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

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

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:

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

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

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

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

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

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 $P \in \mathcal{M}$,

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

where $\psi_P(O;\alpha)$ is a "derivative" of $\mu(\cdot;\alpha)$ at P and $\text{Rem}(P, P^*;\alpha)$ is a "second-order" remainder term which converges to zero as P tends to P^*. 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

$$\underbrace{\mu(\widehat{P};\alpha)}_{\text{Plug-in}} - \mu(P^*;\alpha) = \frac{1}{n}\sum_{i=1}^{n}\psi_{P^*}(O_i;\alpha) - \frac{1}{n}\sum_{i=1}^{n}\psi_{\widehat{P}}(O_i;\alpha) \tag{17}$$

$$+\frac{1}{n}\sum_{i=1}^{n}\{\psi_{\widehat{P}}(O_i;\alpha) - \psi_{P^*}(O_i;\alpha) - E^*[\psi_{\widehat{P}}(O;\alpha) - \psi_{P^*}(O;\alpha)]\} \tag{18}$$

$$+\text{Rem}(\widehat{P}, P^*;\alpha) \tag{19}$$

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., $\alpha(\text{placebo}) = 8$ and $\alpha(\text{Quetaipine}) = -8$) 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

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

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.

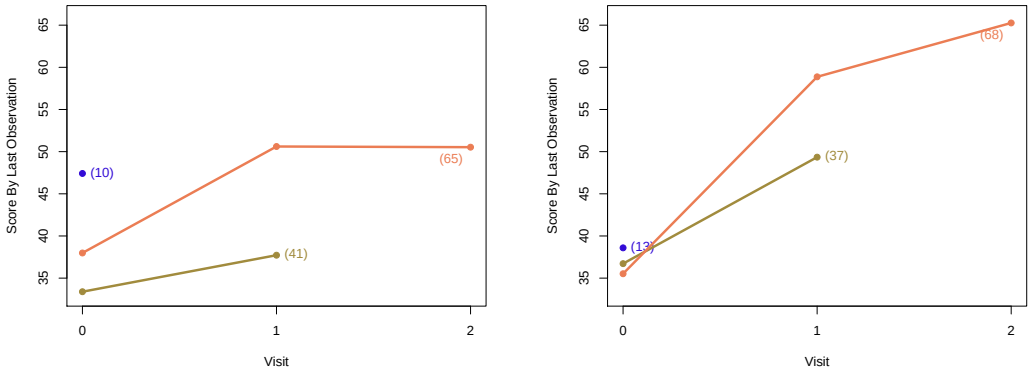

Figure 2: Schematic representation of the global sensitivity analysis framework. Circles represent modeling restrictions placed on the distribution of the observed data, with the outer circle indicating no restrictions and the inner circle indicating type (ii) restrictions. The arrows indicate a mappings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

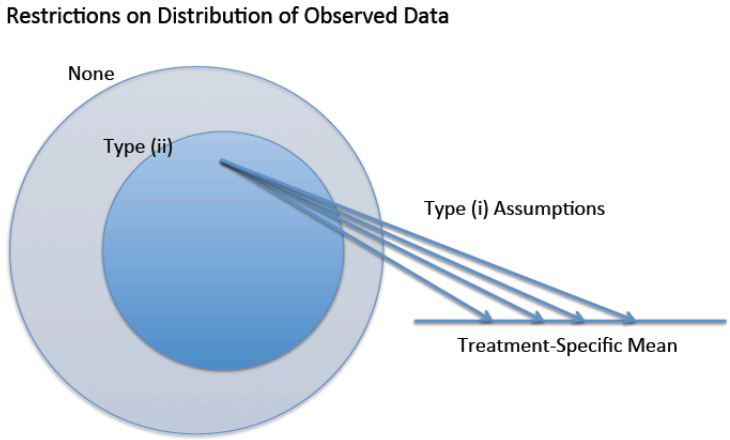

Figure 3: Treatment-specific (left: placebo; right: 600 mg/day Quentiapine) estimates (along with 95% pointwise confidence intervals) of μ^* as a function of α.

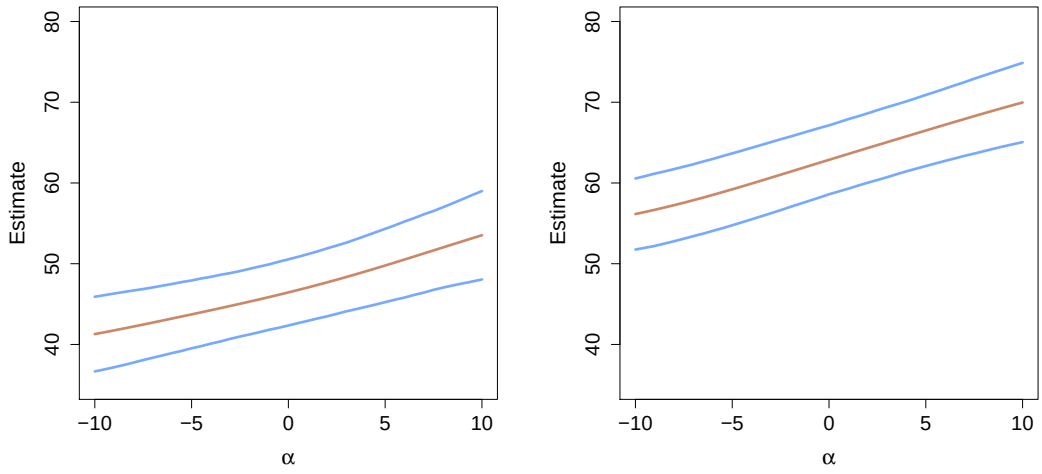

Figure 4: Treatment-specific differences between the estimated mean QLESSF at Visit 2 among non-completers and the estimated mean among completers, as a function of α.

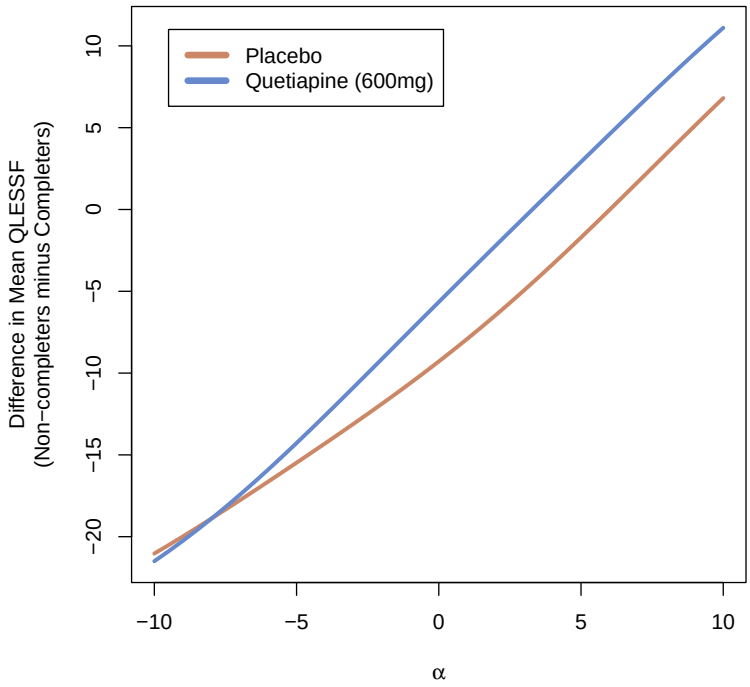

Figure 5: Contour plot of the estimated differences between mean QLESSF at Visit 2 for Quentiapine vs. placebo for various treatment-specific combinations of the sensitivity analysis parameters. The point (0,0) corresponds to the MAR assumption in both treatment arms.