag{6}$$

Under assumption (4), the joint distribution of $(C, \overline{Y}_C)$ (i.e., the monotonized) is identified by a recursive algorithm.

Smoothing Assumptions: We assume fully parametric restrictions on $\rho^*_k(\overline{Y}_{k-1}, \underline{O}_k)$. Specifically, we assume

$$\text{logit}\{\rho^*_k(\overline{Y}_{k-1}, \underline{O}_k)\} = w_k(\overline{Y}_{k-1}, \underline{O}_k; \nu^*_k); \quad k = 1, \ldots, K-1 \tag{7}$$

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

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

1. Set $k = 1$.

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

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

   where $d_k(\overline{Y}_{k-1}, \underline{O}_k; \nu_k^*)$ is the derivative of $w_k(\overline{Y}_{k-1}, \underline{O}_k; \nu_k)$ with respect to ν_k evaluated at ν_k^*.

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

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

   Let $\mathcal{J}_k = \{i : R_{k,i} = 1, M_{k,i} = 0\}$ and $\mathcal{J}'_k = \{i : R_{k,i} = 1, M_{k,i} = 1\}$. For each individual $i \in \mathcal{J}'_k$, impute $Y_{k,i}$ by randomly selecting an element from the set

$$\left\{ Y_{k,l} : l \in \mathcal{J}_k, \widehat{\rho}_k(\overline{Y}_{k-1,l}, \underline{O}_{k,l}) \text{ is "near" } \widehat{\rho}_k(\overline{Y}_{k-1,i}, \underline{O}_{k,i}) \right\} \tag{8}$$

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

The imputation part of this algorithm is similar in spirit to the sequential missing data imputation strategy of [1].

We use this algorithm to create to M monotone missing datasets. The monotone missing data methods discussed above can then be applied to each of these datasets. Overall point estimates can be obtained by averaging across imputed datasets. That is, $\tilde{\mu}_\alpha = \frac{1}{M} \sum_{m=1}^{M} \tilde{\mu}_{\alpha,m}$, where $\tilde{\mu}_{\alpha,m}$ is the one-step estimator of μ^* based on the mth imputed dataset.

To characterize the uncertainty of our estimation procedure, we utilize bootstrapping techniques.

4.3.2 SAS/R Modules and User Documentation

R and SAS versions 3.0 of `SAMON`, posted on 10/29/2016, includes the above imputation method for handling missing data prior to last visit on-study. R and SAS versions 4.0 are scheduled to be posted by 5/31/2017. Each software release includes user documentation.

4.3.3 Case Studies

Two case studies were developed that illustrate the methods developed to handle studies with non-monotone missing data. These include:

1. Randomized trials designed to evaluate the efficacy of different doses of topiramate in reducing pain in patients with diabetic peripheral polyneuropathy. The primary outcome was patient reported pain, measured measured on a 100-mm Visual Analog Scale.

2. A randomized designed to evaluate the effectiveness of a jobs training program. The intervention was designed to teach unemployed workers skills related to searching for employment such as the preparation of job applications and resumes and how to successfully interview. The primary outcome was employment.

4.3.4 Presentations

No presentations have yet been given that describe the proposed method for imputing missing data prior to last visit on-study. The FDA short course planned for 5/8/2017 will describe this methodology.

4.3.5 Manuscripts

Since we changed our strategy for intermittent missing data, we do not view the associated method as a stand alone technical manuscript. Rather, we plan to discuss our strategy as a separate chapter in the book we plan to publish. We originally planned to publish this book in a traditional format via Cambridge University Press. After further consideration, we plan to publish the book via `Leanpub`, as this will allow us to dynamically update it as our methods and software change. The book is currently two-thirds complete and will be submitted for publication by August 31, 2107.

4.3.6 Discussion

The need to develop and test new methods than those proposed in the original application delayed our ability to complete more presentations and manuscripts by the end of the contract period.

4.4 User Feedback

We only received feedback from a summer intern hired by the FDA. The intern make relatively minor suggestions to improve the software. These suggestions will be incorporated into R and SAS versions 4.0 of `SAMON` to be released by 5/31/2017. None of the suggestions necessitated new methodological developments.

5 Auxiliary Projects

During the course of project, we published additional manuscripts about the analysis of randomized trials with missing data.

The first manuscript relates to the trials in which each enrolled subject is expected to undergo a fixed sequence of "pass/fail" tests, one or more test results may be missing, and interest is focused on estimating the distribution of the earliest test at which a subject "passes" ("fails") that and all subsequent tests. This manuscript was motivated by tuberculosis trials. It was published in the *Annals of Applied Statistics*. The citation for the manuscript is:

- Scharfstein DO, Rotnitzky A, Abraham M, McDermott A, Chaisson R and Geiter L (2016): "On the Analysis of Tuberculosis Studies with Intermittent Missing Sputum Data". *Annals of Applied Statistics*. [4]

The second manuscript discusses how to analyze trials in which (1) patients are at high risk of death, (2) functional outcomes are scheduled to be measured on patients who survive to fixed points in time after randomization and (3) there are missing functional outcome data among survivors. This manuscript was motivated by a trial of treatment for late-stage cancer. It was published in *Biometrics*. The citation for the manuscript is:

- Wang CG, Scharfstein DO, Colantuoni E, Girard T and Yan Y(2015): "Inference in Randomized Trials with Death and Missingness". *Biometrics*. [6]

An R package called `idem` has been developed, a translational manuscript is under revision for the *British Medicial Journal* and manuscript that describes the methods and software is in preparation for the *Journal of Statistical Software*.

Finally, we wrote a letter to the editor of *Journal of General Internal Medicine* about the challenge of missing data in randomized trials in which outcomes are scheduled to be collected from electronic health records. The citation for the letter is:

- Kharrazi A, Wang CG and Scharfstein DO (2015): "Prospective HER-Based Clinical Trials: The Challenge of Missing Data". *Journal of General Internal Medicine*. [**?**]

References

[1] Philip W Lavori, Ree Dawson, and David Shera. A multiple imputation strategy for clinical trials with truncation of patient data. *Statistics in medicine*, 14(17):1913–1925, 1995.

[2] R Little, M Cohen, K Dickersin, S Emerson, J Farrar, C Frangakis, JW Hogan, G. Molenberghs, S. Murphy, J. Neaton, A Rotnitzky, DO Scharfstein, W Shih, J Siegel, and H Stern. *The Prevention and Treatment of Missing Data in Clinical Trials*. The National Academies Press, 2010.

[3] D Scharfstein, A McDermott, W Olson, and Wiegand F. Global sensitivity analysis for repeated measures studies with informative drop-out. *Statistics in Biopharmaceutical Research*, 6:338–348, 2014.

[4] Daniel Scharfstein, Andrea Rotnitzky, Maria Abraham, Aidan McDermott, Richard Chaisson, Lawrence Geiter, et al. On the analysis of tuberculosis studies with intermittent missing sputum data. *The Annals of Applied Statistics*, 9(4):2215–2236, 2015.

[5] DO Scharfstein, A McDermott, I Diaz, Carone M, N Lunardon, and I Turkoz. Global sensitivity analysis for repeated measures studies with informative drop-out: A semi-parametric approach. *Biometrics*, 2017.

[6] Chenguang Wang, Daniel O Scharfstein, Elizabeth Colantuoni, Timothy D Girard, and Ying Yan. Inference in randomized trials with death and missingness. *Biometrics*, 2016.

Appendices

A Scharfstein *et al.*, 2017

BIOMETRICS 000, 000–000 DOI: 000
000 0000

Global Sensitivity Analysis for Repeated Measures Studies with Informative Drop-out: A Semi-Parametric Approach

Daniel Scharfstein[1,*], Aidan McDermott[1], Ivan Diaz[2], Marco Carone[3], Nicola Lunardon[4] and Ibrahim Turkoz[5]

[1]Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, U.S.A.

[2]Department of Healthcare Policy & Research, Weill Cornell Medicine, New York, NY, U.S.A.

[3]University of Washington School of Public Health, Seattle, WA, U.S.A.

[4]Ca' Foscari University of Venice, Venice, Italy

[5]Janssen Research and Development, LLC, Titusville, NJ, U.S.A.

**email:* dscharf@jhu.edu

SUMMARY: In practice, both testable and untestable assumptions are generally required to draw inference about the mean outcome measured at the final scheduled visit in a repeated measures study with drop-out. Scharfstein *et al.* (2014) proposed a sensitivity analysis methodology to determine the robustness of conclusions within a class of untestable assumptions. In their approach, the untestable and testable assumptions were guaranteed to be compatible; their testable assumptions were based on a fully parametric model for the distribution of the observable data. While convenient, these parametric assumptions have proven especially restrictive in empirical research. Here, we relax their distributional assumptions and provide a more flexible, semi-parametric approach. We illustrate our proposal in the context of a randomized trial for evaluating a treatment of schizoaffective disorder.

KEY WORDS: Bootstrap; Cross-Validation; Exponential Tilting; Jackknife; Identifiability; One-Step Estimator; Plug-In Estimator; Selection Bias

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

We consider a prospective cohort study design in which outcomes are scheduled to be collected at fixed time-points after enrollment and the parameter of interest is the mean outcome at the last scheduled study visit. We are concerned with drawing inference about this target parameter in the setting where some study participants prematurely stop providing outcome data.

Identifiability of the target parameter requires untestable assumptions about the nature of the process that leads to premature withdrawal. A common benchmark assumption, introduced by Rubin (1976), is that a patient's decision to withdraw between visits k and $k+1$ depends on outcomes through visit k (i.e., past), but not outcomes after visit k (i.e., future). This assumption has been referred to as *missing at random* (MAR). A weaker version of this assumption, termed *sequential ignorability* (SI), posits that the withdrawal decision depends on outcomes through visit k, but not the outcome at the last scheduled study visit (Birmingham et al., 2003). MAR yields identification of the entire joint distribution of the outcomes, while SI only admits identification of the distribution of the outcome at the last scheduled visit. Both parametric (e.g., Schafer, 1997; Little and Rubin, 2014) and semi-parametric (e.g., van der Laan and Robins, 2003; Tsiatis, 2006) approaches have been proposed for drawing inference about the target parameter under these assumptions.

For such untestable assumptions, it is important to conduct a sensitivity analysis to evaluate the robustness of the resulting inferences (e.g., Little et al., 2010; ICH, 1998; CHMP, 2009). As reviewed by Scharfstein et al. (2014), sensitivity analyses can generally be classified as ad-hoc, local and global. Ad-hoc sensitivity analysis involves analyzing the data using a variety of methods and evaluating whether the inferences they yield are consistent with one another. Local sensitivity analysis evaluates how inferences vary in a small neighborhood of the benchmark assumption. In contrast, global sensitivity analysis considers how inferences vary over a much larger neighborhood of the benchmark assumption.

In addition to untestable assumptions, testable restrictions are needed to combat the so-called "curse of dimensionality" (Robins et al., 1997). Scharfstein et al. (2014) developed a global sensitivity analysis approach whereby the untestable and testable assumptions were guaranteed to be compatible. Their testable assumptions were based on a fully parametric model for the distribution of the observable data. In practice, we have found it particularly challenging to posit parametric models that correspond well with the observed data, as we illustrate in Section 4 below. This has motivated the current paper, in which we relax distributional assumptions and develop a more flexible, semi-parametric extension of the Scharfstein et al. (2014) approach. The techniques of Daniels and Hogan (2008) and Linero and Daniels (2015) provide Bayesian solutions to the same problem and also ensure the compatibility of the untestable and testable assumptions. However, the scalability of their approach to settings with a large number of post-baseline assessments has yet to be demonstrated.

In Section 2, we introduce the data structure and define the target parameter of interest. We also review the identification assumptions of Scharfstein *et al.* (2014). In Section 3, we present our inferential approach. In Section 4, we present results from the reanalysis of a clinical trial in which there was substantial premature withdrawal. In Section 5, we describe the results of a simulation study. We provide concluding remarks in Section 6.

2. Data structure, target parameter, assumptions and identifiability

2.1 *Data structure and target parameter*

Let $k = 0, 1, \ldots, K$ refer in chronological order to the scheduled assessment times, with $k = 0$ corresponding to baseline. Let Y_k denote the outcome scheduled to be measured at assessment k. Define R_k to be the indicator that an individual is on-study at assessment k. We assume that all individuals are present at baseline. Furthermore, we assume that individuals do not contribute any further data once they have missed a visit. This pattern is often referred to as monotone drop-out. Let $C = \max\{k : R_k = 1\}$ and note that $C = K$ implies that the individual must have completed the study. For any given vector $z = (z_1, z_2, \ldots, z_K)$, we define $\overline{z}_k = (z_0, z_1, \ldots, z_k)$ and $\underline{z}_k = (z_{k+1}, z_{k+2}, \ldots, z_K)$. For each individual, $O = (C, \overline{Y}_C)$ is drawn from some distribution P^* contained in the non-parametric model $\mathscr{M}$ of distributions. The observed data consist of n independent draws $O_1, O_2, \ldots, O_n$ from P^*. Throughout, the superscript $*$ will be used to denote the true value of the quantity to which it is appended.

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

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

Our 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.

2.2 *Assumptions*

Assumptions are required to draw inference about μ^* based on the available data. We consider a class of assumptions whereby an individual's decision to drop out in the interval between visits k and $k+1$ is not only influenced by past observable outcomes but by the outcome at visit $k+1$. Towards this end, we adopt the following two assumptions introduced in Scharfstein et al. (2014):

Assumption 1: For $k = 0, 1, \ldots, K-2$,

$$P^*\left(Y_K \leqslant y \mid R_{k+1} = 0, R_k = 1, \overline{Y}_{k+1} = \overline{y}_{k+1}\right) \;=\; P^*\left(Y_K \leqslant y \mid R_{k+1} = 1, \overline{Y}_{k+1} = \overline{y}_{k+1}\right).$$

This says that in the cohort of patients who (1) are on-study at assessment k, (2) share the same outcome history through that visit and (3) have the same outcome at assessment $k+1$, the distribution of Y_K is the same for those last seen at assessment k and those still on-study at $k+1$.

Assumption 2: For $k = 0, 1, \ldots, K-1$,

$$dG^*_{k+1}(y_{k+1} \mid \overline{y}_k) \;\propto\; \exp\{\rho_{k+1}(\overline{y}_k, y_{k+1})\} dF^*_{k+1}(y_{k+1} \mid \overline{y}_k) \ ,$$

where $G^*_{k+1}(y_{k+1} \mid \overline{y}_k) := P^*\left(Y_{k+1} \leqslant y_{k+1} \mid R_{k+1} = 0, R_k = 1, \overline{Y}_k = \overline{y}_k\right)$ and $\rho_{k+1}(\overline{y}_k, y_{k+1})$ is a known, pre-specified function of $\overline{y}_k$ and y_{k+1}.

Conditional on any given history $\overline{y}_k$, this assumption relates the distribution of Y_{k+1} for those patients who drop out between assessments k and $k+1$ to those patients who are on study at $k+1$. The special case whereby ρ_{k+1} is constant in y_{k+1} for all k implies that, conditional on the history $\overline{y}_k$, individuals who drop out between assessments k and $k+1$ have the same distribution of Y_{k+1} as those on-study at $k+1$. If instead ρ_{k+1} is an increasing (decreasing) function of y_{k+1} for some k, then individuals who drop-out between assessments k and $k+1$ tend to have higher (lower) values of Y_{k+1} than those who are on-study at $k+1$.

Setting $\ell^*_{k+1}(\overline{y}_k) := \text{logit}\left\{H^*_{k+1}(\overline{y}_k)\right\} - \log\left\{\int \exp\{\rho_{k+1}(\overline{y}_k, u)\} dF^*_{k+1}(u \mid \overline{y}_k)\right\}$, it can be shown that Assumptions 1 and 2 jointly imply that

$$\text{logit}\left\{P^*\left(R_{k+1} = 0 \,\middle|\, R_k = 1, \overline{Y}_{k+1} = \overline{y}_{k+1}, Y_K = y_K\right)\right\} = \ell^*_{k+1}(\overline{y}_k) + \rho_{k+1}(\overline{y}_k, y_{k+1}) \ .$$

We note that since H^*_{k+1} and F^*_{k+1} are identified from the distribution of the observed data, so is $\ell^*_{k+1}(\overline{y}_k)$. Furthermore, we observe that ρ_{k+1} quantifies the influence of Y_{k+1} on the risk of dropping out between assessments k and $k+1$, after controlling for the past history $\overline{y}_k$. In particular, Y_K is seen to not additionally influence this risk. When ρ_{k+1} does not depend on y_{k+1}, we obtain an assumption weaker than MAR but stronger than SI – we refer to it as SI-1. Under SI-1, the decision to withdraw between visits k and $k+1$ depends on outcomes through visit k but not on the outcomes at visits $k+1$ and K. For specified ρ_{k+1}, Assumptions 1 and 2 place no restriction on the distribution of the observed data. As such, ρ_{k+1} is not an empirically verifiable function.

Assumptions 1 and 2 allow the existence of unmeasured common causes of $Y_0, Y_1, \ldots, Y_K$, but does not allow these causes to directly impact, for patients on study at visit k, the decision to drop out before visit $k+1$. This is no different than under MAR or SI. To allow for a direct impact, one could utilize the sensitivity analysis model of Rotnitzky et al. (1998) which specifies

$$\operatorname{logit}\left\{P^*\left(R_{k+1}=0 \,\middle|\, R_k=1, \overline{Y}_k=\overline{y}_k, Y_K=y_K\right)\right\} = h^*_{k+1}(\overline{y}_k) + q_{k+1}(\overline{y}_k, y_K) \ ,$$

where $h^*_{k+1}(\overline{y}_k) := \operatorname{logit}\left\{H^*_{k+1}(\overline{y}_k)\right\} - \log\left\{\int \exp\{\rho_{k+1}(\overline{y}_k, u)\} dF^*_{K,k}(u \mid R_k=1, \overline{y}_k)\right\}$ and $F^*_{K,k}(u \mid R_k = 1, \overline{y}_k) := P^*(Y_K \leqslant u | R_k = 1, \overline{Y}_k = \overline{y}_k)$. Here, $q_{k+1}(\overline{y}_k, y_K)$ quantifies the influence of the outcome scheduled to be measured at the end of the study on the conditional hazard of last being seen at visit k given the observable past $\overline{y}_k$. The key disadvantage of this model is that we have found that it is challenging for scientific experts to articulate how a distal endpoint affects a more proximal event (i.e., drop-out).

2.3 *Identifiability of target parameter*

Under Assumptions 1 and 2 with given ρ_{k+1}, the parameter μ^* is identifiable. To establish identifiability, it suffices to demonstrate that μ^* can be expressed as a functional of the distribution of the observed data. In the current setting, this follows immediately by noting, through repeated applications of the law of iterated expectations, that

$$\mu^* = \mu(P^*) = E^*\left(\frac{R_K Y_K}{\prod_{k=0}^{K-1}[1+\exp\{\ell^*_{k+1}(\overline{Y}_k) + \rho_{k+1}(\overline{Y}_k, Y_{k+1})\}]^{-1}}\right)$$

The functional $\mu(P^*)$ can be equivalently expressed as

$$\int_{y_0}\cdots\int_{y_K} y_K \prod_{k=0}^{K-1}\Bigg\{dF^*_{k+1}(y_{k+1} \mid \overline{y}_k)\left\{1 - H^*_{k+1}(\overline{y}_k)\right\} + \frac{\exp\{\rho_{k+1}(\overline{y}_k, y_{k+1})\} dF^*_{k+1}(y_{k+1} \mid \overline{y}_k)}{\int \exp\{\rho_{k+1}(\overline{y}_k, u)\} dF^*_{k+1}(u \mid \overline{y}_k)} H^*_{k+1}(\overline{y}_k)\Bigg\} dF^*_0(y_0) \ . \quad (1)$$

3. Statistical inference

3.1 *Plug-in estimator*

Given a fixed function ρ_{k+1}, Scharfstein et al. (2014) proposed to estimate μ^* via the plug-in principle. Specifically, they specify parametric models for both F^*_{k+1} and H^*_{k+1}, estimate parameters in these models by maximum likelihood, estimate F^*_0 nonparametrically using the empirical distribution function, and finally, estimate (1) by Monte Carlo integration using repeated draws from the resulting estimates of F^*_{k+1}, H^*_{k+1} and F^*_0. Since (1) is a smooth functional of F^*_0 and of the finite-dimensional parameters of the models for F^*_{k+1} and H^*_{k+1}, the resulting estimator of μ^* is $n^{1/2}$-consistent and, suitably normalized, tends in distribution to a mean-zero Gaussian random variable.

While simple to describe and easy to implement, this approach has a major drawback: the inferences it generates will be sensitive to correct specification of the parametric models imposed on F^*_{k+1} and H^*_{k+1}. Since the fit of these models is empirically verifiable, the plausibility of the models imposed can be scrutinized in any given application. In several instances, we have found it difficult to find models providing an adequate fit to the observed data. This is a serious problem since model misspecification will generally lead to inconsistent inference, which can translate into inappropriate and misleading scientific conclusions. To provide greater robustness, we instead adopt a more flexible modeling approach.

As noted above, the distribution P^* can be represented in terms of $\{(F^*_{k+1}, H^*_{k+1}) : k = 0, 1, \ldots, K-1\}$. Suppose that P^* is contained in the submodel $\mathscr{M}_0 \subset \mathscr{M}$ of distributions that exhibit a first-order Markovian structure in the sense that $F_{k+1}(y_{k+1} \mid \overline{y}_k) = F_{k+1}(y_{k+1} \mid y_k)$ and $H_{k+1}(\overline{y}_k) = H_{k+1}(y_k)$. We can then estimate F^*_{k+1} and H^*_{k+1} by Nadaraya-Watson kernel estimators of the form:

$$\widehat{F}_{k+1,\lambda_F}(y_{k+1} \mid y_k) := \frac{\sum_{i=1}^n R_{k+1,i} I(Y_{k+1,i} \leqslant y_{k+1}) \phi_{\lambda_F}(Y_{k,i} - y_k)}{\sum_{i=1}^n R_{k+1,i} \phi_{\lambda_F}(Y_{k,i} - y_k)} \quad \text{and} \tag{2}$$

$$\widehat{H}_{k+1,\lambda_H}(y_k) := \frac{\sum_{i=1}^n R_{k,i}(1 - R_{k+1,i}) \phi_{\lambda_F}(Y_{k,i} - y_k)}{\sum_{i=1}^n R_{k,i} \phi_{\lambda_F}(Y_{k,i} - y_k)} \ , \tag{3}$$

where ϕ is a symmetric probability density function, ϕ_λ refers to the rescaled density $y \mapsto \phi(y/\lambda)/\lambda$, and (λ_F, λ_H) is a vector of tuning parameters. In practice, the values of these tuning parameters need to be carefully chosen to ensure the resulting estimators of F^*_{k+1} and H^*_{k+1} perform well. As discussed next, we select the tuning parameters via J-fold cross validation.

Writing $F := (F_1, F_2, \ldots, F_K)$ and $H := (H_1, H_2, \ldots, H_K)$, and denoting a typical realization of the prototypical data unit as $o = (c, \overline{y}_c)$, we may define the loss functions

$$\begin{aligned} L_F(F; F^\circ)(o) &:= \sum_{k=0}^{K-1} r_{k+1} \int \{I(y_{k+1} \leqslant u) - F_{k+1}(u \mid y_k)\}^2 \, dF^\circ_{k+1}(u) \ , \\ L_H(H; H^\circ)(o) &:= \sum_{k=0}^{K-1} r_k \left[r_{k+1} - \{1 - H_{k+1}(y_k)\}\right]^2 H^\circ_{k+1} \end{aligned}$$

with $F^\circ := (F^\circ_1, F^\circ_2, \ldots, F^\circ_K)$ and $H^\circ := (H^\circ_1, H^\circ_2, \ldots, H^\circ_K)$ defined by $F^\circ_{k+1}(u) := P(Y_{k+1} \leqslant u \mid R_{k+1} = 1)$ and $H^\circ_{k+1} := P(R_{k+1} = 0 \mid R_k = 1)$. Here, F° and H° represent collections of distributions and probabilities that can be estimated nonparametrically without the need for smoothing. It can be shown that the true risk mappings $F \mapsto E^*\{L_F(F; F^{\circ *})(O)\}$ and $H \mapsto E^*\{L_H(H; H^{\circ *})(O)\}$ are minimized at $F = F^*$ and $H = H^*$, where $F^{\circ *}$ and $H^{\circ *}$ denote the true value of F° and H°, respectively. Given a random partition of the dataset into J validation samples $\{V_1, V_2, \ldots, V_J\}$ with sample sizes $n_1, n_2 \ldots, n_J$, taken to be approximately equal, the oracle selectors for λ_F and λ_H are (van der Vaart et al., 2006)

$$\widetilde{\lambda}_F := \operatorname*{argmin}_{\lambda_F} \frac{1}{J} \sum_{j=1}^{J} E^*\{L_F(\widehat{F}^{(j)}_{\lambda_F}; \widehat{F}^\circ)(O)\} \quad \text{and} \quad \widetilde{\lambda}_H := \operatorname*{argmin}_{\lambda_H} \frac{1}{J} \sum_{j=1}^{J} E^*\{L_H(\widehat{H}^{(j)}_{\lambda_H}; \widehat{H}^\circ)(O)\}.$$

Here, $\widehat{F}^{(j)}_{k+1,\lambda_F}$ and $\widehat{H}^{(j)}_{k+1,\lambda_H}$ are obtained by computing (2) and (3), respectively, on the dataset excluding individuals in V_j. The estimates of nuisance parameter estimators $\widehat{F}^\circ_{k+1}$ and $\widehat{H}^\circ_{k+1}$ are given by the empirical distribution of the observed values of Y_{k+1} within the subset of individuals with $R_{k+1} = 1$ and by the empirical proportion of individuals with $R_{k+1} = 0$ among those with $R_k = 1$, respectively. The quantities $\widetilde{\lambda}_F$ and $\widetilde{\lambda}_H$ cannot be computed in practice since P^* is unknown. Empirical tuning parameter selectors are given by

$$\widehat{\lambda}_F := \operatorname*{argmin}_{\lambda_F} \widehat{\mathcal{R}}_F(\lambda_F) \quad \text{and} \quad \widehat{\lambda}_H := \operatorname*{argmin}_{\lambda_H} \widehat{\mathcal{R}}_H(\lambda_H),$$

where

$$\begin{aligned} \widehat{\mathcal{R}}_F(\lambda_F) &:= \frac{1}{J}\sum_{j=1}^{J}\frac{1}{n_j}\sum_{i\in V_j} L_F(\widehat{F}^{(j)}_{\lambda_F}; \widehat{F}^{\circ})(O_i) \\ &= \frac{1}{J}\sum_{j=1}^{J}\frac{1}{n_j}\sum_{i\in V_j}\sum_{k=0}^{K-1} R_{k+1,i}\left(\frac{\sum_\ell R_{k+1,\ell}\{I(Y_{k+1,i} \leqslant Y_{k+1,l}) - \widehat{F}^{(j)}_{k+1,\lambda_F}(Y_{k+1,l} \mid Y_{k,i})\}^2}{\sum_\ell R_{k+1,\ell}}\right) \end{aligned}$$

and

$$\begin{aligned} \widehat{\mathcal{R}}_H(\lambda_H) &:= \frac{1}{J}\sum_{j=1}^{J}\frac{1}{n_j}\sum_{i\in V_j} L_H(\widehat{H}^{(j)}_{\lambda_H}; \widehat{H}^{\circ})(O_i) \\ &= \frac{1}{J}\sum_{j=1}^{J}\frac{1}{n_j}\sum_{i\in V_j}\sum_{k=0}^{K-1} \frac{R_{k,i}[R_{k+1,i} - \{1 - \widehat{H}^{(j)}_{k+1,\lambda_H}(Y_{k,i})\}]^2 \sum_\ell R_{k,\ell}(1 - R_{k+1,\ell})}{\sum_\ell R_{k,\ell}} . \end{aligned}$$

The naive substitution estimator of μ^* is $\mu(\widehat{P})$, where $\widehat{P}$ is determined by (2) and (3) computed with tuning parameters $(\widehat{\lambda}_F, \widehat{\lambda}_H)$.

3.2 *Generalized Newton-Raphson estimator*

3.2.1 *Preliminaries.* In order to estimate F^*_{k+1} and H^*_{k+1}, smoothing techniques, as used in (2) and (3), must be utilized in order to borrow strength across subgroups of individuals with differing observed outcome histories. These techniques require the selection of tuning parameters governing the extent of smoothing. As in the above procedure, tuning parameters are generally chosen to achieve an optimal finite-sample bias-variance trade-off for the quantity requiring smoothing - here, conditional distribution and probability mass functions. However, this trade-off may be problematic, since the resulting plug-in estimator $\mu(\widehat{P})$ may suffer from excessive and asymptotically nonnegligible bias due to inadequate tuning. This may prevent the plug-in estimator from having regular asymptotic behavior. In particular, 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. Therefore, the plug-in estimator must be regularized in order to serve as an appropriate basis for drawing statistical inference.

If the parameter of interest is a sufficiently smooth functional on the space of possible data-generating distributions, it is sensible to expect a first-order expansion of the form

$$\mu(P) - \mu(P^*) = \int D(P)(o)d(P - P^*)(o) + Rem(P, P^*) \tag{4}$$

to hold, where $D(P)(o)$ is the evaluation at an observation value o of a so-called gradient of μ at P, and $Rem(P, P^*)$ is a second-order remainder term tending to zero as P tends to P^*. In the context of our problem, this is established formally in Lemma 1. Here, the gradient D is an analytic object used to compute, at any given data-generating distribution P, the change in $\mu(P)$ following a slight perturbation of P. Although the gradient is, in general, not uniquely defined, it must have mean zero and finite variance under sampling from P. A discussion on gradients of statistical parameters can be found in Pfanzagl (1982) and in Appendix A.4 of van der Laan and Rose (2011).

Provided (4) holds and for a given estimator $\widehat{P}$ of P^*, algebraic manipulations leads to

$$\begin{aligned}\mu(\widehat{P}) - \mu(P^*) &= \int D(\widehat{P})(o)d(\widehat{P} - P^*)(o) + Rem(\widehat{P}, P^*) \\ &= \frac{1}{n}\sum_{i=1}^{n} D(P^*)(O_i) + \int \{D(\widehat{P})(o) - D(P^*)(o)\}d(P_n - P^*)(o) \\ &\qquad - \frac{1}{n}\sum_{i=1}^{n} D(\widehat{P})(O_i) + Rem(\widehat{P}, P^*)\ ,\end{aligned}$$

where P_n denotes the empirical distribution based on $O_1, O_2, \ldots, O_n$. If $\widehat{P}$ is a sufficiently well-behaved estimator of P^*, it is often the case that the terms $\int\{D(\widehat{P})(o) - D(P^*)(o)\}d(P_n - P^*)(o)$ and $Rem(\widehat{P}, P^*)$ are asymptotically negligible. However, when $\widehat{P}$ involves smoothing, as in this paper, the term $n^{-1}\sum_{i=1}^{n} D(\widehat{P})(O_i)$ generally tends to zero too slowly to allow $\mu(\widehat{P})$ to be an asymptotically linear estimator of μ^*. Nonetheless, the corrected estimator

$$\widehat{\mu} = \mu(\widehat{P}) + \frac{1}{n}\sum_{i=1}^{n} D(\widehat{P})(O_i)$$

is regular and asymptotically linear with influence function $D(P^*)$, provided that the aforementioned terms are asymptotically negligible. Consequently, $\widehat{\mu}$ converges to μ^* in probability and $n^{1/2}(\widehat{\mu}-\mu^*)$ tends in distribution to a zero-mean Gaussian random variable with variance $\sigma^2 := \int D(P^*)(o)^2 dP^*(o)$. This estimator is, in fact, a direct generalization of the one-step Newton-Raphson procedure used in parametric settings to produce an asymptotically efficient estimator. This correction approach was discussed early on by Ibragimov and Khasminskii (1981), Pfanzagl (1982) and Bickel (1982), among others.

An alternative estimation strategy would consist of employing targeted minimum loss-based estimation (TMLE) to reduce bias due to inadequate tuning (van der Laan and Rubin, 2006). TMLE proceeds by modifying the initial estimator $\widehat{P}$ into an estimator $\widetilde{P}$ that preserves the consistency but also satisfies the equation $n^{-1}\sum_{i=1}^{n} D(\widetilde{P})(O_i) = 0$. As such, the TMLE-based estimator $\widetilde{\mu} := \mu(\widetilde{P})$ of μ^* does not require additional correction and is asymptotically efficient. In preliminary simulation studies (not shown here), we found no substantial difference between the TMLE and our proposed one-step estimator $\widehat{\mu}$. In this case, we favor the latter because of its greater ease of implementation.

3.2.2 *Estimator based on canonical gradient: definition and properties.* In our problem, the one-step estimator can be constructed using any gradient D of the parameter μ defined on the model $\mathscr{M}_0$. Efficiency theory motivates the use of the canonical gradient, often called the efficient influence function, in the construction of the above estimator. The resulting estimator is then not only asymptotically linear but also asymptotically efficient relative to model $\mathscr{M}_0$. The canonical gradient can be obtained by projecting any other gradient onto the tangent space, defined at each $P \in \mathscr{M}_0$ as the closure of the linear span of all score functions of regular one-dimensional parametric models through P. A comprehensive treatment of efficiency theory can be found in Pfanzagl (1982) and Bickel et al. (1993).

In our analysis, we restrict our attention to the class of selection bias functions of the form $\rho_{k+1}(\overline{y}_k, y_{k+1}) = \alpha\rho(y_{k+1})$, where ρ is a specified function of y_{k+1} and α is a sensitivity analysis parameter. With this choice, $\alpha = 0$ corresponds SI-1. For the parameter chosen, the canonical gradient $D^\dagger(P)$ relative to $\mathscr{M}_0$, suppressing notational dependence on α, is given by

$$D^\dagger(P)(o) := a_0(y_0) + \sum_{k=0}^{K-1} r_{k+1} b_{k+1}(y_{k+1}, y_k) + \sum_{k=0}^{K-1} r_k\{1 - r_{k+1} - H_{k+1}(y_k)\}c_{k+1}(y_k) \ ,$$

where expressions for $a_0(y_0)$, b_{k+1} and c_{k+1} are given the Appendix. In this paper we suggest the use of the following one-step estimator

$$\widehat{\mu} := \mu(\widehat{P}) + \frac{1}{n}\sum_{i=1}^{n} D^\dagger(\widehat{P})(O_i)$$

which stems from linearization (4), as formalized in the following lemma.

Lemma 1: *For any $P \in \mathscr{M}_0$, the linearization*

$$\mu(P) - \mu(P^*) = \int D^\dagger(P)(o)d(P - P^*)(o) + Rem(P, P^*)$$

holds for a second-order remainder term $Rem(P, P^*)$ defined in Web Appendix B.

In the above lemma, the expression *second-order* refers to the fact that $Rem(P, P^*)$ can be written as a sum of the integral of the product of two error terms each tending to zero as P tends to P^*, that is,

$$Rem(P, P^*) = \sum_{k=0}^{K-1} \int u_k^*(o)\left\{\Psi_k(P)(o) - \Psi_k(P^*)(o)\right\}\left\{\Theta_k(P)(o) - \Theta_k(P^*)(o)\right\} dP^*(o) \quad (5)$$

for certain smooth operators $\Psi_0, \ldots, \Psi_{K-1}, \Theta_0, \ldots, \Theta_{K-1}$ and weight functions $u_0^*, \ldots, u_{K-1}^*$ that possibly depend on P^*. The proof of Lemma 1 follows from the derivations in Web Appendices A and B.

The proposed estimator is asymptotically efficient relative to model $\mathscr{M}_0$ under certain regularity conditions, as outlined below.

Theorem 1: *If (a) $\int\{D^\dagger(\widehat{P})(o) - D^\dagger(P^*)(o)\}d(P_n - P^*)(o) = o_P(n^{-1/2})$ and (b) $Rem(\widehat{P}, P^*) = o_P(n^{-1/2})$, then $\widehat{\mu} = \mu^* + \frac{1}{n}\sum_{i=1}^{n} D^\dagger(P^*)(O_i) + o_P(n^{-1/2})$ and $\widehat{\mu}$ is an asymptotically efficient estimator of μ^* relative to model $\mathscr{M}_0$.*

The proof of this theorem is provided in Web Appendix C. This result not only justifies the use of $\widehat{\mu}$ in practice but also suggests that a Wald-type asymptotic $100\times(1-\gamma)\%$ confidence interval for μ^* can be constructed as

$$\left(\widehat{\mu}-\frac{z_{\gamma/2}\widehat{\sigma}}{\sqrt{n}},\ \widehat{\mu}+\frac{z_{\gamma/2}\widehat{\sigma}}{\sqrt{n}}\right), \tag{6}$$

where $\widehat{\sigma}^2 := \frac{1}{n}\sum_{i=1}^{n} D^\dagger(\widehat{P})(O_i)^2$ is, under mild conditions, a consistent estimator of the asymptotic variance of $n^{1/2}(\widehat{\mu}-\mu^*)$ and $z_{\gamma/2}$ is the $(1-\gamma/2)$-quantile of the standard normal distribution.

Alternative sufficient conditions can be established to guarantee that conditions (a) and (b) of the theorem above hold. For example, a simple application of Lemma 19.24 of van der Vaart (2000) implies that condition (a) holds provided it can be established that

(i) $D^\dagger(\widehat{P})$ is a consistent estimator of $D^\dagger(P^*)$ in the $L_2(P^*)$-norm in the sense that
$$\int\left\{D^\dagger(\widehat{P})(o)-D^\dagger(P^*)(o)\right\}^2 dP^*(o)\xrightarrow{P}0, \text{and}$$

(ii) for some P^*-Donsker class $\mathscr{F}$, $D^\dagger(\widehat{P})$ falls in $\mathscr{F}$ with probability tending to one.

Since our estimator $\widehat{P}$ is based on kernel regression, and is therefore consistent, condition (i) holds by a simple application of the continuous mapping theorem. Condition (ii) is standard in the analysis of estimators based on data-adaptive estimation of nuisance parameters – Giné and Nickl (2008) presents conditions under which it is expected to hold. Condition (b) is satisfied based on the following argument. The use of cross-validation allows the optimal rate $n^{-2/5}$ to be achieved for the estimator $\widehat{P}$ since the latter is constructed using univariate kernel smoothers. By a repeated use of the Cauchy-Schwartz inequality on the various summands of $Rem(\widehat{P},P^*)$ in (5), the continuous mapping theorem allows us to show that, since each term in $Rem(\widehat{P},P^*)$ is a second-order difference involving smooth transformations of components of $\widehat{P}$ and P, $Rem(\widehat{P},P^*)$ tends to zero in probability at a rate faster than $n^{-1/2}$ under very mild conditions, including that the probabilities $\widehat{\pi}(Y_{j-1},Y_j)$ are bounded away from zero with probability tending to one.

3.3 *Practical considerations in confidence interval construction*

As indicated above, an influence function-based asymptotic confidence interval is given by (6). In Section 5, we present the results of a simulation study in which this confidence interval construction results in poor coverage. The poor coverage can be explained in part by the fact that $\widehat{\sigma}^2$ can be severely downward biased in finite samples (Efron and Gong, 1983).

To address this issue, one can consider the jackknife estimator for σ^2,

$$\widehat{\sigma}^2_{JK} := (n-1)\sum_{i=1}^{n}\left(\widehat{\mu}^{(-i)} - \widehat{\mu}^{(\cdot)}\right)^2$$

where $\widehat{\mu}^{(-i)}$ is the estimator of μ^* with the ith individual deleted from the dataset and $\widehat{\mu}^{(\cdot)} := \frac{1}{n}\sum_{i=1}^{n}\widehat{\mu}^{(-i)}$. This estimator is known to be conservative (Efron and Stein, 1981). Using the jackknife, confidence intervals take the form of (6) with $\widehat{\sigma}$ replaced by $\widehat{\sigma}_{JK}$. Our simulation study in Section 5 demonstrates that these intervals perform better than interval (6) although some undercoverage is still present.

Another possible approach would be to utilize the Studentized bootstrap, wherein confidence intervals are formed by choosing cutpoints based on the distribution of

$$\left\{\frac{\widehat{\mu}_{(b)} - \widehat{\mu}}{\widehat{se}(\widehat{\mu}_{(b)})} : b = 1, 2, \ldots, B\right\} \tag{7}$$

where $\widehat{\mu}_{(b)}$ is the estimator of μ^* based on the bth bootstrap dataset and $\widehat{se}(\widehat{\mu}_{(b)})$ is an estimator of the standard error of $\widehat{\mu}_{(b)}$. One can consider standard error estimators based on the influence function or jackknife. An equal-tailed $(1-\gamma)$ confidence interval takes the form $\left\{\widehat{\mu} - t_{1-\gamma/2}\widehat{se}(\widehat{\mu}), \widehat{\mu} - t_{\gamma/2}\widehat{se}(\widehat{\mu})\right\}$, where t_q is the qth quantile of (7). A symmetric $(1-\gamma)$ confidence interval takes the form $\left\{\widehat{\mu} - t^*_{1-\gamma}\widehat{se}(\widehat{\mu}), \widehat{\mu} + t^*_{1-\gamma}\widehat{se}(\widehat{\mu})\right\}$, where $t^*_{1-\gamma}$ is selected so that the sampling distribution of (7) assigns probability mass $1-\gamma$ between $-t^*_{1-\gamma}$ and $t^*_{1-\gamma}$.

Since our analysis depends on estimation of a correctly specified semiparametric model, it appears sensible to use this model to bootstrap the observed data. In our data analysis and simulation study, we use the estimated distribution of the observed data to generate bootstrapped observed datasets. Our simulation study in Section 5 suggests that the symmetric Studentized bootstrap with jackknifed standard errors performs best.

4. SCA-3004 Study

SCA-3004 was a randomized trial designed to evaluate the efficacy and safety of once-monthly, injectable paliperidone palmitate (PP1M), as monotherapy or as an adjunct to pre-study mood stabilizers or antidepressants, relative to placebo (PBO) in delaying the time to relapse in patients with schizoaffective disorder (SCA) (Fu et al., 2014). The study included multiple phases. After initial screening, an open-label phase consisted of a 13-week, flexible-dose, lead-in period and a 12-week, fixed-dose, stabilization period. Stable patients entered a 15-month, double-blind, relapse-prevention phase and were randomized (1:1) to receive either PP1M or placebo injections at baseline (Visit 0) and every 28 days (Visits 1–15). An additional clinic visit (Visit 16) was scheduled 28 days after the last scheduled injection. In the study, 170 and 164 patients were randomized to the PBO and PP1M arms, respectively. One placebo patient was removed because of excessive influence on the analysis.

The main research question was whether or not outcomes in patients with schizoaffective disorder are better maintained if they continued on treatment rather than being withdrawn from treatment and given placebo. Given the explanatory nature of the research question, an ideal study would follow all randomized patients through Visit 16 while maintaining them on their randomized treatment and examine symptomatic and functional outcomes at that time point. Due to ethical considerations, patients who had signs of clinical relapse (determined by symptoms and clinical response to symptoms) were required to be withdrawn from the study. Thus, clinical data were unavailable post-relapse. In addition to this source of missing data, some patients discontinued due to adverse events, withdrew consent or were lost to follow-up. In the trial, 38% and 60% of patients in the PBO and PP1M arms, respectively, were followed through Visit 16 ($p<0.001$).

We focus our analysis on patient function as measured by the Personal and Social Performance (PSP) scale. The PSP, a validated clinician-reported instrument, is scored from 1 to 100, with higher scores indicating better functioning. It has been argued that a clinically meaningful difference in PSP scores is between 7 and 12 points (Patrick et al., 2009).

We seek to estimate, for each treatment group, the mean PSP at Visit 16 in the counterfactual world in which all patients are followed and treated through Visit 16. Since symptoms and function are correlated, the observed PSP data are likely to be a highly biased representation of the counterfactual world of interest. The mean PSP score among completers was 76.53 and 76.96 in the PBO and PP1M arms, respectively; the estimated difference is -0.43 (95% CI: -3.34 to 2.48), indicating a non-significant treatment effect (p=0.77).

In Figure 1, we display the treatment-specific trajectories of mean PSP score, stratified by last visit time. For patients who prematurely terminate the study, it is interesting to notice that there tends to be a worsening of mean PSP scores at the last visit on study.

[Figure 1 about here.]

Before implementing our proposed sensitivity analysis procedure, we implemented the approach of Scharfstein et al. (2014). For each treatment group, we modeled H^*_{k+1} using logistic regression with visit-specific intercepts and a common effect of Y_k. Additionally, we modeled F^*_{k+1} both using beta and truncated normal regression, each with visit-specific intercepts and a common effect of Y_k. Using estimates of the parameters from these models, we simulated 500,000 datasets for each treatment group. We compared the proportion dropping out before visit $k+1$ among those on study at visit k based on the actual and simulated datasets. We also compared the empirical distribution of PSP scores among those on study at visit $k+1$ based on these datasets using the Kolmogorov-Smirnov statistics. The results for the simulations involving the truncated normal regression and beta regression models are shown in the first and second rows of Figure 2, respectively. The figure suggests that these models do not fit the observed data well. For both the truncated normal and beta regression models, inspection of the actual and simulated distribution of PSP scores at each study visit reveals large discrepancies. For the beta regression model, the contrast between the simulated and actual drop-out probabilities for the PP1M arm is particularly poor.

[Figure 2 about here.]

We contrast the fit of these models to the non-parametric smoothing approach proposed in this paper. For estimation of F^*_{k+1} and H^*_{k+1} based on data from the PBO arm, the optimal choices of λ_F and λ_H are 1.81 and 5.18, respectively. The corresponding optimal choices for the PP1M arm were 1.16 and 8.53. Using the estimated F^*_{k+1} and H^*_{k+1} and optimal choices of λ_F and λ_H, we simulated, as before, 500,000 observed datasets for each treatment group. The results of this simulation in comparison to the actual observed data is shown in the bottom row of Figure 2. In sharp contrast to the parametric modeling approach, the results show excellent agreement between the actual and simulated datasets. For each treatment group, inspection of the actual and simulated distribution of PSP scores at the study visit with the largest Kolmogorov-Smirnov statistics reveals only small discrepancies.

Under SI-1, that is, when $\alpha = 0$, the estimated counterfactual means of interest are 73.31 (95% CI: 69.71 to 76.91) and 74.52 (95% CI: 72.28 to 76.75) for the PBO and PP1M arms, respectively. The estimated treatment difference is -1.20 (95% CI: -5.34 to 2.93). Relative to the complete-case analysis, the SI-1 analysis corrects for bias in a direction that is anticipated: the estimated means under SI-1 are lower and, since there is greater drop-out in the PBO arm, there is a larger correction in that arm. As a consequence, the estimated treatment effect is more favorable to PP1M, although the 95% CI still includes 0. For comparative purposes, the plug-in procedure produces estimates of the means that are slightly lower (73.79 and 74.63) and an estimated treatment difference that is slightly larger (-0.84). The logistic-truncated normal and logistic-beta models for the distribution of the observed data produce markedly different results under SI-1. For the logistic-truncated model, the estimated means are 70.62 (95% CI: 67.01 to 74.24) and 74.68 (95% CI: 72.89 to 76.48) with an estimated difference of -4.06 (95% CI: -8.13 to 0.01); for the logistic-beta model, the estimated means are 64.42 (95% CI: 55.15 to 73.69) and 70.55 (95% CI: 67.53 to 73.56) with an estimated difference of -6.13 (95% CI: -15.96 to 3.71).

In our sensitivity analysis, we chose ρ to be the cumulative distribution of a $\texttt{Beta}(6, 11)$ random variable scaled to the interval 1 to 100. The shape of the function was chosen so that when comparing patients on the low end ($\leqslant 30$) and high end ($\geqslant 80$) of the PSP scale there is relatively less difference in the risk of drop-out than when comparing patients in the middle of the PSP scale (30-80). When $\alpha > 0$ ($\alpha < 0$), patients with higher PSP scores are more (less) likely to drop out. Since lower PSP scores represent worse function, it is plausible that $\alpha \leqslant 0$. For completeness, we ranged the treatment-specific α values from -20 to 20.

In Figure 3 (a) and (b), we display the estimated treatment-specific mean PSP at Visit 16 as a function of α along with 95% pointwise confidence intervals. Figure 3 (c) displays a contour plot of the estimated differences between mean PSP at Visit 16 for PBO versus PP1M for various treatment-specific combinations of α. The point (0,0) corresponds to the SI-1 assumption in both treatment arms. There are no treatment-specific combinations of α for which the estimated treatment differences are clinically meaningful or statistically significant (at the 0.05 level). Figure 3 (d) displays the estimated treatment-specific difference in mean PSP at Visit 16 between non-completers and completers as a function of α. For each treatment group and α, the estimated mean among non-completers is back-calculated from the estimated overall mean ($\widehat{\mu}$), the observed mean among completers ($\sum_i R_{K,i} Y_{K,i} / \sum_i R_{K,i}$) and the proportion of completers ($\sum_i R_{K,i}/n$). The differences in the negative range of α are in the clinically meaningful range, suggesting that the considered choices of the sensitivity analysis parameters are reasonable.

[Figure 3 about here.]

5. Simulation study

As in our goodness-of-fit evaluation above, we simulated, using the estimated F_k^* and H_k^* and optimal choices of λ_F and λ_H, 1,000 datasets for each treatment group. For purposes of the simulation study, we treat the best fit to the observed data as the true data generating mechanism. We evaluate the performance of our procedures for various α values ranging from -10 to 10. The target for each α is the mean computed using formula (1).

The results of our simulation study are displayed in Tables 1 and 2. In Table 1, we report for each treatment group and each α the bias and mean-squared error (MSE) for the plug-in estimator $\mu(\widehat{P})$ and the one-step estimator $\widehat{\mu}$. The results show that the one-step estimator has less bias and lower MSE than the plug-in estimator, although the differences are not dramatic. In Table 2, we report, for each treatment group and each α, 95% confidence interval coverage for six confidence interval procedures: (1) normality-based confidence interval with influence function-based standard error estimator (Normal-IF); (2) normality-based confidence interval with jackknife-based standard error estimator (Normal-JK); (3) equal-tailed, Studentized-t bootstrap confidence interval with influence function-based standard error estimator (Bootstrap-IF-ET); (4) equal-tailed, Studentized-t bootstrap confidence interval with jackknife-based standard error estimator (Bootstrap-JK-ET); (5) symmetric, Studentized-t bootstrap confidence interval with influence function-based standard error estimator (Bootstrap-IF-S); (6) symmetric, Studentized-t bootstrap confidence interval with jackknife-based standard error estimator (Bootstrap-JK-S). Bootstrapping was based on 1,000 datasets.

[Table 1 about here.]

[Table 2 about here.]

We found that the normality-based confidence interval with influence function-based standard error estimator underperformed for both treatment groups and all choices of the sensitivity analysis parameters. In general, the confidence interval procedures that used jackknife standard errors performed better than their counterparts that used the influence function-based standard error estimator. The symmetric, Studentized-t bootstrap confidence interval with jackknife-based standard error estimator (Bootstrap-JK-S) exhibited the most consistent performance across treatment groups and sensitivity analysis parameters.

Our simulation studies reveal some evidence of possible residual bias of the one-step estimator in the context considered. The latter is based upon the use of kernel smoothing in order to estimate the various conditional distribution functions required in the evaluation of μ. It may be possible to achieve better small-sample behavior by employing alternative conditional distribution function estimators with better theoretical properties, e.g., Hall et al. (1999). An ensemble learning approach, e.g., van der Laan et al. (2007), may also yield improved function estimators and decrease the residual bias of the resulting one-step estimator. However, the benefits from improved function estimation may possibly be limited by the relatively small sample size investigated in this simulation study. The use of correction procedures based on higher-order asymptotic representations, as described in Robins et al. (2008), van der Vaart et al. (2014), Carone et al. (2014) and Díaz et al. (2016), may lead to improved performance in smaller samples.

6. Discussion

In this paper, we have developed a semi-parametric method for conducting a global sensitivity analysis of repeated measures studies with monotone missing data. We have developed an open-source software package, called `SAMON`, that implements the methods discussed in this paper.

Our approach does not, as of yet, accommodate auxiliary covariates V_k scheduled to be measured at assessment k. Incorporating $\overline{V}_k$ into the conditioning arguments of Assumptions 1 and 2 can serve to increase the plausibility of these assumptions. In particular, $\overline{V}_k$ can be allowed to influence the decision, for patients on study at visit k, to drop out between visits k and $k+1$, and the unmeasured common causes of $Y_0, Y_1, \ldots, Y_K$ can be allowed to indirectly impact the decision to drop out through their relationship with $\overline{V}_k$. In the context of SCA-3004, it would be useful to incorporate the PANSS (Positive and Negative Symptom Scale) and CGI (Clinical Global Impressions) scores as auxiliary covariates as they are related to planned patient withdrawal as well as correlated with PSP. In future work, we plan to extend the methods developed here to accommodate auxiliary covariates. An extension that handles multiple reasons for drop-out is also worthwhile.

In this paper, we imposed a first-order Markovian assumption in modeling the distribution of the observed data. The plausibility of this assumption was considered in the data analysis as we have evaluated the goodness-of-fit of our model, as illustrated in the bottom row of Figure 2. The Markovian assumption can be relaxed by incorporating the past history using (1) a specified function of the past history, (2) semiparametric single index models (Hall and Yao, 2005) or (3) recently developed methods in data adaptive non-parametric function estimation (van der Laan, 2015).

For given α, our estimator of μ^* is essentially an α-specific weighted average of the observed outcomes at visit K. As a result, it does not allow extrapolation outside the support of these outcomes. We found that one patient in the PBO arm who completed the study with the lowest observed PSP score at the final visit had a very large influence on the analysis. Under SI-1 and other values of α, this patient affected the estimated mean in the PBO group by more than 3 points. In contrast to our approach, a mixed modeling approach, which posits a multivariate normal model for the joint distribution of the full data, does allow extrapolation. Inference under this approach is valid under MAR and correct specification of the multivariate normality assumption. We found that this approach provides much more precise inference, yielding a statistically significant treatment effect in favor of PP1M (treatment effect = -4.7, 95% CI: -7.7 to -1.8). Further, this approach was insensitive to the PBO patient that we removed from our analysis. The disadvantages of the mixed model approach are its reliance on normality and the difficulty of incorporating it into global sensitivity analysis.

In SCA-3004 there is a difference, albeit not statistically significant, in baseline PSP score between treatment groups. The PBO arm has a lower baseline mean PSP score than the PP1M arm (71.2 vs. 72.9). Our method can easily address this imbalance by subtracting out this difference from our effect estimates or by formally modeling change from baseline. In either case, the treatment effect estimates would be less favorable to PP1M. It is notable that a mixed model analysis that models change from baseline does yield a statistically significant effect in favor of PP1M. It may also be of interest to adjust the treatment effect estimates for other baseline covariates, either through regression or direct standardization. We will address this issue in future work. We also plan to develop methods for handling intermittent missing outcome data.

7. Supplementary Materials

Web Appendices referenced in Section 3.2.2 are available with this paper at the Biometrics website on Wiley Online Library. The software package `SAMON` can be found at `www.missingdatamatters.org`.

Acknowledgments and Conflicts

This research was sponsored by contracts from the U.S. Food and Drug Administration and the Patient Centered Outcomes Research Institute as well as NIH grant CA183854. The first and second authors (DS and AM) have received compensation from Janssen Research and

Development, LLC for the provision of consulting services; they received no compensation for preparation of this manuscript or the methods contained herein.

References

Bickel, P. J. (1982). On adaptive estimation. *The Annals of Statistics* pages 647–671.

Bickel, P. J., Klaassen, C. A., Bickel, P. J., Ritov, Y., Klaassen, J., Wellner, J. A., and Ritov, Y. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins University Press, Baltimore.

Birmingham, J., Rotnitzky, A., and Fitzmaurice, G. M. (2003). Pattern-mixture and selection models for analysing longitudinal data with monotone missing patterns. *Journal of the Royal Statistical Society: Series B* **65,** 275–297.

Carone, M., Díaz, I., and van der Laan, M. J. (2014). Higher-order targeted minimum loss-based estimation. Technical report, University of California Berkeley, Department of Biostatistics.

CHMP (2009). *Guideline on Missing Data in Confirmatory Clinical Trials.* EMEA, London.

Díaz, I., Carone, M., and van der Laan, M. J. (2016). Second-order inference for the mean of a variable missing at random. *The International Journal of Biostatistics* **12,** 333–349.

Efron, B. and Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician* **37,** 36–48.

Efron, B. and Stein, C. (1981). The jackknife estimate of variance. *The Annals of Statistics* pages 586–596.

Fu, D.-J., Turkoz, I., Simonson, R. B., Walling, D. P., Schooler, N. R., Lindenmayer, J.-P., Canuso, C. M., and Alphs, L. (2014). Paliperidone palmitate once-monthly reduces risk of relapse of psychotic, depressive, and manic symptoms and maintains functioning in a double-blind, randomized study of schizoaffective disorder. *The Journal of Clinical Psychiatry* **76,** 253–262.

Giné, E. and Nickl, R. (2008). Uniform central limit theorems for kernel density estimators. *Probability Theory and Related Fields* **141,** 333–387.

Hall, P., Wolff, R. C., and Yao, Q. (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association* **94,** 154–163.

Hall, P. and Yao, Q. (2005). Approximating conditional distribution functions using dimension reduction. *Annals of Statistics* pages 1404–1421.

Ibragimov, I. A. and Khasminskii, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*. Springer.

ICH (1998). *Statistical Principles for Clinical Trials (E9)*. Geneva.

Linero, A. R. and Daniels, M. J. (2015). A flexible bayesian approach to monotone missing data in longitudinal studies with nonignorable missingness with application to an acute schizophrenia clinical trial. *Journal of the American Statistical Association* **110,** 45–55.

Little, R., Cohen, M., Dickersin, K., Emerson, S., Farrar, J., Frangakis, C., Hogan, J., Molenberghs, G., Murphy, S., Neaton, J., Rotnitzky, A., Scharfstein, D., Shih, W., Siegel, J., and Stern, H. (2010). *The Prevention and Treatment of Missing Data in Clinical Trials*. The National Academies Press.

Little, R. J. and Rubin, D. B. (2014). *Statistical Analysis with Missing Data*. John Wiley & Sons.

Patrick, D. L., Burns, T., Morosini, P., Rothman, M., Gagnon, D. D., Wild, D., and Adriaenssen, I. (2009). Reliability, validity and ability to detect change of the clinician-rated personal and social performance scale in patients with acute symptoms of schizophrenia. *Current Medical Research and Opinion* **25,** 325–338.

Pfanzagl, J. (1982). *Contributions to a General Asymptotic Statistical Theory*. Springer.

Robins, J., Li, L., Tchetgen, E., van der Vaart, A., et al. (2008). Higher-order influence functions and minimax estimation of nonlinear functionals. In *Probability and Statistics:*

Essays in Honor of David A. Freedman, pages 335–421. Institute of Mathematical Statistics.

Robins, J. M., Ritov, Y., et al. (1997). Toward a curse of dimensionality appropriate (coda) asymptotic theory for semi-parametric models. *Statistics in Medicine* **16,** 285–319.

Rotnitzky, A., Robins, J., and Scharfstein, D. (1998). Semiparametric regression for repeated outcomes with non-ignorable non-response. *Journal of the American Statistical Association* **93,** 1321–1339.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63,** 581–592.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data.* CRC Press.

Scharfstein, D., McDermott, A., Olson, W., and F, W. (2014). Global sensitivity analysis for repeated measures studies with informative drop-out. *Statistics in Biopharmaceutical Research* **6,** 338–348.

Tsiatis, A. (2006). *Semiparametric Theory and Missing Data. 2006.* Springer Verlag, New York.

van der Laan, M. (2015). A generally efficient targeted minimum loss based estimator. Technical report, University of California Berkeley, Department of Biostatistics.

van der Laan, M. and Rubin, D. (2006). Targeted maximum likelihood learning. *The International Journal of Biostatistics* **2,** Article 11.

van der Laan, M. J., Polley, E. C., and Hubbard, A. E. (2007). Super learner. *Statistical Applications in Genetics and Molecular Biology* **6,**.

van der Laan, M. J. and Robins, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality.* Springer Science & Business Media.

van der Laan, M. J. and Rose, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data.* Springer Science & Business Media.

van der Vaart, A. et al. (2014). Higher-order tangent spaces and influence functions.

Statistical Science **29,** 679–686.

van der Vaart, A. W. (2000). *Asymptotic Statistics.* Cambridge University Press.

van der Vaart, A. W., Dudoit, S., and van der Laan, M. J. (2006). Oracle inequalities for multi-fold cross validation. *Statistics & Decisions* **24,** 351–371.

Appendix: Canonical Gradient

The derivation of the canonical gradient is provided in Web Appendix A. Here, we present its explicit form. Let $\pi_{k+1}(y_k, y_{k+1}) = [1 + \exp\{\ell_{k+1}(y_k) + \alpha\rho(y_{k+1})\}]^{-1}$, where

$$\ell_{k+1}(y_k) := \text{logit}\,\{H_{k+1}(y_k)\} - \log\left\{\int \exp\{\rho_{k+1}(\overline{y}_k, u)\} dF_{k+1}(u \mid y_k)\right\}.$$

Let $\pi(\overline{y}_K) = \prod_{k=0}^{K-1} \pi_k(y_k, y_{k+1})$,

$$w_{k+1}(y_k) = E\left(\exp\{\alpha\rho(Y_{k+1})\} \mid R_{k+1} = 1, Y_k = y_k\right),$$

and $g_{k+1}(y_{k+1}, y_k) = \{1 - H_{k+1}(y_k)\}w_{k+1}(y_k) + \exp\{\alpha\rho(y_{k+1})\}H_{k+1}(y_k)$.

The canonical gradient is expressed as

$$D^\dagger(P)(o) := a_0(y_0) + \sum_{k=0}^{K-1} r_{k+1} b_{k+1}(y_{k+1}, y_k) + \sum_{k=0}^{K-1} r_k\{1 - r_{k+1} - H_{k+1}(y_k)\}c_{k+1}(y_k)$$

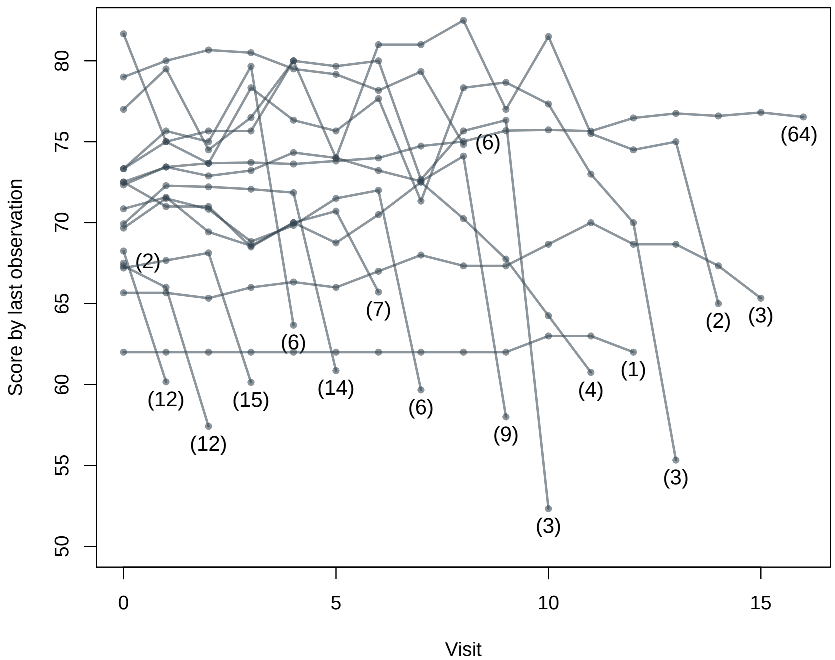
where

$$a_0(y_0) \;=\; E\left(\frac{R_K Y_K}{\pi(\overline{Y}_K)} \middle| Y_0 = y_0\right) - \mu(P)$$

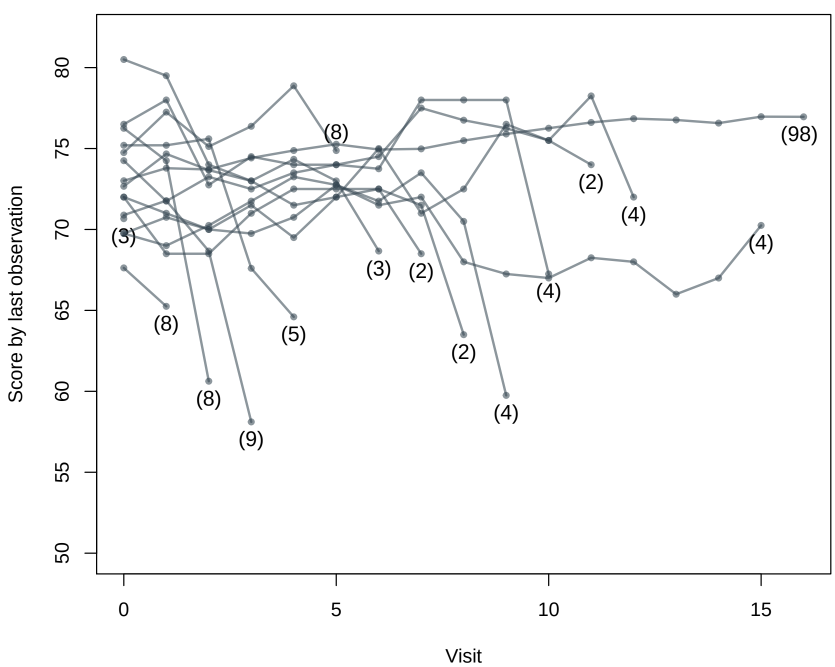
$$\begin{aligned}
& b_{k+1}(y_{k+1}, y_k) \\
&= E\left(\frac{R_K Y_K}{\pi(\overline{Y}_K)} \middle| R_{k+1} = 1, Y_{k+1} = y_{y+1}, Y_k = y_k\right) - E\left(\frac{R_K Y_K}{\pi(\overline{Y}_K)} \middle| R_{k+1} = 1, Y_k = y_k\right) \\
&\quad + E\left(\frac{R_K Y_K}{\pi(\overline{Y}_K)} \left(\frac{\exp\{\alpha\rho(Y_{k+1})\}}{g_{k+1}(Y_{k+1}, Y_k)}\right) \middle| R_{k+1} = 1, Y_k = y_k\right) H_{k+1}(y_k) \left(1 - \frac{\exp\{\alpha\rho(y_{k+1})\}}{w_{k+1}(y_k)}\right)
\end{aligned}$$

$$\begin{aligned}
& c_{k+1}(y_k) \\
&= \; E\left(\frac{R_K Y_K}{\pi(\overline{Y}_K)} \left(\frac{\exp\{\alpha\rho(Y_{k+1})\}}{g_{k+1}(Y_{k+1}, Y_k)}\right) \middle| R_k = 1, Y_k = y_k\right) \\
&\quad - E\left(\frac{R_K Y_K}{\pi(\overline{Y}_K)} \left(\frac{1}{g_{k+1}(Y_{k+1}, Y_k)}\right) \middle| R_k = 1, Y_k = y_k\right) w_{k+1}(y_k)
\end{aligned}$$

Figure 1: Treatment-specific trajectories of mean PSP scores, stratified by last visit time.



(a) Placebo



(b) PP1M

Figure 2: Left column: Comparison of the proportion dropping out before visit $k+1$ among those on study at visit k based on the actual and simulated datasets. Right column: Comparison, using the Kolmogorov-Smirnov statistics, of the empirical distribution of PSP scores among those on study at visit $k+1$ based on the actual and simulated datasets. First row: Logistic regression for conditional probabilities of drop-out and truncated normal regressions for outcomes; Second row: Logistic regression for conditional probabilities of drop-out and beta regressions for outcomes; Third row: Non-parametric smoothing for conditional probabilities of drop-out and for outcomes.

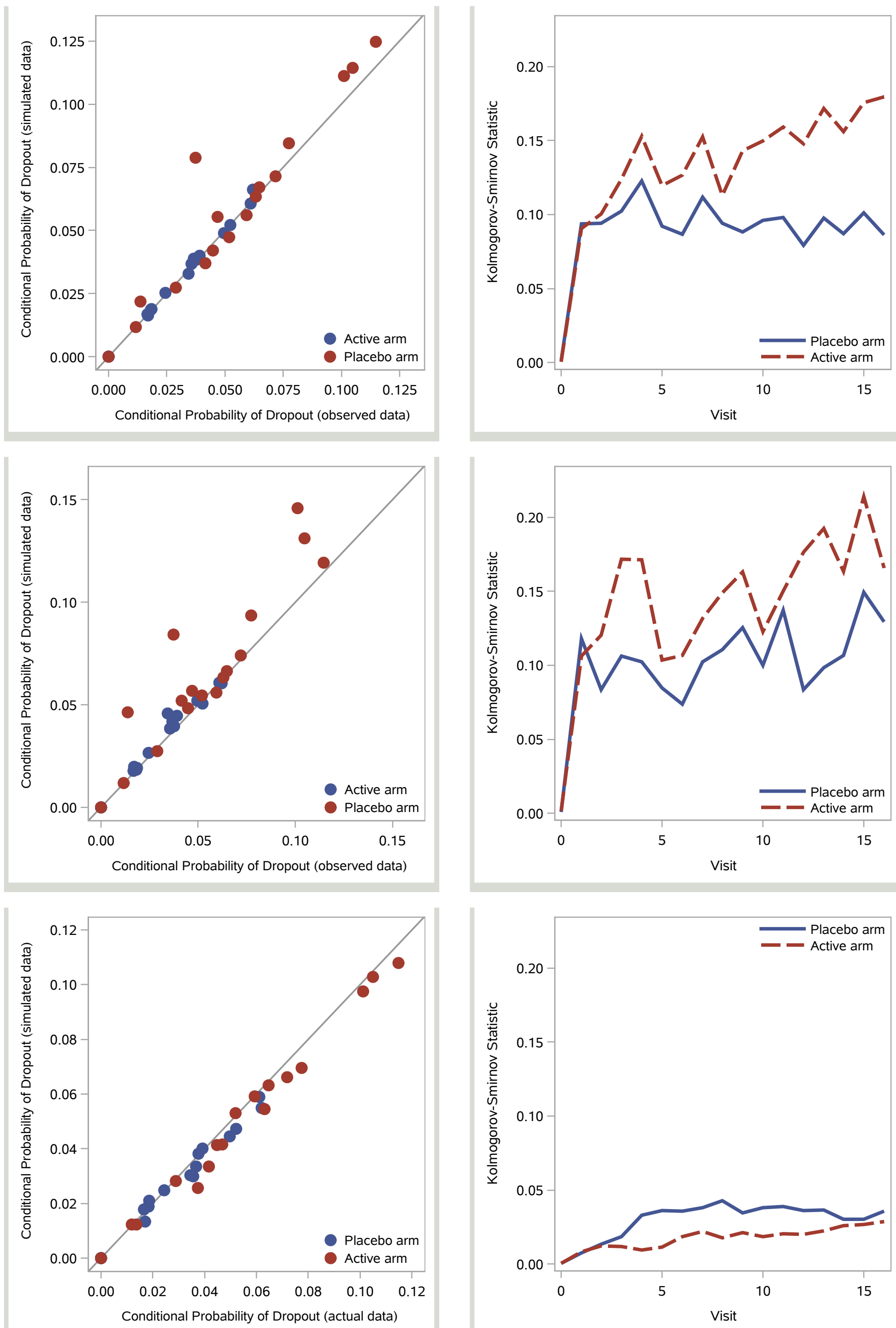
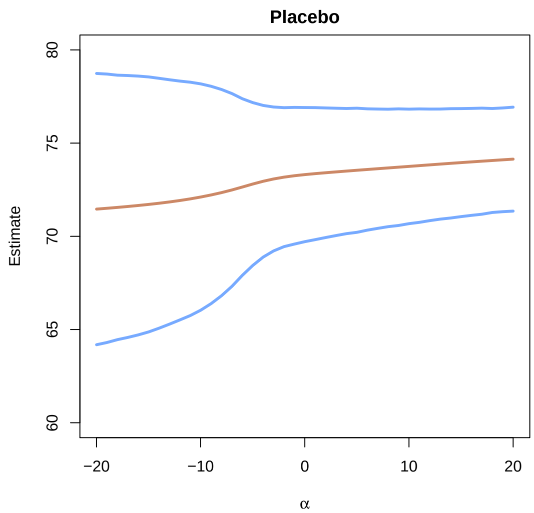
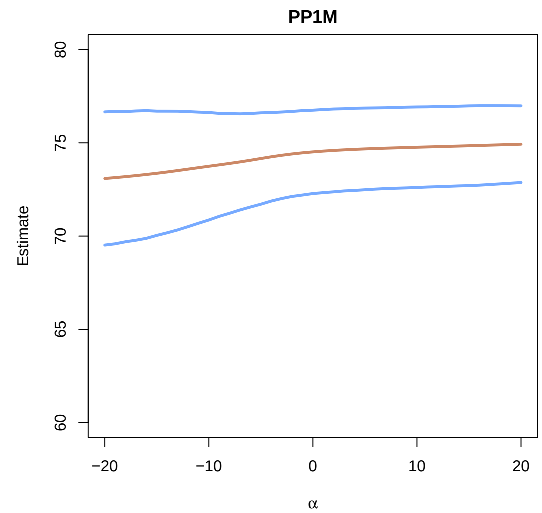


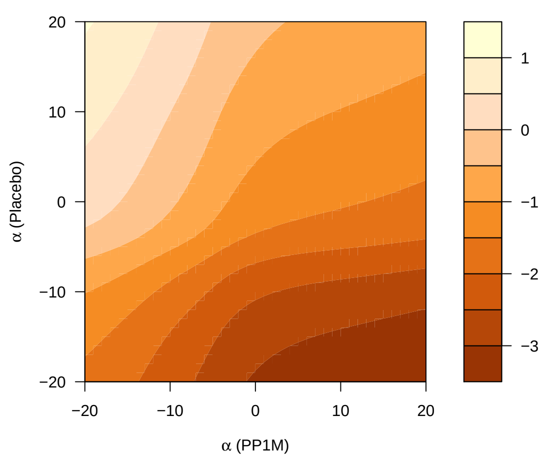
Figure 3: (a) and (b): Treatment-specific mean PSP at Visit 16 as a function of α, along with 95% pointwise confidence intervals; (c): Contour plot of the estimated differences between mean PSP at Visit 16 for PBO vs. PP1M for various treatment-specific combinations of α; (d): Treatment-specific differences between the mean PSP for non-completers and completers, as a function of α.



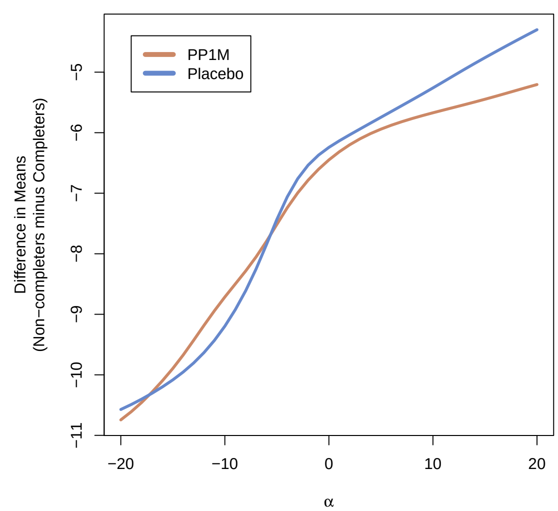
(a)



(b)



(c)



(d)

Table 1: Treatment-specific simulation results: Bias and mean-squared error (MSE) for the plug-in ($\mu(\widehat{P})$) and one-step ($\widehat{\mu}$) estimators, for various choices of α.

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

Table 2: Treatment-specific simulation results: Coverage for (1) normality-based confidence interval with influence function-based standard error estimator (Normal-IF); (2) normality-based confidence interval with jackknife-based standard error estimator (Normal-JK); (3) equal-tailed, Studentized-t bootstrap confidence interval with influence function-based standard error estimator (Bootstrap-IF-ET); (4) equal-tailed, Studentized-t bootstrap confidence interval with jackknife-based standard error estimator (Bootstrap-JK-ET); (5) symmetric, Studentized-t bootstrap confidence interval with influence function-based standard error estimator (Bootstrap-IF-S); (6) symmetric, Studentized-t bootstrap confidence interval with jackknife-based standard error estimator (Bootstrap-JK-S), for various choices of α.

		PBO	PP1M
α	Procedure	Coverage (%)	Coverage (%)
-10	Normal-IF	86.1	88.6
	Normal-JK	92.1	92.6
	Bootstrap-IF-ET	90.2	91.9
	Bootstap-JK-ET	92.4	93.7
	Bootstap-IF-S	92.3	92.7
	Bootstap-JK-S	93.9	94.3
-5	Normal-IF	89.0	91.7
	Normal-JK	94.1	94.2
	Bootstrap-IF-ET	91.7	92.6
	Bootstap-JK-ET	93.6	94.9
	Bootstap-IF-S	94.1	94.2
	Bootstap-JK-S	95.1	95.1
-1	Normal-IF	90.8	93.4
	Normal-JK	94.9	94.8
	Bootstrap-IF-ET	91.0	94.0
	Bootstap-JK-ET	92.8	94.9
	Bootstap-IF-S	94.4	94.7
	Bootstap-JK-S	95.0	95.3
0	Normal-IF	90.7	93.5
	Normal-JK	95.0	94.9
	Bootstrap-IF-ET	92.8	93.9
	Bootstap-JK-ET	94.3	95.0
	Bootstap-IF-S	95.3	94.7
	Bootstap-JK-S	96.0	95.1
1	Normal-IF	90.9	93.5
	Normal-JK	94.9	94.8
	Bootstrap-IF-ET	92.8	93.5
	Bootstap-JK-ET	94.2	95.0
	Bootstap-IF-S	95.3	94.6
	Bootstap-JK-S	96.0	95.2
5	Normal-IF	91.5	93.7
	Normal-JK	94.6	95.1
	Bootstrap-IF-ET	92.6	93.8
	Bootstap-JK-ET	93.8	94.7
	Bootstap-IF-S	94.9	95.1
	Bootstap-JK-S	96.0	95.5
10	Normal-IF	92.1	93.4
	Normal-JK	94.8	95.0
	Bootstrap-IF-ET	92.9	93.8
	Bootstap-JK-ET	93.9	94.8
	Bootstap-IF-S	94.7	95.0
	Bootstap-JK-S	95.6	95.4

Web-based Supplementary Materials for
Global Sensitivity Analysis for Repeated Measures Studies with Informative Drop-out: A Semi-Parametric Approach by
Scharfstein, McDermott, Diaz, Carone, Lunardon and Turkoz

Appendix A: Derivation of Canonical Gradient

In this section, we derive the efficient influence function in the nonparametric model $\mathscr{M}$ (EIF) and in the Markov-restricted model $\mathscr{M}_0$ (EIF_0). To find EIF, we use the fact that the canonical gradient of target parameter is the efficient influence function in model $\mathscr{M}$. To find the EIF_0, we project EIF onto to tangent space for the $\mathscr{M}_0$.

Let P denote a distribution in $\mathscr{M}$, characterized by $P_k(\overline{y}_{k-1}) = P(R_k = 1|R_{k-1} = 0, \overline{Y}_{k-1} = \overline{y}_{kk-1})$, $F_k(y_k|\overline{y}_{k-1}) = P(Y_k \leq y_k|R_k = 1, \overline{Y}_{k-1} = \overline{y}_{k-1})$ and $F_0(y_0) = P(Y_0 \leq y_0)$. In what follows, expectations are taken with respect to P. Let $\{P_\eta : \eta\}$ denote a parametric submodel of $\mathscr{M}$ passing through P (i.e., $P_{\eta=0} = P$). Let $s(O)$ be the score for η evaluated at $\eta = 0$. Let $\mathcal{T}$ denote the tangent space of $\mathscr{M}$. The canonical gradient is defined as the unique element $D \in \mathcal{T}$ that satisfies

$$\frac{\partial}{\partial \eta}\mu(P_\eta)\big|_{\eta=0} = E\{s(O)D(O)\}.$$

We consider parametric submodels, indexed by $\eta = \{\epsilon_0, \varepsilon_k, \upsilon_k : k = 1, \ldots, K\}$, characterized by

$$dF_{0,\eta_0} = dF_0(y_0)\left\{1 + \varepsilon_0 h_0(y_0)\right\} :\ E\{h_0(Y_0)\} = 0$$

$$dF_{k,\eta_k}(y_k|\overline{y}_{k-1}) = dF_k(y_k|\overline{y}_{k-1})\left\{1 + \varepsilon_k h_k(\overline{y}_k)\right\} :\ E(h_k(\overline{Y}_k)|R_k = 1, \overline{Y}_{k-1}) = 0$$

$$P_{k,\upsilon_k}(\overline{y}_{k-1}) = \frac{P_k(\overline{y}_{k-1})\exp\{\upsilon_k l_k(\overline{y}_{k-1})\}}{P_k(\overline{y}_{k-1})\exp\{\upsilon_k l_k(\overline{y}_{k-1})\} + 1 - P_k(\overline{y}_{k-1})} :\ l_k(\cdot) \text{ is any function of } \overline{y}_{k-1}$$

The associated score functions evaluated at $\eta = 0$ are $h_0(Y_0)$, $R_k h_k(\overline{Y}_k)$ and $R_{k-1}\{R_k - P_k(\overline{Y}_{k-1})\}l_k(\overline{Y}_{k-1})$.

The target parameter as a functional of P_η is

$$\mu(P_\eta) = \int \cdots \int y_K \prod_{j=1}^{K} \Bigg(dF_j(y_j|\overline{y}_{j-1})\left\{1 + \varepsilon_j h_j(\overline{y}_j)\right\} \left(\frac{P_j(\overline{y}_{j-1})\exp\{\upsilon_j l_j(\overline{y}_{j-1})\}}{P_j(\overline{y}_{j-1})\exp\{\upsilon_j l_j(\overline{y}_{j-1})\} + 1 - P_j(\overline{y}_{j-1})} \right)$$
$$+ \frac{dF_j(y_j|\overline{y}_{j-1})\exp\{\alpha r(y_j)\}\left\{1 + \varepsilon_j h_j(\overline{y}_j)\right\}\left(\frac{1 - P_j(\overline{y}_{j-1})}{P_j(\overline{y}_{j-1})\exp\{\upsilon_j l_j(\overline{y}_{j-1})\} + 1 - P_j(\overline{y}_{j-1})}\right)}{\int \exp\{\alpha r(y_j)\} dF_j(y_j|\overline{y}_{j-1})\left\{1 + \varepsilon_j h_j(\overline{y}_j)\right\}} \Bigg) dF_0(y_0)\left\{1 + \varepsilon_0 h_0(y_0)\right\}$$

In what follows, we represent $P_k(\overline{y}_{k-1})$, $dF_k(y_k|\overline{y}_{k-1})$, $dF_0(y_0)$, $\alpha r(y_k)$, $h_k(\overline{y}_k)$ and $l_k(\overline{y}_{k-1})$ by P_k, Q_k, Q_0, αr_k, h_k and l_k, respectively. The derivative with respect to ε_0 (evaluated at $\eta = 0$) is $d\varepsilon_0(h_0)$ equal to

$$\int \cdots \int y_K \prod_{j=1}^{K} \left(Q_j P_j + \frac{Q_j \exp(\alpha r_j)(1 - P_j)}{\int \exp(\alpha r_j) Q_j} \right) Q_0 h_0$$

The derivative with respect to ε_k (evaluated at $\eta = 0$) is $d\varepsilon_k(h_k)$ equal to

$$\int \cdots \int y_K \prod_{j \neq k} \left(Q_j P_j + \frac{Q_j \exp(\alpha r_j)(1 - P_j)}{\int \exp(\alpha r_j) Q_j} \right)$$
$$\times \left\{ Q_k P_k h_k + \frac{\left(\int \exp(\alpha r_k) Q_k\right)\exp(\alpha r_k) Q_k h_k - Q_k \exp(\alpha r_k) \int \exp(\alpha r_k) Q_k h_k}{\left(\int \exp(\alpha r_k) Q_k\right)^2}(1 - P_k) \right\} Q_0$$

The derivative with respect to υ_k (evaluated at $\eta = 0$) is $d\upsilon_k(l_k)$ equal to

$$\int \cdots \int y_K \prod_{j \neq k}^{K} \left(Q_j P_j + \frac{Q_j \exp(\alpha r_j)(1 - P_j)}{\int \exp(\alpha r_j) Q_j} \right) \left(Q_k (P_k(1 - P_k) l_k) - \frac{Q_k \exp(r_k)(P_k(1 - P_k) l_k)}{\int \exp(\alpha r_k) Q_k} \right) Q_0$$

Any element of can be expressed as $\mathcal{T}$ can be expressed as

$$a(Y_0) + \sum_{k=1}^{K} R_k b_k(\overline{Y}_k) + \sum_{k=1}^{K} R_{k-1}(R_k - P_k) c_k(\overline{Y}_{k-1})$$

where $E\{a(Y_0)\} = 0$, $E(b_j(\overline{Y}_j)|R_j = 1, \overline{Y}_{j-1}) = 0$ and $c_j(\cdot)$ is any function of $\overline{Y}_{j-1}$. We need to find functions $a(Y_0)$,$b_k(\overline{Y}_k)$ and $c_k(\overline{Y}_{k-1})$ such that

$$E\{a(Y_0)h_0(Y_0)\} = d\varepsilon_0(h_0)$$
$$E\{R_k b_k(\overline{Y}_k)h_k(\overline{Y}_k)\} = d\varepsilon_k(h_k)$$
$$E\{R_{k-1}(R_k - P_k)^2 c_k(\overline{Y}_{k-1})l_k(\overline{Y}_{k-1})\} = d\nu_k(l_k)$$

First, notice that

$$E\{a_0(Y_0)h_0(Y_0)\} = \int_{y_0} a_0(y_0)h_0(y_0)Q_0$$

and

$$d\varepsilon_0(h_0) = \int_{y_0} \left\{ \int \cdots \int y_K \prod_{j=1}^{K} \left(Q_j P_j + \frac{Q_j \exp(\alpha r_j)(1-P_j)}{\int \exp(\alpha r_j) Q_j} \right) \right\} h_0 Q_0$$

Thus, $E\{a_0^*(Y_0)h_0(Y_0)\} = d\varepsilon_0(h_0)$ where

$$a_0^*(Y_0) = \int_{y_1} \cdots \int_{y_K} y_K \frac{\prod_{j=1}^K \left(Q_j P_j + \frac{Q_j \exp(\alpha r_j)(1-P_j)}{\int \exp(\alpha r_j) Q_j} \right)}{\prod_{j=1}^K Q_j P_j} \prod_{j=1}^K Q_j P_j = E\left(\frac{R_K Y_K}{\prod_{j=1}^K \left[1 + \exp\left\{ g_j(\overline{Y}_{j-1}) + \alpha r(Y_j) \right\}\right]^{-1}} \middle| Y_0 \right)$$

with $g_k = \log\left\{(1-P_k)/P_k\right\} - \log\left(\int \exp(r_k)Q_k\right)$. Note that $a_0^*(Y_0)$ does not have mean zero; it actually has mean μ. We can substract out its mean to obtain $a_0(Y_0) = a_0^*(Y_0) - \mu$; note that $E\{a_0(Y_0)h_0(Y_0)\} = d\varepsilon_0(h_0)$.

Second, notice that

$$E\left\{R_k b_k(\overline{Y}_k)h_k(\overline{Y}_k)\right\} = \int_{y_0} \cdots \int_{y_k} b_k(\overline{y}_k)h_k(\overline{y}_k) \left(\prod_{j=1}^k Q_j P_j \right) Q_0$$

and

$$\begin{aligned}
&d\varepsilon_k(h_k) \\
&= \int_{y_0} \cdots \int_{y_k} \int_{y_{k+1}} \cdots \int_{y_K} \left(y_K \frac{\prod_{j=1}^K \left(Q_j P_j + \frac{Q_j \exp(\alpha r_j)(1-P_j)}{\int \exp(\alpha r_j) Q_j} \right)}{\prod_{j=1}^K Q_j P_j} \right) \left(\prod_{j=k+1}^K Q_j P_j \right) \\
&\qquad \left(h_k - \frac{\exp(\alpha r_k)(1-P_k) \int_{y_k^*} \exp(\alpha r_k^*) Q_k^* h_k^*}{P_k \left(\int \exp(\alpha r_k) Q_k\right)^2 + \exp(\alpha r_k)(1-P_k) \int \exp(\alpha r_k) Q_k} \right) \left(\prod_{j=1}^k Q_j P_j \right) Q_0 \\
&= \int_{y_0} \cdots \int_{y_k} \int_{y_{k+1}} \cdots \int_{y_K} \left(y_K \frac{\prod_{j=1}^K \left(Q_j P_j + \frac{Q_j \exp(\alpha r_j)(1-P_j)}{\int \exp(\alpha r_j) Q_j} \right)}{\prod_{j=1}^K Q_j P_j} \right\} \left(\prod_{j=k+1}^K Q_j P_j \right) h_k \left(\prod_{j=1}^k Q_j P_j \right) Q_0 - \\
&\quad \int_{y_0} \cdots \int_{y_{k-1}} \int_{y_k} \int_{y_{k+1}} \cdots \int_{y_K} \left(y_K \frac{\prod_{j=1}^K \left(Q_j P_j + \frac{Q_j \exp(\alpha r_j)(1-P_j)}{\int \exp(\alpha r_j) Q_j} \right)}{\prod_{j=1}^K Q_j P_j} \right) \left(Q_k \prod_{j=k+1}^K Q_j P_j \right) \\
&\qquad \left(\frac{\exp(\alpha r_k)(1-P_k) \int_{y_k^*} \exp(\alpha r_k^*) Q_k^* h_k^*}{P_k \left(\int \exp(\alpha r_k) Q_k\right)^2 + \exp(\alpha r_k)(1-P_k) \int \exp(\alpha r_k) Q_k} \right) \left(P_k \prod_{j=1}^{k-1} Q_j P_j \right) Q_0 \\
&= \int_{y_0} \cdots \int_{y_k} \left\{ \int_{y_{k+1}} \cdots \int_{y_K} \left(y_K \frac{\prod_{j=1}^K \left(Q_j P_j + \frac{Q_j \exp(\alpha r_j)(1-P_j)}{\int \exp(\alpha r_j) Q_j} \right)}{\prod_{j=1}^K Q_j P_j} \right) \left(\prod_{j=k+1}^K Q_j P_j \right) \right\} h_k \left(\prod_{j=1}^k Q_j P_j \right) Q_0 - \\
&\quad \int_{y_0} \cdots \int_{y_{k-1}} \int_{y_k^*} \left\{ \int_{y_k} \int_{y_{k+1}} \cdots \int_{y_K} \left(y_K \frac{\prod_{j=1}^K \left(Q_j P_j + \frac{Q_j \exp)\alpha r_j)(1-P_j)}{\int \exp(\alpha r_j) Q_j} \right)}{\prod_{j=1}^K Q_j P_j} \right\} \left(Q_k \prod_{j=k+1}^K Q_j P_j \right) \\
&\qquad \left(\frac{\exp(\alpha r_k)(1-P_k)}{P_k \left(\int \exp(\alpha r_k) Q_k\right)^2 + \exp(\alpha r_k)(1-P_k) \int \exp(\alpha r_k) Q_k} \right) \right\} \exp(\alpha r_k^*) h_k^* \left(Q_k^* P_k \prod_{j=1}^{k-1} Q_j P_j \right) Q_0 \\
&= \int_{y_0} \cdots \int_{y_k} E\left(\frac{R_K Y_K}{\prod_{j=1}^K \left[1 + \exp\left\{ g_j(\overline{Y}_{j-1}) + \alpha r(Y_j) \right\}\right]^{-1}} \middle| R_k = 1, \overline{Y}_k = \overline{y}_k \right) h_k \left(\prod_{j=1}^k Q_j P_j \right) Q_0 -
\end{aligned}$$

$$\int_{y_0} \cdots \int_{y_k} E\left(\frac{R_K Y_K}{\prod_{j=1}^K \left[1+\exp\left\{g_j(\overline{Y}_{j-1})+\alpha r(Y_j)\right\}\right]^{-1}} \right.$$
$$\left.\left(\frac{\exp(\alpha r_k)\,(1-P_k)}{P_k\left(\int \exp(\alpha r_k)Q_k\right)^2+\exp(\alpha r_k)(1-P_k)\int \exp(\alpha r_k)Q_k} \right) \middle| \, R_k=1, \overline{Y}_{k-1}=\overline{y}_{k-1} \right) \exp(\alpha r_k) h_k \left(\prod_{j=1}^k Q_j P_j\right) Q_0$$

Thus $E\left\{R_k b_k^*(\overline{Y}_k) h_k(\overline{Y}_k)\right\} = d\varepsilon_k(h_k)$, where

$$b_k^*(\overline{Y}_k)$$
$$= E\left(\frac{R_K Y_K}{\prod_{j=1}^K \left[1+\exp\left\{g_j(\overline{Y}_{j-1})+\alpha r(Y_j)\right\}\right]^{-1}} \middle| R_k=1, \overline{Y}_k \right) -$$
$$E\left(\frac{R_K Y_K}{\prod_{j=1}^K \left[1+\exp\left\{g_j(\overline{Y}_{j-1})+\alpha r(Y_j)\right\}\right]^{-1}} \left(\frac{\exp(r_k)\,(1-P_k)}{P_k\left(\int \exp(\alpha r_k)Q_k\right)^2+\exp(\alpha r_k)(1-P_k)\int \exp(\alpha r_k)Q_k} \right) \middle| \, R_k=1, \overline{Y}_{k-1} \right) \times$$
$$\exp(\alpha r_k)$$

Note that $b_k^*(\overline{Y}_k)$ does not have mean 0 given $R_k=1$ and $\overline{Y}_{k-1}$. We can substract out $E(b_k^*(\overline{Y}_k)|R_k=1,\overline{Y}_{k-1})$ to obtain

$$b_k(\overline{Y}_k)$$
$$= E\left(\frac{R_K Y_K}{\prod_{j=1}^K \left[1+\exp\left\{g_j(\bar{Y}_{j-1})+\alpha r(Y_j)\right\}\right]^{-1}} | R_k=1, \overline{Y}_k \right) - E\left(\frac{R_K Y_K}{\prod_{j=1}^K \left[1+\exp\left\{g_j(\overline{Y}_{j-1})+\alpha r(Y_j)\right\}\right]^{-1}} | R_k=1, \overline{Y}_{k-1} \right) -$$
$$E\left(\frac{R_K Y_K}{\prod_{j=1}^K \left[1+\exp\left\{g_j(\overline{Y}_{j-1})+\alpha r(Y_j)\right\}\right]^{-1}} \left(\frac{\exp(\alpha r_k)\,(1-P_k)}{P_k\left(\int \exp(\alpha r_k)Q_k\right)^2+\exp(\alpha r_k)(1-P_k)\int \exp(\alpha r_k)Q_k} \right) | R_k=1, \overline{Y}_{k-1} \right) \times$$
$$\exp(\alpha r_k)+$$
$$E\left(\frac{R_K Y_K}{\prod_{j=1}^K \left[1+\exp\left\{g_j(\overline{Y}_{j-1})+\alpha r(Y_j)\right\}\right]^{-1}} \left(\frac{\exp(\alpha r_k)\,(1-P_k)}{P_k\left\{\int \exp(\alpha r_k)Q_k\right)^2+\exp(\alpha r_k)(1-P_k)\int \exp(\alpha r_k)Q_k} \right) | R_k=1, \overline{Y}_{k-1} \right) \times$$
$$E\left(\exp(\alpha r_k) | R_k=1, \overline{Y}_{k-1}\right)$$

Note that $E\left\{R_k b_k(\overline{Y}_k) h_k(\overline{Y}_k)\right\} = d\varepsilon_k(h_k)$ since $E\left(h(Y_k)|R_k=1,\overline{Y}_{k-1}\right)=0$.

Third, notice that

$$E\{R_{k-1}(R_k-P_k)^2 c_k(\overline{Y}_{k-1}) l_k(\overline{Y}_{k-1})\} = \int_{y_0} \cdots \int_{y_{k-1}} c_k(\overline{y}_{k-1}) P_k(1-P_k) l_k(\overline{y}_{k-1}) \left(\prod_{j=1}^{k-1} Q_j P_j\right) Q_0$$

and

$$d\upsilon_k(l_k)$$
$$= \int_{y_0} \cdots \int_{y_{k-1}} \left\{ \int_{y_k} \cdots \int_{y_K} y_K \frac{\prod_{j=1}^K \left(Q_j P_j + \frac{Q_j \exp(\alpha r_j)(1-P_j)}{\int \exp(\alpha r_j) Q_j}\right)}{\prod_{j=1}^K Q_j P_j} \frac{Q_k - \frac{Q_k \exp(\alpha r_k)}{\int \exp(\alpha r_k) Q_k}}{Q_k P_k + \frac{Q_k \exp(\alpha r_k)(1-P_k)}{\int \exp(\alpha r_k) Q_k}} \left(\prod_{j=k}^K Q_j P_j\right) \right\} \times$$
$$P_k(1-P_k) l_k \left(\prod_{j=1}^{k-1} Q_j P_j\right) Q_0$$

Thus,

$$c_k(\overline{Y}_{k-1}) = E\left(\frac{R_K Y_K}{\prod_{j=1}^K \left[1+\exp\left\{g_j(\overline{Y}_{j-1})+\alpha r(Y_j)\right\}\right]^{-1}} \left(\frac{1-\frac{\exp(\alpha r_k)}{\int \exp(\alpha r_k) Q_k}}{P_k + \frac{\exp(\alpha r_k)(1-P_k)}{\int \exp(\alpha r_k) Q_k}} \right) \middle| \, R_{k-1}=1, \overline{Y}_{k-1} \right)$$

This completes the derivation of EIF.

The tangent space for $\mathscr{M}_0$, $\mathcal{T}_0$, has elements of the form:

$$\tilde{a}(Y_0) + \sum_{k=1}^K R_k \tilde{b}_k(Y_k, Y_{k-1}) + \sum_{k=1}^K R_{k-1}(R_k-P_k)\tilde{c}_k(Y_{k-1})$$

where $E\{\tilde{a}(Y_0)\} = 0$ and $E(\tilde{b}_k(Y_k, Y_{k-1})|R_k = 1, Y_{k-1}) = 0$. The projection of EIF onto $\mathcal{T}_0$ has $\tilde{a}(Y_0) = a(Y_0)$, $\tilde{b}_k(Y_k, Y_{k-1}) = E(b_k(\overline{Y}_k)|R_k = 1, Y_k, Y_{k-1})$ and $\tilde{c}_k(Y_{k-1}) = E(c_k(\overline{Y}_{k-1})|R_{k-1} = 1, Y_{k-1})$. This completes the derivation of EIF_0

Appendix B: Remainder Term - Explicit Form and Derivation

The remainder term has the following explicit form:

$$\begin{aligned} Rem(P, P^*) &= \mu(P) - \mu(P^*) + \int D^\dagger(P)(o)dP^*(o) \\ &= \sum_{k=0}^{K-1} Rem_{1,k}(P, P^*) + \sum_{k=1}^{K-1} Rem_{2,k}(P, P^*) + \sum_{k=2}^{K-1} Rem_{3,k}(P, P^*) \ , \end{aligned}$$

where

$$\begin{aligned} Rem_{1,k}(P, P^*) &:= E^*\left\{R_k E^*\left(R_{k+1}\exp\{\alpha r(Y_{k+1})\}|R_k = 1, Y_k\right) Rem_{1,k,1}(P, P^*)(O) Rem_{1,k,2}(P, P^*)(O)\right\}, \\ Rem_{1,k,1}(P, P^*)(O) &:= \frac{E\left(\frac{R_K Y_K \exp\{\alpha r(Y_{k+1})\}}{\prod_{j\neq k+1}\pi_j(Y_{j-1}, Y_j)} \,\middle|\, R_k = 1, Y_k\right)}{E(R_{k+1}\exp\{\alpha r(Y_{k+1})\}\,|\,R_k = 1, Y_k)} - \frac{E^*\left(\frac{R_K Y_K \exp\{\alpha r(Y_{k+1})\}}{\prod_{j=1}^{k}\pi_j(Y_{j-1}, Y_j)\prod_{j=k+2}^{K}\pi_j^*(Y_{j-1}, Y_j)} \,\middle|\, R_k = 1, Y_k\right)}{E^*(R_{k+1}\exp\{\alpha r(Y_{k+1})\}\,|\,R_k = 1, Y_k)} \ , \\ Rem_{1,k,2}(P, P^*)(O) &:= \frac{H^*_{k+1}(Y_k)}{E^*(R_{k+1}\exp\{\alpha r(Y_{k+1})\}\,|\,R_k = 1, Y_k)} - \frac{H_{k+1}(Y_k)}{E\left(R_{k+1}\exp\{\alpha r(Y_{k+1})\}\,|\,R_k = 1, Y_k\right)} \ , \\ Rem_{2,k}(P, P^*) &:= E^*\left\{R_k Rem_{2,k,1}(P, P^*)(O) Rem_{2,k,2}(P, P^*)(O)\right\} \ , \\ Rem_{2,k,1}(P, P^*)(O) &:= E^*\left(\frac{R_K Y_K}{\prod_{j=k+1}^{K}\pi_j(Y_{j-1}, Y_j)} \,\middle|\, R_k = 1, Y_k\right) - E\left(\frac{R_K Y_K}{\prod_{j=k+1}^{K}\pi_j(Y_{j-1}, Y_j)} \,\middle|\, R_k = 1, Y_k\right) \ , \\ Rem_{2,k,2}(P, P^*)(O) &:= E\left(\frac{1}{\prod_{j=1}^{k}\pi_j(Y_{j-1}, Y_j)} \,\middle|\, R_k = 1, Y_k\right) - E^*\left(\frac{1}{\prod_{j=1}^{k}\pi_j(Y_{j-1}, Y_j)} \,\middle|\, R_k = 1, Y_k\right) \ , \\ Rem_{3,k}(P, P^*) &:= E^*\left\{R_k Rem_{3,k,1}(P, P^*)(O) Rem_{3,k,2}(P, P^*)(O)\right\} \ , \\ Rem_{3,k,1}(P, P^*)(O) &:= E^*\left(\frac{R_K Y_K}{\prod_{j=k+1}^{K}\pi_j(Y_{j-1}, Y_j)} \,\middle|\, R_k = 1, Y_k\right) - E\left(\frac{R_K Y_K}{\prod_{j=k+1}^{K}\pi_j(Y_{j-1}, Y_j)} \,\middle|\, R_k = 1, Y_k\right) \\ Rem_{3,k,2}(P, P^*)(O) &:= E\left(\frac{1}{\prod_{j=1}^{k}\pi_j(Y_{j-1}, Y_j)} \,\middle|\, R_k = 1, Y_k, Y_{k-1}\right) - E^*\left(\frac{1}{\prod_{j=1}^{k}\pi_j(Y_{j-1}, Y_j)} \,\middle|\, R_k = 1, Y_k, Y_{k-1}\right) \ . \end{aligned}$$

To derive this expression, we start by writing

$$\mu(P^*) = \sum_{k=1}^{K} E^*\left\{\left(\frac{1}{\pi_k^*(Y_{k-1}, Y_k)} - \frac{1}{\pi_k(Y_{k-1}, Y_k)}\right)\frac{R_K Y_K}{\prod_{l=1}^{k-1}\pi_l(Y_{l-1}, Y_l)\prod_{l=k+1}^{K}\pi_l^*(Y_{l-1}, Y_l)}\right\} + E^*\left[\frac{R_K Y_K}{\prod_{l=1}^{K}\pi_l(Y_{l-1}, Y_l)}\right]$$

Using this expression, we can write

$$\begin{aligned} Rem(P, P^*) = &-\sum_{k=1}^{K} E^*\left\{\left(\frac{1}{\pi_k^*(Y_{k-1}, Y_k)} - \frac{1}{\pi_k(Y_{k-1}, Y_k)}\right)\frac{R_K Y_K}{\prod_{l=1}^{k-1}\pi_l(Y_{l-1}, Y_l)\prod_{l=k+1}^{K}\pi_l^*(Y_{l-1}, Y_l)}\right\} - \\ &E^*\left(\frac{R_K Y_K}{\prod_{l=1}^{K}\pi_l(Y_{l-1}, Y_l)}\right) + E^*\left\{E\left(\frac{R_K Y_K}{\prod_{l=1}^{K}\pi_l(Y_{l-1}, Y_l)} \,\middle|\, Y_0\right)\right\} + \\ &\sum_{k=1}^{K} E^*\left\{R_k E\left(\frac{R_K Y_K}{\prod_{l=1}^{K}\pi_l(Y_{l-1}, Y_l)} \,\middle|\, R_k = 1, Y_k, Y_{k-1}\right)\right\} - \sum_{k=1}^{K} E^*\left\{R_k E\left(\frac{R_K Y_K}{\prod_{l=1}^{K}\pi_l(Y_{l-1}, Y_l)} \,\middle|\, R_k = 1, Y_{k-1}\right)\right\} + \\ &\sum_{k=1}^{K} E^*\left\{R_k E\left(\frac{R_K Y_K}{\prod_{l=1}^{K}\pi_l(Y_{l-1}, Y_l)}\left(\frac{\exp\{\alpha r(Y_k)\}}{g_k(Y_k, Y_{k-1})}\right) \,\middle|\, R_k = 1, Y_{k-1}\right) H_k(Y_{k-1})\right\} - \\ &\sum_{k=1}^{K} E^*\left(R_k E\left(\frac{R_K Y_K}{\prod_{l=1}^{K}\pi_l(Y_{l-1}, Y_l)}\left(\frac{\exp\{\alpha r(Y_k)\}}{g_k(Y_k, Y_{k-1})}\right) \,\middle|\, R_k = 1, Y_{k-1}\right) H_k(Y_{k-1})\frac{\exp\{\alpha r(Y_k)\}}{w_k(Y_{k-1})}\right) + \\ &\sum_{k=1}^{K} E^*\left\{R_{k-1}\{1 - R_k - H_k(Y_{k-1})\}E\left(\frac{R_K Y_K}{\prod_{l=1}^{K}\pi_l(Y_{l-1}, Y_l)}\left(\frac{\exp\{\alpha r(Y_k)\}}{g_k(Y_k, Y_{k-1})}\right) \,\middle|\, R_{k-1} = 1, Y_{k-1}\right)\right\} - \\ &\sum_{k=1}^{K} E^*\left\{R_{k-1}\{1 - R_k - H_k(Y_{k-1})\}E\left(\frac{R_K Y_K}{\prod_{l=1}^{K}\pi_l(Y_{l-1}, Y_l)}\left(\frac{1}{g_k(Y_k, Y_{k-1})}\right) \,\middle|\, R_{k-1} = 1, Y_{k-1}\right) w_k(Y_{k-1})\right\} \end{aligned}$$

Let $E_k(Y_{k-1}) = E\left(R_k \exp\{\alpha r(Y_k)\}\,|\,R_{k-1} = 1, Y_{k-1}\right)$. Through the properties of conditional expectations, we can write

$$Rem(P, P^*) = -\sum_{k=1}^{K} E^*\left\{R_{k-1}\left(\frac{H_k^*(Y_{k-1})}{E_k^*(Y_{k-1})} - \frac{H_k(Y_{k-1})}{E_k(Y_{k-1})}\right) E^*\left(\frac{R_K Y_K \exp\{\alpha r(Y_k)\}}{\prod_{l=1}^{k-1}\pi_l(Y_{l-1}, Y_l)\prod_{l=k+1}^{K}\pi_l^*(Y_{l-1}, Y_l)} \,\middle|\, R_{k-1} = 1, Y_{k-1}\right)\right\} -$$

$$
\begin{aligned}
& E^*\left(\frac{R_K Y_K}{\prod_{l=1}^K \pi_l(Y_{l-1},Y_l)}\right) + E^*\left\{E\left(\frac{R_K Y_K}{\prod_{l=1}^K \pi_l(Y_{l-1},Y_l)}\,\middle|\, Y_0\right)\right\} + \\
& \sum_{k=1}^K E^*\left\{R_k E\left(\frac{R_K Y_K}{\prod_{l=1}^K \pi_l(Y_{l-1},Y_l)}\,\middle|\, R_k=1,Y_k,Y_{k-1}\right)\right\} - \\
& \sum_{k=1}^K E^*\left\{R_{k-1}\left(\frac{1-H_k^*(Y_{k-1})}{1-H_k(Y_{k-1})}\right) E\left(\frac{R_K Y_K}{\prod_{l=1}^K \pi_l(Y_{l-1},Y_l)}\,\middle|\, R_{k-1}=1,Y_{k-1}\right)\right\} + \\
& \sum_{k=1}^K E^*\left\{R_{k-1}\left(\frac{1-H_k^*(Y_{k-1})}{1-H_k(Y_{k-1})}\right) E\left(\frac{R_K Y_K}{\prod_{l=1}^K \pi_l(Y_{l-1},Y_l)}\left(\frac{\exp\{\alpha r(Y_k)\}}{g_k(Y_k,Y_{k-1})}\right)\,\middle|\, R_{k-1}=1,Y_{k-1}\right) H_k(Y_{k-1})\right\} - \\
& \sum_{k=1}^K E^*\left(R_{k-1} E\left(\frac{R_K Y_K}{\prod_{l=1}^K \pi_l(Y_{l-1},Y_l)}\left(\frac{\exp\{\alpha r(Y_k)\}}{g_k(Y_k,Y_{k-1})}\right)\,\middle|\, R_{k-1}=1,Y_{k-1}\right) H_k(Y_{k-1})\frac{E_k^*(Y_{k-1})}{E_k(Y_{k-1})}\right) + \\
& \sum_{k=1}^K E^*\left\{R_{k-1}\left(\frac{H_k^*(Y_{k-1})-H_k(Y_{k-1})}{H_k(Y_{k-1})}\right) E\left(\frac{R_K Y_K}{\prod_{l=1}^K \pi_l(Y_{l-1},Y_l)}\left(\frac{\exp\{\alpha r(Y_k)\}}{g_k(Y_k,Y_{k-1})}\right)\,\middle|\, R_{k-1}=1,Y_{k-1}\right) H_k(Y_{k-1})\right\} - \\
& \sum_{k=1}^K E^*\left\{R_{k-1}\left(\frac{H_k^*(Y_{k-1})-H_k(Y_{k-1})}{1-H_k(Y_{k-1})}\right) E\left(\frac{R_K Y_K}{\prod_{l=1}^K \pi_l(Y_{l-1},Y_l)}\left(\frac{1}{g_k(Y_k,Y_{k-1})}\right)\,\middle|\, R_{k-1}=1,Y_{k-1}\right) E_k(Y_{k-1})\right\}
\end{aligned}
$$

Using the fact that $\frac{1}{\pi_k(Y_{k-1},Y_k)} = 1 + \frac{H_k(Y_{k-1})}{E_k(Y_{k-1})}\exp\{\alpha r(Y_k)\}$, we can write

$$
\begin{aligned}
& Rem(P,P^*) \\
&= -\sum_{k=1}^K E^*\left\{R_{k-1}\left(\frac{H_k^*(Y_{k-1})}{E_k^*(Y_{k-1})} - \frac{H_k(Y_{k-1})}{E_k(Y_{k-1})}\right) E^*\left(\frac{R_K Y_K \exp\{\alpha r(Y_k)\}}{\prod_{l=1}^{k-1}\pi_l(Y_{l-1},Y_l)\prod_{l=k+1}^K \pi_l^*(Y_{l-1},Y_l)}\,\middle|\, R_{k-1}=1,Y_{k-1}\right)\right\} - \\
&\quad E^*\left(\frac{R_K Y_K}{\prod_{l=1}^K \pi_l(Y_{l-1},Y_l)}\right) + E^*\left\{E\left(\frac{R_K Y_K}{\prod_{l=2}^K \pi_l(Y_{l-1},Y_l)}\,\middle|\, Y_0\right)\right\} + E^*\left(E\left(\frac{R_K Y_K \exp\{\alpha r(Y_1)\}}{\prod_{l=2}^K \pi_l(Y_{l-1},Y_l)}\,\middle|\, Y_0\right)\frac{H_1(Y_0)}{E_1(Y_0)}\right) + \\
&\quad \sum_{k=1}^K E^*\left\{R_k E\left(\frac{R_K Y_K}{\prod_{l=1}^K \pi_l(Y_{l-1},Y_l)}\,\middle|\, R_k=1,Y_k,Y_{k-1}\right)\right\} - \\
&\quad \sum_{k=1}^K E^*\left\{R_{k-1}\left(\frac{1-H_k^*(Y_{k-1})}{1-H_k(Y_{k-1})}\right) E\left(\frac{R_K Y_K}{\prod_{l=1}^{k-1}\pi_l(Y_{l-1},Y_l)\prod_{l=k+1}^K \pi_l(Y_{l-1},Y_l)}\,\middle|\, R_{k-1}=1,Y_{k-1}\right)\right\} - \\
&\quad \sum_{k=1}^K E^*\left(R_{k-1}\left(\frac{1-H_k^*(Y_{k-1})}{1-H_k(Y_{k-1})}\right) E\left(\frac{R_K Y_K \exp\{\alpha r(Y_k)\}}{\prod_{l=1}^{k-1}\pi_l(Y_{l-1},Y_l)\prod_{l=k+1}^K \pi_l(Y_{l-1},Y_l)}\,\middle|\, R_{k-1}=1,Y_{k-1}\right)\frac{H_k(Y_{k-1})}{E_k(Y_{k-1})}\right) + \\
&\quad \sum_{k=1}^K E^*\left(R_{k-1}\left(\frac{1-H_k^*(Y_{k-1})}{1-H_k(Y_{k-1})}\right) E\left(\frac{R_K Y_K \exp\{\alpha r(Y_k)\}}{\prod_{l=1}^{k-1}\pi_l(Y_{l-1},Y_l)\prod_{l=k+1}^K \pi_l(Y_{l-1},Y_l)}\,\middle|\, R_{k-1}=1,Y_{k-1}\right)\frac{H_k(Y_{k-1})}{E_k(Y_{k-1})}\right) - \\
&\quad \sum_{k=1}^K E^*\left(R_{k-1} E\left(\frac{R_K Y_K \exp\{\alpha r(Y_k)\}}{\prod_{l=1}^{k-1}\pi_l(Y_{l-1},Y_l)\prod_{l=k+1}^K \pi_l(Y_{l-1},Y_l)}\,\middle|\, R_{k-1}=1,Y_{k-1}\right)\frac{H_k(Y_{k-1})}{E_k(Y_{k-1})}\frac{E_k^*(Y_{k-1})}{E_k(Y_{k-1})}\right) + \\
&\quad \sum_{k=1}^K E^*\left(R_{k-1}\left(\frac{H_k^*(Y_{k-1})-H_k(Y_{k-1})}{H_k(Y_{k-1})}\right) E\left(\frac{R_K Y_K \exp\{\alpha r(Y_k)\}}{\prod_{l=1}^{k-1}\pi_l(Y_{l-1},Y_l)\prod_{l=k+1}^K \pi_l(Y_{l-1},Y_l)}\,\middle|\, R_{k-1}=1,Y_{k-1}\right)\frac{H_k(Y_{k-1})}{E_k(Y_{k-1})}\right) - \\
&\quad \sum_{k=1}^K E^*\left\{R_{k-1}\left(\frac{H_k^*(Y_{k-1})-H_k(Y_{k-1})}{1-H_k(Y_{k-1})}\right) E\left(\frac{R_K Y_K}{\prod_{l=1}^{k-1}\pi_l(Y_{l-1},Y_l)\prod_{l=k+1}^K \pi_l(Y_{l-1},Y_l)}\,\middle|\, R_{k-1}=1,Y_{k-1}\right)\right\}
\end{aligned}
$$

Cancelling and combining terms, we obtain

$$
\begin{aligned}
& Rem(P,P^*) \\
&= -\sum_{k=1}^K E^*\left\{R_{k-1}\left(\frac{H_k^*(Y_{k-1})}{E_k^*(Y_{k-1})} - \frac{H_k(Y_{k-1})}{E_k(Y_{k-1})}\right) E^*\left(\frac{R_K Y_K \exp\{\alpha r(Y_k)\}}{\prod_{l=1}^{k-1}\pi_l(Y_{l-1},Y_l)\prod_{l=k+1}^K \pi_l^*(Y_{l-1},Y_l)}\,\middle|\, R_{k-1}=1,Y_{k-1}\right)\right\} - \\
&\quad E^*\left(\frac{R_K Y_K}{\prod_{l=1}^K \pi_l(Y_{l-1},Y_l)}\right) + E^*\left\{E\left(\frac{R_K Y_K}{\prod_{l=2}^K \pi_l(Y_{l-1},Y_l)}\,\middle|\, Y_0\right)\right\} + E^*\left[E\left[\frac{R_K Y_K \exp\{\alpha r(Y_1)\}}{\prod_{l=2}^K \pi_l(Y_{l-1},Y_l)}\,\middle|\, Y_0\right]\frac{H_1(Y_0)}{E_1(Y_0)}\right] + \\
&\quad \sum_{k=1}^K E^*\left(R_k E\left(\frac{R_K Y_K}{\prod_{l=1}^K \pi_l(Y_{l-1},Y_l)}\,\middle|\, R_k=1,Y_k,Y_{k-1}\right)\right) - \\
&\quad \sum_{k=1}^K E^*\left\{R_{k-1} E\left(\frac{R_K Y_K}{\prod_{l=1}^{k-1}\pi_l(Y_{l-1},Y_l)\prod_{l=k+1}^K \pi_l(Y_{l-1},Y_l)}\,\middle|\, R_{k-1}=1,Y_{k-1}\right)\right\} - \\
&\quad \sum_{k=1}^K E^*\left(R_{k-1} E\left(\frac{R_K Y_K \exp\{\alpha r(Y_k)\}}{\prod_{l=1}^{k-1}\pi_l(Y_{l-1},Y_l)\prod_{l=k+1}^K \pi_l(Y_{l-1},Y_l)}\,\middle|\, R_{k-1}=1,Y_{k-1}\right)\frac{H_k(Y_{k-1})}{E_k(Y_{k-1})}\frac{E_k^*(Y_{k-1})}{E_k(Y_{k-1})}\right) +
\end{aligned}
$$

$$\sum_{k=1}^{K} E^*\left(R_{k-1}\left(\frac{H_k^*(Y_{k-1})-H_k(Y_{k-1})}{H_k(Y_{k-1})}\right) E\left(\frac{R_K Y_K \exp\{\alpha r(Y_k)\}}{\prod_{l=1}^{k-1}\pi_l(Y_{l-1},Y_l)\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_{k-1}=1, Y_{k-1}\right)\frac{H_k(Y_{k-1})}{E_k(Y_{k-1})}\right)$$

Through further algebraic manipulation, we obtain that $Rem(P,P^*) = Rem_1(P,P^*) + Rem_2(P,P^*)$, where

$$\begin{aligned}
&Rem_1(P,P^*)\\
&= -\sum_{k=1}^{K} E^*\Bigg\{R_{k-1}E_k^*(Y_{k-1})\left(\frac{H_k^*(Y_{k-1})}{E_k^*(Y_{k-1})}-\frac{H_k(Y_{k-1})}{E_k(Y_{k-1})}\right)\\
&\quad\left(\frac{E^*\left(\frac{R_K Y_K \exp\{\alpha r(Y_k)\}}{\prod_{l=1}^{k-1}\pi_l(Y_{l-1},Y_l)\prod_{l=k+1}^{K}\pi_l^*(Y_{l-1},Y_l)}\middle| R_{k-1}=1,Y_{k-1}\right)}{E_k^*(Y_{k-1})}-\frac{E\left(\frac{R_K Y_K \exp\{\alpha r(Y_k)\}}{\prod_{l=1}^{k-1}\pi_l(Y_{l-1},Y_l)\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_{k-1}=1,Y_{k-1}\right)}{E_k(Y_{k-1})}\right)\Bigg\}
\end{aligned}$$

and

$$Rem_2(P,P^*) = -E^*\left(\frac{R_K Y_K}{\prod_{l=1}^{K}\pi_l(Y_{l-1},Y_l)}\right)+\sum_{k=1}^{K} E^*\left\{R_k E\left(\frac{R_K Y_K}{\prod_{l=1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k,Y_{k-1}\right)\right\}-\sum_{k=1}^{K-1} E^*\left\{R_k E\left(\frac{R_K Y_K}{\prod_{l=1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\}$$

Notice that $Rem_1(P,P^*)$ is second order. It remains to show that $Rem_2(P,P^*)$ is second order. In our derivation, we use the fact that, for $k=1,\ldots,K-1$,

$$E\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k,Y_{k-1}\right) = E\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)$$

and

$$\begin{aligned}
&E^*\left\{R_k E\left(\frac{1}{\prod_{l=1}^{k}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)E^*\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\}\\
&= E^*\left\{R_{k+1} E\left(\frac{1}{\prod_{l=1}^{k}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)E^*\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_{k+1}=1,Y_{k+1},Y_k\right)\right\}\\
&= E^*\left\{\frac{R_{k+1}}{\pi_{k+1}(Y_k,Y_{k+1})} E\left(\frac{1}{\prod_{l=1}^{k}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)E^*\left(\frac{R_K Y_K}{\prod_{l=k+2}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_{k+1}=1,Y_{k+1}\right)\right\}\\
&= E^*\left\{R_{k+1} E\left(\frac{1}{\prod_{l=1}^{k+1}\pi_l(Y_{l-1},Y_l)}\middle| R_{k+1}=1,Y_{k+},Y_k\right)E^*\left(\frac{R_K Y_K}{\prod_{l=k+2}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_{k+1}=1,Y_{k+1}\right)\right\}
\end{aligned}$$

We can write

$$\begin{aligned}
Rem_2(P,P^*) = &-E^*\left\{R_1 E^*\left(\frac{1}{\pi_1(Y_1,Y_0)}\middle| R_1=1,Y_1\right)E^*\left(\frac{R_K Y_K}{\prod_{l=2}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_1=1,Y_1\right)\right\}+\\
&E^*\left\{R_1 E^*\left(\frac{1}{\pi_1(Y_1,Y_0)}\middle| R_1=1,Y_1\right)E\left(\frac{R_K Y_K}{\prod_{l=2}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_1=1,Y_1\right)\right\}-\\
&E^*\left\{R_1 E\left(\frac{1}{\pi_1(Y_1,Y_0)}\middle| R_1=1,Y_1\right)E\left(\frac{R_K Y_K}{\prod_{l=2}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_1=1,Y_1\right)\right\}+\\
&\sum_{k=2}^{K} E^*\left\{R_k E\left(\frac{1}{\prod_{l=1}^{k}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k,Y_{k-1}\right)E\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\}-\\
&\sum_{k=2}^{K-1} E^*\left\{R_k E\left(\frac{1}{\prod_{l=1}^{k}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)E\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\}
\end{aligned}$$

We add the following zero terms to $Rem_2(P,P^*)$:

$$\begin{aligned}
A(P,P^*) = \sum_{k=1}^{K-1}\Bigg[&E^*\left\{R_k E\left(\frac{1}{\prod_{l=1}^{k}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)E^*\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\}-\\
&E^*\left\{R_k E\left(\frac{1}{\prod_{l=1}^{k}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)E^*\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\}\Bigg]
\end{aligned}$$

$$= \sum_{k=1}^{K-1} E^*\left\{R_k E\left(\frac{1}{\prod_{l=1}^{k}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right) E^*\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\} -$$
$$\sum_{k=2}^{K} E^*\left\{R_k E\left(\frac{1}{\prod_{l=1}^{k}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k,Y_{k-1}\right) E^*\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\}$$

$$B(P,P^*) = \sum_{k=2}^{K-1}\left[E^*\left\{R_k E^*\left(\frac{1}{\prod_{l=1}^{k}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k,Y_{k-1}\right) E^*\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\} -\right.$$
$$\left. E^*\left\{R_k E^*\left(\frac{1}{\prod_{l=1}^{k}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k,Y_{k-1}\right) E^*\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\}\right] +$$
$$\left[E^*\left\{R_k E^*\left(\frac{1}{\prod_{l=1}^{k}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k,Y_{k-1}\right) E\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\} -\right.$$
$$\left. E^*\left\{R_k E^*\left(\frac{1}{\prod_{l=1}^{k}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k,Y_{k-1}\right) E\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\}\right]$$

So,

$$Rem_2(P,P^*) = -E^*\left\{R_1 E^*\left(\frac{1}{\pi_1(Y_1,Y_0)}\middle| R_1=1,Y_1\right) E^*\left(\frac{R_K Y_K}{\prod_{l=2}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_1=1,Y_1\right)\right\} +$$
$$E^*\left\{R_1 E^*\left(\frac{1}{\pi_1(Y_1,Y_0)}\middle| R_1=1,Y_1\right) E\left(\frac{R_K Y_K}{\prod_{l=2}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_1=1,Y_1\right)\right\} -$$
$$E^*\left\{R_1 E\left(\frac{1}{\pi_1(Y_1,Y_0)}\middle| R_1=1,Y_1\right) E\left(\frac{R_K Y_K}{\prod_{l=2}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_1=1,Y_1\right)\right\} +$$
$$E^*\left\{R_1 E\left(\frac{1}{\pi_1(Y_1,Y_0)}\middle| R_1=1,Y_1\right) E^*\left(\frac{R_K Y_K}{\prod_{l=2}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_1=1,Y_1\right)\right\} +$$
$$\sum_{k=2}^{K} E^*\left\{R_k E\left(\frac{1}{\prod_{l=1}^{k}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k,Y_{k-1}\right) E\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\} -$$
$$\sum_{k=2}^{K-1} E^*\left\{R_k E\left(\frac{1}{\prod_{l=1}^{k}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right) E\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\} +$$
$$\sum_{k=2}^{K-1} E^*\left\{R_k E\left(\frac{1}{\prod_{l=1}^{k}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right) E^*\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\} -$$
$$\sum_{k=2}^{K} E^*\left\{R_k E\left(\frac{1}{\prod_{l=1}^{k}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k,Y_{k-1}\right) E^*\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\} +$$
$$\sum_{k=2}^{K-1}\left[E^*\left\{R_k E^*\left(\frac{1}{\prod_{l=1}^{k}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k,Y_{k-1}\right) E^*\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\} -\right.$$
$$\left. E^*\left\{R_k E^*\left(\frac{1}{\prod_{l=1}^{k}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k,Y_{k-1}\right) E^*\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\}\right] +$$
$$\left[E^*\left\{R_k E^*\left(\frac{1}{\prod_{l=1}^{k}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k,Y_{k-1}\right) E\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\} -\right.$$
$$\left. E^*\left\{R_k E^*\left(\frac{1}{\prod_{l=1}^{k}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k,Y_{k-1}\right) E\left(\frac{R_K Y_K}{\prod_{l=k+1}^{K}\pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\}\right]$$

Through algebra,

$$
\begin{aligned}
Rem_2(P,P^*) = &- E^*\left[R_1\left\{E^*\left(\frac{1}{\pi_1(Y_1,Y_0)}\middle| R_1=1,Y_1\right) - E\left(\frac{1}{\pi_1(Y_1,Y_0)}\middle| R_1=1,Y_1\right)\right\}\right.\\
&\left.\left\{E^*\left(\frac{R_KY_K}{\prod_{l=2}^K \pi_l(Y_{l-1},Y_l)}\middle| R_1=1,Y_1\right) - E\left(\frac{R_KY_K}{\prod_{l=2}^K \pi_l(Y_{l-1},Y_l)}\middle| R_1=1,Y_1\right)\right\}\right] +\\
&\sum_{k=2}^{K-1} E^*\left[R_k\left\{E^*\left(\frac{1}{\prod_{l=1}^k \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k,Y_{k-1}\right) - E\left(\frac{1}{\prod_{l=1}^k \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k,Y_{k-1}\right)\right\}\right.\\
&\left.\left\{E^*\left(\frac{R_KY_K}{\prod_{l=k+1}^K \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right) - E\left(\frac{R_KY_K}{\prod_{l=k+1}^K \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\}\right] -\\
&\sum_{k=2}^{K-1} E^*\left\{R_kE\left(\frac{1}{\prod_{l=1}^k \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)E\left(\frac{R_KY_K}{\prod_{l=k+1}^K \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\} +\\
&\sum_{k=2}^{K-1} E^*\left\{R_kE\left(\frac{1}{\prod_{l=1}^k \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)E^*\left(\frac{R_KY_K}{\prod_{l=k+1}^K \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\} -\\
&\sum_{k=2}^{K-1} E^*\left\{R_kE^*\left(\frac{1}{\prod_{l=1}^k \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k,Y_{k-1}\right)E^*\left(\frac{R_KY_K}{\prod_{l=k+1}^K \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\} +\\
&\sum_{k=2}^{K-1} E^*\left\{R_kE^*\left(\frac{1}{\prod_{l=1}^k \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k,Y_{k-1}\right)E\left(\frac{R_KY_K}{\prod_{l=k+1}^K \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\}
\end{aligned}
$$

We now use the fact that, for all $k=2,\ldots,K-1$ and $f_k(Y_k)$,

$$
E^*\left\{R_kE^*\left(\frac{1}{\prod_{l=1}^k \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k,Y_{k-1}\right)f_k(Y_k)\right\} = E^*\left\{R_kE^*\left(\frac{1}{\prod_{l=1}^k \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)f_k(Y_k)\right\}
$$

to conclude that

$$
\begin{aligned}
Rem_2(P,P^*) = &- \sum_{k=1]}^{K-1} E^*\left[R_k\left\{E^*\left(\frac{1}{\prod_{l=1}^k \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right) - E\left(\frac{1}{\prod_{l=1}^k \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\}\right.\\
&\left.\left\{E^*\left(\frac{R_KY_K}{\prod_{l=k+1}^K \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right) - E\left(\frac{R_KY_K}{\prod_{l=k+1}^K \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\}\right] +\\
&\sum_{k=2}^{K-1} E^*\left[R_k\left\{E^*\left(\frac{1}{\prod_{l=1}^k \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k,Y_{k-1}\right) - E\left(\frac{1}{\prod_{l=1}^k \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k,Y_{k-1}\right)\right\}\right.\\
&\left.\left\{E^*\left(\frac{R_KY_K}{\prod_{l=k+1}^K \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right) - E\left(\frac{R_KY_K}{\prod_{l=k+1}^K \pi_l(Y_{l-1},Y_l)}\middle| R_k=1,Y_k\right)\right\}\right]
\end{aligned}
$$

In this form, it is easy to see that $Rem_2(P,P^*)$ is second order.

Appendix C: Proof of Theorem 1

We can write that

$$
\begin{aligned}
\widehat{\mu}-\mu^* &= \mu(\widehat{P})-\mu(P^*)+\frac{1}{n}\sum_{i=1}^n D^\dagger(\widehat{P})(O_i)\\
&= -\int D^\dagger(\widehat{P})(o)dP^*(o) + Rem(\widehat{P},P^*) + \frac{1}{n}\sum_{i=1}^n D^\dagger(\widehat{P})(O_i)\\
&= \frac{1}{n}\sum_{i=1}^n D^\dagger(P^*)(O_i) + \int\left\{D^\dagger(\widehat{P})(o)-D^\dagger(P^*)(o)\right\}d(P_n-P^*)(o) + Rem(\widehat{P},P^*).
\end{aligned}
$$

Under conditions (a) and (b), we obtain that $\widehat{\mu}$ is an asymptotically linear estimator of μ^* with influence function $D^\dagger(P^*)$. Since $D^\dagger(P^*)$ is the canonical gradient of μ at P^* relative to $\mathscr{M}_0$, we conclude that $\widehat{\mu}$ is asymptotically efficient relative to $\mathscr{M}_0$.

B Scharfstein and McDermott, 2017

Global Sensitivity Analysis of Clinical Trials with Missing Patient Reported Outcomes

Daniel O. Scharfstein*[1] and Aidan McDermott†[1]

[1]Johns Hopkins Bloomberg School of Public Health, Baltimore, MD U.S.A.

April 21, 2017

Abstract

Randomized trials with patient reported outcomes are commonly plagued by missing data. The analysis of such trials relies on untestable assumptions about the missing data mechanism. To address this issue, it has been recommended that the sensitivity of the trial results to assumptions should be a mandatory reporting requirement. In this paper, we discuss a recently developed methodology (Scharfstein *et al.*, *Biometrics*, 2017) for conducting sensitivity analysis of randomized trials in which outcomes are scheduled to be measured at fixed points in time after randomization and some subjects prematurely withdraw from study participation. The methodology is explicated in the context of a placebo-controlled randomized trial designed to evaluate a treatment for bipolar disorder. We present a comprehensive data analysis and a simulation study to evaluate the performance of the method. A software package entitled SAMON (R and SAS versions) that implements our methods is available at `www.missingdatamatters.org`.

*dscharf@jhu.edu
†amcderm1@jhu.edu

1 Introduction

Missing outcome data are a widespread problem in clinical trials, including those with patient-reported outcomes. Since such outcomes require active engagement of patients and patients, while encouraged, are not required to remain or provide data while on-study, high rates of missing data can be expected.

To understand the magnitude of this issue, we reviewed all randomized trials [1] reporting five major patient-reported outcomes (SF-36, SF-12, Patient Health Questionnaire-9, Kansas City Cardiomyopathy Questionnaire, Minnesota Living with Heart Failure Questionnaire) published in five leading general medical journals (*New England Journal of Medicine*, *Journal of the American Medical Association*, *Lancet*, *British Medical Journal*, *PLoS One*) between January 1, 2008 and January 31, 2017. We identified 145 studies, which are summarized in Table 1. There is large variation in the percentages of missing data, with 78.6% of studies reporting percentages greater than 10%, 43.4% greater than 20% and 24.8% greater than 30%. Fielding *et al.* [46] conducted a similar review of clinical trials reporting quality of life outcomes in four of these journals during 2005/6 and found a comparable distribution of missing data percentages. Given the quality of these journals, it is likely that the percentages reported in Fielding *et al.* [46] and in Table 1 are an optimistic representation of percentages of missing data across the universe of clinical trials with patient-reported outcomes published in the medical literature.

Missing outcome data complicates the inferences that can be drawn about treatment effects. While unbiased estimates of treatment effects can be obtained from trials with no missing data, this is no longer true when data are missing on some patients. The essential problem is that inference about treatment effects relies on *unverifiable* assumptions about the nature of the mechanism that generates the missing data. While we may 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

[1] We focused on randomized trials in which patients in each treatment group were scheduled to be interviewed at a common set of post baseline assessment times. We excluded crossover trials, 10 trials in which patients were at high risk of death during the scheduled follow-up period, and 6 studies which did not report follow-up rates at the assessment times.

and whether differences in these distributions can be explained by the observed data.

It is widely recognized that the way to address the problem caused by missing outcome data is to posit varying assumptions about the missing data mechanism and evaluate how inference about treatment effects is affected by these assumptions. Such an approach is called "sensitivity analysis." A 2010 National Research Council (NRC) report entitled "The Prevention and Treatment of Missing Data in Clinical Trials" [90] and a follow-up manuscript published in the *New England Journal of Medicine* [91] recommends:

> 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.

Li *et al.* [89] echoed this recommendation (see Standard 8) in their PCORI sponsored report entitled "Minimal Standards in the Prevention and Handling of Missing Data in Observational and Experimental Patient Centered Outcomes Research".

The set of possible assumptions about the missing data mechanism is very large and cannot be fully explored. As discussed in Scharfstein *et al.* [136], there are three main approaches to sensitivity analysis: ad-hoc, local and global.

- Ad-hoc sensitivity analysis involves analyzing data using a few different analytic methods (e.g., last or baseline observation carried forward, complete or available case analysis, mixed models, imputation) and evaluating whether the resulting inferences are consistent. The problem with this approach is that consistency of inferences across the various methods does not imply that there are no reasonable assumptions under which the inference about the treatment effect is different.

- Local sensitivity analysis [94, 156, 153, 27] evaluates whether inferences are robust in a small neighborhood around a reasonable benchmark assumption, such as the classic missing at random assumption [92]. Unfortunately, this approach does not address whether the inferences are robust to plausible assumptions outside of the local neighborhood.

- Global sensitivity analysis [131, 138, 133, 134, 31, 136, 137] emphasized in Chapter 5 of the NRC report [90], evaluates 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. From this analysis, it can be determined how much deviation from the benchmark assumption is required in order for the inferences to change. If the deviation is 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. Some researchers have dubbed this approach "tipping point analysis" [169, 18].

In this paper, we consider randomized clinical trials in which patient-reported outcomes are scheduled to be measured at baseline (prior to randomization) and at a fixed number of post-baseline assessment times. We assume that some patients discontinue participation prior to the final assessment time and that all outcomes are observed while the patients are on-study. This assumption implies that there is no intermittent missing outcome data. We discuss a recently developed methodology [137] for conducting global sensitivity analysis of such trials. We explicate the methodology in the context of a randomized trial designed to evaluate the efficacy of quetiapine fumarate for the treatment of patients with bipolar disorder.

2 Quetiapine Bipolar Trial

The Quetiapine Bipolar trial was a multi-center, placebo-controlled, double-dummy study in which patients with bipolar disorder were randomized equally to one of three treatment arms: placebo, Quetiapine 300 mg/day or Quetiapine 600 mg/day [17]. Randomization was stratified by type of bipolar disorder: 1 or 2. A key secondary patient-reported endpoint was the short-form version of the Quality of Life Enjoyment Satisfaction Questionnaire (QLESSF, [41]), which was scheduled to be measured at baseline, week 4 and week 8 [2].

In this paper, we will 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 [3] We seek to compare the mean QLESSF outcomes at week 8 between these two treatment groups, in a world in which there are no missing outcomes. Unfortunately, this comparison is complicated because patients prematurely withdrew from the study. Figure 1 displays the treatment-specific trajectories of mean QLESSF scores, stratified by last available measurement. Notice that only 65 patients (56%) in placebo arm and 68 patients (58%) in the 600mg/day arm had a complete set of QLESSF scores. Further, the patients with complete data tend to have higher average QLESSF scores, suggesting that a complete-case analysis could be biased.

3 Global Sensitivity Analysis

Chapter 5 of the NRC report [90] lays out a general framework for global sensitivity analysis. In this framework, inference about treatment effects requires two types of assumptions: (i) untestable assumptions about the distribution of outcomes among those with missing data and (ii) testable assumptions that serve to increase the efficiency of estimation (see Figure 2). Type (i) assumptions are required to "identify" parameters of interest: identification means that one can mathematically express parameters of interest (e.g., treatment arm-specific means, treatment effects) in terms of the distribution of the observed data. In other words, if one were given the distribution of the observed data and given a type (i) assumption, then one could compute the value of the parameter of interest (see arrows in Figure 2). In the absence of identification, one cannot learn the value of the parameter

[2] Data were abstracted from the clinical study report available at http://psychrights.org/research/Digest/NLPs/Seroquel/UnsealedSeroquelStudies/. The number of patients that were abstracted does not exactly match the number of patients reported in Calabrese *et al.*, [17]

[3] These sample sizes exclude three randomized patients - one from placebo and two from 600 mg/day Quetiapine. From each group, one patient was removed because of undue influence on the analysis. In the 600 mg/day Quetiapine arm, one patient had incomplete questionaire data at baseline.

of interest based only on knowledge of the distribution of the observed data. Identification implies that the parameters of interest can, *in theory*, be estimated if the sample size is large enough.

There are an infinite number of ways of positing type (i) assumptions. It is impossible to consider all such assumptions. A reasonable way of positing these assumptions is to

(a) stratify individuals with missing outcomes according to the data that were able to be collected on them and the occasions at which the data were collected, and

(b) separately for each stratum, hypothesize a connection (or link) between the distribution of the missing outcomes with the distribution of these outcomes for patients who share the same recorded data and for whom the distribution is identified.

The connection that is posited in (b) is a type (i) assumption. The problem with this approach is that the stratum of people who share the same recorded data will typically be very small (e.g., the number of patients who share exactly the same baseline data will be very small). As a result, it is necessary to draw strength across strata by "smoothing." Smoothing is required because, *in practice*, we are not working with large enough sample sizes. Without smoothing, the data analysis will not be informative because the uncertainty (i.e., standard errors) of the parameters of interest will be too large to be of substantive use. Thus, it is necessary to impose type (ii) smoothing assumptions (represented by the inner circle in Figure 2). Type (ii) assumptions are testable (i.e., place restrictions on the distribution of the observed data) and should be scrutinized via model checking.

The global sensitivity framework proceeds by parameterizing (i.e., indexing) the connections (i.e., type (i) assumptions) in (b) above 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 the last part of this section, 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.

3.1 Notation and Data Structure

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

The observed data for an individual are $O = (Y_0, R_1, Y_1^{obs}, R_2, Y_2^{obs})$, which is drawn from some distribution P^* contained within a set of distributions $\mathcal{M}$ (to be discussed later). Throughout, 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: $f(Y_0)$, $P[R_1 = 1|Y_0]$, $f(Y_1|R_1 = 1, Y_0)$, $P[R_2 = 1|R_1 = 1, Y_1, Y_0]$ and $f(Y_2|R_2 = 1, Y_1, Y_0)$.

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.

3.2 Benchmark Assumption (Missing at Random)

Missing at random [92] is a widely used assumption for analyzing longitudinal studies with missing outcome data. To understand this assumption, we define the following strata:

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

- $B_2(y_1, y_0)$: patients on-study at visit 2 with $Y_1 = y_1$ and $Y_0 = y_0$.

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

- For all y_0, the distribution of Y_1 and Y_2 for patients in strata $A_0(y_0)$ is the same as the distribution of Y_1 and Y_2 for patients in strata $B_1(y_0)$
- For all y_0, y_1, the distribution of Y_2 for patients in strata $A_1(y_1, y_0)$ is the same as the distribution of Y_2 for patients in strata $B_2(y_1, y_0)$

Mathematically, we can express these assumptions as follows:

$$f^*(Y_1, Y_2|\underbrace{R_1 = 0, Y_0 = y_0}_{A_0(y_0)}) = f^*(Y_1, Y_2|\underbrace{R_1 = 1, Y_0 = y_0}_{B_1(y_0)}) \text{ for all } y_0 \tag{1}$$

and

$$f^*(Y_2|\underbrace{R_2 = 0, R_1 = 1, Y_1 = y_1, Y_0 = y_0}_{A_1(y_1,y_0)}) = f^*(Y_2|\underbrace{R_2 = 1, Y_1 = y_1, Y_0 = y_0}_{B_2(y_1,y_0)}) \text{ for all } y_1, y_0 \tag{2}$$

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

$$P^*[R_1 = 0|Y_2 = y_2, Y_1 = y_1, Y_0 = y_0] = P^*[R_1|Y_0 = y_0] \tag{3}$$

and

$$P^*[R_2 = 0|R_1 = 1, Y_2 = y_2, Y_1 = y_1, Y_0 = y_0] = P^*[R_2 = 0|R_1 = 1, Y_1 = y_1, Y_0 = y_0] \tag{4}$$

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

- 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.

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

$$\mu^* = \mu(P^*) = \int_{y_0} \int_{y_1} \int_{y_2} y_2 dF_2^*(y_2|y_1, y_0) dF_1^*(y_1|y_0) dF_0^*(y_0) \tag{5}$$

where $F_2^*(y_2|y_1, y_0) = P^*[Y_2 \leq y_2 | R_2 = 1, Y_1 = y_1, Y_0 = y_0]$, $F_1^*(y_1|y_0) = P^*[Y_1 \leq y_1 | R_1 = 1, Y_0 = y_0]$ and $F_0^*(y_0) = P^*[Y_0 \leq y_0]$.

Before proceeding to the issue of estimation, we will build a class of assumptions around the missing at random assumption using a modeling device called exponential tilting [7].

3.3 Missing Not at Random and Exponential Tilting

To build a class of missing not at random assumptions, consider Equation (1) of the missing at random assumption. This equation is equivalent to the following two assumptions:

$$f^*(Y_2|\underbrace{R_1 = 0, Y_1 = y_1, Y_0 = y_0}_{A_0(y_1, y_0)}) = f^*(Y_2|\underbrace{R_1 = 1, Y_1 = y_1, Y_0 = y_0}_{B_1(y_1, y_0)}) \text{ for all } y_0, y_1 \tag{6}$$

and

$$f^*(Y_1|\underbrace{R_1 = 0, , Y_0 = y_0}_{A_0(y_0)}) = f^*(Y_1|\underbrace{R_1 = 1, Y_0 = y_0}_{B_1(y_0)}) \text{ for all } y_0 \tag{7}$$

where

- $A_0(y_1, y_0) \subset A_0(y_0)$: patients last seen at visit 0 with $Y_0 = y_0$ and $Y_1 = y_1$.
- $B_1(y_1, y_0) \subset B_1(y_0)$: patients on-study at visit 1 with $Y_0 = y_0$ and $Y_1 = y_1$.

Equation (6) posits the following type (i) "linking" assumption:

- For all y_0 and y_1, the distribution of Y_2 for patients in strata $A_0(y_1, y_0)$ is the same as the distribution of Y_2 for patients in strata $B_1(y_1, y_0)$

It has been referred to as the "non-future" dependence assumption [32] because it implies that R_1 (i.e., the decision to drop-out 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). We will retain this assumption.

Next, we impose the following exponential tilting "linking" assumptions:

$$f^*(Y_1|\underbrace{R_1 = 0, Y_0 = y_0}_{A_0(y_0)}) \propto f^*(Y_1|\underbrace{R_1 = 1, Y_0 = y_0}_{B_1(y_0)}) \exp\{\alpha r(Y_1)\} \text{ for all } y_0 \tag{8}$$

$$f^*(Y_2|\underbrace{R_2 = 0, R_1 = 1, Y_1 = y_1, Y_0 = y_0}_{A_1(y_1,y_0)}) \propto f^*(Y_2|\underbrace{R_2 = 1, Y_1 = y_1, Y_0 = y_0}_{B_2(y_1,y_0)}) \exp\{\alpha r(Y_2)\} \text{ for all } y_0, y_1 \tag{9}$$

where $r(\cdot)$ is a specified function which we will assume to be an increasing function of its argument and α is a sensitivity analysis parameter. The missing not at random class of assumptions that we propose involves Equations (6), (8) and (9), where $r(\cdot)$ is considered fixed and α is a sensitivity analysis parameter that serves as the class index. Importantly, notice how (8) reduces to (7) and (9) reduces to (2) when $\alpha = 0$. Thus, when $\alpha = 0$, the MAR assumption is obtained. When $\alpha > 0$ (< 0), notice that (8) and (9) imply

- For all y_0, the distribution of Y_1 for patients in strata $A_0(y_0)$ is weighted more heavily (i.e., tilted) to higher (lower) values than the distribution of Y_1 for patients in strata $B_1(y_0)$

- For all y_0, y_1, the distribution of Y_2 for patients in strata $A_1(y_1, y_0)$ is weighted more heavily weighted (i.e., tilted) to higher (lower) values than the distribution of Y_2 for patients in strata $B_2(y_1, y_0)$

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

Using Bayes' rule, we can re-write expressions (6), (8) and (9) succinctly as:

$$\text{logit } P^*[R_1 = 0|Y_2 = y_2, Y_1 = y_1, Y_0 = y_0] = l_1^*(y_0) + \alpha r(y_1) \tag{10}$$

and

$$\text{logit } P^*[R_2 = 0|R_1 = 1, Y_2 = y_2, Y_1 = y_1, Y_0 = y_0] = l_2^*(y_1, y_0) + \alpha r(y_2) \tag{11}$$

where

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

and

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

Written in this way, the drop-out process is stochastic with the following properties:

- 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 1 and 0 *and*, in a specified way, the value of the outcome at visit 2.

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

$$\int_{y_0}\int_{y_1}\int_{y_2} y_2 \left\{ dF_2^*(y_2|y_1,y_0)\{1-H_2^*(y_1,y_0)\} + \frac{dF_2^*(y_2|y_1,y_0)\exp\{\alpha r(y_2)\}}{\int_{y_2'} dF_2^*(y_2'|y_1,y_0)\exp\{\alpha r(y_2')\}} H_2^*(y_1,y_0) \right\} \times$$
$$\left\{ dF_1^*(y_1|y_0)\{1-H_1^*(y_0)\} + \frac{dF_1^*(y_1|y_0)\exp\{\alpha r(y_1)\}}{\int_{y_1'} dF_1^*(y_1'|y_0)\exp\{\alpha r(y_1')\}} H_1^*(y_0) \right\} dF_0^*(y_0) \qquad (12)$$

where $H_2^*(y_1, y_0) = P^*[R_2 = 0|R_1 = 1, Y_1 = y_1, Y_0 = y_0]$ and $H_1^*(y_0) = P^*[R_1 = 0|Y_0 = y_0]$

4 Inference

For given α, formula (12) shows that μ^* depends on $F_2^*(y_2|y_1, y_0)$, $F_1^*(y_1|y_0)$, $H_2^*(y_1, y_0)$ and $H_1^*(y_0)$. Thus, it is natural to consider estimating μ^* by "plugging in" estimators of $F_2^*(y_2|y_1, y_0)$, $F_1^*(y_1|y_0)$, $F_0^*(y_0)$, $H_2^*(y_1, y_0)$ and $H_1^*(y_0)$ into (12). 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. For example, a non-parametric estimate of $F_2^*(y_2|y_1, y_0)$ would take the form:

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

This estimator will perform very poorly (i.e., have high levels of uncertainly in moderate sample sizes) because the number of subjects who complete the study (i.e., $R_2 = 1$) and are observed to have outcomes at visits 1 and 0 exactly equal to y_1 and y_0 will be very small and can only be expected to grow very slowly as the sample size increases. As a result, a a plug-in estimator of μ^* that uses such non-parametric estimators will perform poorly. We address this problem in three ways.

4.1 Testable Assumptions

First we make the estimation task slightly easier by assuming that

$$F_2^*(y_2|y_1, y_0) = F_2^*(y_2|y_1) \tag{13}$$

and

$$H_2^*(y_1, y_0) = H_2^*(y_1) \tag{14}$$

That is, (13) states that, among subjects who complete the study, information about Y_0 does not provide any information about the distribution of Y_2 above and beyond information about Y_1 and (14) states that, among subjects on-study at visit 1, information about Y_0 does not influence of the risk of dropping out before visit 2 above and beyond information about Y_1. These assumptions are, with large enough samples, testable from the observed data. As such, we distinguish them from type (i) assumptions and refer to them as type (ii) assumptions.

4.2 Kernel Smoothing with Cross-Validation

Second we estimate $F_2^*(y_2|y_1)$, $F_1^*(y_1|y_0)$, $H_2^*(y_1)$ and $H_1^*(y_0)$ using kernel smoothing techniques. To motivate this idea, consider the following non-parametric estimate of $F_2^*(y_2|y_1)$

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

This estimator will still perform poorly, although better than $\widehat{F}_2(y_2|y_1, y_0)$, since there will be at least as many completers with Y_1 values equal to y_1 than completers with Y_1 *and* Y_0 values equal to y_1 and y_0, respectively. To improve its performance, we replace $I(Y_{1,i} = y_1)$ by $\phi\left(\frac{Y_{1,i} - y_1}{\lambda_{F_2}}\right)$, where $\phi(\cdot)$ is the density function for a standard normal random variable and λ_{F_2} is a tuning parameter. For fixed λ_{F_2}, let

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

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 λ_{F_2}, the larger the influence of values of Y_1 further from y_1 on the estimator. As $\lambda_{F_2} \to \infty$, the contribution of each completer to the estimator becomes equal, yielding bias but low variance. As $\lambda_{F_2} \to 0$, only completers with Y_1 values equal to y_1 contribute, yielding low bias but high variance.

To address the bias-variance trade-off, cross validation [62] is typically used to select λ_{F_2}. In cross validation, the dataset is randomly divided into J (typically, 10) approximately equal parts. Each part is called a validation set. Let V_j be the indices of the subjects in the jth validation set. Let n_j be the associated number of subjects. Let $\widehat{F}_2^{(j)}(y_2|y_1;\lambda_{F_2})$ be the estimator of $F_2^*(y_2|y_1)$ based on the dataset that excludes the jth validation set (referred to as the jth training set). If λ_{F_2} is a good choice then one would expect

$$CV_{F_2^*(\cdot|\cdot)}(\lambda_{F_2}) = \frac{1}{J}\sum_{j=1}^{J}\left\{\frac{1}{n_j}\sum_{i\in V_j} R_{2,i} \underbrace{\int\left\{I(Y_{2,i}\leq y_2) - \widehat{F}_2^{(j)}(y_2|Y_{1,i};\lambda_{F_2})\right\}^2 d\widehat{F}_2^{\circ}(y_2)}_{\text{Distance for } i\in V_j}\right\} \tag{15}$$

will be small, where $\widehat{F}_2^{\circ}(y_2)$ is the empirical distribution of Y_2 among subjects on-study at visit 2. In (15), the quantity in the vertical braces is a measure of how well the estimator of $F_2(y_2|y_1)$ based on the jth training set "performs" on the jth validation set. For each individual i in the jth validation set with an observed outcome at visit 2, we measure, by the quantity above the horizontal brace in (15), the distance (or loss) between the collection of indicator variables $\{I(Y_{2,i} \leq y_2) : d\widehat{F}_2^{\circ}(y_2) > 0\}$ and the corresponding collection of predicted values $\{\widehat{F}_2^{(j)}(y_2|Y_{1,i};\lambda_{F_2}) : d\widehat{F}_2^{\circ}(y_2) > 0\}$. The distance for each of these individuals are then summed and divided by the number of subjects in the jth validation set. Finally, an average across the J validation/training sets is computed. We can then estimate $F_2^*(y_2|y_1)$ by $\widehat{F}_2(y_2|y_1;\widehat{\lambda}_{F_2})$, where $\widehat{\lambda}_{F_2} = \text{argmin } CV_{F_2^*(\cdot|\cdot)}(\lambda_{F_2})$.

Using this idea, we can estimate $F_1^*(y_1|y_0)$ by

$$\widehat{F}_1(y_1|y_0;\widehat{\lambda}_{F_1}) = \frac{\sum_{i=1}^n R_{1,i} I(Y_{1,i} \le y_1)\phi\left(\frac{Y_{0,i}-y_0}{\widehat{\lambda}_{F_1}}\right)}{\sum_{i=1}^n R_{1,i}\phi\left(\frac{Y_{0,i}-y_0}{\widehat{\lambda}_{F_1}}\right)}$$

where $\widehat{\lambda}_{F_1}$ is the minimizer of

$$CV_{F_1^*(\cdot|\cdot)}(\lambda_{F_1}) = \frac{1}{J}\sum_{j=1}^J \left\{ \frac{1}{n_j}\sum_{i\in V_j} R_{1,i}\int \left\{ I(Y_{1,i}\le y_1) - \widehat{F}_1^{(j)}(y_1|Y_{0,i};\lambda_{F_1})\right\}^2 d\widehat{F}_1^{\circ}(y_1)\right\}$$

and $\widehat{F}_1^{\circ}(y_1)$ is the empirical distribution of Y_1 among subjects on-study at visit 1. Further, we estimate $H_k^*(y_{k-1})$ $(k = 1, 2)$ by

$$\widehat{H}_k(y_{k-1};\widehat{\lambda}_{H_k}) = \frac{\sum_{i=1}^n R_{k-1,i}(1-R_{k,i})\phi\left(\frac{Y_{k-1,i}-y_{k-1}}{\widehat{\lambda}_{H_k}}\right)}{\sum_{i=1}^n R_{k-1,i}\phi\left(\frac{Y_{k-1,i}-y_{k-1}}{\widehat{\lambda}_{H_k}}\right)}$$

where $\widehat{\lambda}_{H_k}$ is the minimizer of

$$CV_{H_k^*(\cdot)}(\lambda_{H_k}) = \frac{1}{J}\sum_{j=1}^J \left\{ \frac{1}{n_j}\sum_{i\in V_j} R_{k-1,i}\{1 - R_{k,i} - \widehat{H}_k^{(j)}(Y_{k-1,i};\widehat{\lambda}_{H_k})\}\widehat{H}_k^{\circ}\right\}$$

and $\widehat{H}_k^{\circ}$ is the proportion of individual with drop out between visits $k-1$ and k among those on-study at visit $k-1$.

4.3 Correction Procedure

The cross-validation procedure for selecting tuning parameters achieves optimal finite-sample bias-variance trade-off for the quantities requiring smoothing, i.e., the conditional distribution functions $F_k^*(y_k|y_{k-1})$ and probability mass functions $H_k^*(y_{k-1})$. This optimal trade-off is usually not optimal for estimating μ^*. In fact, the plug-in estimator of μ^* could possibly suffer from excessive and asymptotically non-negligible bias due to inadequate tuning. This may prevent the plug-in estimator from enjoying regular asymptotic behavior, upon which statistical inference is generally based. In particular, 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. Thus, our third move is to "correct" the plug-in estimator. Specifically, the goal is to construct an estimator that is "asymptotically linear" (i.e., can be expressed as the average of i.i.d. random variables plus a remainder term that is asymptotically negligible).

We now motivate the correction procedure. Let $\mathcal{M}$ be the class of distributions for the observed data O that satisfy constraints (13) and (14). It can be shown that, for $P \in \mathcal{M}$,

$$\mu(P;\alpha) - \mu(P^*;\alpha) = -E^*[\psi_P(O;\alpha) - \psi_{P^*}(O;\alpha)] + \mathrm{Rem}(P, P^*;\alpha), \tag{16}$$

where $\psi_P(O;\alpha)$ is a "derivative" of $\mu(\cdot;\alpha)$ at P and $\mathrm{Rem}(P, P^*;\alpha)$ is a "second-order" remainder term which converges to zero as P tends to P^*. This 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 (16) plus some simple algebraic manipulation teaches us that

$$\begin{aligned}
\underbrace{\mu(\widehat{P};\alpha)}_{\text{Plug-in}} - \mu(P^*;\alpha) &= \frac{1}{n}\sum_{i=1}^{n}\psi_{P^*}(O_i;\alpha) - \frac{1}{n}\sum_{i=1}^{n}\psi_{\widehat{P}}(O_i;\alpha) && (17)\\
&+\frac{1}{n}\sum_{i=1}^{n}\{\psi_{\widehat{P}}(O_i;\alpha)-\psi_{P^*}(O_i;\alpha)-E^*[\psi_{\widehat{P}}(O;\alpha)-\psi_{P^*}(O;\alpha)]\} && (18)\\
&+\mathrm{Rem}(\widehat{P},P^*;\alpha) && (19)
\end{aligned}$$

where $\widehat{P}$ is the estimated distribution of P^* discussed in the previous section. Under smoothness and boundedness conditions, term (18) will be $o_{P^*}(n^{-1/2})$ (i.e., will converge in probabity to zero even when it is multipled by $\sqrt{n}$). Provided $\widehat{P}$ converges to P^* at a reasonably fast rate, term (19) will also be $o_{P^*}(n^{-1/2})$. The second term in (17) 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. However, by adding the second term in (17) to the plug-in estimator, we can construct a "corrected" estimator that does have this representation. Formally, the corrected estimator is

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

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

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

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

The efficient influence function in model $\mathcal{M}$ is presented in Appendix A.

4.4 Confidence interval construction

For given α, there are many ways to construct confidence intervals for μ^*. Above, we discussed the Wald-based technique. In Section 6, we present the results of a simulation study in which this technqiue results in poor coverage in moderately sized samples. The poor coverage can be explained in part due to the fact that $\tilde{\sigma}(\alpha)^2$ can be severely downward biased in finite samples [37].

Resampling-based procedures may be used to improve performance. A first idea is to consider the jackknife estimator for σ^2_α:

$$\tilde{\sigma}^2_{JK,\alpha} = (n-1)\sum_{i=1}^{n}\{\tilde{\mu}_\alpha^{(-i)} - \tilde{\mu}_\alpha^{(\cdot)}\}^2$$

where $\tilde{\mu}_\alpha^{(-i)}$ is the estimator of μ^* with the ith individual deleted from the dataset and $\tilde{\mu}_\alpha^{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n}\tilde{\mu}_\alpha^{(-i)}$. This estimator is known to be conservative [38], but is the "method of choice if one does not want to do bootstrap computations" [37]. Using the jackknife estimator of the variance, one can construct a Wald confidence interval with $\tilde{\sigma}_\alpha$ replaced by $\tilde{\sigma}_{JK,\alpha}$. Our simulation study in Section 6 demonstrates that these latter intervals perform better, but still have coverage lower than desired.

Another idea is to use studentized-t bootstrap. Here, confidence intervals are formed by choosing cutpoints based on the distribution of

$$\left\{ \frac{\tilde{\mu}_\alpha^{(b)} - \tilde{\mu}_\alpha}{\tilde{se}\left(\tilde{\mu}_\alpha^{(b)}\right)} : b = 1, 2, \ldots, B \right\} \tag{20}$$

where $\tilde{\mu}_\alpha^{(b)}$ is the estimator of μ^* based on the bth bootstrap dataset and $\tilde{se}\left(\tilde{\mu}_\alpha^{(b)}\right)$ is an estimator of the standard error of $\tilde{\mu}_\alpha^{(b)}$ (e.g., $\tilde{\sigma}_\alpha/\sqrt{n}$ or $\tilde{\sigma}_{JK,\alpha}/\sqrt{n}$) . An equal-tailed confidence interval takes the form:

$$\left\{ \tilde{\mu}_\alpha - t_{1-\gamma/2}\tilde{se}\left(\tilde{\mu}_\alpha^{(b)}\right), \tilde{\mu}_\alpha - t_{\gamma/2}\tilde{se}\left(\tilde{\mu}_\alpha^{(b)}\right) \right\},$$

where t_q is the qth quantile of (20). A symmetric confidence interval takes the form:

$$\left\{\tilde{\mu}_\alpha - t^*_{1-\gamma}\tilde{se}\left(\tilde{\mu}^{(b)}_\alpha\right), \tilde{\mu}_\alpha + t^*_{1-\gamma}\tilde{se}\left(\tilde{\mu}^{(b)}_\alpha\right)\right\},$$

where $t^*_{1-\gamma}$ is selected so that $(1-\gamma)$ of the distribution of (20) is between $-t^*_{1-\gamma}$ and $t^*_{1-\gamma}$.

In terms of bootstrapping, there are two main choices: non-parametric and parametric. The advantage of non-parametric bootstrap is that it does not require a model for the distribution of the observed data. Since our analysis depends on correct specification and on estimation of such a model, it makes sense to use this model to bootstrap observed datasets. In our data analysis and simulation study, we use the estimated distribution of the observed data to generate bootstrapped observed datasets.

Our simulation study in Section 6 shows that the symmetric studentized-t bootstrap with jackknife standard errors performs best. We used this procedure in our data analysis.

5 Analysis of Quetiapine Trial

The first step of the analysis is to estimate the smoothing parameters and assess the goodness of fit of our models for H^*_j (drop-out) and F^*_j (outcome). We assumed a common smoothing parameter for the H^*_j $(j = 1, 2)$ models and a common smoothing parameter for F^*_j $(j = 1, 2)$ models; F^*_0 was estimated by its empirical distribution. The estimated smoothing parameters for the drop-out (outcome) model are 11.54 (6.34) and 9.82 (8.05) for the placebo and 600 mg arms, respectively. 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%, respectively. Estimates derived from the estimated model for the distribution of the observed data are 7.99% and 38.19%, respectively. 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, respectively. 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 missing at random, the estimated values of μ^* are 46.45 (95% CI: 42.35,50.54) and 62.87 (95% CI: 58.60,67.14) for the placebo and 600 mg arms, respectively. The estimated difference between 600 mg and placebo is 16.42 (95% 10.34, 22.51), which represents both a statistically and clinically significant improvement in quality of life in favor of Quetiapine. [4]

In our sensitivity analysis, we set $r(y) = y$ and ranged the sensitivity analysis parameter from -10 and 10 in each treatment arm.[5] Figure 3 presents treatment-specific estimates (along with 95% pointwise confidence intervals) of μ^* as a function of α. To help interpret the sensitivity analysis parameter, Figure 4 displays treatment-specific differences between the estimated mean QLESSF at Visit 2 among non-completers and the estimated mean among completers, as a function of α. For example, when $\alpha = -10$ non-completers are estimated to have more than 20 points lower quality of life than completers; this holds for both treatment arms. In contrast, when $\alpha = 10$ non-completers are estimated to have 6 and 11 points higher quality of life than completers in the placebo and Quetiapine arms, respectively. The plausibility of α can be judged with respect the plausibility of these differences. In this setting, it may be considered unreasonable that completers are worse off in terms of quality of life than non-completers, in which case α should be restricted to be less than 6 in the placebo arm and less than 3 in the Quentiapine arm.

Figure 5 displays a contour plot of the estimated differences between mean QLESSF at Visit 2 for Quentiapine vs. placebo for various treatment-specific combinations of the sensitivity analysis parameters. The point (0,0) corresponds to the MAR assumption in both treatment arms. The figure shows that the differences are statistically significant (represented by dots) in favor of Quetiapine at almost all combinations of the sensitivity analysis parameters. Only when the sensitivity analysis

[4] All confidence intervals are symmetric studentized-t bootstrap with jackknife standard errors.

[5] According to Dr. Dennis Rivicki and Dr. Jean Endicott, 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.

are highly differential (e.g., α(placebo) = 8 and α(Quetaipine) = -8) are the differences no longer statistically significant. This figure shows that conclusions under MAR are highly robust.

6 Simulation Study

To evaluate the statistical properties of our proposed procedure, we conducted a realistic simulation study that mimics the data structure in the Quetiapine study. We generated 2500 placebo and Quetiapine datasets using the estimated distributions of the observed data from the Quentiapine study as the true data generating mechanisms. For given treatment-specific α, these true data generating mechanisms can be mapped to a true value of μ^*. For each dataset, the sample size was to set to 116 and 118 in the placebo and Quetiapine arms, respectively.

Table 2 reports bias and mean-squared error for the plug-in and corrected estimators, as a function of α. The bias tends to be low for both estimators and the mean-squared error is lower for the corrected estimators, except at extreme values of α.

Table 3 reports the coverage properties of six difference methods for constructing confidence intervals: (1) Wald with influence function standard errors (Wald-IF), (2) Wald with jackknife standard errors (Wald-JK), (3) equal-tailed studentized parametric bootstrap with influence function standard errors (Bootstrap-IF-ET), (4) equal-tailed studentized parametric bootstrap with jackknife standard errors (Bootstrap-JK-ET), (5) symmetric studentized parametric bootstrap with influence function standard errors (Bootstrap-IF-S) and (6) symmetric studentized parametric bootstrap with jackknife standard errors (Bootstrap-JK-S); 2000 parametric bootstraps were used. The results demonstrate that using jackknife standard errors is superior to influence function standard errors. In this simulation, the best performing procedures are Wald with jackknife standard errors and symmetric studentized parametric bootstrap with jackknife standard errors, with the latter experiencing, for some values of α, coverages 1-2% higher than nominal levels. In other simulations (reported elsewhere), we have found that Wald with jacknife standard errors can have lower than nominal levels of coverage. Thus, we recommend using symmetric studentized parametric bootstrap with jackknife standard errors.

7 Discussion

Our review of leading medical journals demonstrated that missing data are a common occurrence in randomized trials with patient-reported outcomes. As per the 2010 NRC report [90], it is essential to evaluate the robustness of trial results to untestable assumptions about the underlying missing data mechanism. In this paper, we have presented a methodology [137] for conducting global (as opposed to ad-hoc or local) sensitivity analysis of trials in which (1) outcomes are scheduled to be measured at fixed points after randomization and (2) missing data are monotone. While we developed our method in the context of a motivating example with two post-baseline measurements, it naturally generalizes to studies with more measurements [137]. Our sensitivity analysis is anchored around the commonly used missing at random assumption. We have developed a software package called SAMON to implement our procedure. R and SAS versions of the software are available at `www.missingdatamatters.org`.

We have found that our procedure can be sensitive to outliers. In fact, we discarded two patients (one from each treatment arm) from the Quetiapine Study because of their undue influence. In the placebo arm, the patient was a completer and had baseline, visit 1 and visit 2 raw scores of 17, 26 and 48, respectively. At $\alpha = 10$, the scaled absolute DFBETA for this observation was 2.75 with the next largest absolute DFBETA being 1.13. In the Quetiapine arm, the patient was a completer and had baseline, visit 1 and visit 2 raw scores of 31, 29 and 18, respectively. At $\alpha = -10$, the scaled absolute DFBETA for this observation was 3.20 with the next largest absolute DFBETA being 0.52. One way to address the issue of outliers would be the robustify the influence function using ideas from the robust statistics literature [69].

Our procedure does not currently handle intermittent missing data. In many randomized trials, intermittent missing data is usually a second order concern. We propose imputing intermittent observations, under a reasonable assumption (see, for example, [130]) to create a monotone data structure and then apply the methods outlined in this paper with proper accounting for uncertainty in the imputation process.

We believe that the methods and software that we have developed should be applied to all trials with missing outcome data, including but limited to those that are patient-reported. Trial results that are sensitive to untestable assumptions about the missing data mechanism should be viewed with skepticism, while greater credence should be given those that exhibit robustness. Our methods are not a substitute for study designs and procedures that minimize missing data.

Acknowledgments

This research was sponsored by contracts from the U.S. Food and Drug Administration and the Patient Centered Outcomes Research Institute as well as NIH grant CA183854.

References

[1] L Adamsen, M Quist, C Andersen, T Moller, J Herrstedt, D Kronborg, MT Baadsgaard, K Vistisen, J Midtgaard, B Christiansen, M Stage, MT Kronborg, and M Rorth. Effect of a multimodal high intensity exercise intervention in cancer patients undergoing chemotherapy: randomised controlled trial. *BMJ*, 339:b3410, 2009.

[2] UC Adler, S Kruger, M Teut, R Ludtke, L Schutzler, F Martins, SN Willich, K Linde, and CM Witt. Homeopathy for depression: A randomized, partially double-blind, placebo-controlled, four-armed study (DEP-HOM). *PloS One*, 8(9):e74537, 2013.

[3] AA Ahimastos, PJ Walker, C Askew, A Leicht, E Pappas, P Blombery, CM Reid, J Golledge, and BA Kingwell. Effect of ramipril on walking times and quality of life among patients with peripheral artery disease and intermittent claudication: a randomized controlled trial. *JAMA*, 309(5):453–60, 2013.

[4] VA Andreeva, C Latarche, S Hercberg, S Briancon, P Galan, and E Kesse-Guyot. B vitamin and/or n-3 fatty acid supplementation and health-related quality of life: ancillary findings from the su.fol.om3 randomized trial. *PLoS One*, 9(1):e84844, 2014.

[5] Joaquin A Anguera, Joshua T Jordan, Diego Castaneda, Adam Gazzaley, and Patricia A Areán. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. *BMJ innovations*, 2(1):14–21, 2016.

[6] J Arnold, S Goodacre, P Bath, and J Price. Information sheets for patients with acute chest pain: randomised controlled trial. *BMJ*, 338:b541, 2009.

[7] O Barndorff-Nielsen and David R Cox. Edgeworth and saddle-point approximations with statistical applications. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 279–312, 1979.

[8] KJ Barnhoorn, H van de Meent, RT van Dongen, FP Klomp, H Groenewoud, H Samwel, MW Nijhuis-van der Sanden, JP Frolke, and JB Staal. Pain exposure physical therapy (PEPT) compared to conventional treatment in complex regional pain syndrome type 1: a randomised controlled trial. *BMJ Open*, 5(12):e008283, 2015.

[9] DB Bekelman, ME Plomondon, EP Carey, MD Sullivan, KM Nelson, B Hattler, CF McBryde, KG Lehmann, K Gianola, PA Heidenreich, and JS Rumsfeld. Primary results of the patient-centered disease management (PCDM) for heart failure study: A randomized clinical trial. *JAMA Intern Med*, 175(5):725–32, 2015.

[10] NM Benda, JP Seeger, GG Stevens, BT Hijmans-Kersten, AP van Dijk, L Bellersen, EJ Lamfers, MT Hopman, and DH Thijssen. Effects of high-intensity interval training versus continuous training on physical fitness, cardiovascular function and quality of life in heart failure patients. *PLoS One*, 10(10):e0141256, 2015.

[11] Anneleen Berende, Hadewych JM ter Hofstede, Fidel J Vos, Henriët van Middendorp, Michiel L Vogelaar, Mirjam Tromp, Frank H van den Hoogen, A Rogier T Donders, Andrea WM Evers, and Bart Jan Kullberg. Randomized trial of longer-term therapy for symptoms attributed to lyme disease. *New England Journal of Medicine*, 374(13):1209–1220, 2016.

[12] N Bergmann, S Ballegaard, P Bech, A Hjalmarson, J Krogh, F Gyntelberg, and J Faber. The effect of daily self-measurement of pressure pain sensitivity followed by acupressure on depression and quality of life versus treatment as usual in ischemic heart disease: a randomized clinical trial. *PLoS One*, 9(5):e97553, 2014.

[13] JL Berk, OB Suhr, L Obici, Y Sekijima, SR Zeldenrust, T Yamashita, MA Heneghan, PD Gorevic, WJ Litchy, JF Wiesman, E Nordh, M Corato, A Lozza, A Cortese, J Robinson-Papp, T Colton, DV Rybin, AB Bisbee, Y Ando, S Ikeda, DC Seldin, G Merlini, M Skinner, JW Kelly, and PJ Dyck. Repurposing diflunisal for familial amyloid polyneuropathy: a randomized clinical trial. *JAMA*, 310(24):2658–67, 2013.

[14] H Bruhn, CM Bond, AM Elliott, PC Hannaford, AJ Lee, P McNamee, BH Smith, MC Watson, R Holland, and D Wright. Pharmacist-led management of chronic pain in primary care: results from a randomised controlled exploratory trial. *BMJ Open*, 3(4), 2013.

[15] C Burton, D Weller, W Marsden, A Worth, and M Sharpe. A primary care symptoms clinic for patients with medically unexplained symptoms: pilot randomised trial. *BMJ Open*, 2:e000513, 2012.

[16] Jason W Busse, Mohit Bhandari, Thomas A Einhorn, Emil Schemitsch, James D Heckman, Paul Tornetta, Kwok-Sui Leung, Diane Heels-Ansdell, Sun Makosso-Kallyth, Gregory J Della Rocca, et al. Re-evaluation of low intensity pulsed ultrasound in treatment of tibial fractures (trust): randomized clinical trial. *bmj*, 355:i5351, 2016.

[17] Joseph R Calabrese, Paul E Keck Jr, Wayne Macfadden, Margaret Minkwitz, Terence A Ketter, Richard H Weisler, Andrew J Cutler, Robin McCoy, Ellis Wilson, Jamie Mullen, et al. A randomized, double-blind, placebo-controlled trial of quetiapine in the treatment of bipolar i or ii depression. *American Journal of Psychiatry*, 2005.

[18] Gregory Campbell, Gene Pennello, and Lilly Yue. Missing data in the regulation of medical devices. *Journal of biopharmaceutical statistics*, 21(2):180–195, 2011.

[19] M Cartwright, SP Hirani, L Rixon, M Beynon, H Doll, P Bower, M Bardsley, A Steventon, M Knapp, C Henderson, A Rogers, C Sanders, R Fitzpatrick, J Barlow, and SP Newman. Effect of telehealth on quality of life and psychological outcomes over 12 months (Whole Systems Demonstrator telehealth questionnaire study): nested study of patient reported outcomes in a pragmatic, cluster randomised controlled trial. *BMJ*, 346:f653, 2013.

[20] T Chalder, KA Goldsmith, PD White, M Sharpe, and AR Pickles. Rehabilitative therapies for chronic fatigue syndrome: a secondary mediation analysis of the PACE trial. *Lancet Psychiatry*, 2(2):141–52, 2015.

[21] Dixon Chibanda, Helen A Weiss, Ruth Verhey, Victoria Simms, Ronald Munjoma, Simbarashe Rusakaniko, Alfred Chingono, Epiphania Munetsi, Tarisai Bere, Ethel Manda, et al. Effect of a primary care–based psychological intervention on symptoms of common mental disorders in zimbabwe: A randomized clinical trial. *JAMA*, 316(24):2618–2626, 2016.

[22] H Christensen, PJ Batterham, JA Gosling, LM Ritterband, KM Griffiths, FP Thorndike, N Glozier, B O'Dea, IB Hickie, and AJ Mackinnon. Effectiveness of an online insomnia program (SHUTi) for prevention of depressive episodes (the GoodNight Study): a randomised controlled trial. *Lancet Psychiatry*, 2016.

[23] DJ Cohen, B Van Hout, PW Serruys, FW Mohr, C Macaya, P den Heijer, MM Vrakking, K Wang, EM Mahoney, S Audi, K Leadley, KD Dawkins, and AP Kappetein. Quality of life after

pci with drug-eluting stents or coronary-artery bypass surgery. *N Engl J Med*, 364(11):1016–26, 2011.

[24] SP Cohen, SA Strassels, L Foster, J Marvel, K Williams, M Crooks, A Gross, C Kurihara, C Nguyen, and N Williams. Comparison of fluoroscopically guided and blind corticosteroid injections for greater trochanteric pain syndrome: multicentre randomised controlled trial. *BMJ*, 338:b1088, 2009.

[25] Lisa Conboy, Travis Gerke, Kai-Yin Hsu, Meredith St John, Marc Goldstein, and Rosa Schnyer. The effectiveness of individualized acupuncture protocols in the treatment of gulf war illness: A pragmatic randomized clinical trial. *PloS one*, 11(3):e0149161, 2016.

[26] K Cooley, O Szczurko, D Perri, EJ Mills, B Bernhardt, Q Zhou, and D Seely. Naturopathic care for anxiety: a randomized controlled trial. *PLoS One*, 4(8):e6628, 2009.

[27] J. Copas and S. Eguchi. Local sensitivity approximations for selectivity bias. *Journal of the Royal Statistical Society, Series B*, 63(871-895), 2001.

[28] P Coventry, K Lovell, C Dickens, P Bower, C Chew-Graham, D McElvenny, M Hann, A Cherrington, C Garrett, CJ Gibbons, C Baguley, K Roughley, I Adeyemi, D Reeves, W Waheed, and L Gask. Integrated primary care for patients with mental and physical multimorbidity: cluster randomised controlled trial of collaborative care for patients with depression comorbid with diabetes or cardiovascular disease. *BMJ*, 350:h638, 2015.

[29] JR Curtis, AL Back, DW Ford, L Downey, SE Shannon, AZ Doorenbos, EK Kross, LF Reinke, LC Feemster, B Edlund, RW Arnold, K O'Connor, and RA Engelberg. Effect of communication skills training for residents and nurse practitioners on quality of communication with patients with serious illness: a randomized trial. *JAMA*, 310(21):2271–81, 2013.

[30] BH Cuthbertson, J Rattray, MK Campbell, M Gager, S Roughton, A Smith, A Hull, S Breeman,

J Norrie, D Jenkinson, R Hernandez, M Johnston, E Wilson, and C Waldmann. The practical study of nurse led, intensive care follow-up programmes for improving long term outcomes from critical illness: a pragmatic randomised controlled trial. *BMJ*, 339:b3723, 2009.

[31] MJ Daniels and JW Hogan. *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. CRC Press, 2008.

[32] P. Diggle and M.G. Kenward. Informative drop-out in longitudinal data analysis. *Applied Statistics*, 43:49–93, 1994.

[33] A Dijk-De Vries, MA Bokhoven, B Winkens, B Terluin, JA Knottnerus, T Weijden, and JThM van Eijk. Lessons learnt from a cluster-randomised trial evaluating the effectiveness of Self-Management Support (SMS) delivered by practice nurses in routine diabetes care. *BMJ Open*, 5(6), 2015.

[34] JB Dixon, LM Schachter, PE O'Brien, K Jones, M Grima, G Lambert, W Brown, M Bailey, and MT Naughton. Surgical vs conventional therapy for weight loss treatment of obstructive sleep apnea: a randomized controlled trial. *JAMA*, 308(11):1142–9, 2012.

[35] SK Dobscha, K Corson, NA Perrin, GC Hanson, RQ Leibowitz, MN Doak, KC Dickinson, MD Sullivan, and MS Gerrity. Collaborative care for chronic pain in primary care: a cluster randomized trial. *JAMA*, 301(12):1242–52, 2009.

[36] JC Dumville, G Worthy, JM Bland, N Cullum, C Dowson, C Iglesias, JL Mitchell, EA Nelson, MO Soares, and DJ Torgerson. Larval therapy for leg ulcers (VenUS II): randomised controlled trial. *BMJ*, 338:b773, 2009.

[37] Bradley Efron and Gail Gong. A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37(1):36–48, 1983.

[38] Bradley Efron and Charles Stein. The jackknife estimate of variance. *The Annals of Statistics*, pages 586–596, 1981.

[39] F El-Khoury, B Cassou, A Latouche, P Aegerter, MA Charles, and P Dargent-Molina. Effectiveness of two year balance training programme on prevention of fall induced injuries in at risk women aged 75-85 living in community: Ossebo randomised controlled trial. *BMJ*, 351:h3830, 2015.

[40] MH Emmelot-Vonk, HJ Verhaar, HR Nakhai Pour, A Aleman, TM Lock, JL Bosch, DE Grobbee, and YT Schouw. Effect of testosterone supplementation on functional mobility, cognition, and other parameters in older men: a randomized controlled trial. *JAMA : the Journal of the American Medical Association*, 299(1):39–52, 2008.

[41] Jean Endicott, J Nee, W Harrison, and R Blumenthal. Quality of life enjoyment and satisfaction questionnaire. *Psychopharmacol Bull*, 29(2):321–326, 1993.

[42] Charles C Engel, Lisa H Jaycox, Michael C Freed, Robert M Bray, Donald Brambilla, Douglas Zatzick, Brett Litz, Terri Tanielian, Laura A Novak, Marian E Lane, et al. Centrally assisted collaborative telecare for posttraumatic stress disorder and depression among military personnel attending primary care: A randomized clinical trial. *JAMA internal medicine*, 176(7):948–956, 2016.

[43] F Fakhry, S Spronk, L van der Laan, JJ Wever, JA Teijink, WH Hoffmann, TM Smits, JP van Brussel, GN Stultiens, A Derom, PT den Hoed, GH Ho, LC van Dijk, N Verhofstad, M Orsini, A van Petersen, K Woltman, I Hulst, MR van Sambeek, D Rizopoulos, EV Rouwet, and MG Hunink. Endovascular revascularization and supervised exercise for peripheral artery disease and intermittent claudication: A randomized clinical trial. *JAMA*, 314(18):1936–44, 2015.

[44] B Favrat, K Balck, C Breymann, M Hedenus, T Keller, A Mezzacasa, and C Gasche. Eval-

uation of a single dose of ferric carboxymaltose in fatigued, iron-deficient women–PREFER a randomized, placebo-controlled study. *PLoS One*, 9(4):e94217, 2014.

[45] LE Fernandez-Rhodes, AD Kokkinis, MJ White, CA Watts, S Auh, NO Jeffries, JA Shrader, TJ Lehky, L Li, JE Ryder, EW Levy, BI Solomon, MO Harris-Love, A La Pean, AB Schindler, C Chen, NA Di Prospero, and KH Fischbeck. Efficacy and safety of dutasteride in patients with spinal and bulbar muscular atrophy: a randomised placebo-controlled trial. *Lancet Neurol*, 10(2):140–7, 2011 Feb.

[46] Shona Fielding, Graeme Maclennan, Jonathan A Cook, and Craig R Ramsay. A review of rcts in four medical journals to assess the use of imputation to overcome missing data in quality of life outcomes. *Trials*, 9(1):51, 2008.

[47] J Fisher, H Rowe, K Wynter, T Tran, P Lorgelly, LH Amir, J Proimos, S Ranasinha, H Hiscock, J Bayer, and W Cann. Gender-informed, psychoeducational programme for couples to prevent postnatal common mental disorders among primiparous women: cluster randomised controlled trial. *BMJ Open*, 6(3):e009396, 2016.

[48] KE Flynn, IL Pina, DJ Whellan, L Lin, JA Blumenthal, SJ Ellis, LJ Fine, JG Howlett, SJ Keteyian, DW Kitzman, WE Kraus, NH Miller, KA Schulman, JA Spertus, CM O'Connor, and KP Weinfurt. Effects of exercise training on health status in patients with chronic heart failure: HF-ACTION randomized controlled trial. *JAMA*, 301(14):1451–9, 2009.

[49] C Francois, N Rahhali, Y Chalem, P Sorensen, A Luquiens, and HJ Aubin. The effects of as-needed nalmefene on patient-reported outcomes and quality of life in relation to a reduction in alcohol consumption in alcohol-dependent patients. *PLoS One*, 10(6):e0129289, 2015.

[50] Samuel Frank, Claudia M Testa, David Stamler, Elise Kayson, Charles Davis, Mary C Edmondson, Shari Kinel, Blair Leavitt, David Oakes, Christine O'neill, et al. Effect of deutetrabenazine

on chorea among patients with huntington disease: a randomized clinical trial. *Jama*, 316(1):40–50, 2016.

[51] RB Frobell, EM Roos, HP Roos, J Ranstam, and LS Lohmander. A randomized trial of treatment for acute anterior cruciate ligament tears. *N Engl J Med*, 363(4):331–42, 2010.

[52] RB Frobell, HP Roos, EM Roos, FW Roemer, J Ranstam, and LS Lohmander. Treatment for acute anterior cruciate ligament tear: five year outcome of randomised trial. *BMJ*, 346:f232, 2013.

[53] PA Ganz, RS Cecchini, TB Julian, RG Margolese, JP Costantino, LA Vallow, KS Albain, PW Whitworth, ME Cianfrocca, AM Brufsky, HM Gross, GS Soori, JO Hopkins, L Fehrenbacher, K Sturtz, TF Wozniak, TE Seay, EP Mamounas, and N Wolmark. Patient-reported outcomes with anastrozole versus tamoxifen for postmenopausal patients with ductal carcinoma in situ treated with lumpectomy plus radiotherapy (NSABP B-35): a randomised, double-blind, phase 3 clinical trial. *Lancet*, 2015.

[54] MB Gavi, DV Vassalo, FT Amaral, DC Macedo, PL Gava, EM Dantas, and V Valim. Strengthening exercises improve symptoms and quality of life but do not change autonomic modulation in fibromyalgia: a randomized clinical trial. *PLoS One*, 9(3):e90767, 2014.

[55] Zoher Ghogawala, James Dziura, William E Butler, Feng Dai, Norma Terrin, Subu N Magge, Jean-Valery CE Coumans, J Fred Harrington, Sepideh Amin-Hanjani, J Sanford Schwartz, et al. Laminectomy plus fusion versus laminectomy alone for lumbar spondylolisthesis. *New England Journal of Medicine*, 374(15):1424–1434, 2016.

[56] S Gilbody, E Littlewood, C Hewitt, G Brierley, P Tharmanathan, R Araya, M Barkham, P Bower, C Cooper, L Gask, D Kessler, H Lester, K Lovell, G Parry, DA Richards, P Andersen, S Brabyn, S Knowles, C Shepherd, D Tallon, and D White. Computerised cognitive

behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. *BMJ*, 351:h5627, 2015.

[57] M Gine-Garriga, C Martin-Borras, A Puig-Ribera, C Martin-Cantera, M Sola, and A Cuesta-Vargas. The effect of a physical activity program on the total number of primary care visits in inactive patients: A 15-month randomized controlled trial. *PLoS One*, 8(6):e66392, 2013.

[58] N Glozier, H Christensen, S Naismith, N Cockayne, L Donkin, B Neal, A Mackinnon, and I Hickie. Internet-delivered cognitive behavioural therapy for adults with mild to moderate depression and high cardiovascular disease risks: a randomised attention-controlled trial. *PLoS One*, 8(3):e59139, 2013.

[59] H Goldberg, W Firtch, M Tyburski, A Pressman, L Ackerson, L Hamilton, W Smith, R Carver, A Maratukulam, LA Won, E Carragee, and AL Avins. Oral steroids for acute radiculopathy due to a herniated lumbar disk: a randomized clinical trial. *JAMA*, 313(19):1915–23, 2015.

[60] BJ Goudie, AR andLipworth, PJ Hopkinson, L Wei, and AD Struthers. Tadalafil in patients with chronic obstructive pulmonary disease: a randomised, double-blind, parallel-group, placebo-controlled trial. *Lancet Respir Med*, 2(4):293–300, 2014.

[61] GE Grande, L Austin, G Ewing, N O'Leary, and C Roberts. Assessing the impact of a Carer Support Needs Assessment Tool (CSNAT) intervention in palliative home care: a stepped wedge cluster trial. *BMJ Support Palliat Care*, 2015 Dec 30.

[62] Peter Hall, Jeff Racine, and Qi Li. Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association*, 99:1015–1026, 2004.

[63] F Halperin, SA Ding, DC Simonson, J Panosian, A Goebel-Fabbri, M Wewalka, O Hamdy, M Abrahamson, K Clancy, K Foster, D Lautz, A Vernon, and AB Goldfine. Roux-en-Y gastric

bypass surgery or lifestyle with intensive medical management in patients with type 2 diabetes: feasibility and 1-year results of a randomized clinical trial. *JAMA Surg*, 149(7):716–26, 2014.

[64] J. M. Hare, J. E. Fishman, G. Gerstenblith, D. L. DiFede Velazquez, J. P. Zambrano, V. Y. Suncion, M. Tracy, E. Ghersin, P. V. Johnston, J. A. Brinker, E. Breton, J. Davis-Sproul, I. H. Schulman, J. Byrnes, A. M. Mendizabal, M. H. Lowery, D. Rouy, P. Altman, C. Wong Po Foo, P. Ruiz, A. Amador, J. Da Silva, I. K. McNiece, and A. W. Heldman. Comparison of allogeneic vs autologous bone marrow-derived mesenchymal stem cells delivered by transendocardial injection in patients with ischemic cardiomyopathy: the POSEIDON randomized trial. *JAMA*, 308(22):2369–2379, Dec 2012.

[65] K Hegarty, L O'Doherty, A Taft, P Chondros, S Brown, J Valpied, J Astbury, A Taket, L Gold, G Feder, and J Gunn. Screening and counselling in the primary care setting for women who have experienced intimate partner violence (WEAVE): a cluster randomised controlled trial. *Lancet*, 382(9888):249–58, 2013.

[66] C Hellum, LG Johnsen, K Storheim, OP Nygaard, JI Brox, I Rossvoll, M Ro, L Sandvik, and O Grundnes. Surgery with disc prosthesis versus rehabilitation in patients with low back pain and degenerative disc: two year follow-up of randomised study. *BMJ*, 342:d2786, 2011.

[67] Lars P Hölzel, Zivile Ries, Levente Kriston, Jörg Dirmaier, Jördis M Zill, Christine Rummel-Kluge, Wilhelm Niebling, Isaac Bermejo, and Martin Härter. Effects of culture-sensitive adaptation of patient information material on usefulness in migrants: a multicentre, blinded randomised controlled trial. *BMJ open*, 6(11):e012008, 2016.

[68] WC Hsu, TL Wang, YJ Lin, LF Hsieh, CM Tsai, and KH Huang. Addition of lidocaine injection immediately before physiotherapy for frozen shoulder: a randomized controlled trial. *PLoS One*, 10(2):e0118217, 2015.

[69] Peter J Huber and EM Ronchetti. *Robust Statistics*. Wiley, 2009.

[70] JC Huffman, CA Mastromauro, SR Beach, CM Celano, CM DuBois, BC Healy, L Suarez, BL Rollman, and JL Januzzi. Collaborative care for depression and anxiety disorders in patients with recent cardiac events: the Management of Sadness and Anxiety in Cardiology (MOSAIC) randomized clinical trial. *JAMA Intern Med*, 174(6):927–35, 2014 Jun.

[71] CM Jenkinson, M Doherty, AJ Avery, A Read, MA Taylor, TH Sach, P Silcocks, and KR Muir. Effects of dietary intervention and quadriceps strengthening exercises on pain and function in overweight people with knee pain: randomised controlled trial. *BMJ*, 339:b3170, 2009.

[72] AA Khalafallah, AE Dennis, K Ogden, I Robertson, RH Charlton, JM Bellette, JL Shady, N Blesingk, and M Ball. Three-year follow-up of a randomised clinical trial of intravenous versus oral iron for anaemia in pregnancy. *BMJ Open*, 2(5), 2012.

[73] MN Khan, P Jais, J Cummings, L Di Biase, P Sanders, DO Martin, J Kautzner, S Hao, S Themistoclakis, R Fanelli, D Potenza, R Massaro, O Wazni, R Schweikert, W Saliba, P Wang, A Al-Ahmad, S Beheiry, P Santarelli, RC Starling, A Dello Russo, G Pelargonio, J Brachmann, V Schibgilla, A Bonso, M Casella, A Raviele, M Haissaguerre, and A Natale. Pulmonary-vein isolation for atrial fibrillation in patients with heart failure. *N Engl J Med*, 359(17):1778–85, 2008.

[74] TH Kim, KH Kim, JW Kang, M Lee, KW Kang, JE Kim, JH Kim, S Lee, MS Shin, SY Jung, AR Kim, HJ Park, HJ Jung, HS Song, HJ Kim, JB Choi, KE Hong, and SM Choi. Moxibustion treatment for knee osteoarthritis: a multi-centre, non-blinded, randomised controlled trial on the effectiveness and safety of the moxibustion treatment versus usual care in knee osteoarthritis patients. *PLoS One*, 9(7):e101973, 2014.

[75] A Kirkley, TB Birmingham, RB Litchfield, JR Giffin, KR Willits, CJ Wong, BG Feagan, A Donner, SH Griffin, LM D'Ascanio, JE Pope, and PJ Fowler. A randomized trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med*, 359(11):1097–107, 2008.

[76] DW Kitzman, P Brubaker, T Morgan, M Haykowsky, G Hundley, WE Kraus, J Eggebeen, and BJ Nicklas. Effect of caloric restriction or aerobic exercise training on peak oxygen consumption and quality of life in obese older patients with heart failure with preserved ejection fraction: A randomized clinical trial. *JAMA*, 315(1):36–46, 2016 Jan 5.

[77] J Klevens, R Kee, W Trick, D Garcia, FR Angulo, R Jones, and LS Sadowski. Effect of screening for partner violence on women's quality of life: a randomized controlled trial. *JAMA*, 308(7):681–9, 2012 Aug 15.

[78] MB Koek, E Buskens, H van Weelden, PH Steegmans, CA Bruijnzeel-Koomen, and V Sigurdsson. Home versus outpatient ultraviolet B phototherapy for mild to severe psoriasis: pragmatic multicentre randomised controlled non-inferiority trial (PLUTO study). *BMJ*, 338:b1542, 2009.

[79] A Kogure, K Kotani, S Katada, H Takagi, M Kamikozuru, T Isaji, and S Hakata. A randomized, single-blind, placebo-controlled study on the efficacy of the arthrokinematic approach-hakata method in patients with chronic nonspecific low back pain. *PLoS One*, 10(12):e0144325, 2015.

[80] RL Kravitz, P Franks, MD Feldman, DJ Tancredi, CA Slee, RM Epstein, PR Duberstein, RA Bell, M Jackson-Triche, DA Paterniti, C Cipri, AM Iosif, S Olson, S Kelly-Reif, A Hudnut, S Dvorak, C Turner, and A Jerant. Patient engagement programs for recognition and initial treatment of depression in primary care: a randomized trial. *JAMA*, 310(17):1818–28, 2013.

[81] K Kroenke, MJ Bair, TM Damush, J Wu, S Hoke, J Sutherland, and W Tu. Optimized antidepressant therapy and pain self-management in primary care patients with depression and musculoskeletal pain: a randomized controlled trial. *JAMA*, 301(20):2099–110, 2009.

[82] K. Kroenke, D. Theobald, J. Wu, K. Norton, G. Morrison, J. Carpenter, and W. Tu. Effect of telecare management on pain and depression in patients with cancer: a randomized trial. *JAMA*, 304(2):163–71, 2010.

[83] Saba M Lambert, Digafe T Alembo, Shimelis D Nigusse, Lawrence K Yamuah, Stephen L Walker, and Diana NJ Lockwood. A randomized controlled double blind trial of ciclosporin versus prednisolone in the management of leprosy patients with new type 1 reaction, in ethiopia. *PLoS Negl Trop Dis*, 10(4):e0004502, 2016.

[84] C Lau, R Yu, and J Woo. Effects of a 12-week Hatha yoga intervention on metabolic risk and quality of life in Hong Kong Chinese adults with and without metabolic syndrome. *PLoS One*, 10(6):e0130731, 2015.

[85] NT Lautenschlager, KL Cox, L Flicker, JK Foster, FM van Bockxmeer, J Xiao, KR Greenop, and OP Almeida. Effect of physical activity on cognitive function in older adults at risk for Alzheimer disease: a randomized trial. *JAMA*, 300(9):1027–37, 2008.

[86] BA Lawton, SB Rose, CR Elley, AC Dowell, A Fenton, and SA Moyes. Exercise on prescription for women aged 40-74 recruited through primary care: two year randomised controlled trial. *BMJ*, 337:a2509, 2008.

[87] A LeBlanc, J Herrin, MD Williams, JW Inselman, ME Branda, ND Shah, EM Heim, SR Dick, M Linzer, DH Boehm, KM Dall-Winther, MR Matthews, KJ Yost, KK Shepel, and VM Montori. Shared decision making for antidepressants in primary care: A cluster randomized trial. *JAMA Intern Med*, 175(11):1761–70, 2015.

[88] Hong Lei, Nima Toosizadeh, Michael Schwenk, Scott Sherman, Stephan Karp, Esther Sternberg, and Bijan Najafi. A pilot clinical trial to objectively assess the efficacy of electroacupuncture on gait in patients with parkinson's disease using body worn sensors. *PloS one*, 11(5):e0155613, 2016.

[89] Tianjing Li, Susan Hutfless, Daniel O Scharfstein, Michael J Daniels, Joseph W Hogan, Roderick JA Little, Jason A Roy, Andrew H Law, and Kay Dickersin. Standards should be applied in

the prevention and handling of missing data for patient-centered outcomes research: a systematic review and expert consensus. *Journal of clinical epidemiology*, 67(1):15–32, 2014.

[90] R Little, M Cohen, K Dickersin, S Emerson, J Farrar, C Frangakis, JW Hogan, G. Molenberghs, S. Murphy, J. Neaton, A Rotnitzky, DO Scharfstein, W Shih, J Siegel, and H Stern. *The Prevention and Treatment of Missing Data in Clinical Trials*. The National Academies Press, 2010.

[91] R. J. Little, R. D'Agostino, M. L. Cohen, K. Dickersin, S. S. Emerson, J. T. Farrar, C. Frangakis, J. W. Hogan, G. Molenberghs, S. A. Murphy, J. D. Neaton, A. Rotnitzky, D. Scharfstein, W. J. Shih, J. P. Siegel, and H. Stern. The prevention and treatment of missing data in clinical trials. *N. Engl. J. Med.*, 367(14):1355–1360, Oct 2012.

[92] Roderick JA Little and Donald B Rubin. *Statistical Analysis with Missing Data*. John Wiley & Sons, 2014.

[93] KH Ly, A Truschel, L Jarl, S Magnusson, T Windahl, R Johansson, P Carlbring, and G Andersson. Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. *BMJ Open*, 4(1):e003440, 2014.

[94] G. Ma, A.B. Toxel, and D.F. Heitjan. An index of local sensitivity to nonignorable drop-out in longitudinal modelling. *Statistics in Medicine*, 24:2129–2150, 2005.

[95] S MacPherson, H abd Richmond, M Bland, S Brealey, R Gabe, A Hopton, A Keding, H Lansdown, S Perren, M Sculpher, E Spackman, D Torgerson, and I Watt. Acupuncture and counselling for depression in primary care: a randomised controlled trial. *PLoS Med*, 10(9):e1001518, 2013.

[96] K Mansikkamaki, J Raitanen, CH Nygard, E Tomas, R Rutanen, and R Luoto. Long-term

effect of physical activity on health-related quality of life among menopausal women: a 4-year follow-up study to a randomised controlled trial. *BMJ Open*, 5(9):e008232, 2015.

[97] DB Mark, W Pan, NE Clapp-Channing, KJ Anstrom, JR Ross, RS Fox, GP Devlin, CE Martin, C Adlbrecht, PA Cowper, LD Ray, EA Cohen, GA Lamas, and JS Hochman. Quality of life after late invasive therapy for occluded arteries. *N Engl J Med*, 360(8):774–83, 2009.

[98] M Marklund, B Carlberg, L Forsgren, T Olsson, H Stenlund, and KA Franklin. Oral appliance therapy in patients with daytime sleepiness and snoring or mild to moderate sleep apnea: A randomized clinical trial. *JAMA Intern Med*, 175(8):1278–85, 2015.

[99] Corby K Martin, Manju Bhapkar, Anastassios G Pittas, Carl F Pieper, Sai Krupa Das, Donald A Williamson, Tammy Scott, Leanne M Redman, Richard Stein, Cheryl H Gilhooly, et al. Effect of calorie restriction on mood, quality of life, sleep, and sexual function in healthy nonobese adults: The calerie 2 randomized clinical trial. *JAMA internal medicine*, 176(6):743–752, 2016.

[100] CM McClellan, F Cramp, J Powell, and JR Benger. A randomised trial comparing the clinical effectiveness of different emergency department healthcare professionals in soft tissue injury management. *BMJ Open*, 2(6), 2012.

[101] CJ McDermott, PJ Shaw, CL Cooper, S Dixon, WO Baird, MJ Bradburn, P Fitzgerald, C Maguire, SK Baxter, T Williams, SV Baudouin, D Karat, K Talbot, J Stradling, N Maynard, M Turner, A Sarela, S Bianchi, R Ackroyd, SC Bourke, J Ealing, H Hamdalla, C Young, A Bentley, S Galloway, RW Orrell, W Wedzicha, M Elliot, P Hughes, R Berrisford, CO Hanemann, I Imam, AK Simonds, L Taylor, R Leek, N Leigh, M Dewey, and A Radunovic. Safety and efficacy of diaphragm pacing in patients with respiratory insufficiency due to amyotrophic lateral sclerosis (DiPALS): A multicentre, open-label, randomised controlled trial. *The Lancet Neurology*, 14(9):883–92, 2015.

[102] M. M. McDermott, P. Ades, J. M. Guralnik, A. Dyer, L. Ferrucci, K. Liu, M. Nelson, D. Lloyd-Jones, L. Van Horn, D. Garside, M. Kibbe, K. Domanchuk, J. H. Stein, Y. Liao, H. Tao, D. Green, W. H. Pearce, J. R. Schneider, D. McPherson, S. T. Laing, W. J. McCarthy, A. Shroff, and M. H. Criqui. Treadmill exercise and resistance training in patients with peripheral arterial disease with and without intermittent claudication: a randomized controlled trial. *JAMA*, 301(2):165–74, 2009.

[103] MM McDermott, K Liu, JM Guralnik, MH Criqui, B Spring, L Tian, K Domanchuk, L Ferrucci, D Lloyd-Jones, M Kibbe, H Tao, L Zhao, Y Liao, and WJ Rejeski. Home-based walking exercise intervention in peripheral artery disease: a randomized clinical trial. *JAMA*, 310(1):57–65, 2013.

[104] M McFall, AJ Saxon, CA Malte, B Chow, S Bailey, DG Baker, JC Beckham, KD Boardman, TP Carmody, AM Joseph, MW Smith, MC Shih, Y Lu, M Holodniy, and PW Lavori. Integrating tobacco cessation into mental health care for posttraumatic stress disorder: a randomized controlled trial. *JAMA*, 304(22):2485–93, 2010.

[105] A McMillan, DJ Bratton, R Faria, M Laskawiec-Szkonter, S Griffin, RJ Davies, AJ Nunn, JR Stradling, RL Riha, and MJ Morrell. Continuous positive airway pressure in older people with obstructive sleep apnoea syndrome (PREDICT): A 12-month, multicentre, randomised trial. *The Lancet Respiratory Medicine*, 2(10):804–12, 2014.

[106] GE Mead, C Graham, P Dorman, SK Bruins, SC Lewis, MS Dennis, and PA Sandercock. Fatigue after stroke: baseline predictors and influence on survival. analysis of data from uk patients recruited in the international stroke trial. *PLoS One*, 6(3):e16988, 2011.

[107] Dafna Merom, Erin Mathieu, Ester Cerin, Rachael L Morton, Judy M Simpson, Chris Rissel, Kaarin J Anstey, Catherine Sherrington, Stephen R Lord, and Robert G Cumming. Social dancing and incidence of falls in older adults: a cluster randomised controlled trial. *PLoS Med*, 13(8):e1002112, 2016.

[108] S Middleton, P McElduff, J Ward, JM Grimshaw, S Dale, C D'Este, P Drury, R Griffiths, NW Cheung, C Quinn, M Evans, D Cadilhac, and C Levi. Implementation of evidence-based treatment protocols to manage fever, hyperglycaemia, and swallowing dysfunction in acute stroke (QASC): a cluster randomised controlled trial. *Lancet*, 378(9804):1699–706, 2011.

[109] T Miyagawa, H Kawamura, M Obuchi, A Ikesaki, A Ozaki, K Tokunaga, Y Inoue, and M Honda. Effects of oral l-carnitine administration in narcolepsy patients: a randomized, double-blind, cross-over and placebo-controlled trial. *PLoS One*, 8(1):e53707, 2013.

[110] D. C. Mohr, J. Ho, J. Duffecy, D. Reifler, L. Sokol, M. N. Burns, L. Jin, and J Siddique. Effect of telephone-administered vs face-to-face cognitive behavioral therapy on adherence to therapy and depression outcomes among primary care patients: a randomized trial. *JAMA*, 307(21):2278–85, 2012.

[111] DC Mohr, J Duffecy, J Ho, M Kwasny, X Cai, MN Burns, and M Begale. A randomized controlled trial evaluating a manualized telecoaching protocol for improving adherence to a web-based intervention for the treatment of depression. *PLoS One*, 8(8):e70086, 2013.

[112] Xavier Montalban, Stephen L Hauser, Ludwig Kappos, Douglas L Arnold, Amit Bar-Or, Giancarlo Comi, Jérôme de Seze, Gavin Giovannoni, Hans-Peter Hartung, Bernhard Hemmer, et al. Ocrelizumab versus placebo in primary progressive multiple sclerosis. *New England Journal of Medicine*, 2016.

[113] M Mordin, C Masaquel, C Abbott, and C Copley-Merriman. Factors affecting the health-related quality of life of patients with cervical dystonia and impact of treatment with abobotulinumtoxinA (dysport): results from a randomised, double-blind, placebo-controlled study. *BMJ Open*, 4(10):e005150, 2014.

[114] MC Morey, DC Snyder, R Sloane, HJ Cohen, B Peterson, TJ Hartman, P Miller, DC Mitchell, and W Demark-Wahnefried. Effects of home-based diet and exercise on functional outcomes

among older, overweight long-term cancer survivors: RENEW: a randomized controlled trial. *JAMA*, 301(18):1883–91, 2009.

[115] AJ Morgan, AF Jorm, and AJ Mackinnon. Self-help for depression via e-mail: A randomised controlled trial of effects on depression and self-help behaviour. *PLoS One*, 8(6):e66537, 2013.

[116] CJ Morrell, P Slade, R Warner, G Paley, S Dixon, SJ Walters, T Brugha, M Barkham, GJ Parry, and J Nicholl. Clinical effectiveness of health visitor training in psychologically informed approaches for depression in postnatal women: pragmatic cluster randomised trial in primary care. *BMJ*, 338:a3045, 2009.

[117] AW Murphy, ME Cupples, SM Smith, M Byrne, MC Byrne, and J Newell. Effect of tailored practice and patient care plans on secondary prevention of heart disease in general practice: cluster randomised controlled trial. *BMJ*, 339:b4220, 2009.

[118] P Musiat, P Conrod, J Treasure, A Tylee, C Williams, and U Schmidt. Targeted prevention of common mental health disorders in university students: randomised controlled trial of a transdiagnostic trait-focused web-based intervention. *PLoS One*, 9(4):e93621, 2014.

[119] H Nagayama, K Tomori, K Ohno, K Takahashi, K Ogahara, T Sawada, S Uezu, R Nagatani, and K Yamauchi. Effectiveness and cost-effectiveness of occupation-based occupational therapy using the aid for decision making in occupation choice (ADOC) for older residents: Pilot cluster randomized controlled trial. *PLoS One*, 11(3):e0150374, 2016.

[120] B Oerkild, M Frederiksen, JF Hansen, and E Prescott. Home-based cardiac rehabilitation is an attractive alternative to no cardiac rehabilitation for elderly patients with coronary heart disease: results from a randomised clinical trial. *BMJ Open*, 2(6), 2012.

[121] D Pareyson, MM Reilly, A Schenone, GM Fabrizi, T Cavallaro, L Santoro, G Vita, A Quattrone, L Padua, F Gemignani, F Visioli, M Laura, D Radice, D Calabrese, RA Hughes, and A Solari.

Ascorbic acid in Charcot-Marie-Tooth disease type 1A (CMT-TRIAAL and CMT-TRAUK): a double-blind randomised trial. *Lancet Neurol*, 10(4):320–8, 2011.

[122] A Patel, M Buszewicz, J Beecham, M Griffin, G Rait, I Nazareth, A Atkinson, J Barlow, and A Haines. Economic evaluation of arthritis self management in primary care. *BMJ*, 339:b3532, 2009.

[123] Vikram Patel, Benedict Weobong, Helen A Weiss, Arpita Anand, Bhargav Bhat, Basavraj Katti, Sona Dimidjian, Ricardo Araya, Steve D Hollon, Michael King, et al. The healthy activity program (hap), a lay counsellor-delivered brief psychological treatment for severe depression, in primary care in india: a randomised controlled trial. *The Lancet*, 389(10065):176–185, 2017.

[124] J Poole, K Mavromatis, JN Binongo, A Khan, Q Li, M Khayata, E Rocco, M Topel, X Zhang, C Brown, MA Corriere, J Murrow, S Sher, S Clement, K Ashraf, A Rashed, T Kabbany, R Neuman, A Morris, A Ali, S Hayek, J Oshinski, YS Yoon, EK Waller, and AA Quyyumi. Effect of progenitor cell mobilization with granulocyte-macrophage colony-stimulating factor in patients with peripheral artery disease: a randomized clinical trial. *JAMA*, 310(24):2631–9, 2013.

[125] Atif Rahman, Syed Usman Hamdani, Naila Riaz Awan, Richard A Bryant, Katie S Dawson, Muhammad Firaz Khan, Mian Mukhtar-Ul-Haq Azeemi, Parveen Akhtar, Huma Nazir, Anna Chiumento, et al. Effect of a multicomponent behavioral intervention in adults impaired by psychological distress in a conflict-affected area of pakistan: a randomized clinical trial. *JAMA*, 316(24):2609–2617, 2016.

[126] M Ramly, MF Ming, K Chinna, S Suboh, and R Pendek. Effect of vitamin D supplementation on cardiometabolic risks and health-related quality of life among urban premenopausal women in a tropical country–a randomized controlled trial. *PLoS One*, 9(10):e110476, 2014.

[127] DA Richards, JJ Hill, L Gask, K Lovell, C Chew-Graham, P Bower, J Cape, S Pilling, R Araya, D Kessler, JM Bland, C Green, S Gilbody, G Lewis, C Manning, A Hughes-Morley, and

M Barkham. Clinical effectiveness of collaborative care for depression in UK primary care (CADET): cluster randomised controlled trial. *BMJ*, 347:f4913, 2013.

[128] David A Richards, David Ekers, Dean McMillan, Rod S Taylor, Sarah Byford, Fiona C Warren, Barbara Barrett, Paul A Farrand, Simon Gilbody, Willem Kuyken, et al. Cost and outcome of behavioural activation versus cognitive behavioural therapy for depression (cobra): a randomised, controlled, non-inferiority trial. *The Lancet*, 388(10047):871–880, 2016.

[129] LP Richardson, E Ludman, E McCauley, J Lindenbaum, C Larison, C Zhou, G Clarke, D Brent, and W Katon. Collaborative care for adolescents with depression in primary care: a randomized clinical trial. *JAMA*, 312(8):809–16, 2014.

[130] James M Robins. Non-response models for the analysis of non-monotone non-ignorable missing data. *Statistics in Medicine*, 16(1):21–37, 1997.

[131] JM Robins, A Rotnitzky, and DO Scharfstein. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In E. Halloran, editor, *Statistical Models for Epidemiology*, pages 1–94. Springer-Verlag, 2000.

[132] BL Rollman, BH Belnap, MS LeMenager, S Mazumdar, PR Houck, PJ Counihan, WN Kapoor, HC Schulberg, and CF Reynolds. Telephone-delivered collaborative care for treating post-CABG depression: a randomized controlled trial. *JAMA*, 302(19):2095–103, 2009.

[133] A Rotnitzky, JM Robins, and DO Scharfstein. Semiparametric regression for repeated outcomes with non-ignorable non-response. *Journal of the American Statistical Association*, 93:1321–1339, 1998.

[134] A Rotnitzky, DO Scharfstein, TL Su, and JM Robins. A sensitivity analysis methodology for randomized trials with potentially non-ignorable cause-specific censoring. *Biometrics*, 57:103–113, 2001.

[135] Chris Salisbury, Alicia O'Cathain, Louisa Edwards, Clare Thomas, Daisy Gaunt, Sandra Hollinghurst, Jon Nicholl, Shirley Large, Lucy Yardley, Glyn Lewis, et al. Effectiveness of an integrated telehealth service for patients with depression: a pragmatic randomised controlled trial of a complex intervention. *The Lancet Psychiatry*, 3(6):515–525, 2016.

[136] D Scharfstein, A McDermott, W Olson, and Wiegand F. Global sensitivity analysis for repeated measures studies with informative drop-out. *Statistics in Biopharmaceutical Research*, 6:338–348, 2014.

[137] DO Scharfstein, A McDermott, I Diaz, Carone M, N Lunardon, and I Turkoz. Global sensitivity analysis for repeated measures studies with informative drop-out: A semi-parametric approach. *Biometrics*, 2017.

[138] DO Scharfstein, A Rotnitzky, and JM Robins. Adjusting for non-ignorable drop-out using semi-parametric non-response models (with discussion). *Journal of the American Statistical Association*, 94:1096–1146, 1999.

[139] M Sharpe, KA Goldsmith, AL Johnson, T Chalder, J Walker, and PD White. Rehabilitative treatments for chronic fatigue syndrome: long-term follow-up from the PACE trial. *Lancet Psychiatry*, 2(12):1067–74, 2015.

[140] DE Simkiss, HA Snooks, N Stallard, PK Kimani, B Sewell, D Fitzsimmons, R Anthony, S Winstanley, L Wilson, CJ Phillips, and S Stewart-Brown. Effectiveness and cost-effectiveness of a universal parenting skills programme in deprived communities: multicentre randomised controlled trial. *BMJ Open*, 3(8), 2013.

[141] R Small, L Watson, J Gunn, C Mitchell, and S Brown. Improving population-level maternal health: a hard nut to crack? long term findings and reflections on a 16-community randomised trial in Australia to improve maternal emotional and physical health after birth. *PLoS One*, 9(2):e88457, 2014.

[142] NL Stanley, MA andWilson, DM Novy, HM Rhoades, PD Wagener, AJ Greisinger, JA Cully, and ME Kunik. Cognitive behavior therapy for generalized anxiety disorder among older adults in primary care: a randomized clinical trial. *JAMA*, 301(14):1460–7, 2009.

[143] DR Strayer, WA Carter, BC Stouch, SR Stevens, L Bateman, PJ Cimoch, CW Lapp, DL Peterson, and WM Mitchell. A double-blind, placebo-controlled, randomized, clinical trial of the TLR-3 agonist rintatolimod in severe cases of chronic fatigue syndrome. *PLoS One*, 7(3):e31334, 2012.

[144] FM Stuby, S Dobele, SD Schaffer, S Mueller, A Ateschrang, M Baumann, and D Zieker. Early functional postoperative therapy of distal radius fracture with a dynamic orthosis: results of a prospective randomized cross-over comparative study. *PLoS One*, 10(3):e0117720, 2015.

[145] MD Sullivan, WJ Katon, LC Lovato, ME Miller, AM Murray, KR Horowitz, RN Bryan, HC Gerstein, S Marcovina, BE Akpunonu, J Johnson, JF Yale, J Williamson, and LJ Launer. Association of depression with accelerated cognitive decline among patients with type 2 diabetes in the ACCORD-MIND trial. *JAMA Psychiatry*, 70(10):1041–7, 2013.

[146] JS Temel, JA Greer, A Muzikansky, ER Gallagher, S Admane, VA Jackson, CM Dahlin, CD Blinderman, J Jacobsen, WF Pirl, JA Billings, and TJ Lynch. Early palliative care for patients with metastatic non-small-cell lung cancer. *N Engl J Med*, 363(8):733–42, 2010.

[147] SP Therkelsen, G Hetland, T Lyberg, I Lygren, and E Johnson. Effect of a medicinal agaricus blazei murill-based mushroom extract, AndoSan, on symptoms, fatigue and quality of life in patients with ulcerative colitis in a randomized single-blinded placebo controlled study. *PLoS One*, 11(3):e0150191, 2016.

[148] Stig Palm Therkelsen, Geir Hetland, Torstein Lyberg, Idar Lygren, and Egil Johnson. Effect of the medicinal agaricus blazei murill-based mushroom extract, andosan tm, on symptoms,

fatigue and quality of life in patients with crohn's disease in a randomized single-blinded placebo controlled study. *PloS one*, 11(7):e0159288, 2016.

[149] N Titov, G Andrews, M Davies, K McIntyre, E Robinson, and K Solley. Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. *PLoS One*, 5(6):e10939, 2010.

[150] N Titov, BF Dear, L Johnston, C Lorian, J Zou, B Wootton, J Spence, PM McEvoy, and RM Rapee. Improving adherence and clinical outcomes in self-guided internet treatment for anxiety and depression: randomised controlled trial. *PLoS One*, 8(7):e62873, 2013.

[151] N Titov, BF Dear, L Johnston, PM McEvoy, B Wootton, MD Terides, M Gandy, V Fogliati, R Kayrouz, and RM Rapee. Improving adherence and clinical outcomes in self-guided internet treatment for anxiety and depression: a 12-month follow-up of a randomised controlled trial. *PLoS One*, 9(2):e89591, 2014.

[152] A Tiwari, DY Fong, KH Yuen, H Yuk, P Pang, J Humphreys, and L Bullock. Effect of an advocacy intervention on mental health in Chinese women survivors of intimate partner violence: a randomized controlled trial. *JAMA*, 304(5):536–43, 2010.

[153] A.B. Troxel, G. Ma, and D.F. Heitjan. An index of local sensitivity to nonignorability. *Statistica Sinica*, 14:1221–1237, 2004.

[154] AA Tsiatis. *Semiparametric Theory and Missing Data. 2006.* Springer Verlag, New York, 2006.

[155] WA van Gemert, J van der Palen, EM Monninkhof, A Rozeboom, R Peters, H Wittink, AJ Schuit, and PH Peeters. Quality of life after diet or exercise-induced weight loss in overweight to obese postmenopausal women: The SHAPE-2 randomised controlled trial. *PLoS One*, 10(6):e0127520, 2015.

[156] G. Verbeke, G. Molenberghs, H. Thijs, E. Lesaffre, and M.G. Kenward. Sensitivity analysis for nonrandom dropout: A local influence approach. *Biometrics*, 57:7–14, 2001.

[157] M Wall, MP McDermott, KD Kieburtz, JJ Corbett, SE Feldon, DI Friedman, DM Katz, JL Keltner, EB Schron, and MJ Kupersmith. Effect of acetazolamide on visual function in patients with idiopathic intracranial hypertension and mild visual loss: the idiopathic intracranial hypertension treatment trial. *JAMA*, 311(16):1641–51, 2014 Apr 23-30.

[158] TS Walsh, LG Salisbury, JL Merriweather, JA Boyd, DM Griffith, G Huby, S Kean, SJ Mackenzie, A Krishan, SC Lewis, GD Murray, JF Forbes, J Smith, JE Rattray, AM Hull, and P Ramsay. Increased hospital-based physical rehabilitation and information provision after intensive care unit discharge: The RECOVER randomized clinical trial. *JAMA Intern Med*, 175(6):901–10, 2015.

[159] J Walters, H Cameron-Tucker, K Wills, N Schuz, J Scott, A Robinson, M Nelson, P Turner, R Wood-Baker, and EH Walters. Effects of telephone health mentoring in community-recruited chronic obstructive pulmonary disease on self-management capacity, quality of life and psychological morbidity: a randomised controlled trial. *BMJ Open*, 3(9):e003097, 2013.

[160] C Wang, CH Schmid, R Rones, R Kalish, J Yinh, DL Goldenberg, Y Lee, and T McAlindon. A randomized trial of tai chi for fibromyalgia. *N Engl J Med*, 363(8):743–54, 2010.

[161] D Wardlaw, SR Cummings, J Van Meirhaeghe, L Bastian, JB Tillman, J Ranstam, R Eastell, P Shabe, K Talmadge, and S Boonen. Efficacy and safety of balloon kyphoplasty compared with non-surgical care for vertebral compression fracture (FREE): a randomised controlled trial. *Lancet*, 373(9668):1016–24, 2009.

[162] JN Weinstein, TD Tosteson, JD Lurie, AN Tosteson, E Blood, B Hanscom, H Herkowitz, F Cammisa, T Albert, SD Boden, A Hilibrand, H Goldberg, S Berven, and H An. Surgical versus nonsurgical therapy for lumbar spinal stenosis. *N Engl J Med*, 358(8):794–810, 2008.

[163] Constance M Weisner, Felicia W Chi, Yun Lu, Thekla B Ross, Sabrina B Wood, Agatha Hinman, David Pating, Derek Satre, and Stacy A Sterling. Examination of the effects of an intervention aiming to link patients receiving addiction treatment with health care: the linkage clinical trial. *Jama psychiatry*, 73(8):804–814, 2016.

[164] DM Weiss, RJ Casten, BE Leiby, LA Hark, AP Murchison, D Johnson, S Stratford, J Henderer, BW Rovner, and JA Haller. Effect of behavioral intervention on dilated fundus examination rates in older african american individuals with diabetes mellitus: A randomized clinical trial. *JAMA Ophthalmol*, 133(9):1005–12, 2015.

[165] PD White, KA Goldsmith, AL Johnson, L Potts, R Walwyn, JC DeCesare, HL Baber, M Burgess, LV Clark, DL Cox, J Bavinton, BJ Angus, G Murphy, M Murphy, H O'Dowd, D Wilks, P McCrone, T Chalder, and M Sharpe. Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): a randomised trial. *Lancet*, 377(9768):823–36, 2011.

[166] A Wilkins, H Mossop, I Syndikus, V Khoo, D Bloomfield, C Parker, J Logue, C Scrase, H Patterson, A Birtle, J Staffurth, Z Malik, M Panades, C Eswar, J Graham, M Russell, P Kirkbride, JM O'Sullivan, A Gao, C Cruickshank, C Griffin, D Dearnaley, and E Hall. Hypofractionated radiotherapy versus conventionally fractionated radiotherapy for patients with intermediate-risk localised prostate cancer: 2-year patient-reported outcomes of the randomised, non-inferiority, phase 3 CHHiP trial. *Lancet Oncol*, 16(16):1605–16, 2015.

[167] J Williams, T Russell, D Durai, WY Cheung, A Farrin, K Bloor, S Coulton, and G Richardson. Effectiveness of nurse delivered endoscopy: Findings from randomised multi-institution nurse endoscopy trial (MlNuET). *BMJ (Clinical Research Ed.)*, 338(7693):b231, 2009.

[168] K Witt, C Daniels, J Reiff, P Krack, J Volkmann, MO Pinsker, M Krause, V Tronnier, M Kloss, A Schnitzler, L Wojtecki, K Botzel, A Danek, R Hilker, V Sturm, A Kupsch, E Karner, and

G Deuschl. Neuropsychological and psychiatric changes after deep brain stimulation for parkinson's disease: a randomised, multicentre study. *Lancet Neurol*, 7(7):605–14, 2008.

[169] X. Yan, S. Lee, and N. Li. Missing data handling methods in medical device clinical trials. *Journal of Biopharmaceutical Statistics*, 19:1085–1098, 2009.

[170] JO Younge, MF Wery, RA Gotink, EM Utens, M Michels, D Rizopoulos, EF van Rossum, MG Hunink, and JW Roos-Hesselink. Web-based mindfulness intervention in heart disease: A randomized controlled trial. *PLoS One*, 10(12):e0143843, 2015.

[171] LN Zonneveld, YR van Rood, R Timman, CG Kooiman, A Van't Spijker, and JJ Busschbach. Effective group training for patients with unexplained physical symptoms: a randomized controlled trial with a non-randomized one-year follow-up. *PLoS One*, 7(8):e42629, 2012.

Appendix A: Influence Function

Let

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

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

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

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

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

Using semiparametric theory [154], the efficient influence function in model $\mathcal{M}$ can be computed as:

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

where

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

Table 1: List of Studies

Study	Indication	Journal	Endpoint	n	Follow-Up	Missing Data (%)
Berende (2016)	Lyme Disease	NEJM	SF-36	280	14 wks.	6.8%
Cohen (2011)	Cardiac Surgey	NEJM	SF-36	1800	1,6,12 mos.	9.5%-9.7%
Frobell (2010)	ACL Injury	NEJM	SF-36	141	3,6,12,24 mos.	14.2%-14.9%
Ghogawala (2016)	Lumbar Spondylolisthesis	NEJM	SF-36	66	1.5, 3, 6, 12, 24, 36, 48 mos.	12.1% - 31.8%
Khan (2008)	Heart Failure	NEJM	MLHFQ	81	6 mos.	0.0%
Kirkley (2008)	Oseteoarthritis	NEJM	SF-36	188	3,6,12,18, 24 mos.	9.6%-21.3%
Mark (2009)	Myocardial Infarction	NEJM	SF-36	951	4,12,24 mos.	12.4%-18.7%
Montalban (2016)	Multiple Sclerosis	NEJM	SF-36	732	120 wks.	21.3%
Temel (2010)	Metastatic Lung Cancer	NEJM	PHQ-9	151	12 wks.	31.1%
Wang (2010)	Fibromyalgia	NEJM	SF-36	66	12,24 wks.	7.6%-10.6%
Weinstein (2008)	Spinal Stenosis	NEJM	SF-36	289	1.5,3,6,12,24 mos.	11.8%-23.5%
Chalder (2015)	Chronic Fatigue Syndrome	Lancet-P	SF-36	641	52 wks.	14.0%
Christensen (2016)	Insomnia/Depression	Lancet-P	PHQ-9	1149	6 wks., 6 mos.	49.4%-56.1%
Fernandez-Rhodes (2011)	Spinal & Bulbar Muscular Atrophy	Lancet-N	SF-36	50	24 mos.	14.0%
Ganz (2015)	Ductal Carcinoma In Situ	Lancet	SF-12	1193	Every 6 mos. thru 54 mos.	4.9%-35.2%
Goudie (2014)	COPD	Lancet-RM	SF-36	120	12 wks.	5.8%
Hegarty (2013)	Intimate Partner Violence	Lancet	SF-12	272	6,12 mos.	30.9%-32.0%
McMillan (2014)	Sleep Apnoea	Lancet-RM	SF-36	278	3,12 mos.	11.9%-16.9%
Middelton (2011)	Stroke	Lancet	SF-36	1126	90 days	10.4%
Pareyson (2011)	Charcot-Marie-Tooth Disease	Lancet-N	SF-36	277	24 mos.	20.2%
Patel (2016)	Depression	Lancet	PHQ-9	495	3 mos.	5.9%
Richards (2016)	Depression	Lancet	PHQ-9	440	6, 12, 18 mos.	13.6% - 19.1%
Sharpe (2015)	Chronic Fatigue Syndrome	Lancet-P	SF-36	481	12,24,52,134 wks.	25.0%-26.1%
Salisbury (2016)	Depression	Lancet-P	PHQ-9	609	4,8,12 mos.	13.8%-15.4%
Wardlaw (2009)	Vertebral Fracture	Lancet	SF-36	300	1,3,6,12 mos.	13.0%-25.0%
White (2011)	Chronic Fatigue Syndrome	Lancet	SF-36	641	12, 24, 52 wks.	4.4%-5.6%
Wilkins (2015)	Localized Prostate Cancer	Lancet-O	SF-36	2100	24 mos.	31.2%
Witt (2008)	Parkinson's	Lancet-N	SF-36	156	6 mos.	21.2%
Ahimastos (2013)	Peripheral Artery Disease	JAMA	SF-36	212	6 mos.	5.7%
Bekelman (2015)	Heart Failure	JAMA-IM	KCCQ	392	3,6,12 mos.	10.2%-15.6%
Berk (2013)	Familial Amyloid Polyneuropathy	JAMA	SF-36	130	1,2 yrs.	32.3%-47.7%
Chibanda (2016)	Mental Disorders	JAMA	PHQ-9	573	6 mos.	9.1%
Curtis (2013)	Quality of Communication	JAMA	SF-12	472	10 mos.	58.9%
Dixon (2012)	Obstructive Sleep Apnea	JAMA	SF-36	60	2 yrs.	13.3%
Dobscha (2009)	Musculoskeletal Pain	JAMA	PHQ-9	401	3,6,12 mos.	3.0%-9.7%
Emmelot-Vonk (2008)	Low Testosterone	JAMA	SF-36	237	3,6 mos.	5.1%-12.7%
Engel (2016)	PTSD/Depression	JAMA-IM	SF-12	660	3,6,12 mos.	6.4%-12.1%
Fakhry (2015)	Intermittent Claudication	JAMA	SF-36	212	12 mos.	8.0%
Flynn (2009)	Heart Failure	JAMA	KCCQ	2331	3,6,9,12,24,36 mos.	12.6%-75.4%
Frank (2016)	Huntington Disease	JAMA	SF-36	90	12 wks.	<10%
Goldberg (2015)	Acute Sciatica	JAMA	SF-36	269	3,52 wks.	0.7%-13.0%

Continued on next page

Table 1 – *Continued from previous page*

Study	**Indication**	**Journal**	**Endpoint**	n	**Follow-Up**	**Missing Data (%)**
Halperin (2014)	Diabetes	JAMA-S	SF-36	43	1 yr.	11.6%
Hare (2012)	Ischemic Cardiomyopathy	JAMA	MLHFQ	31	3,6,12 mos.	9.7%-22.6%
Huffman (2014)	Depression/Anxiety	JAMA-IM	SF-12	183	24 wks.	6.0%
Kitzman (2016)	Heart Failure	JAMA	MLHFQ	100	20 wks.	8.0%
Klevens (2012)	Intimate Partner Violence	JAMA	SF-12	2700	1 yr.	12.4%
Kravitz (2013)	Depression	JAMA	SF-12	603	12 wks.	22.6%
Kroenke (2009)	Pain and Depression	JAMA	SF-36	250	1,3,6,12 mos.	4.0%-18.0%
Kroenke (2010)	Depression	JAMA	SF-36	405	1,3,6,12 mos.	12.6%-33.6%
Lautenschlager (2008)	Alzheimer's Disease	JAMA	SF-36	170	18 mos.	21.8%
LeBlanc (2015)	Depression	JAMA-IM	PHQ-9	301	3,6 mos.	60.8%-62.5%
Lenze (2009)	Anxiety	JAMA	SF-36	177	12 wks.	22.6%
Marklund (2015)	Sleep	JAMA-IM	SF-36	96	4 mos.	5.2%
Martin (2016)	Weight Loss	JAMA-IM	SF-36	220	12, 24 mos.	9.1%-13.6%
McDermott (2009)	Peripheral Artery Disease	JAMA	SF-36	156	6 mos.	19.2%
McDermott (2013)	Peripheral Artery Disease	JAMA	SF-36	194	6 mos.	8.2%
McFall (2010)	PTSD	JAMA	PHQ-9	943	3,6,9,12,15,18 mos.	12.4%-21.4%
Mohr (2012)	Depression	JAMA	PHQ-9	325	4,9, 14,18 wks.	9.2%-13.2%
Morey (2009)	Weight Control	JAMA	SF-36	641	12 mos.	12.9%
Poole (2013)	Peripheral Artery Disease	JAMA	SF-36	159	3,6 mos.	6.9%-18.2%
Rahman (2016)	Psychological Distress	JAMA	PHQ-9	346	3 mos.	12.4%
Richardson (2014)	Depression	JAMA	PHQ-9	101	6,12 mos.	18.8%-20.8%
Rollman (2009)	Depression	JAMA	SF-36	302	2,4,8 mos.	14.6%-16.6%
Stanley (2009)	Anxiety	JAMA	SF-12	134	3,6,9,12,15 mos.	14.2%-31.3%
Sullivan (2013)	Diabetes	JAMA-P	PHQ-9	2977	20,40 mos.	6.8%-11.1%
Tiwari (2010)	Initimate Partner Violence	JAMA	SF-12	200	3,9 mos.	0.0%
Wall (2014)	Intracranial Hypertension	JAMA-N	SF-36	165	6 mos.	23.6%
Walsh (2015)	Physical Rehabilitation	JAMA - IM	SF-12	240	3,6,12 mos.	17.9%-35.4%
Weisner (2016)	Addiction	JAMA-P	PHQ-9	503	6 mos.	9,.5%
Weiss (2015)	Diabetic Retinopathy Prevention	JAMA-O	PHQ-9	206	6 mos.	13.1%
Adamsen (2009)	Cancer	BMJ	SF-36	269	6 wks.	12.6%
Anguera (2016)	Depression	BMJ-I	PHQ-9	626	4,8,12 wks.	55.4%-69.8%
Arnold (2009)	Chest Pain	BMJ	SF-36	700	1 mo.	29.4%
Barnhoorn (2015)	Pain	BMJ-O	SF-36	56	3,6,9 mos.	3.6%-5.4%
Bruhn (2013)	Chronic Pain	BMJ-O	SF-12	196	6 mos.	33.7%-34.2%
Burton (2012)	Unexplained Symptoms	BMJ-O	PHQ-9	32	12 wks.	18.8%
Busse (2016)	Tibial Fractures	BMJ	SF-36	501	6,12,18,26,38,52 wks.	5.2% - 39.9%
Cartwright (2013)	Chronic Conditions	BMJ	SF-12	1573	4,12 mos.	37.3%-38.1%
Cohen (2009)	Trochanteric Pain	BMJ	SF-36	65	1,3 mos.	4.6%-46.2%
Coventry (2015)	Chronic Conditions	BMJ	PHQ-9	387	4 mos.	16.0%
Cuthbertson (2009)	Trauma	BMJ	SF-36	286	6,12 mos.	25.9%-34.6%
Dijk-De Vries (2015)	Diabetes Care	BMJ-O	264	SF-12	4,12 mos.	11.7%-15.5%

Continued on next page

Table 1 – *Continued from previous page*

Study	Indication	Journal	Endpoint	n	Follow-Up	Missing Data (%)
Dumville (2009)	Leg Ulcers	BMJ	SF-12	267	12 mos.	47.9%
El-Khoury (2015)	Fall Prevention	BMJ	SF-36	706	12,24 mos.	15.2%-19.5%
Fisher (2015)	Postpartum Mental Disorders	BMJ-O	400	SF-36	26 wks.	9.0%
Frobell (2013)	ACL Injury	BMJ	SF-36	121	5 yrs.	0.8%
Gilbody (2015)	Depression	BMJ	PHQ-9	691	4,12,24 mos.	23.9%-33.3%
Grande (2015)	Care Giving	BMJ S & PC	SF-12	681	4.5 mos.	1.8%
Griffin (2014)	Fractures	BMJ	SF-36	151	2 yrs.	23.2%
Hellum (2011)	Back Pain	BMJ	SF-36	179	1.5,3,6,12,24 mos.	7.8%-22.3%
Holzel (2016)	Depression/Back Pain	BMJ-O	PHQ-9	435	2 mos.	33.8%
Jenkinson (2009)	Knee Pain	BMJ	SF-36	389	24 mos.	18.8%
Khalafallah (2012)	Pregnancy	BMJ-O	SF-36	196	4 wks.	35.7%
Koek (2009)	Psoriasis	BMJ	SF-36	196	End of Therapy	6.1%
Lawton (2008)	Inactive Women	BMJ	SF-36	1089	12,24 mos.	7.4%-10.6%
Ly (2014)	Depression	BMJ-O	PHQ-9	81	2,6 mos.	11.1%-14.8%
Mansikkamaki (2015)	Menopause	BMJ-O	SF-36	176	0.5, 2.5, 4 yrs.	15.3% - 46.0%
McClellan (2012)	Soft Tissue Injury	BMJ-O	SF-12	372	2,8 wks.	40.1%-42.7%
Mordin (2014)	Cervical Dystonia	BMJ-O	SF-36	116	8 wks.	28.4%
Morrell (2009)	Postnatal Depression	BMJ	SF-12	4084	1.5,6,12 mos.	36.2%-58.9%
Murphy (2009)	Heart Disease	BMJ	SF-12	903	18 mos.	28.1%
Oerkild (2012)	Coronary Heart Disease	BMJ-O	SF-12	40	3,6,12 mos.	5.0%-10.0%
Patel (2009)	Osteoarthritis	BMJ	SF-36	812	4,12 mos.	38.2%-40.5%
Richards (2013)	Depression	BMJ	PHQ-9	581	4,12 mos.	13.7%-14.7%
Simkiss (2013)	Parenting Skills	BMJ-O	SF-12	286	9 mos.	19.2%
Walters (2013)	COPD	BMJ-O	SF-36	182	6,12 mos.	13.7%-15.4%
Williams (2009)	Gastrointestinal Endoscopy	BMJ	SF-36	1888	1, 30, 365 days	23.3%-32.7%
Adler (2013)	Depression	PLoS	SF-12	44	6 wks.	15.9%
Andreeva (2014)	Cardiovascular Disease	PLoS	SF-36	2501	3 yrs.	21.0%
Benda (2015)	Heart Failure	PLoS	SF-36	24	12 wks.	29.2%
Bergmann (2014)	Ischemic Heart Disease	PLoS	SF-36	213	3 mos.	15.0%
Conboy (2016)	Gulf War Illness	PLoS	SF-36	104	2,4,6 mos.	13.6% - 19.4%
Cooley (2009)	Anxiety	PLoS	SF-36	87	12 wks.	17.2%
Favrat (2014)	Iron Deficiency	PLoS	SF-12	294	56 days	3.7%
Francois (2015)	Alcohol Dependence	PLoS	SF-36	667	12,24 wks.	18.6%-39.7%
Gavi (2014)	Fibromyalgia	PLoS	SF-36	80	16 wks.	17.5%
Gine-Garriga (2013)	Chronic Conditions	PLoS	SF-12	362	3,6,12 mos.	12.7%-16.0%
Glozier (2013)	Depression, Cardiovascular Disease	PLoS	PHQ-9	562	12 wks.	4.3%
Hsu (2015)	Frozen Shoulder	PLoS	SF-36	72	6 mos.	8.3%
Kenealy (2015)	Chronic Conditions	PLoS	SF-36	171	6 mos.	11.7%
Kim (2014)	Chronic Knee Osteoarthritis	PLoS	SF-36	212	5 wks.	8.5%
Kogure (2015)	Back Pain	PLoS	SF-36	186	6 mos.	3.8%
Lambert (2016)	Leprosy	PLoS - NTD	SF-36	73	28 wks.	20.5%

Continued on next page

Table 1 – *Continued from previous page*

Study	**Indication**	**Journal**	**Endpoint**	n	**Follow-Up**	**Missing Data (%)**
Lau (2015)	Metabolic Syndrome	PLoS	SF-36	173	12 wks.	11.0%
MacPherson (2013)	Depression/Co-Morbid Pain	PLoS-M	PHQ-9	755	3,6,9,12 mos.	18.7%-24.6%
Lei (2016)	Parkinson's Disease	PLoS	SF-12	15	3 wks.	0.0%
Mead (2011)	Stroke	PLoS	SF-36	1400	64 wks.	22.9%
Merom (2016)	Falls	PLoS	SF-12	530	12 mos.	21.9%
Miyagawa (2013)	Narcolepsy	PLoS	SF-36	30	16 wks.	6.7%
Mohr (2013)	Depression	PLoS	PHQ-9	102	12 wks.	13.7%
Morgan (2013)	Depression	PLoS	PHQ-9	1736	3,6 wks.	55.5%-66.9%
Musiat (2014)	Mental Health	PLoS	PHQ-9	1047	6,12 wks.	50.3%-61.7%
Nagayama (2016)	Aging	PLoS	SF-36	54	4 mos.	18.5%
Ramly (2014)	Vitamin D Deficiency	PLoS	SF-36	192	6,12 mos.	6.8%-10.9%
Small (2014)	Postpartum Health	PLoS	SF-36	18424	2 yrs.	62.9%
Strayer (2012)	Chronic Fatigue	PLoS	SF-36	234	40 wks.	17.1%
Stuby (2015)	Distal Radius Fracture	PLoS	SF-36	29	3 mos.	0.0%
Therkelsen (2016)	Ulcerative Colitis	PLoS	SF-36	62	3 wks.	19.4%
Therkelsen (2016)	Crohn's Disease	PLoS	SF-36	76	3 wks.	34.2%
Titov (2010)	Depression	PLoS	PHQ-9	141	Post Tx., 4 mos.	17.0%-29.2%
Titov (2013)	Depression	PLoS	PHQ-9	274	3 mos.	40.1%
Titov (2014)	Depression	PLoS	PHQ-9	274	12 mos.	42.7%
van Gemert (2015)	Weight Control	PLoS	SF-36	243	4 mos.	11.1%
Younge (2015)	Heart Disease	PLoS	SF-36	324	3 mos.	20.1%
Zonneveld (2012)	Unexplained Symptoms	PLoS	SF-36	162	3 mos, 3,12 mos Post Tx.	17.9%-47.3%

Table 2: Treatment- and α-specific simulation results: Bias and mean-squared error (MSE) for the plug-in ($\mu(\widehat{P};\alpha)$) and corrected ($\widetilde{\mu}_\alpha$) estimators, for various choices of α.

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

Table 3: Treatment- and α-specific simulation results: Confidence interval coverage for (1) Wald with influence function standard errors (Wald-IF), (2) Wald with jackknife standard errors (Wald-JK), (3) equal-tailed studentized parametric bootstrap with influence function standard errors (Bootstrap-IF-ET), (4) equal-tailed studentized parametric bootstrap with jackknife standard errors (Bootstrap-JK-ET), (5) symmetric studentized parametric bootstrap with influence function standard errors (Bootstrap-IF-S) and (6) symmetric studentized parametric bootstrap with jackknife standard errors (Bootstrap-JK-S); 2000 parametric bootstraps were used.

		Placebo	Quetiapine
α	Procedure	Coverage	Coverage
-10	Wald-IF	91.5%	90.5%
	Wald-JK	95.0%	94.6%
	Bootstrap-IF-ET	94.3%	93.8%
	Bootstap-JK-ET	94.4%	93.4%
	Bootstap-IF-S	95.2%	94.6%
	Bootstap-JK-S	95.0%	94.6%
-5	Wald-IF	93.5%	92.9%
	Wald-JK	95.0%	94.8%
	Bootstrap-IF-ET	95.2%	94.6%
	Bootstap-JK-ET	94.8%	94.6%
	Bootstap-IF-S	95.4%	95.2%
	Bootstap-JK-S	95.1%	95.2%
-1	Wald-IF	93.9%	94.2%
	Wald-JK	94.9%	95.4%
	Bootstrap-IF-ET	95.1%	94.8%
	Bootstap-JK-ET	95.1%	94.6%
	Bootstap-IF-S	95.3%	96.4%
	Bootstap-JK-S	95.1%	96.3%
0	Wald-IF	93.8%	94.0%
	Wald-JK	95.0%	95.4%
	Bootstrap-IF-ET	94.6%	94.5%
	Bootstap-JK-ET	94.6%	94.6%
	Bootstap-IF-S	95.5%	96.6%
	Bootstap-JK-S	95.2%	96.7%
1	Wald-IF	93.3%	93.7%
	Wald-JK	95.1%	95.5%
	Bootstrap-IF-ET	94.6%	94.6%
	Bootstap-JK-ET	94.6%	94.6%
	Bootstap-IF-S	95.5%	96.5%
	Bootstap-JK-S	95.2%	96.5%
5	Wald-IF	90.8%	91.3%
	Wald-JK	95.3%	95.7%
	Bootstrap-IF-ET	93.2%	91.6%
	Bootstap-JK-ET	93.8%	93.0%
	Bootstap-IF-S	95.5%	95.4%
	Bootstap-JK-S	95.8%	96.4%
10	Wald-IF	85.4%	87.8%
	Wald-JK	94.9%	94.5%
	Bootstrap-IF-ET	88.2%	87.0%
	Bootstap-JK-ET	92.2%	89.7%
	Bootstap-IF-S	94.6%	93.9%
	Bootstap-JK-S	95.5%	95.1%

Figure 1: Treatment-specific (left: placebo; right: 600 mg/day Quetiapine) trajectories of mean QLESSF scores, stratified by last available measurement. Blue, brown and orange represent the trajectories of patients last seen at visits 0, 1 and 2, respectively. The number in parentheses at the end of each trajectory represents the number of associated patients.

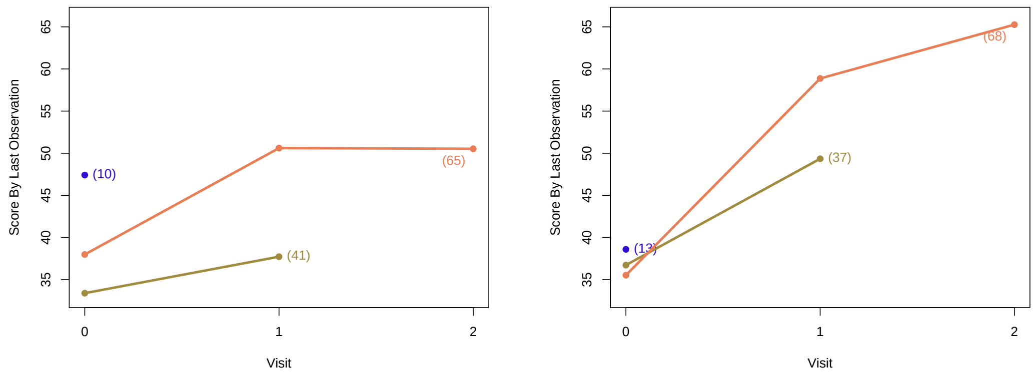


Figure 2: Schematic representation of the global sensitivity analysis framework. Circles represent modeling restrictions placed on the distribution of the observed data, with the outer circle indicating no restrictions and the inner circle indicating type (ii) restrictions. The arrows indicate a mappings from the distribution of the observed data to the true mean, which depends on the type (i) assumptions.

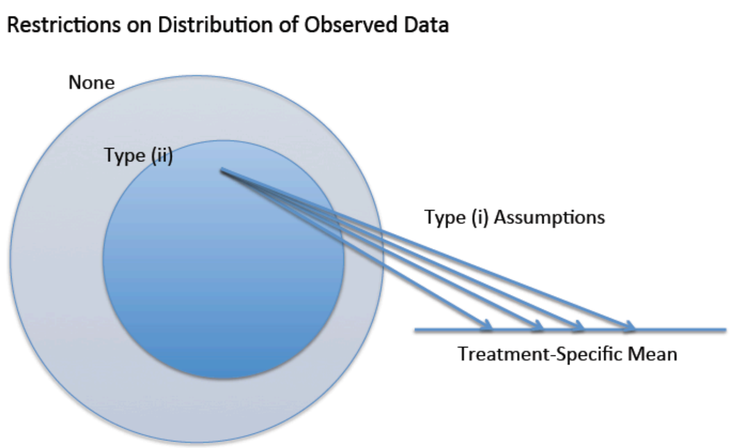


Figure 3: Treatment-specific (left: placebo; right: 600 mg/day Quentiapine) estimates (along with 95% pointwise confidence intervals) of μ^* as a function of α.

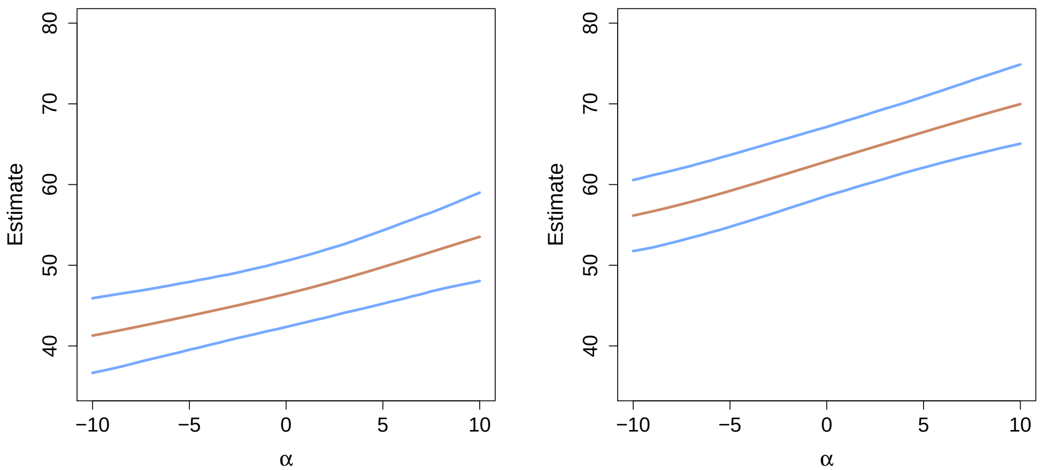


Figure 4: Treatment-specific differences between the estimated mean QLESSF at Visit 2 among non-completers and the estimated mean among completers, as a function of α.

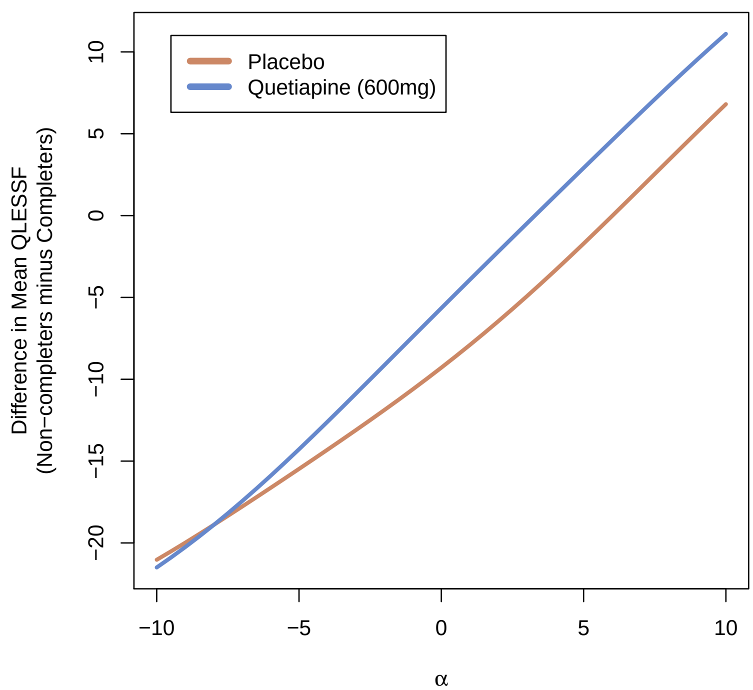


Figure 5: Contour plot of the estimated differences between mean QLESSF at Visit 2 for Quentiapine vs. placebo for various treatment-specific combinations of the sensitivity analysis parameters. The point (0,0) corresponds to the MAR assumption in both treatment arms.