Defining, Evaluating, and Removing Bias Induced by Linear Imputation in Longitudinal Clinical Trials with MNAR Missing Data

Ronald W. Helms[a,b], **Laura Helms-Reece**[a], **Russell W. Helms**[a], and **Mary W. Helms**[a]

[a] Rho, Inc., Chapel Hill

[b] Department of Biostatistics, University of North Carolina

Abstract

M*issing not at random* (MNAR) post-dropout missing data from a longitudinal clinical trial result in the collection of "biased data", which leads to biased estimators and tests of corrupted hypotheses. In a full rank linear model analysis the model equation, $E[Y] = X\beta$, leads to the definition of the primary parameter $\beta = (X'X)^{-1}X'E[Y]$, and the definition of linear secondary parameters of the form $\theta = L\beta = L(X'X)^{-1}X'E[Y]$, including for example, a parameter representing a "treatment effect". These parameters depend explicitly on $E[Y]$, which raises the questions: what is $E[Y]$ when some elements of the incomplete random vector Y are not observed and MNAR, or when such a Y is "completed" via imputation? We develop a rigorous, readily interpretable definition of $E[Y]$ in this context that leads directly to definitions of β, $\text{Bias}\left(\hat{\beta}\right) = E\left[\hat{\beta}\right] - \beta$, $\text{Bias}\left(\hat{\theta}\right) = E\left[\hat{\theta}\right] - L\beta$, and the extent of hypothesis corruption. These definitions provide a basis for evaluating, comparing, and removing biases induced by various linear imputation methods for MNAR incomplete data from longitudinal clinical trials. Linear imputation methods use earlier data from a subject to impute values for post-dropout missing values and include "Last Observation Carried Forward" (LOCF) and "Baseline Observation Carried Forward" (BOCF), among others. We illustrate the methods of evaluating, comparing, and removing biases and the effects of testing corresponding corrupted hypotheses via a hypothetical, but very realistic longitudinal analgesic clinical trial.

Keywords

imputation; bias; clinical trial; longitudinal; dropout; corrupted hypothesis; parameter definition; MAR; missing at random; MNAR; missing not at random; BOCF; LOCF

1. INTRODUCTION

Many clinical trials utilize multiple, scheduled, longitudinal outcome assessments, often including some or all of: screening period evaluations, baseline evaluations, and post-randomization or "on treatment" evaluations. Typically some subjects withdraw from the study prior to completing all scheduled evaluations; some of these "dropouts" withdraw for reasons that are not stochastically independent of their assessment values, such as lack of efficacy ("LOE"), side effects of a treatment ("intolerability"), or similar. Missing data from post-dropout scheduled evaluations are generally regarded as "missing not at random"

Corresponding author contact information: Ronald W. Helms, Ph.D., Rho, Inc., 6330 Quadrangle Drive, Chapel Hill NC 27517, Ron_Helms@RhoWorld.com, 919-408-8000 ext 6222.

("MNAR", Little and Rubin, 2002, Chapter 15). When likelihood or quasi-likelihood estimation procedures are applied to MNAR data, as might happen, for example when a mixed (random effects) model is used for analysis, the estimators are generally biased and the direction and magnitude of the MNAR-induced bias may be difficult to assess (Little and Rubin, 2002). "Selection models" and "pattern mixture models" have been posed to cope with MNAR-induced difficulties, but the practical application of these methods typically require sample sizes that far exceed the number of subjects in a typical clinical trial.

Prior to the ready availability of mixed model methods statisticians often used a univariate (one value per subject) change-from-baseline primary outcome variable that was the evaluation of the primary outcome at the final scheduled evaluation occasion minus the corresponding value at baseline (pre-randomization, pre-treatment). Although multiple post-randomization evaluations were available, these were generally correlated and before mixed models were available the analysis was problematic, especially when the data were incomplete. Dropouts led to a quandary. They could not be ignored as that would violate the important intention-to-treat principle. Dropouts could not be included directly in the analysis – no data were available from the final scheduled evaluation occasion. In desperation statisticians created data: they imputed data for the final scheduled evaluation.

Demonstrating that imputation introduces bias into the data seems to be easy, as we will illustrate in the next section. The illustration will also make the point that the bias "demonstration" is based on assumptions about expected responses of subjects after they drop out, assumptions that are not universally accepted. One FDA reviewer humorously and effectively articulated this point by referring to the assumptions as "… what might have happened in some parallel universe, but not in this universe" (Helms, 2009a). Nonetheless, in an attempt to be conservative, some statisticians, including some in regulatory agencies, have sometimes insisted that researchers use imputation methods that appeared to bias the treatment effect toward the null hypothesis (usually: smaller treatment effect), sometimes drastically so. In some cases simulations have demonstrated that the imputation-induced bias from the specified method appeared to be substantially larger than the actual effect of any known drug, while other imputation methods could have introduced less bias (Helms, 2009b). The tools presented in this paper allow direct calculation of imputation-induced bias, eliminating the need for simulations.

Section 2 of this paper sets the stage for subsequent development by introducing a hypothetical, realistic example longitudinal clinical trial afflicted with MNAR post-dropout missing data. Section 3 replaces the assumptions from "some parallel universe" with mathematically sound derivations of expected values of available data E[Y], in the context of post-dropout, MNAR missing data. Section 4 introduces a typical mixed model for analysis of a continuous outcome from a longitudinal clinical trial, gives a definition of the primary parameter, β, as a linear transformation of E[Y], and defines a general linear secondary parameter and a special case secondary parameter representing a treatment effect. Bias is then defined in terms of these definitions and a straightforward procedure is shown for eliminating the bias, leading to estimators without imputation-induced bias and to corresponding uncorrupted hypotheses. We will discuss several types of linear imputation in this context and demonstrate the evaluation of bias introduced by each. A realistic example illustrates the use of the methods.

2. HYPOTHETICAL ILLUSTRATION OF IMPUTATION-INDUCED BIAS

We use a hypothetical – but very realistic – example to introduce the "real world" nature of imputation-induced bias.

2.1. A Realistic Clinical Trial Example

Consider a superiority clinical trial conducted to demonstrate that an analgesic, the "active treatment," is superior to placebo for treatment of chronic pain. (The type of pain, e.g., from osteoarthritis, from diabetic neuropathy, etc., would be specified in the protocol; we do not specify the pain type here.) The trial has typical characteristics: multicenter, randomized, placebo-controlled, doubly-masked, and longitudinal. The primary outcome is a pain score. A score of 0 represents "no pain" and a score of 10 represents "the most horrible pain you can imagine". Intermediate scores represent intermediate levels of pain. (We omit numerous details that are not important here.)

Medical histories are used to screen potential subjects. Selected, willing patients give informed consent, at which point they are enrolled and become subjects, and the clinical staff obtains baseline information (demographics, medical history, etc.), including a baseline pain evaluation score.

A baseline pain score of at least 5 is an inclusion criterion; subjects who meet this and other inclusion/exclusion criteria proceed to randomization; others are withdrawn. Immediately after the baseline evaluation, eligible subjects are randomly assigned to placebo or active treatment. The clinical trial material along with appropriate instructions are given to the subject, and the subject is scheduled to return for 12 consecutive weekly post-randomization, on-treatment clinic visits (visits 2–13) during which pain score evaluations are made. The intention-to-treat group is the set of all randomized subjects who received at least one dose of clinical trial material and completed at least one post-randomization pain evaluation.

The statistical mixed model will treat all pain scores, including baseline, as random variables. The notation[1] is

Y_{ctsv} = pain score at visit v, from subject s who was enrolled in clinic c and assigned to treatment t (t="C" for placebo control, t="A" for active treatment).

This clinical trial has:

c = 1, 2, …, 20 = C = number of clinics;

s=1, 2, …, N = number of subjects;

N_C = 148 subjects in the control group and

N_A = 151 subjects in the active treatment group;

v=1, 2, …, 13=V visits;

v=1 for the baseline visit.

The population mean parameters are

$\mu_{tv} = E[Y_{stv}]$, subject to restrictions that clinics are "blocks" in experimental design terminology, that is, the restrictions are that clinic main effects are random and all interactions involving clinics are zero.

The population treatment effect parameter is

---

[1]Typographic notation: A random variable is represented by an italicized capital letter (e.g., *Y*) and its realized value is represented by the corresponding lower case non-italicized letter (e.g., y). A non-italicized Latin letter indicates a known constant (e.g., "V"); a non-italicized Greek letter indicates an unknown constant (e.g., μ). **Bold** indicates a vector (e.g., β, *Y*) or matrix; non-bold indicates a scalar. A lower case subscript (e.g., "v") is an index variable; the same letter in upper case is its maximum value (e.g., v=1, 2, …, 13=V); "s" and "N" are exceptions (s=1, 2, …, N).

$$\tau = (\mu_{A,13} - \mu_{A,1}) - (\mu_{C,13} - \mu_{C,1}).$$

In the definition of τ, $(\mu_{A,13} - \mu_{A,1})$ is the (population) mean pain score change-from-baseline to final scheduled visit for the active treatment group and $(\mu_{C,13} - \mu_{C,1})$ is the corresponding (population) mean pain score change-from-baseline for the placebo treatment group. The primary null hypothesis is H_0: $\tau \geq 0$ and the alternative hypothesis is H_A: $\tau < 0$. (An effective treatment will lead to a greater reduction in mean pain score than placebo.) We will use this hypothetical-but-realistic example clinical trial throughout the paper.

2.2. The Parallel Universe Issue

When there are MNAR dropouts the model equation, which implies some assumptions, namely $\mu_{tv} = E[Y_{stv}]$, has been controversial. (See the "parallel universe" comment above). The issue can be briefly summarized as follows. Consider a specific dropout subject, say, subject s=1 who received treatment C and who dropped out after visit v=3 for lack of efficacy. We assume all data from subsequent scheduled visits are MNAR missing. The random variable for visit 4, $Y_{1,C,1,4}$, was never realized. What does $E[Y_{1,C,1,4}]$ mean? That is, what is the expected value of a random variable that could never be realized because the subject dropped out before the random variable could be realized? This topic will be addressed in a subsequent section.

2.3. LOCF Imputation Illustration

In this type of study a subject who receives treatment (active or placebo) will typically experience a decline in pain scores over several weeks, typically stabilizing around a value somewhat below the baseline value, as illustrated in Figure 1, which displays a situation in which the active treatment is more effective in reducing pain than placebo. The numeric values used to generate Figure 1 are shown in Table 1.

Consider a scenario in which placebo subjects tend to drop out earlier than active-treatment subjects. Some placebo subjects find the treatment to be less effective than what they had been taking before entering the study and decide to withdraw, often relatively early in the study, for lack of efficacy (LOE). In Figure 1, hypothetical control "Subject C" drops out after visit 3. Active treatment subjects may tend to experience intolerable side effects and drop out somewhat later than LOE subjects, illustrated in Figure 1 by Subject A2, who dropped out after visit 9. From baseline to the point of dropout each of these two hypothetical subjects tended to track slightly above their treatment group's population means.

The open circles and squares in the Figure illustrate how last-observation-carried-forward ("LOCF") imputation would work for the hypothetical placebo and active-treatment subjects, respectively. These two cases illustrate how, but do not demonstrate that, in this case LOCF imputation introduces a bias away from the null hypothesis. When placebo subjects tend to drop out earlier, they have experienced less placebo effect, i.e., less pain relief, than placebo subjects who complete the study. LOCF tends to impute relatively high pain scores for such subjects, well above the population means (large solid dots). An active-treatment subject who drops out later (for intolerability) tends to have experienced a decline in pain scores before dropout. As illustrated by the hypothetical active-treatment subject A2 in Figure 1, LOCF imputation tends to impute lower pain score values. These two subjects' LOCF imputation values illustrate a larger "treatment effect" (dotted vertical arrow) than the difference between the population means at the last visit (solid vertical arrow), a bias toward

larger treatment effect, which is a bias away from the null hypothesis. If LOCF imputation were used in this situation the Type I error rate might be inflated.

This subject A2 scenario also illustrates what some FDA reviewers call "imputing a good result [a big improvement in pain score, compared to baseline] to a bad outcome" [dropout for intolerability], which they find objectionable. The objection has face validity even though it is not based on statistical principles. On the other hand, minimizing bias is a widely accepted statistical principle, which one might use effectively to argue against some types of imputation in such a case.

3. EXPECTED VALUES AND DEFINITIONS OF PARAMETERS

3.1. Expected Values of Informatively Censored Data Variables

Our goal is to resolve the "parallel universe" issue by deriving interpretable definitions of $E[Y]$ subject to MNAR post-dropout missing data. We will explain the derivations via a device that Einstein and others called a "thought experiment" (or "Gedankenversuch"; see, e.g., Brown, 1993).

Although one can "see" the LOCF-induced bias in Figure 1 we will need some notation and derivations to manage these issues objectively. We will use a variety of types of variables and parameters and will use the following typographical conventions.

Following notation in Scheffè's 1959 classic text, we use mnemonic letters for index variables, e.g, v for "visit". An index variable and its upper limit are the same Latin letter in lower case and upper case, respectively, e.g., for visit, $v = 1, 2, \ldots, V$. Other typographical conventions are described in a footnote.

3.1.1. Thought Experiment 1: Complete Longitudinal Data—We conduct a thought experiment, different from but related to the clinical trial, that produces realizations y_{1v}, $v = 1, 2, \ldots, V$, of V longitudinal random variables Y_{1v}, $v = 1, 2, \ldots, V$, representing values of a continuous "score" from one subject on a fixed, finite scale (e.g., $0 \leq Y_{1v} \leq 10$), with no missing data. (The notation in this section differs from the notation for the hypothetical clinical trial. The first subscript, 1, specifies that these variables are from Thought Experiment 1.) The y_{1v} are not just realized values of the Y_{1v}, but the realized values have also been rounded to a small number (e.g., 1 or 2) of significant digits. We define the vectors

$$\mathbf{y}_1 = [y_{11}, y_{12}, \ldots, y_{1v}]' \text{ and }$$
$$\mathbf{Y}_1 = [Y_{11}, Y_{12}, \ldots, Y_{1v}]'.$$

We let

$$\boldsymbol{\mu}_1 = (\mu_{11}, \mu_{12}, \ldots, \mu_{1v})' = E[\mathbf{Y}_1] \text{ and}$$
$$\textstyle\sum_1 = [\sigma_{1,ij}] = V[\mathbf{Y}_1],$$

all elements of which are assumed to be finite, a reasonable assumption here and for most clinical trial outcome variables.

Although most of us have an intuitive understanding of $\boldsymbol{\mu}_1 = E[\mathbf{Y}_1]$, it is useful to develop a frequentist interpretation of $\boldsymbol{\mu}_1$ for use in our subsequent Thought Experiments. We replicate $\mathbf{Y}_1$ independently R times, creating a sequence of random vectors

$$\boldsymbol{Y}_1^{(r)} = (Y_{11}^{(r)}, Y_{12}^{(r)}, \ldots, Y_{1v}^{(r)})', \; r=1,2,\ldots,R.$$

and realizations

$$\boldsymbol{y}_1^{(r)} = (y_{11}^{(r)}, y_{12}^{(r)}, \ldots, y_{1v}^{(r)})', \; r=1,2,\ldots,R.$$

After the R-th replication, we compute the mean vector,

$$\overline{\boldsymbol{Y}}_1^{(r)} = \left(\frac{1}{r}\right) \sum_{r=1}^{r} \boldsymbol{Y}_1^{(r)}.$$

For each v, $V[Y_{1v}^{(r)}] = \sigma_{1vv}$ is a finite constant for every r. Thus

$$\sum_{r=1}^{\infty} r^{-2} Var(Y_{1v}^{(r)}) = \sigma_{1,vv} \pi^2/6 < \infty.$$

By the Kolmogorov strong law of large numbers (Sen and Singer, 1994, Theorem 2.3.10) for each v = 1,2, …, V, $\overline{Y}_{1v}^{(r)} \to E[Y_{1v}] = \mu_{1v}$ almost surely (a.s.) and, in vector form, $\overline{\boldsymbol{Y}}_1^{(r)} \to E[\boldsymbol{Y}_1] = \boldsymbol{\mu}_1$ a.s.

Thus, the frequentist interpretation of $\boldsymbol{\mu}_1$ in this complete-data thought experiment: if we were to repeat the original one-subject "experiment" many times, e.g., enroll "many" subjects with the same mean vector and covariance matrix, the sample mean vector would converge strongly to the population mean vector, $\boldsymbol{\mu}_1$.

3.1.2. Thought Experiment 2: MAR Post-Dropout Incomplete Longitudinal Data

—We repeat Thought Experiment 1 with the following modification. When $\boldsymbol{Y}_{1v}^{(r)}$ is realized its value, $\boldsymbol{y}_{1v}^{(r)}$, is not revealed to humans but rather both are submitted to "Missos", the mythological goddess of missing data. Missos uses the following procedure to simulate the deletion of data when a subject drops out of a study. Missos uses a random number generator that generates a random last visit number, $V^{(r)} \in \{1, 2, …, V\}$. $V^{(r)}$ is stochastically independent of $\boldsymbol{Y}_1^{(r)}$ and functionally independent of $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$, although the parameters of the distribution of $V^{(r)}$ may depend upon $\boldsymbol{\mu}_1$ and/or $\boldsymbol{\Sigma}_1$. To avoid pathological cases of no interest here we assume $0 < Pr[V^{(r)} = v] < 1$, v = 1, 2, …, V. The realized value of the random variable $V^{(r)}$ is $v^{(r)}$, r = 1, 2, …. Missos then creates $\boldsymbol{Y}_2^{(r)}$ and $\boldsymbol{y}_2^{(r)}$ by deleting all $Y_{1v}^{(r)}$ with v > $v^{(r)}$ and deleting all $y_{1v}^{(r)}$ with v > $v^{(r)}$, respectively, and reports the remaining elements of $y_{1v}^{(r)}$ to the experimenter as $\boldsymbol{y}_2^{(r)}$. Clearly the missing values are Missing at Random ("MAR").

The values that Missos reveals to the experimenter are the realizations of the conditional random variables

$$Y_{2v}^{(r)} = (Y_{1v}^{(r)} | v \leq v^{(r)}), \; v=1, 2, \ldots, v^{(r)}.$$

We define the corresponding random vector with $v^{(r)}$ elements,

$$\boldsymbol{Y}_2^{(r)} = (Y_{21}^{(r)}, Y_{22}^{(r)}, \ldots, Y_{2,v^{(r)}}^{(r)}).$$

$\mathbf{y}_2^{(r)}$ is the realization of $\boldsymbol{Y}_2^{(r)}$, v = 1, 2, ..., $v^{(r)}$.

3.1.3. Reminder about Conditional Random Variables—To facilitate understanding of these random variables please recall the following from the multivariate normal distribution and a subsequent conditional normal distribution. Let $\boldsymbol{Z} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where p > 1 and the mean vector and covariance matrix, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are not related to the similarly-named parameters above. Consider the familiar conditional random variable that is the first element of $\boldsymbol{Z}$ conditional on the observed values of all the other elements. The "full name" of the conditional random variable is $(Z_1 | \boldsymbol{Z}_2 = \mathbf{z}_2)$ and the typical abbreviated name is $(Z_1 | \mathbf{z}_2)$. The distribution of the conditional random variable is

$$(Z_1 | \boldsymbol{Z}_2 = \mathbf{z}_2)$$
$$\sim N_1[\mu_1 + \sum\nolimits_{12} \sum\nolimits_{22}^{-1} (\mathbf{z}_2 - \boldsymbol{\mu}_2), \sigma_{11} - \sum\nolimits_{12} \sum\nolimits_{22}^{-1} \sum\nolimits_{21}].$$

Now, a priori we all know that $\Pr[\boldsymbol{Z}_2 = \mathbf{z}_2] = 0$, i.e., realizing $\boldsymbol{Z}_2 = \mathbf{z}_2$ is essentially an impossibility, but virtually all of us have no difficulty using this conditional random variable, $(Z_1 | \boldsymbol{Z}_2 = \mathbf{z}_2)$, as the dependent variable in a multiple regression analysis, regressing the dependent variable Z_1 on the vector $\mathbf{z}_2$ of realized values of the elements of $\boldsymbol{Z}_2$. In this case we simply utilize the conditional random variable; we do not overly concern ourselves with hypothetical conditional random variables that would have been observed if a different value of $\boldsymbol{Z}_2$ had been realized.

3.1.4. Thought Experiment 2, Continued—Back in Thought Experiment 2, we have

the conditional random variables $Y_{2v}^{(r)} = \left(Y_{1v}^{(r)} | v \leq v^{(r)} \right)$, r = 1, 2, ... These conditional random variables are somewhat simpler to grasp than the multivariate normal because here the conditioning event has $\Pr[V \quad v^{(r)}] > 0$ (i.e., not an "impossibility"). Nonetheless, we cannot simply ignore the hypothetical random variables that would have been observed if a larger value of $v^{(r)}$ had been observed because we must explicitly exclude the non-realized missing values from analyses. This becomes even more important when, as in studies of analgesics for chronic pain for example, the hypothesis of primary interest involves mean pain at the last few scheduled visits, which tend to be the visits with the most missing

values. Nonetheless, we have demonstrated that in the MAR case $E[Y_{2v}^{(r)}]$ exists for all v = 1, 2, ..., V and r=1, 2, ...

Because $V^{(r)}$ is independent of $Y_{1v}^{(r)}$ and the missing data are MAR, we assume the existence of, define, and evaluate

$$\mu_{2v} = E\left[Y_{2v}^{(r)}\right] = E\left[\left(Y_{1v}^{(r)}|v \le v^{(r)}\right)\right] = E\left[Y_{1v}^{(r)}\right] = \mu_{1v}, \; v=1,2,\ldots,v, \; r=1,2,\ldots$$

By analogy with Thought Experiment 1, for each v=1, 2, …, V, we compute the mean of all nonmissing random variables through the R-th iteration: $\overline{Y}_{2v}^{(r)} = Mean(Y_{2v}^{(r)}, \; r=1,2,\ldots,r)$ where the mean is computed over the available (non-missing) values of $Y_v^{(r)}$, and $\overline{Y}_2^{(r)} = (\overline{Y}_{21}^{(r)}, \overline{Y}_{22}^{(r)}, \ldots, y_{2V}^{(r)})'$. Some of the means, $\overline{Y}_{2v}^{(r)}$, may be missing when R is not large, but when R is sufficiently large our assumptions guarantee that all such means will be available.

As before, under these circumstances by the Kolmogorov strong law of large numbers (Sen and Singer, 1994, Theorem 2.3.10 applied to each element of the vector) we conclude that $\overline{Y}_2^{(r)} \to \mu_1 \; a. \, s.$

It would be nice to be able to have a series of random vectors $Y_2^{(r)}$ so that we could write $\overline{Y}_2^{(r)} = \left(\frac{1}{r}\right)\sum_1^r Y_2^{(r)}$. Unfortunately, because the number of elements in $Y_2^{(r)} = (Y_{21}^{(r)}, Y_{21}^{(r)}, \ldots, Y_{2,v^{(r)}}^{(r)})$ varies with r, such a simple definition of $\overline{Y}_2^{(R)}$ does not exist. However, for each value of v = 1, 2, …, V, $\overline{Y}_{2v}^{(r)} \to \mu_{1v} \; a. \, s.$, which ensures that the vector sequence $\overline{Y}_2^{(r)} \to \mu_1 \; a. \, s.$

3.1.5. Thought Experiment 3: MNAR Post-Dropout Incomplete Longitudinal Data—We repeat Thought Experiment 2 with the following modification: Missos uses a specified algorithm for generating $V^{(r)}$ that leads to MNAR missing data. The algorithm is the same for all replications, i.e., for all values of r, but of course, $V^{(r)}$ will typically vary from one rep to another. The algorithm may include the use of random variables that are either stochastically independent of all $Y^{(r)}$, as in Thought Experiment 2, or stochastically dependent on some or all elements of $Y^{(r)}$. However, $Y^{(r)}$ is independent of $Y^{(q)}$. for r   q and $V^{(r)}$ is independent of $V^{(q)}$ for r   q. The algorithm could involve functions of the elements of $Y^{(r)}$, e.g., one part of the algorithm might be: $V^{(r)}$ is either V, or the smallest value of v such that $y_{1v}^{(r)} - y_{11}^{(r)} > 2$, whichever is smaller. (The example stems from a clinical trial of an analgesic in which $y_1^{(r)}$ is a baseline pain score value, $y_v^{(r)}$ is a post-baseline pain score (v>1), and the subject drops out when the level of pain increases by at least 2 pain score units.) We retain the assumption that $0 < Pr[V^{(r)} = v] < 1$, v = 1, 2, …, V, to avoid pathological cases. Our intent is to generate $V^{(r)}$ so that the missing values (when $V^{(r)} < V$) are Missing Not at Random (MNAR).

This process produces the conditional random variables

$$Y_{3v}^{(r)} = (Y_{1v}^{(r)}|v \le v^{(r)}), \; v=1,2,\ldots,v^{(r)}$$

and the corresponding random vector with $v^{(r)}$ elements,

$$\boldsymbol{Y}_3^{(r)} = (Y_{31}^{(r)}, Y_{32}^{(r)}, \ldots, Y_{3,v^{(r)}}^{(r)}).$$

The realizations of the random variables are $y_{3v}^{(r)}$, v = 1, 2, …, v$^{(r)}$, and the realization of the random vector is $\mathbf{y}_3^{(r)} = (y_{31}^{(r)}, y_{32}^{(r)}, \ldots, y_{3,v^{(r)}}^{(r)})'$.

As before, after R reps, for each v we compute the mean of the available scalar random variables, $\overline{Y}_{3v}^{(r)} = Mean(Y_{3v}^{(r)}, r = 1, 2, \ldots, r)$, where the mean is computed over the available (non-missing) values of $Y_{3v}^{(r)}$, and we have the mean vector, $\overline{\boldsymbol{Y}}_3^{(r)} = (\overline{Y}_{31}^{(r)}, \overline{Y}_{32}^{(r)}, \ldots, \overline{Y}_{3v}^{(r)})'$. Again, when R is sufficiently large our assumptions guarantee that all elements of $\overline{\boldsymbol{Y}}_3^{(r)}$ will be available. As the variances of all the random variables are all finite and random variables in separate reps are independent, the Kolmogorov strong law of large numbers applies to each series $Y_{3v}^{(r)}$, v = 1, 2, …, V, and we can conclude that

$$\overline{\boldsymbol{Y}}_3^{(r)} \rightarrow \boldsymbol{\mu}_3 = \boldsymbol{\mu}_1 + \boldsymbol{\lambda}_3 \ a.s.$$

where

$$\lambda_{3v} = \left(\lim_{R \rightarrow \infty} Y_{3v}^{(r)}\right) - \mu_{1v} = \mu_{3v} - \mu_{1v}, \ v = 1, 2, \ldots, v,$$

is the bias at visit v stemming from the MNAR process.

As in Thought Experiment 2 it would be nice to be able to have a series of random vectors $\boldsymbol{Y}_3^{(r)}$ so that we could write $\overline{\boldsymbol{Y}}_3^{(r)} = \left(\frac{1}{r}\right) \sum_1^r \boldsymbol{Y}_3^{(r)}$. As before, because the number of elements in $\boldsymbol{Y}_3^{(r)} = (Y_{31}^{(r)}, Y_{32}^{(r)}, \ldots, Y_{3,v^{(r)}}^{(r)})$ varies with r, such a simple definition of $\overline{\boldsymbol{Y}}_3^{(r)}$ does not exist. However, for each value of v = 1, 2, …, V, $\overline{Y}_{3v}^{(r)} \rightarrow \mu_{1v} + \lambda_{3v} \ a. s.$, which ensures that the vector sequence $\overline{\boldsymbol{Y}}_3^{(r)} \rightarrow \boldsymbol{\mu}_1 + \boldsymbol{\lambda}_3 \ a. s.$

The point of Thought Experiment 3 is that even in the case of MNAR dropouts, $E\left[Y_{3v}^{(\mathbf{r})}\right]$ exists for each v = 1, 2, …, V and has essentially the same frequentist interpretation (although perhaps not the same value) as in the complete case. This particular set of $E\left[Y_{3v}^{(\mathbf{r})}\right]$ values exists and has a sensible interpretation in this universe, not just in some "parallel universe".

Statisticians may disagree about whether the equation $E\left[\overline{Y}_{3v}^{(r)}\right] = \mu_{3v} = \mu_{1v} + \lambda_{3v}$ is meaningful or useful because, as a practical matter, $\boldsymbol{\lambda}_3$ can not be estimated in most clinical trials, which possibly makes the argument irrelevant. However, we have answered an important question: even when there are MNAR dropouts, the random variables whose values may be unavailable because of an MNAR missing value process, are from distributions that have actual means.

3.2. Covariance Matrix

Mixed model analyses of longitudinal data require that the covariances exist. Having demonstrated the existence of the expected values we can turn to the covariances of the elements of $\boldsymbol{Y}_3^{(r)}$, which, when the equations are meaningful, are defined by:

$$\sigma_{3,\mathrm{uv}}^{(r)} = E\left[\left(Y_{3,u}^{(r)} - \mu_{3u}\right)\left(Y_{3,v}^{(r)} - \mu_{3v}\right)\right].$$

We assume, as is reasonable for many clinical trial outcome variables, that the expectation in the preceding equation exists and is finite for each rep, r, and u ≤ v$^{(r)}$ and v ≤ v$^{(r)}$. For each u=1, 2, …, v; v=1, 2, …, v$^{(r)}$, let

$$\widehat{\sigma}_{3,\mathrm{uv}}^{(r)} = \left(Y_{3,u}^{(r)} - \mu_{3u}\right)\left(Y_{3,v}^{(r)} - \mu_{3v}\right)$$

and

$$\overline{\sigma}_{3,\mathrm{uv}}^{(r)} = \mathrm{Mean}\left(\widehat{\sigma}_{3,\mathrm{uv}}^{(r)}, r=1, 2, \ldots, r\right)$$

where the mean is taken over all existing values of $\widehat{\sigma}_{3,\mathrm{uv}}^{(r)}$, u=1, 2, …, v; v=1, 2, …, V. Assuming the fourth moments of $Y_{3v}^{(r)}$ exist, which is reasonable for response variables in many clinical trials, the Kolmogorov strong law of large numbers applies to each of the covariances, $\widehat{\sigma}_{3,\mathrm{uv}}^{(r)}$, u = 1, 2, …, v; v = 1, 2, …, v$^{(r)}$, and we can conclude that $\overline{\sigma}_{3,\mathrm{uv}}^{(r)} \to \sigma_{3,\mathrm{uv}}$ a. s., u = 1, 2, …, v; v = 1, 2, …, v$^{(r)}$, which defines $\sigma_{3,\mathrm{uv}}$. We define $\sigma_{3,\mathrm{vu}} = \sigma_{3,\mathrm{uv}}$ and the symmetric V × V matrix $\Sigma_3 = [\sigma_{3,\mathrm{uv}}]$. Unfortunately, the equation $\sum_3 = V[\boldsymbol{Y}_3^{(r)}]$ is not meaningful because the number of elements in $\boldsymbol{Y}_3^{(r)}$ varies with r. Rather, we use the interpretation $\sigma_{3,\mathrm{uv}} = \mathrm{Cov}\,(Y_{3,\mathrm{u}}, Y_{3,\mathrm{v}})$ when both are realized.

The conclusion is that under assumptions that are reasonable for many clinical trials the covariances of the conditional random variables Y_{3v}, v = 1, 2, …, V exist and have the usual interpretations.

3.3. Extension to Covariates

In the developments above we used constant parameters) μ_1, λ_1 and Σ_1. The development can be extended, in a straightforward manner, to situations in which the elements of these parameters are linear functions $\mu_{1v}(\boldsymbol{x}_v) = \boldsymbol{x}_v\boldsymbol{\beta}$, $\lambda_{1v}(\boldsymbol{x}_v) = \boldsymbol{x}_v\boldsymbol{\xi}$, of a vector, $\boldsymbol{x}_v$, of covariates. The covariance matrix Σ_1 can also be modeled, as in a mixed model; the only requirement for extending the Kolmogorov strong law of large numbers is that the first four moments of the distribution of $\boldsymbol{Y}_1$ are finite, which is reasonable for the outcome variables in many clinical trials.

Sen and Singer (1994), in comments following the Khintchine strong law of large numbers (p. 71) describe how to extend the results above to the vector situation, which includes mixed models.

4. EVALUATION OF IMPUTATION-INDUCED ESTIMATOR BIAS

Now that we have established the existence of the first two moments of the conditional random variables that are realized in a clinical trial with MNAR post-dropout missing data, we turn to investigating imputation-induced bias. To simplify the exposition and notation we consider the case in which all missing data are post-dropout. In practice the results and methods are easily applied to datasets with "intermittent" missing data.

We now need a different, but related, set of notation for data from an actual clinical trial, in contrast to data from the thought experiments.

4.1. Data Notation

We use the following variables to index (identify) subjects, treatments and visits:

s indexes subjects, i.e., s denotes a subject's number, s = 1, 2, …, N. Subjects are numbered without regard to treatment group, clinic, etc.

t denotes a treatment group, t = "A" for the "active" treatment, t = "C" for the control. Note that after randomization, given s the value of t is redundant. That is, one can look at a table of random assignments to determine the treatment group to which subject s was assigned. Nonetheless it is often convenient to use t as a subscript.

v is a variable that indexes clinic visits or evaluation occasions. In our example v = 1, 2, …, V, visit 1 is a baseline visit, and visit V is the last scheduled visit for all subjects.

We use the following notation for dropout information.

V_s = the number of the last visit completed by subject s. V_s is a random variable with values in {1, 2, …, V}. V_s may depend upon (be not independent of) one or more primary outcomes.

v_s = the number of the last visit completed by subject s. v_s is the realized (known, constant) value of the random variable V_{st}; with values in {1, 2, …, V}.

Following the notions in the thought experiments we hypothesize a set of underlying random outcome variables, some of which are realized and some of which are scheduled to be realized but follow a dropout event and are never realized.

Let $Y_{1,st}$ denote the complete V × 1 vector of underlying outcome random variables at all visits, for subject s, who received treatment t. We assume that at least the first four moments of $Y_{1,st}$ are all finite, a reasonable assumption for many clinical trial outcome variables. This assumption is sufficient for application of the Kolmogorov strong law of large numbers to imply the existence of the means, and, variances and covariances of the realized, pre-dropout conditional random variables.

We have the pre-dropout random variables and their realizations:

Y_{stv} denotes the *conditional* random variable that is the primary outcome variable (pain score in the illustration) from subject s, who is randomly assigned to treatment group t, at visit v, for those combinations of t, s, and v for which the outcome variable is realized. In Thought Experiment 3 above, in the r-th rep Y_{stv} was denoted

$$Y_{3v}^{(r)} = (Y_{1v}^{(r)} | v \leq v^{(r)}), v = 1, 2, …, v^{(r)}.$$

y_{stv} denotes the realized value of Y_{stv} in those cases when Y_{stv} is realized, rounded to a small number of significant digits.

The variables above may be collected into convenient vectors:

$Y_s = (Y_{st1}, Y_{st2}, \ldots, Y_{stv_s})'$ is the vector of conditional random variables for subject s.

$\mathbf{y}_s = (y_{st1}, y_{st2}, \ldots, y_{stv_s})'$ is the vector of realized values of the conditional random variables for subject s.

$Y = (Y'_1, Y'_2, \ldots, Y'_N)'$ is the vector of all conditional random variables for all subjects.

$\mathbf{y} = (Y'_1, Y'_2, \ldots, Y'_N)'$ is the vector of realized values of all the conditional random variables for all subjects.

4.2. A Mixed Model, Parameters, and Hypotheses

The following paragraphs define notation and specify some of the assumptions for a mixed (random effects) model for data from the hypothetical-but-realistic clinical trial described above. A similar model can be defined for data from many longitudinal clinical trials.

A mixed model has two components that may be called the "Expected Value (part of the) Model," which specifies assumptions about E[Y], and the "Covariance (part of the) Model," which specifies assumptions about V[Y], the covariance matrix of Y. Typically variation in the Covariance Model has no effect on the definition of bias. We shall therefore assume only that the Covariance Model specifies that random vectors from distinct subjects are stochastically independent and that V[Y] is nonsingular. We focus on the Expected Value Model.

The generic mixed model Expected Value Model equation is E[Y] = $\mathbf{X}\boldsymbol{\beta}$. We will use this general formulation, but for simplicity of exposition of the illustration we also define a cell mean model, a special case of the general model that, in this case, has no covariates:

μ_{tv} = E[Y_{stv}] for t $\in$ {"A", "C"}, v = 1, 2, ... V, and for each combination of s, t, and v for which at least one Y_{stv} was realized.

Note the implicit model assumption that for a given treatment group, t, and visit number, v, the expected value is the same for all subjects.

We collect the means into convenient vectors:

$$\boldsymbol{\mu}_t = (\mu_{t1}, \mu_{t2}, \ldots, \mu_{tV})', \ t \in \{\text{"A"}, \text{"C"}\},$$

the vector of treatment group population means. The mixed model $\boldsymbol{\beta}$ is partitioned as:

$$\boldsymbol{\beta} = (\boldsymbol{\mu}'_C, \boldsymbol{\mu}'_t)' = (\boldsymbol{\beta}'_C, \boldsymbol{\beta}'_t)' = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2)'.$$

4.2.1. Baseline Variables: Dependent Variables, Covariates, Change-from-Baseline—This model may be somewhat non-traditional in that the model treats each pre-randomization primary outcome variable as a random variable, on the dependent variable ("Y") side of the model rather than as a fixed covariate on the "X" side of the model.

A model with baseline values as covariates is based on the conditional distribution of post-randomization variables conditioned on the realized values of one or more baseline variables. The covariates help to achieve two objectives: (1) adjust for baseline imbalances between treatment groups and (2) reduce standard errors of estimators and consequently increase power of hypothesis tests. In principle, the use of the conditional distribution

restricts the "inference space," i.e., the population of patients to which the results of the study may be extrapolated. This point is often ignored in practice.

Some other models use a change-from-baseline outcome variable, $Y_{stv}^* = Y_{stv} - Y_{st1}$, or similar, as the model's dependent variable. These models also, to some extent, accomplish (1) and (2) and are not based on the conditional distribution. The structure of $V[Y_s^*]$ is typically more complicated than the structure of $V[Y_s]$.

The present model accomplishes (1) in the definition of the treatment effect, below, and (2) via the typically positive correlations between post-baseline measurements and baseline measurement. The methods we describe here can also be used in a straightforward manner for both of the other types of models.

4.2.2. Definition of β in terms of E[Y]—In the generic Expected Value Model equation, above, one can think of E[Y] as a fundamental quantity and β as the Expected Value Model's "primary parameter." (We will define interesting "secondary parameters" in the general linear form $\theta = \mathbf{L}\beta$.) It is useful to solve the Expected Value model equation for β, which produces a definition of β as a linear transformation of the more primitive parameter vector, E[Y]. Without loss of generality we assume **X** has full column rank; the solution of the generic equation is:

$$\beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[Y].$$

We will use this definition equation *per se* and also apply it generically to other linear models below.

4.2.3. Treatment Effect Parameters and Hypotheses—We define a general linear secondary parameter of the form $\theta = \mathbf{L}\beta$. The treatment effect is a special case.

There are multiple approaches to defining "treatment effect" in a longitudinal clinical trial. For the present purposes we adopt just one method. Within each treatment group we compute the average of the population means from the last 4 visits and subtract the baseline population mean:

$$\left(\left(\frac{1}{4}\right)\left[\sum_{j=V-3}^{V}\mu_{A,j}\right] - \mu_{A,1}\right) \text{ and } \left(\left(\frac{1}{4}\right)\left[\sum_{j=V-3}^{V}\mu_{C,j}\right] - \mu_{C,1}\right)$$

We define the treatment effect, τ, as the difference of these two "change from baseline" parameters:

$$\tau = \mathbf{L}\beta = \left(-\mu_{A,1} + \left(\frac{1}{4}\right)\sum_{j=V-3}^{V}\mu_{A,j}\right) - \left(-\mu_{C,1} + \left(\frac{1}{4}\right)\sum_{j=V-3}^{V}\mu_{C,j}\right)$$

where the $1 \times (2V)$ vector **L** has elements that take the specified linear combination of the elements of β. We use the mnemonic symbol τ in place of the generic θ for this "special" parameter.

In this pain study $\tau < 0$ corresponds to the active treatment arm experiencing greater efficacy (lowering mean pain scores by a larger amount) than the control treatment arm.

The corresponding null and alternative hypotheses are, in the traditional notation: $H_0: \tau \quad 0$ *vs.* $H_A: \tau < 0$. The example study, like many pharmaceutical clinical trials, is a superiority clinical trial: the objective is to demonstrate that the "active" treatment has superior efficacy in comparison to the control treatment. Consequently, H_A is a "one-tail" hypothesis. A "two-tail" H_A would be appropriate for some clinical trials, such as for example, a head-to-head comparison of two active treatments in which the objective is to demonstrate that one (either one) of the two treatments is superior to the other.

4.3. Imputation Notation

Consider an imputation method in which we "complete" a vector for each dropout subject by imputing values for unrealized variables that are linear combinations of the available data from that subject, i.e., can be written in the following form for subject s:

$$\underset{v \times 1}{\boldsymbol{Y}_{I,s}} = \underset{v \times v_s}{\mathbf{A}_s} \cdot \underset{v_s \times 1}{\boldsymbol{Y}_s}.$$

When subject s has incomplete data, $\boldsymbol{Y}_{I,st}$ is the imputation-completed vector of outcome variables for subject s, hence the subscript I. However, we define a $\boldsymbol{Y}_{I,s}$ vector for every subject; when subject s has complete data, $\boldsymbol{Y}_{I,s} = \boldsymbol{Y}_s$. Baseline-observation-carried-forward (BOCF), LOCF, and some other imputation methods can be written in this form. Table 2 illustrates the $\mathbf{A}_s$ matrices for LOCF and BOCF for "subject C" in the example in Table 1 (s is not specified), where t="C" and $v_{sC}=3$.

Some clinical trials include substantial investigation and documentation of the reasons why subjects drop out. With FDA's approval, in some cases, sponsors of such studies have used a "hybrid" imputation scheme. For example, BOCF may be used for a subject who drops out for lack of efficacy, LOCF might be used for a subject who drops out for intolerability and the hybrid method might use no imputation for a specific subject whose post-dropout missing data are reasonably believed to be MAR. All of these methods can be implemented using the approach above by using an appropriate $\mathbf{A}_s$ matrix. When a subject has complete data ($v_s=V$) or if no imputation is being performed for subject s, $\mathbf{A}_s = I_{v_s}$, a $v_s \times v_s$ identity matrix.

Let $\mathbf{X}_{I,s}$ denote the mixed model "design matrix" for subject s, where the subscript "$\mathbf{I}$" indicates this is the design matrix for a dependent vector $\boldsymbol{Y}_{I,s}$ that is either complete (subject did not drop out) or contains imputed elements. $\mathbf{X}_{I,s}$ is the complete design matrix one would have for subject s if subject s had complete data. In the special case of the cell mean model above:

$$\underset{v \times 2v}{\mathbf{X}_{I,s}} = \begin{cases} [\mathbf{I}_v, \mathbf{0}], & \text{if subject } s \text{ received treatment } t{=}C \text{ (control)} \\ [\mathbf{0}, \mathbf{I}_v], & \text{if subject } s \text{ received treatment } t{=}A \text{ (active)} \end{cases}.$$

The boldface $\mathbf{I}_V$ denotes a $V \times V$ identity matrix. An $\mathbf{X}_{I,s}$ matrix is defined for every subject, whether the subject's data required imputation or not.

Let $\mathbf{X}_s$ denote the $v_s \times 2V$ design matrix for a mixed model for the available data, $\boldsymbol{Y}_s$, from subject s (who is in treatment group t). An $\mathbf{X}_s$ matrix is defined for every subject. In the special case of the cell mean model above:

$$\mathbf{X}_s \atop v_s \times 2v = \begin{cases} [(\mathbf{I}_{v_s}, \mathbf{0}), \mathbf{0}], \ \text{if subject } s \text{ received treatment } t{=}C \ (\text{control}) \\ [\mathbf{0}, (\mathbf{I}_{v_s}, \mathbf{0})], \ \text{if subject } s \text{ received treatment } t{=}A \ (\text{active}) \end{cases} .$$

The $(\mathbf{I}_{v_s}, \mathbf{0})$ matrices, and the $\mathbf{0}$ matrices not in parentheses, are $v_s \times V$. We assume at least one subject in each treatment group has complete outcome data through the final scheduled visit, visit V; this insures that no columns of $\mathbf{X}$ (the vertical concatenation of all the $\mathbf{X}_s$ matrices) are zero and that $\mathbf{X}$ has full column rank.

The notation above leads to

$$E[\mathbf{Y}_{I,s}] = \mathbf{A}_s E[\mathbf{Y}_s] = \mathbf{A}_s \mathbf{X}_s \boldsymbol{\beta} = \mathbf{X}_{I,s} \boldsymbol{\beta}_\mathbf{I},$$

where $\boldsymbol{\beta}_\mathbf{I} \quad \boldsymbol{\beta}$ and when subject s has incomplete data, $\mathbf{X}_{I,s} \quad \mathbf{A}_s \mathbf{X}_s$.

We combine Y-vectors and A- and X-matrices from all subjects:

$$\mathbf{A} = \text{BlockDiag} \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \dots \\ \mathbf{A}_N \end{bmatrix}, \ \mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \dots \\ \mathbf{Y}_N \end{bmatrix}, \ \mathbf{Y}_I = \begin{bmatrix} \mathbf{Y}_{I,1} \\ \mathbf{Y}_{I,2} \\ \dots \\ \mathbf{Y}_{I,N} \end{bmatrix} = \mathbf{AY}, \mathbf{X}_I = \begin{bmatrix} \mathbf{X}_{I,1} \\ \mathbf{X}_{I,2} \\ \dots \\ \mathbf{X}_{I,N} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \dots \\ \mathbf{X}_N \end{bmatrix}.$$

4.4. Imputation-Induced Estimator Bias

The Expected Value Model of the form $E[\mathbf{Y}_\mathbf{I}] = \mathbf{X}_\mathbf{I}\boldsymbol{\beta}_\mathbf{I}$, when solved for $\boldsymbol{\beta}_\mathbf{I}$ produces a definition of $\boldsymbol{\beta}_\mathbf{I}$:

$$\boldsymbol{\beta}_I = (\mathbf{X}_I' \mathbf{X}_I)^{-1} \mathbf{X}_I' E[\mathbf{Y}_I] = (\mathbf{X}_I' \mathbf{X}_I)^{-1} \mathbf{X}_I' \mathbf{AX}\boldsymbol{\beta} = \mathbf{T}\boldsymbol{\beta}.$$

Fitting this model produces the unbiased estimator $\hat{\boldsymbol{\beta}_I}$ for $\boldsymbol{\beta}_\mathbf{I}$, i.e., $E[\hat{\boldsymbol{\beta}_I}] = \boldsymbol{\beta}_I$. However, because $\boldsymbol{\beta}_\mathbf{I} \quad \boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}_I}$ is a biased estimator of $\boldsymbol{\beta}$. The imputation-induced bias in $\hat{\boldsymbol{\beta}_I}$ (as an estimator of $\boldsymbol{\beta}$) is:

$$\mathbf{Bias}(\hat{\boldsymbol{\beta}}_I) = \boldsymbol{\beta}_\mathbf{I} - \boldsymbol{\beta} = [(\mathbf{X}_I' \mathbf{X}_I)^{-1} \mathbf{X}_I' \mathbf{AX} - \mathbf{I}] = [\mathbf{T} - \mathbf{I}]\boldsymbol{\beta} = \mathbf{C}_{\text{BCM}}\boldsymbol{\beta}.$$

The "bias coefficient matrix," $\mathbf{C}_{\text{BCM}}$, conveniently summarizes all the calculations vis-à-vis $\mathbf{Bias}(\hat{\boldsymbol{\beta}_I})$ and the

$$\mathbf{T} = (\mathbf{X}_I' \mathbf{X}_I)^{-1} \mathbf{X}_I' \mathbf{AX}$$

matrix represents the relationship between $\boldsymbol{\beta}_\mathbf{I}$ and $\boldsymbol{\beta}$:

$$\boldsymbol{\beta}_I = [(\mathbf{X}_I' \mathbf{X}_I)^{-1} \mathbf{X}_I' \mathbf{AX}]\boldsymbol{\beta} = \mathbf{T}\boldsymbol{\beta}.$$

The treatment effect (parameter) from fitting the model to the imputation-completed data is $\tau_I = \mathbf{L}\boldsymbol{\beta}_I$, with estimator $\hat{\tau}_I = \mathbf{L}\hat{\boldsymbol{\beta}}_I$, which has imputation-induced bias

$$\text{Bias}(\hat{\tau}_I) = \tau_I - \tau = \mathbf{L}(\boldsymbol{\beta}_I - \boldsymbol{\beta}) = \mathbf{L}\,Bias(\boldsymbol{\beta}_I) = \mathbf{L}[\mathbf{T} - \mathbf{I}]\boldsymbol{\beta} = \mathbf{L}_{\text{BCV}}\boldsymbol{\beta}.$$

The "bias coefficient vector"

$$\mathbf{L}_{\text{BCV}} = \mathbf{L}[(\mathbf{X}_I'\mathbf{X}_I)^{-1}\mathbf{X}_I'\mathbf{A}\mathbf{X} - \mathbf{I}] = \mathbf{L}[\mathbf{T} - \mathbf{I}]$$

conveniently summarizes all the calculations vis-à-vis $\text{Bias}(\tau_I)$.

The matrices in the equations above can become quite large. Here are some convenient computational forms:

$$(\mathbf{X}_I'\mathbf{X}_I) = \sum_{s=1}^{N}\mathbf{X}_{I,s}'\mathbf{X}_{I,s}, \quad (\mathbf{X}_I'A\mathbf{X}) = \sum_{s=1}^{N}\mathbf{X}_{I,s}'\mathbf{A}_s\mathbf{X}_s$$

4.5. Imputation-Induced Hypothesis Corruption

The null and alternative hypotheses that are tested in the mixed model for imputation-completed data are "corrupted hypotheses," the "corruption" arising from the bias in $\hat{\tau}_I$ as an estimator of τ. We write the original hypotheses above as $H_0(\tau): \tau \geq 0$ and $H_A(\tau): \tau < 0$. The corrupted hypotheses are:

$$H_0(\tau_I): \tau_I \geq 0 \iff \tau + \text{Bias}(\hat{\tau}_I) \geq 0 \iff \tau \geq -\text{Bias}(\hat{\tau}_I)$$
$$H_A(\tau_I): \tau_I < 0 \iff \tau + \text{Bias}(\hat{\tau}_I) < 0 \iff \tau < -Bias(\hat{\tau}_I)$$

Suppose, for example, that H_A is true, $\tau = -0.7$ and $\text{Bias}(\hat{\tau}_I) = 0.5$. Then the corrupted null and alternative hypotheses tested in the model for the imputation-completed data are:

$$H_0: \tau \geq -0.5 \text{ and } H_A(\tau): \tau < -0.5.$$

The bias reduces the "effective treatment effect" from $\tau = -0.7$ to $\tau_I = -0.7 - (-0.5) = -0.2$. The much smaller "effective treatment effect" means the power is much smaller than would have been the case if uncorrupted hypotheses had been tested.

4.6. Removing Imputation-Induced Bias and Hypothesis Corruption

Note that $\boldsymbol{\beta}_I = \mathbf{T}\boldsymbol{\beta}$ and, as $\mathbf{T}$ is generally nonsingular (because we are using full rank models), $\boldsymbol{\beta} = \mathbf{T}^{-1}\boldsymbol{\beta}_I$. Therefore $\hat{\boldsymbol{\beta}}_U = \mathbf{T}^{-1}\hat{\boldsymbol{\beta}}_I$ and $\hat{\tau}_U = \mathbf{L}\hat{\boldsymbol{\beta}}_U$ are unbiased estimators of $\boldsymbol{\beta}$ and τ, respectively, from the imputation-completed data. Using these estimators (and *appropriate* covariance matrix and variance estimates, respectively) in the calculation of test statistics leads to tests of hypotheses that are not corrupted by imputation-induced bias. Thus one can use the imputation-completed data and also use estimators that are free of imputation-induced bias and corresponding uncorrupted hypothesis tests.

4.7. Illustration of Imputation-Induced Estimator Bias and Hypothesis Corruption

We applied the methodology above to the hypothetical clinical trial. The trial had a target of 150 randomized subjects per treatment group; the actual numbers were 148 in the control (placebo) group and 151 in the active treatment group; N=199 total. The dropout patterns came from an actual clinical trial and are characterized in Table 3.

The "Number of Subjects for whom this was the Last Visit" columns display the numbers of subjects whose last visit was the visit number in the first column. For example, in the Control group, 7 subjects dropped out after visit 1 and before visit 2. The row for visit 13 displays the numbers of subjects who completed the study, as their last visit was the last scheduled visit. The "Percent of Subjects for whom this was the Last Visit" columns show the percentages of subjects whose last visit was the visit number in the first column. The percentages add to 100% in each of the two columns. These percentages are displayed in Figure 2. The information in Table 2 came from an actual clinical trial; that both treatment arms had approximately the same (66.2%) completion rate is coincidental.

The "Number of Subjects Remaining in Study after Visit" columns of Table 3 display the numbers of subjects who had not yet dropped out by the visit number specified in the first column. The first row displays the total numbers of subjects who completed visit 1 and were randomized. The row for visit 13 displays the numbers of subjects who completed all visits.

The imputation-induced bias in $\hat{\beta_I}$ is a vector-function of β which is, of course, unknown. To apply the method one evaluates the coefficient matrix $\mathbf{C}_{BCM}$ and vector $\mathbf{L}_{BCV}$ defined above and multiplies each by a hypothesized β. To illustrate the calculations we used the matrices for the cell mean model and the values of β from Table 1, as if the "population means" in the table were the population means of the conditional random variables, Y_{stv}. Separate matrices and vectors were computed for BOCF and LOCF. Each $\mathbf{C}_{BCM}$ is 26×26, too large to display here.

The BOCF-induced bias in $\hat{\beta_I}$ is displayed in Table 4 and Figure 3. As one can see in the lower part of the figure, the bias is 0 at visit 1 and increases steadily through visit 13. By visit 13 the bias in the active treatment group mean, 1.86, is almost as big as the actual coefficient, 2.0, while the bias in the control group is 1.18.

The LOCF-induced bias in $\hat{\beta_I}$ is displayed in Table 5. As with the BOCF bias, the LOCF bias is 0 at visit 1 and increases steadily through visit 13, with a maximum for both treatment groups just under 1.0. In the initial discussion of this hypothetical example we hypothesized that LOE subjects, mostly in the control (placebo) group, might tend to drop out earlier than subjects withdrawing due to side effects, resulting in LOCF imputation introducing an anticonservative bias toward H_A. The reasons for dropout are not included in the present data but one can see that, in this case, LOCF introduced a conservative bias toward the null hypothesis.

The elements of the treatment effect bias coefficient vectors, $\mathbf{L}_{BCV,}$ for BOCF and LOCF are shown in Table 6. The zero BOCF coefficients for visits 2–9 stem from the fact that the imputation carried forward baseline values. The non-zero BOCF coefficients for visits 10–13 stem from the comparison of the mean of the last 4 visits to baseline.

Evaluating the treatment effect biases from BOCF and LOCF directly as $\mathrm{Bias}(\hat{\tau_I}) = \mathbf{L}_{BCV}\beta$ gives values of $\tau_I - \tau = 0.49$ and 0.13, respectively.

Table 7 displays a more intuitively appealing calculation of these biases. The first row shows the baseline population means, which are all 7.5 in this hypothetical example. The

second row displays the means of the population means, averaged over visits 10–13. The columns headed "β" show calculations using β, the (hypothesized) population means. The columns headed "β_I" show calculations using BOCF and LOCF β_I, respectively. Row 3 shows the changes from baseline to mean from visits 10–13 for each treatment group. The treatment effect, τ_I, in each pair of columns (row 4) is the differences in the changes-from-baseline. The biases in row 5 are values of $\tau_I - \tau$ for the two imputation methods. The equation $\mathrm{Bias}(\hat{\tau_I}) = \mathbf{L}_{BCV}\beta$ is easy to program in software such as SAS PROC IML, but the table calculations may be more intuitive and are easy to "program" in spreadsheet software.

The BOCF treatment effect bias of 0.49 is approximately the same magnitude (opposite direction) as a small but "clinically meaningful" treatment effect in some pain studies. Our objective here is to illustrate the application of the method, not examine the characteristics of imputation-induced bias in this hypothetical study. If we were examining BOCF bias for this study we could modify β to create a treatment effect of about 0.49 and recalculate $\mathrm{Bias}(\tau_I) = \mathbf{L}_{BCV}\beta$ the calculations are simple because $\mathbf{L}_{BCV}$ does not change. [Indeed, in just under 10 minutes the author modified the spreadsheet used to create many of the tables to change β to represent a treatment effect of −0.49. (Only active treatment means were changed.) The BOCF treatment effect bias is unchanged at 0.49, exactly cancelling the treatment effect.]

In this example using BOCF the imputation-corrupted hypotheses are

$$H_0(\tau_I){:}\tau_I \geq 0 \iff \tau + \mathrm{Bias}(\hat{\tau}_I) \geq 0 \iff \tau + 0.49 \geq 0 \iff \tau \geq -0.49$$
$$H_A(\tau_I){:}\tau_I < 0 \iff \tau + \mathrm{Bias}(\hat{\tau}_I) < 0 \iff \tau + 0.49 < 0 \iff \tau < -0.49.$$

In some cases a treatment effect of −0.49 is clinically meaningful; here the corrupted *null hypothesis* would include an actual, clinically meaningful treatment effect. Thus if it turned out that the active treatment produced a clinically meaningful effect, 0.49 unit larger decrease in mean pain score relative to the control ($\tau = -0.49$), using the imputed data the corrupted *null* hypothesis would be true! One can show that the bias is intrinsic: increasing the sample size would not make this problem go away.

The point can also be made by examining the test statistic. When this is a *t*-statistic, for example, performing the test with $\hat{\tau}_I$ is as follows:

$$t_I = \frac{\hat{\tau}_I}{s.E.(\hat{\tau}_I)} = \frac{[\hat{\tau} - (-0.49)]}{s.E.(\hat{\tau}_I)} = \frac{[\hat{\tau} + 0.49]}{s.E.(\hat{\tau}_I)} \cong t + \frac{0.49}{s.E.(\hat{\tau}_I)}.$$

Here, of course, *t* and $\hat{\tau}$ are computed from available data, as for example, by fitting a mixed model to the available data with no imputation. As H_0 is rejected for large *negative t* values, the rightmost term in the equation above represents, approximately, a penalty induced by the imputation. If the bias does not change with total sample size and s. e. ($\hat{\tau}_I$) decreases with total sample size, conducting a larger study – increasing the sample size – would *increase* the imputation-induced bias penalty in the test statistic.

4.8. Comparing Biases of Various Linear Imputation Methods and Non-Imputation

$\mathrm{Bias}(\hat{\tau_I})$ is a function of τ, whose value is unknowable. An analogy is useful: power is also a function of unknowable τ, and power calculations are an important component of the design of clinical trials. As in the case of power calculations, we investigate the effect of $\mathrm{Bias}(\hat{\tau_I})$ by calculating $\mathrm{Bias}(\hat{\tau_I})$ for an "interesting" range of τ values, including $\tau = 0$, the H_0 value,

and for imputation methods of interest. One can graph $\text{Bias}(\hat{\tau_I})$ vs. τ, including multiple "curves," one for each method. In our experience such a graph typically displays clear differences in the bias characteristics of the imputation methods.

This type of comparison does not evaluate MNAR-induced bias in the data, which afflicts all the methods, with or without imputation, but which Siddiqui, Hung, and O'Neill (2009) found to be quite small in most of the clinical trials they considered.

5. DISCUSSION

On both the execution side and the regulatory side, our objectives as public health professionals and as drug development clinical trialists include: (1) to protect the safety of clinical trial subjects, (2) to kill bad drugs – ensure that they are not marketed – and do so as quickly and inexpensively as practical, (3) to bring good drugs to market as quickly as practical while minimizing the required investments – consistent with the other objectives. And yes, objective (3) *is* an appropriate objective for regulatory authorities.

Most clinical trial statisticians seem to agree: In general, bias is bad and corrupted hypotheses are bad. Typically we are not choosing between zero bias and non-zero bias, but rather we are choosing among several biased procedures, each with its own characteristics. We want to minimize estimator bias and hypothesis corruption, consistent with our objectives.

A bias toward H_A is undesirable because it inflates the Type I error rate and increases the probability of either not killing a bad drug or continuing to invest research resources and calendar time in a bad drug (not killing it quickly).

A bias toward H_0 is undesirable because: (a) to achieve acceptable power a sponsor must expose more subjects to experimental procedures (contrary to the objective of protecting safety of clinical trial subjects), (b) it increases calendar time in clinical trials, which leads to: (c) delaying making good treatments available, (d) killing more good drugs, (e) spending more money and other resources, and (f) reducing the amount of time between marketing approval and expiration of patent, all of which are inconsistent with objective (3).

We have focused on methodology for evaluating linear-imputation-induced estimator bias and hypothesis corruption. These methods include the widely-used BOCF and LOCF but not some other important methods, such as mixed-model-based methods, that make use of all available data from all subjects. We can use the tools in this paper to evaluate alternative linear imputation methods (including no imputation at all) in an attempt to minimize this part of the total bias, which is under our control.

We must not minimize the importance of MNAR-induced bias in the data, as discussed in the first part of the paper. Unlike imputation-induced bias, some level of MNAR-induced bias may be inevitable but it is important to understand that the bias is in the *data*, not the statistical method. Under current ethical standards (that we support) subjects must be permitted to withdraw from clinical trials when they wish, regardless of the reason. In many therapeutic areas, including analgesics for chronic pain, most dropout subjects' post-dropout data must be presumed to be MNAR.

One can assert that the MNAR bias is irrelevant because the expectations of the conditional outcome variables are the most appropriate means for our models. Consider a non-research situation when a physician and a patient decide to try a new (to the subject) therapy, such as a new analgesic. After an appropriate discussion (that depends upon the patient's ability to comprehend the information) of the product's advantages and disadvantages, the physician

issues a prescription (perhaps giving the patient some "samples"), the patient acquires the product and begins taking it in a manner more-or-less similar to how subjects take the product in a clinical trial. In some cases – real life and clinical trial – titration to a therapeutically effective dose is required. Both patient and subjects may experience relief and/or side effects. When the relief is inadequate or the side effects intolerable both the patient and subject will typically return to the physician and decide whether to terminate that treatment or continue for a while. Over a time interval similar to the duration of treatment in a clinical trial, a patient either discontinues that therapy or perseveres, much as in the clinical trial. The overall point is that subject dropout behavior may closely approximate real life patient dropout behavior. To the extent that that is the case, the conditional distribution of available data in a clinical trial is similar to the distribution that would be obtained in "real life" and an unbiased analysis of the available data would be substantially preferable to imputation-based alternatives. Of course, reasonable people may disagree on the extent to which the conditional clinical trial distributions are equivalent to "real life" distributions.

These facts should energize clinical trialists, not discourage us. The situation is not totally beyond our control. Reducing the proportion of dropout subjects leads to less MNAR bias and less imputation-induced bias. We control study design; the use of enriched designs and related methods, albeit not a panacea, can substantially reduce the proportion of dropout subjects. Study execution can affect dropout rates. We can train and incent clinic staff members on the importance of subject retention; experience has shown that properly trained and incentivized clinic staff members produce better subject retention rates. Retraining during the course of a long study can be helpful, partially when clinical staff turnover rates are high. In addition, we can use a variety of ethical methods to encourage and incent subjects to "hang in there" and complete the entire protocol.

In addition, some post-dropout missing data are MAR ("My girlfriend is moving to Anchorage and so am I.") and analyzing the available data from such a subject (e.g., using a mixed model) does not introduce bias. We have found that carefully designed case report forms that require careful investigation and recording of comprehensive information on the reason a subject drops out, can form the basis for an external, expert, treatment-masked adjudication of whether a subject's post-dropout data are MAR. When the analysis uses all available data (e.g., a mixed model analysis) using imputation for a subject with MAR post-dropout missing data would introduce unnecessary bias into the analysis. The number of subjects in this category in a typical longitudinal clinical trial is small (about 5% or less in our experience), but when BOCF is being used the method can be cost effective. For example, suppose a trial has 10 subjects in this category. If BOCF were used for these 10 subjects they would typically contribute a bias toward the null hypothesis. Using the 10 subjects' actual data contributes unbiased information that, if H_A is correct, increases power by an amount corresponding to 10 subjects.

We can paraphrase the philosopher Desiderius Erasmus, "Biases: can't live with them, can't live without them." Longitudinal clinical trials are almost always afflicted by MNAR missing data, leading to data with biased expected values. Evaluating those biases can be very difficult. If we use linear imputation (BOCF, LOCF, others) to cope, this paper presents methods that are useful for evaluating the imputation-induced biases and consequently corrupted hypotheses.

Acknowledgments

References

Brown, JR. The Laboratory of the Mind: Thought Experiments in the Natural Sciences. Routledge; London: 1993.

Helms, Ronald W. (2009a). This utterance was made in a private meeting between FDA and a pharmaceutical company. The FDA biostatistician requested that we not reveal his or her identity.

Helms, Ronald W. (2009b). This refers to two separate reports prepared for two pharmaceutical companies that used simulations to estimate imputation-induced bias from multiple forms of imputation and also from mixed model analyses of available data without imputation. The reports are subject to confidentiality agreements between the author's employer and the pharmaceutical companies and cannot be made public. This paper contains very similar results based on direct computations, not simulations.

Little, RJA.; Rubin, DB. Statistical Analysis with Missing Data. 2nd ed.. New York: John Wiley; 2002.

Scheffè; Henry. The Analysis of Variance. New York: John Wiley; 1959.

Sen, PK.; Singer, J. Large Sample Methods in Statistics: An Introduction with Applications. New York: Chapman and Hall/CRC; 1994.

Siddiqui, Ohidul; Hung, HM James; O'Neill, Robert. MMRM vs. LOCF: A Comprehensive Comparison Based on Simulation Study and 25 NDA Datasets. Journal of Biopharmaceutical Statistics. 2009; 19:2, 227–246.
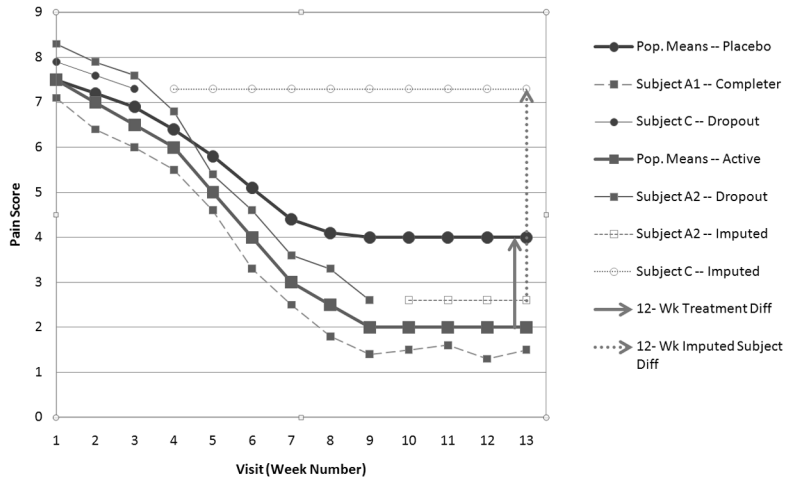
Figure 1.
Illustration of LOCF Imputation-Induced Bias from a Hypothetical Longitudinal Clinical Trial
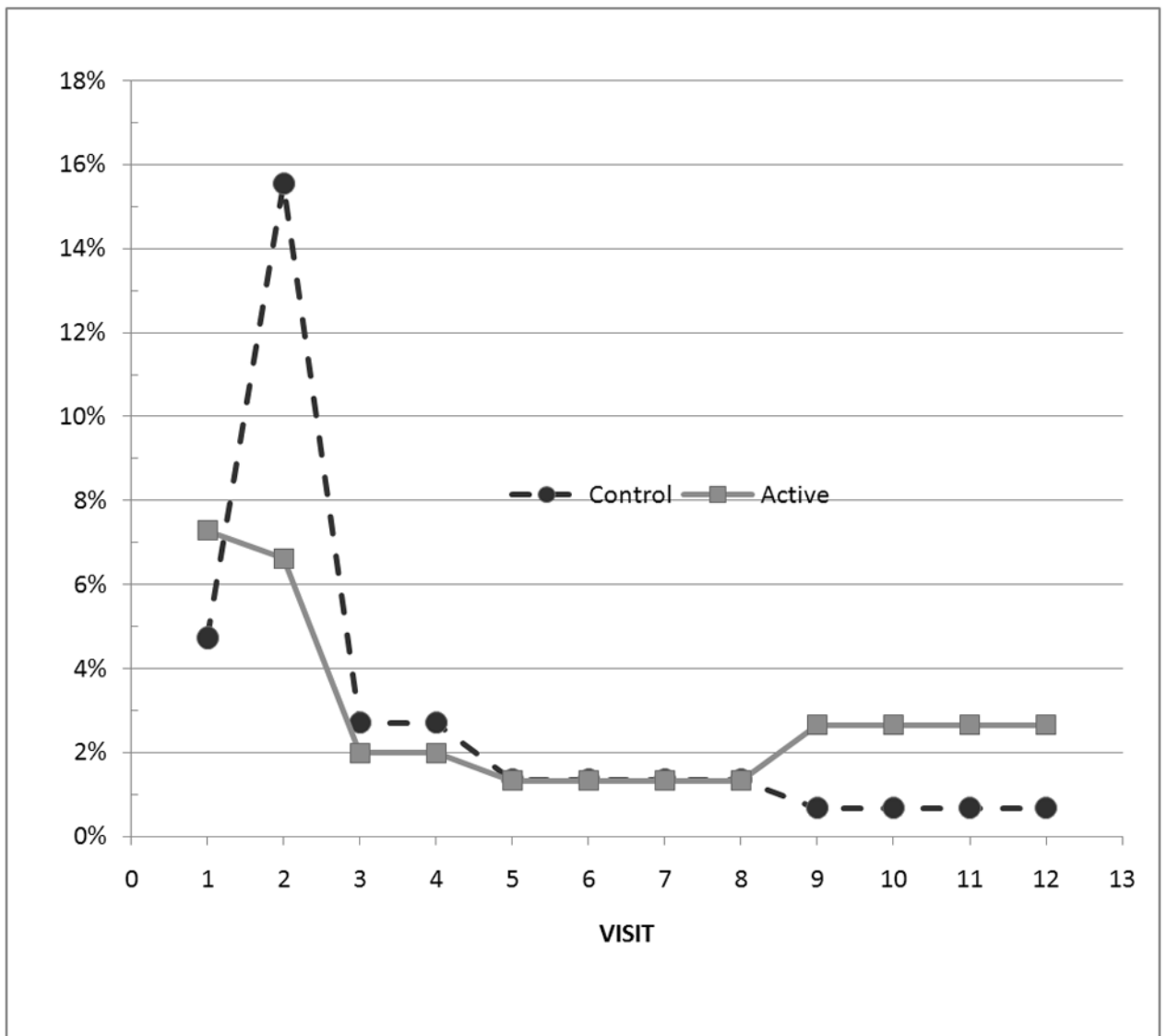
Figure 2.
Percentages of Subjects Dropping Out After Each Visit, by Treatment Group
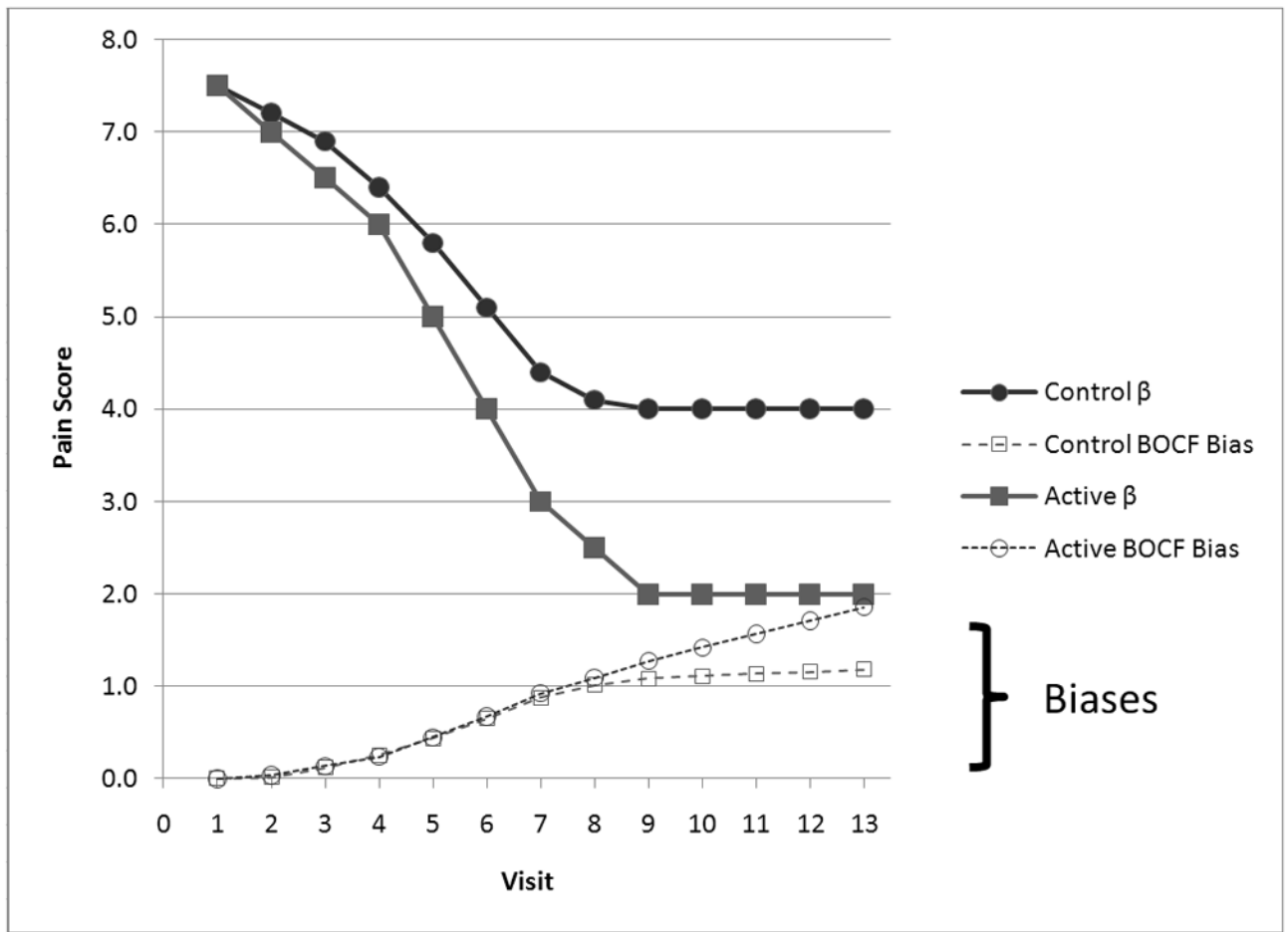
Figure 3.

BOCF Induced Bias in $\widehat{\partial\hat{I}\hat{I}\acute{s}}_I$, Hypothetical Longitudinal Clinical Trial

Table 1

Hypothetical Treatment Group Means and Hypothetical Subject Data

Visit	Population Means – Control (Placebo)	Subject C – Dropout	Subject C – Imputed	Pop. Means – Active	Subject A1 – Completer	Subject A2 – Dropout	Subject A2 -- Imputed
1	7.5	7.9		7.5	7.1	8.3	
2	7.2	7.6		7.0	6.4	7.9	
3	6.9	7.3		6.5	6.0	7.6	
4	6.4		7.3	6.0	5.5	6.8	
5	5.8		7.3	5.0	4.6	5.4	
6	5.1		7.3	4.0	3.3	4.6	
7	4.4		7.3	3.0	2.5	3.6	
8	4.1		7.3	2.5	1.8	3.3	
9	4		7.3	2.0	1.4	2.6	
10	4		7.3	2.0	1.5		2.6
11	4		7.3	2.0	1.6		2.6
12	4		7.3	2.0	1.3		2.6
13	4		7.3	2.0	1.5		2.6

Table 2

Illustration of LOCF and BOCF as Linear Transformations of Available Data

Visit	LOCF* Y_{LsC}	BOCF* Y_{LsC}	LOCF Y_{LsC}	BOCF Y_{LsC}	LOCF A_{sC}			BOCF A_{sC}			y_{sC}	Y_{sC}
1	Y_{sC1}	Y_{sC1}	7.9	7.9	1	0	0	1	0	0	7.9	Y_{sC1}
2	Y_{sC2}	Y_{sC2}	7.6	7.6	0	1	0	0	1	0	7.6	Y_{sC2}
3	Y_{sC3}	Y_{sC3}	7.3	7.3	0	0	1	0	0	1	7.3	Y_{sC3}
4	$Y_{sC3} \rightarrow Y_{sC4}$	$Y_{sC1} \rightarrow Y_{sC4}$	7.3	7.9	0	0	1	1	0	0		
5	$Y_{sC3} \rightarrow Y_{sC5}$	$Y_{sC1} \rightarrow Y_{sC5}$	7.3	7.9	0	0	1	1	0	0		
6	$Y_{sC3} \rightarrow Y_{sC6}$	$Y_{sC1} \rightarrow Y_{sC6}$	7.3	7.9	0	0	1	1	0	0		
7	$Y_{sC3} \rightarrow Y_{sC7}$	$Y_{sC1} \rightarrow Y_{sC7}$	7.3	7.9	0	0	1	1	0	0		
8	$Y_{sC3} \rightarrow Y_{sC8}$	$Y_{sC1} \rightarrow Y_{sC8}$	7.3	7.9	0	0	1	1	0	0		
9	$Y_{sC3} \rightarrow Y_{sC9}$	$Y_{sC1} \rightarrow Y_{sC9}$	7.3	7.9	0	0	1	1	0	0		
10	$Y_{sC3} \rightarrow Y_{sC10}$	$Y_{sC1} \rightarrow Y_{sC10}$	7.3	7.9	0	0	1	1	0	0		
11	$Y_{sC3} \rightarrow Y_{sC11}$	$Y_{sC1} \rightarrow Y_{sC11}$	7.3	7.9	0	0	1	1	0	0		
12	$Y_{sC3} \rightarrow Y_{sC12}$	$Y_{sC1} \rightarrow Y_{sC12}$	7.3	7.9	0	0	1	1	0	0		
13	$Y_{sC3} \rightarrow Y_{sC13}$	$Y_{sC1} \rightarrow Y_{sC13}$	7.3	7.9	0	0	1	1	0	0		

*The symbol "→" means "replaces".

Table 3

Dropout Patterns in the Hypothetical Study

Visit Number	Number of Subjects for whom this was the Last Visit		Percent of Subjects for whom this was the Last Visit		Number of Subjects Remaining in Study after Visit	
	Control	Active	Control	Active	Control	Active
1 (Baseline)	7	11	4.7%	7.3%	148	151
2	23	10	15.5%	6.6%	141	140
3	4	3	2.7%	2.0%	118	130
4	4	3	2.7%	2.0%	114	127
5	2	2	1.4%	1.3%	110	124
6	2	2	1.4%	1.3%	108	122
7	2	2	1.4%	1.3%	106	120
8	2	2	1.4%	1.3%	104	118
9	1	4	0.7%	2.6%	102	116
10	1	4	0.7%	2.6%	101	112
11	1	4	0.7%	2.6%	100	108
12	1	4	0.7%	2.6%	99	104
13 (Last Scheduled Visit; "Completers")	98	100	66.2%	66.2%	98	100
Totals	148	151	100.0%	100.0%	NA	NA

Table 4

BOCF Induced Bias in $\hat{\beta_I}$

Visit	Control Treatment Group			Active Treatment Group		
	β	BOCF β_I	BOCF Bias	β	BOCF β_I	BOCF Bias
1	7.5	7.50	0	7.5	7.50	0
2	7.2	7.21	0.01	7.0	7.04	0.04
3	6.9	7.02	0.12	6.5	6.64	0.14
4	6.4	6.65	0.25	6.0	6.24	0.24
5	5.8	6.24	0.44	5.0	5.45	0.45
6	5.1	5.75	0.65	4.0	4.67	0.67
7	4.4	5.28	0.88	3.0	3.92	0.92
8	4.1	5.11	1.01	2.5	3.59	1.09
9	4.0	5.09	1.09	2.0	3.27	1.27
10	4.0	5.11	1.11	2.0	3.42	1.42
11	4.0	5.14	1.14	2.0	3.57	1.57
12	4.0	5.16	1.16	2.0	3.71	1.71
13	4.0	5.18	1.18	2.0	3.86	1.86

Table 5

LOCF Induced Bias in $\widehat{\partial I_i \acute{s}}_I$

Visit	Control Treatment Group			Active Treatment Group		
	β	LOCF $β_I$	LOCF Bias	β	LOCF $β_I$	LOCF Bias
1	7.5	7.50	0	7.5	7.50	0
2	7.2	7.21	0.01	7.0	7.04	0.04
3	6.9	6.98	0.08	6.5	6.61	0.11
4	6.4	6.59	0.19	6.0	6.19	0.19
5	5.8	6.14	0.34	5.0	5.36	0.36
6	5.1	5.63	0.53	4.0	4.56	0.56
7	4.4	5.13	0.73	3.0	3.76	0.76
8	4.1	4.92	0.82	2.5	3.37	0.87
9	4.0	4.85	0.85	2.0	2.99	0.99
10	4.0	4.85	0.85	2.0	2.99	0.99
11	4.0	4.85	0.85	2.0	2.99	0.99
12	4.0	4.85	0.85	2.0	2.99	0.99
13	4.0	4.85	0.85	2.0	2.99	0.99

Table 6

Treatment Effect Bias Coefficient Vectors

	BOCF		LOCF	
Visit	Control	Active	Control	Active
1	−0.3277	0.2980	−0.0473	0.0728
2	0	0	−0.1554	0.0662
3	0	0	−0.0270	0.0199
4	0	0	−0.0270	0.0199
5	0	0	−0.0135	0.0132
6	0	0	−0.0135	0.0132
7	0	0	−0.0135	0.0132
8	0	0	−0.0135	0.0132
9	0	0	−0.0068	0.0265
10	0.0794	−0.0646	0.0743	−0.0447
11	0.0811	−0.0712	0.0777	−0.0579
12	0.0828	−0.0778	0.0811	−0.0712
13	0.0845	−0.0844	0.0845	−0.0844

Table 7

Computing Bias in $\hat{\tau_I}$

	β		BOCF β_I		LOCF β_I	
	Control	Active	Control	Active	Control	Active
Visit1	7.50	7.50	7.50	7.50	7.50	7.50
Mean, Visits 10–13	4.00	2.00	5.15	3.64	4.85	2.99
Change	−3.50	−5.50	−2.35	−3.86	−2.65	−4.51
Trt. Effect	$\tau = -2.00$		$\tau_I = -1.51$		$\tau_I = -1.87$	
Bias in $\hat{\tau_I}$			$\tau_I - \tau = 0.49$		$\tau_I - \tau = 0.13$	