A Comparison of Last Observation Carried Forward and Multiple Imputation in a Longitudinal Clinical Trial

THESIS

Presented in Partial Fulfillment of the Requirements for the Degree Master of Science in the Graduate School of The Ohio State University

By

Tara Lynn Carmack, BS
Graduate Program in Public Health

The Ohio State University
2012

Master's Examination Committee:

Professor Rebecca Andridge, Advisor
Professor Abigail Shoben
Abstract

In randomized clinical trials, the presence of missing data presents challenges in determining the actual treatment effect of the study. It is particularly problematic in longitudinal studies when patients followed over time withdrawal from the study. Although it is important to anticipate and attempt to prevent these drop-outs in the study design, it is still likely that a significant amount of missingness will be present in the final data. It is important to have statistical methods that effectively analyze data which contains missing values, and produce unbiased results. This study compares several methods for handling missing data in longitudinal trials. The focus is on the single imputation method of last observation carried forward, and compares it to complete case analysis, multiple imputation and two additional versions of multiple imputation where everyone was imputed as if they were actually in the control group (placebo-imputation). We simulated a randomized control trial with a treatment and placebo group and two time points. After creating the data, we imparted missingness in the follow-up time point. We considered three mechanisms for the missing data: missing completely at random (MCAR), missing at random (MAR) and not missing at random (NMAR). The results indicated that in all situations, last observation carried forward produced extremely biased estimates of treatment effect. Both placebo imputations produced similarly biased estimates. Complete case analysis was only valid in the situation where the data was
MCAR. Traditional multiple imputation using regression performed the best of all the methods.
Dedication

This document is dedicated to my family.
Vita

May 2002 ..................................................Trevor G Browne High School

2006 ..........................................................B.S. Molecular Bioscience and

Biotechnology, Arizona State University

2012 ..........................................................M.S. Biostatistics, The Ohio State

University

2011 to present ........................................Graduate Teaching Associate, Department

of Biostatistics, The Ohio State University

Fields of Study

Major Field: Public Health
# Table of Contents

Abstract .................................................................................................................................................. ii

Dedication ................................................................................................................................................. iv

Vita ............................................................................................................................................................. v

List of Tables ............................................................................................................................................ viii

Chapter 1: Introduction ................................................................................................................................. 1

  Missing Data Mechanisms ............................................................................................................................ 2

  Methods for Dealing with Missing Data ...................................................................................................... 4

Chapter 2: Literature Review ....................................................................................................................... 7

Chapter 3: Methods .................................................................................................................................... 11

  Models ....................................................................................................................................................... 12

  Parameters ............................................................................................................................................... 14

  Imputation and Analyses ............................................................................................................................ 16

Chapter 4: Results ...................................................................................................................................... 18

Chapter 5: Discussion .................................................................................................................................. 20

  Conclusion ............................................................................................................................................... 22

References .................................................................................................................................................... 23
List of Tables

Table 1. Estimated treatment effect bias and power for Scenario 1 .................................. 19

Table 2. Estimated treatment effect bias and power for Scenario 2 ................................. 19
Chapter 1: Introduction

Randomized clinical trials are an important means of evaluating the safety and effectiveness of new treatments. It is of utmost importance that the results of the trials produce valid, statistically sound results for numerous reasons. If results are not valid, it can result in drugs or treatments being approved that in reality have ineffective or dangerous results. Conversely, it would also be harmful to determine a drug is not beneficial when in reality it is, as this would eliminate a therapy that could potentially help people. Additionally, billions of dollars are spent annually on clinical studies, so it is important that they produce valid results so that the time and resources spent on them are not wasted.

The occurrence of missing data is a major issue in clinical trials. It is particularly common when observations are made repeatedly on the same subject in a longitudinal study. As a majority of the large-scale trials involve a longitudinal design, it is important to continually evaluate and improve the methods developed to deal with the missing data. Although careful consideration should be done when designing the trial to reduce the likelihood of missing data, it is likely that even the best planned trial will have missing data, particularly for later time-points.

Advances in computer technology and availability have allowed for sophisticated methods to be developed to handle missing data. The methods might attempt to reach valid results from the data that are available, or might attempt to impute values into those
that are missing to create a full dataset that can then be analyzed. This paper focuses on the imputation method of last observation carried forward, which is an old method that does not require sophisticated software, but is still used regularly. We will simulate clinical trials with repeated measurements and impart different types of missingness on the observations. We will then use last observation carried forward to impute the missing values, and compare these results to other methods of handling missing data to determine which produce the most accurate estimates of treatment effect.

Missing Data Mechanisms

Before attempting to analyze data that contain missing values, it is important to consider the mechanisms that lead to the missing data. There are multiple ways to deal with missing data, which will be discussed later, but the properties of these methods strongly depend on the missing data mechanism. Rubin was the first to formalize the role of the mechanism in the analysis of data with missing values (Little and Rubin, 2002). He proposed treating the missing data indicators as random variables and assigning them a distribution. To discuss these mechanisms, let us define the complete data \( Y = (y_{ij}) \), where \( y_{ij} \) is the value of variable \( Y_j \) for subject \( i \), and the missing data indicator matrix \( M = (M_{ij}) \) such that:

\[
M_{ij} = \begin{cases} 
1 & \text{if } y_{ij} \text{ is missing} \\
0 & \text{if } y_{ij} \text{ is observed}
\end{cases}
\]

We further define \( \phi \) as a collection of unknown parameters.

The first mechanism to consider is missing completely at random (MCAR). If the missingness does not depend on any of the values of the observed or unobserved data, and occurs purely by chance, then it is classified as MCAR. That is:
In general, this is a strong assumption, and is unlikely to be the missing data mechanism in a clinical trial (NRC, 2010). Examples of causes of missingness in clinical trials that can be classified as MCAR include patients withdrawing from studies due to relocation to non-study areas, or patients who die or become sick from causes unrelated to the trial.

The next mechanism is missing at random (MAR). This occurs when missingness depends on some component of the observed data ($Y_{obs}$), but does not depend on the missing data ($Y_{mis}$). That is:

$$f(M|Y, \phi) = f(M| Y_{obs}, \phi)$$

In general, a MAR mechanism is a more probable cause of missing data in clinical studies than a MCAR mechanism. The assumptions of MAR are less restrictive than those of MCAR. An example of missingness that is MAR would be patients who drop out of a longitudinal clinical trial for weight loss because after the first few visits they have not lost sufficient weight. The missing data thus depends on the observed weights at the visits with observed data.

The final mechanism to consider is called not missing at random (NMAR). This occurs when, after adjusting for the observed data, the missingness still depends on the unobserved values. That is, the equation:

$$f(M|Y, \phi) = f(M| Y_{obs}, Y_{mis}, \phi)$$

cannot be reduced. Data that are NMAR are substantially more difficult to handle than data that are MCAR or MAR. An example that could occur in a clinical study would be a study whose outcome is some sort of lab value, where the machine performing the
analysis is only able to report values of a particular range, say 0.5 – 250 mg/dL. Patients whose values are too high or too low would not have recorded values.

Methods for Dealing with Missing Data

The first thing to consider when attempting to analyze data with missing values is whether the missing data can be classified as ignorable or non-ignorable. Ignorable missing data result in estimates that are typically unbiased, if the correct analysis methods are used. Non-ignorable missing data results in estimates that are biased. In general, MCAR and MAR mechanisms result in ignorable missing data, and NMAR results in non-ignorable missing data, though it is possible for MAR data to be non-ignorable. Most methods for dealing with missing data apply to ignorable missing data. Methods for non-ignorable missing data (NMAR) are more complex and will not be considered in this paper. Although it is often possible to determine if a particular missing data mechanism is MCAR or MAR, it is not possible to distinguish between MAR and NMAR (NRC, 2010).

The first method is to disregard any observations with missing data and to perform the analyses on the completely recorded observations. This is known as complete case analysis (CCA). This method is easy to perform, but is only realistic in situations where the amount of missing data is small, and the missing data mechanism is MCAR. If the missing data are not MCAR, or if a large percent of observations have missing values, then CCA is likely to result in biased estimates.

A second method for dealing with missing data is single imputation, where missing values are imputed with the best available guess for the missing value. The
imputation could be based on the observed values and covariates of the observation with missing data, or could be based on the observed values of the entire dataset. The final result of using single imputation is a completed dataset, which can then be analyzed using standard methods. A major limitation of using single imputation is that standard variance formulas applied to the filled-in dataset will underestimate the variance of the estimates (Rubin and Little, 2002). This can result in confidence intervals that are too narrow and significance tests with higher probabilities of Type I errors. There are several methods of single imputation, but this paper focuses on the method of last observation carried forward (LOCF). This method can be used in longitudinal studies, where participants drop out at some time point. It replaces the missing values with the latest observed value of that participant. This method is typically only valid if the missing data mechanism is MCAR. Furthermore, even if the mechanism is MCAR it is likely to result in biased results, unless the outcome variable is assumed to remain constant over time, which typically is not a valid assumption in clinical studies. Despite its limitations, many investigators choose to use LOCF because of the justification that it usually results in conservative results. This generally results in the under-estimation of treatment effects and the decreased likelihood of obtaining statistically significant results.

An extension of single imputation is multiple imputation (MI). This involves repeating the imputation process $D$ times so that $D$ data sets are created with imputed missing values. This results in $D$ sets of parameter estimates, which are then combined to form the final estimates. An assumption of MI is that the imputation methods results in slightly different estimates for each of the created data sets. The benefit of performing
multiple imputation over single imputation is that it accounts for the variability associated with imputing the missing values. This results in better estimates of standard errors, and thus more accurate confidence intervals and statistical tests. Depending on the amount of missingness, as little as 5 to 10 imputations are enough to reach valid results (Yuan, 2000).
Given the importance of properly handling missing data in clinical trials, many studies and subsequent journal articles have been done to study and compare the effects of the different methods of analyzing data sets that contain missing values. Of particular interest for this paper are those that focus on the effects of using last observation carried forward in a clinical trial setting.

Gadbury et al. (2003) studied the effect of imputation methods on a simulated longitudinal clinical trial with control and treatment arms. The trial simulated a weight loss study with four time points. The ultimate parameter of interest was the mean difference between the two groups at the final measurement. Data were drawn from a multivariate normal distribution, with a mean vector for the treatment group equal to $\mu_T = [92,88,85,82]$ and for the control group equal to $\mu_C = [92,90,90,89]$. Both were assumed to have standard deviations equal to $[14.3, 14.0, 14.2, 13.9]$. The expected treatment effect was a 7 kg additional weight loss in the treatment group.

For the missing data pattern, all subjects had measurements for the first two visits, and then missingness was imparted in a monotone pattern – once subjects have a missing value, they do not return for subsequent measurements. For the MCAR mechanism, all subjects had the same probability of missing visits 3 and 4. For the MAR mechanism, those who had gained weight in the previous measurement were more likely to be missing – missingness depends on the observed data. For the NMAR mechanism,
subjects who gained weight during visit 3 or 4 were more likely to miss that visit – missingness depends on the unobserved data.

The results showed that CCA worked well for estimating the treatment effect when MCAR, but underestimated the effect when MAR or NMAR, with MAR performing worse than NMAR. LOCF underestimated the effect in all situations, with MCAR performing the worst, and NMAR the best. Multiple imputation produced unbiased results for all three mechanisms. The mixed effects model slightly overestimated the effect in all three, but the estimates were less than 3% higher for each mechanism. The results of the study indicate that CCA is only a viable option for data that are MCAR, and that LOCF consistently produced biased results.

The work of Lane (2008) also attempted to study the effects of using LOCF in a longitudinal trial. He also used simulated data, but the datasets were simulated to match six actual trials. He then imposed several drop-out mechanisms and compared the results of LOCF with a mixed model for repeated measurements (MMRM) imputation method by comparing the bias in the treatment difference and the power of treatment comparison. His results indicated that with equal drop-out rates in the treatment and placebo arms, LOCF generally underestimated the treatment effect. But with unequal drop-out rates, the bias generally had a much larger absolute magnitude, and would often overestimate the treatment effect. The MMRM generally produced much smaller biases than LOCF, and performed adequately in most of the scenarios except when the missingness was NMAR and when there were substantially different drop-out rates between the two groups. In terms of power loss, MMRM rarely resulted in a loss of power of more than
20%, whereas LOCF caused a loss of more than 20% power in approximately half of the simulations.

In 2010, the Panel on the Handling of Missing Data in Clinical Trials, which was formed by the National Research Council, released a report with recommendations related to missing data in clinical trials (NRC, 2010). The panel concluded that it is possible to have more structured approaches to designing and analyzing clinical trials to reduce the effect of missing data. The report focused on two elements of missing data. The first emphasizes the importance of carefully planning the studies to reduce the likelihood of missing data. They emphasized the importance of predicting which patients are likely to withdrawal and providing incentives for them to remain in the study, such as the use of rescue medications. They also recommend that investigators continue to collect information of subjects who discontinue the treatment. The second element of the report focused on the actual analysis of data with missing values. They emphasized the importance of stating in the protocol which method for handling missing data would be used in the trial. They also concluded that “single imputation methods like last observation carried forward and baseline observation carried forward should not be used as the primary approach to the treatment of missing data unless the assumptions that underlie them are scientifically justified.” The panel concludes with the recommendation that more research and funding be devoted to developing new methods to handle missing data, as well as to ensure that analysts are continually trained on implementing the new methods.
Despite the known problems with LOCF, it is still the imputation method used in many clinical trials. A simple search of clinical trial articles with repeated measurements and a continuous outcome indicated that of the reports that explicitly specified how missing data was handled, over one third used LOCF. Woolley et al. (2009) attempted to determine the actual prevalence of LOCF in clinical trials. They reviewed the results of 352 antidepressant studies published between 1965 and 2004. Trials were limited to those with randomized designs studying depression where the publication was the first report of the trial. The results indicated that the count and percent of trials using LOCF have increased over time. Overall, 52% of the reports acknowledged that they used LOCF. The study also found that the LOCF trials had significantly more subjects enrolled per trial than the non-LOCF trials, and also had larger proportion of subjects drop-out per trial.
Chapter 3: Methods

A simulation study was run in SAS Version 9.2 to compare the methods of analysis using complete case analysis, LOCF, regular multiple imputation using regression, and two versions of multiple imputation using regression where everyone was imputed as if they were in the control group, a process known as placebo-imputation. This was all accomplished by creating a macro which uses the various parameters described in the next section.

The first portion of the macro creates the data, which includes a treatment indicator, baseline observation, follow-up observation and an unspecified covariate X. The second portion creates missing indicators for the follow-up visit, and creates a new follow-up observation that is recoded as missing based on the missing indicator for that subject. The next portion carries out the four imputation methods mentioned in the previous section. The macro then analyzes the datasets created by the imputation methods to determine the estimated treatment effect using linear regression. It also performs a complete case analysis, as well as an analysis of the pre-deletion data to determine the actual effect seen in the generated data. Finally, it determines the bias observed in the parameter estimate by comparing it to the expected results.

Each macro results in the creation of 1000 datasets, and thus 1000 estimates of the treatment effect parameter. The macro stores the results of each analysis, resulting in a dataset that contains the estimates for each of the 6 analysis methods. The average bias
was calculated for each analysis, as well as the proportion of times the analyses resulted in a statistically significant treatment effect.

Models

The simulated study was a randomized controlled trial with a control group and treatment group, both having an equal number of subjects. Each subject had two observations: a baseline value and a follow-up value. The baseline data were fully observed. Missingness was generated for the follow-up data.

The variables included:

\[ y_{ijt} = \text{outcome for the } i\text{th subject in the } j\text{th group at time } t, \text{ where} \]

\[ i = 1, \ldots, n \]

\[ n = \text{number of subjects in each group} \]

\[ j = \{0,1\} \]

\[ t = \{1,2\} \]

\[ z_j = \text{treatment indicator} = \begin{cases} 
0 & \text{control} (j = 0) \\
1 & \text{treatment} (j = 1) 
\end{cases} \]

\[ x_{ij} \sim \text{Binomial}(p) = \text{covariate} \]

\[ m_{ijt} = \text{missing data indicator} = \begin{cases} 
0 & y_{ijt} \text{ observed} \\
1 & y_{ijt} \text{ missing} 
\end{cases} \]

Both the baseline value and X were assumed to be independent of treatment.

The baseline value was simulated to follow a normal distribution, such that

\[ y_{ij1} \sim N(\mu, \sigma_1^2) \]

The follow-up values were simulated based on \( Y_1, Z \) and X.

For the control group (Z=0):
\[ y_{i02} | y_{i01} \sim N(y_{i01} - \Delta_c, \sigma_2^2) \]

where \( \Delta_c \) = average response for the control group.

For the treatment group (\( Z=1 \)):

\[ y_{i12} | y_{i11} \sim N(y_{i11} - x_{i1}\Delta_R - (1 - x_{i1})\Delta_{NR}, \sigma_2^2) \]

where \( \Delta_R \) = average response for the treatment responders \( (x_{i1} = 1) \)

\( \Delta_{NR} \) = average response for the treatment non-responders \( (x_{i1} = 0) \)

The expected results of the simulations are:

\[ E[Y_1] = \mu \]
\[ E[Y_2 | Z = 0] = \mu - \Delta_c \]
\[ E[Y_2 | Z = 1, X=0] = \mu - \Delta_{NR} \]
\[ E[Y_2 | Z = 1, X=1] = \mu - \Delta_R \]
\[ E[Y_2 | Z=1] = \mu - p\Delta_R - (1 - p)\Delta_{NR} \]
\[ E[Y_2] = \mu - \Delta_c/2 - p\Delta_R/2 - (1 - p)\Delta_{NR}/2 \]

Once the treatment indicator, \( X \), baseline value and follow-up value were simulated, missing data indicators were created to determine which observations had missing data at the follow-up value.

MCAR – 40% missingness

MAR – missingness depends on \( X \)

\[ \logit[P(M = 1 | Y_2, X)] = \beta_0 + \beta_1 X \]

NMAR – missingness depends on \( Y_2 \)

\[ \logit[P(M = 1 | Y_2, X)] = \alpha_0 + \alpha_1 Y_2 \]
For the MCAR missing data, the probability of missing is independent of all variables. Thus it is expected that all four combinations of $Z$ and $X$ will have approximately equal amounts of missing data. In the MAR model, those with a value of $X=1$ will have a higher probability of being missing. This should not affect the mean observed response in the control group, but in the treatment group this will result in more responders having missing $Y_2$ values than non-responders, which would lower the average observed response. In the NMAR model, those who have a lower observed response and thus a higher relative $Y_2$ value will have an increased chance of missing data. This will cause more missing data to occur in the control group and non-responsive treatment group than in the responsive treatment group.

Parameters

The parameters defined in the previous section were entered as followed into SAS:

Sample size: $n = 200$ per group

Covariate $X$: $p = 0.5$

Baseline outcome:

$\mu = 0$

$\sigma^2_{Y_1} = 1$

Average response:

Scenario 1: $\Delta_C = 0$

$\Delta_{NR} = 0.25$

$\Delta_R = 0.5$
Scenario 2: \[ \Delta_C = 0.25 \]
\[ \Delta_{NR} = 0.25 \]
\[ \Delta_R = 0.5 \]
\[ \sigma^2 = 1 \]

Missing data models:

MCAR: 40% missing

MAR: \[ \beta_1 = 0.7 \]
\[ \beta_0 = -0.76 \]

NMAR: \[ \alpha_1 = 1.0 \]
\[ a_0 = \begin{cases} -0.1 & \text{Scenario 1} \\ 0 & \text{Scenario 2} \end{cases} \]

The parameters for the missing data models result in the following expected amounts of missing data:

MCAR: 40% missing

MAR: 31.9% when \( X = 0 \)
\[ 48.5\% \text{ when } X = 1 \]

NMAR:

Scenario 1: 47.5% when \( Z = 0 \)
\[ 41.3\% \text{ when } Z = 1, \ X = 0 \]
\[ 35.4\% \text{ when } Z = 1, \ X = 1 \]

Scenario 2: 43.8% when \( Z = 0 \)
\[ 43.8\% \text{ when } Z = 1, \ X = 0 \]
\[ 37.8\% \text{ when } Z = 1, \ X = 1 \]
Imputation and Analyses

For all analyses, the effect of treatment was estimated using linear regression, with the observed and/or imputed $Y_2$ values being the dependent variable and the treatment level $Z$ being the independent variable. From the expected values listed in the Models section of this chapter, the expected treatment effect can be calculated as:

$$E[Y_2|Z=1] - E[Y_2|Z = 0] = -p\Delta_R - (1 - p)\Delta_{NR} + \Delta_C$$

which should be equal to the estimate of the slope coefficient for treatment in the regression model. The percent bias can thus be calculated as the difference in the observed and expected effects divided by the expected effect.

The first imputation method was last observation carried forward (LOCF). All missing $Y_2$ values were imputed as the observed $Y_1$ values. Once a complete data set was obtained using the imputed values, the regression model was fit, and the relative bias was calculated using the estimate of the treatment effect coefficient.

The second method was regular multiple imputation. This was completed in three steps. In the first step, PROC MI was used to create 10 multiply imputed datasets. The REG option was used so that the $Y_2$ values were imputed based on $Y_1$, $X$ and $Z$. In the second step, the regression model was fit by imputation (performed separately for each of the 10 imputed datasets), and the parameter and standard error estimates were saved into a data set. Finally, PROC MIANALYZE was used to combine the 10 estimated
parameters and standard errors into the final estimates. These final estimates were then used to calculate the relative bias.

The next two methods involved performing multiple imputation where everyone was treated as if they were in the placebo group (“placebo-imputation”). The imputation steps are identical to those of the regular multiple imputation, with the exception of the model used to impute the values in step one. In the first version, a new treatment indicator variable \( Z^* \) was created such that:

\[
Z^* = \begin{cases} 
0 & Z = 0 \ (placebo) \\
1 & Z = 1, M = 0 \ (treatment, observed) \\
0 & Z = 1, M = 1 \ (treatment \ missing) 
\end{cases}
\]

\( Y_2 \) was then imputed based on \( Y_1, X \) and \( Z^* \). In the second version of placebo-imputation, a new variable \( Y_2^* \) was created such that:

\[
Y_2^* = \begin{cases} 
Y_2 & Z = 0 \\
missing & Z = 1 
\end{cases}
\]

This results in \( Y_2^* \) consisting only of the observed control values for \( Y_2 \). \( Y_2^* \) was then imputed based on \( Y_1 \) and \( X \). This results in imputation for the missing \( Y_2 \) values for both groups, as well as the non-missing values for the treatment group. Once the imputed values were obtained after step one, they were replaced with the values for \( Y_2 \) that were actually observed.

The final two analyses were the complete case analysis, where the regression model was fit using the observed \( Y_2 \) values with missing values included, and the pre-deletion analysis where the model was fit using the \( Y_2 \) values that were created prior to adding missingness.
Chapter 4: Results

The results for scenarios 1 and 2 are listed in Table 1 and 2, respectively. For the MCAR data, both scenarios resulted in unbiased estimates for the pre-deletion analysis, CCA, and the regular regression MI. In both scenarios, the bias for LOCF and both placebo imputations ranged between -40.3% and -41.0%. For the MAR data, both scenarios produced unbiased pre-deletion data. The CCA and regular MI produced slightly biased results, and were similar within each scenario, with scenario 2 producing a larger relative bias. LOCF and both placebo imputation methods again produced nearly identical results within each scenario, though scenario 2 produced slightly larger biases. For the NMAR data, both scenarios produced relatively unbiased pre-deletion data (-1.8 and -3.9%). The CCA, LOCF, and regular MI analyses for scenario 1 had slightly higher biases than scenario 2 [(-22.8, -43.0, -10.0%) vs (-19.3, -39.7, -8.6%)]. Both placebo-imputations were similar within and between each scenario. Unlike the MAR and MCAR simulations where the LOCF and placebo-imputation were nearly identical, the placebo-imputation in the NMAR simulation had larger biases than the LOCF, particularly in scenario 2.

The average power to detect a significant treatment effect for the pre-deletion data was 86.8% in scenario 1, and 17.4% in scenario 2. All of the imputation methods and CCA resulted in a loss of power. The order of power loss from least to greatest was the
same for all six simulations: regular MI, CCA, LOCF, and placebo-imputation (both 
versions roughly tied).

<table>
<thead>
<tr>
<th></th>
<th>Pre-deletion</th>
<th>CCA</th>
<th>LOCF</th>
<th>Regular MI</th>
<th>Placebo Imputation (Version 1)</th>
<th>Placebo Imputation (Version 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MCAR</strong></td>
<td>% Bias</td>
<td>-0.1</td>
<td>0.1</td>
<td>-40.3</td>
<td>-0.2</td>
<td>-40.3</td>
</tr>
<tr>
<td></td>
<td>% Reject</td>
<td>86.7</td>
<td>67.3</td>
<td>48.9</td>
<td>78.5</td>
<td>37.3</td>
</tr>
<tr>
<td><strong>MAR</strong></td>
<td>% Bias</td>
<td>-0.3</td>
<td>-6.4</td>
<td>-43.5</td>
<td>-5.5</td>
<td>-43.5</td>
</tr>
<tr>
<td></td>
<td>% Reject</td>
<td>87.5</td>
<td>58.4</td>
<td>44.8</td>
<td>74.8</td>
<td>30.8</td>
</tr>
<tr>
<td><strong>NMAR</strong></td>
<td>% Bias</td>
<td>-1.8</td>
<td>-22.8</td>
<td>-43.0</td>
<td>-10.0</td>
<td>-47.4</td>
</tr>
<tr>
<td></td>
<td>% Reject</td>
<td>86.2</td>
<td>50.4</td>
<td>44.6</td>
<td>73.0</td>
<td>32.0</td>
</tr>
</tbody>
</table>

Table 1. Estimated treatment effect bias and power for Scenario 1

<table>
<thead>
<tr>
<th></th>
<th>Pre-deletion</th>
<th>CCA</th>
<th>LOCF</th>
<th>Regular MI</th>
<th>Placebo Imputation (Version 1)</th>
<th>Placebo Imputation (Version 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MCAR</strong></td>
<td>% Bias</td>
<td>-1.0</td>
<td>-3.2</td>
<td>-40.6</td>
<td>-2.3</td>
<td>-40.5</td>
</tr>
<tr>
<td></td>
<td>% Reject</td>
<td>17.4</td>
<td>11.8</td>
<td>9.1</td>
<td>13.5</td>
<td>5.0</td>
</tr>
<tr>
<td><strong>MAR</strong></td>
<td>% Bias</td>
<td>1.1</td>
<td>-9.1</td>
<td>-47.9</td>
<td>-11.8</td>
<td>-47.5</td>
</tr>
<tr>
<td></td>
<td>% Reject</td>
<td>17.7</td>
<td>11.0</td>
<td>9.5</td>
<td>13.7</td>
<td>5.3</td>
</tr>
<tr>
<td><strong>NMAR</strong></td>
<td>% Bias</td>
<td>-3.9</td>
<td>-19.3</td>
<td>-39.7</td>
<td>-8.6</td>
<td>-48.0</td>
</tr>
<tr>
<td></td>
<td>% Reject</td>
<td>17.2</td>
<td>11.2</td>
<td>10.3</td>
<td>13.7</td>
<td>4.8</td>
</tr>
</tbody>
</table>

Table 2. Estimated treatment effect bias and power for Scenario 2
Chapter 5: Discussion

The main difference between scenario 1 and 2 was that in scenario 1, the controls were not expected to receive any benefit from the trial, whereas in scenario 1 they were expected to have the same response as the non-responsive treatment subjects. This results in a substantially different expected treatment effect. In scenario 1, those receiving treatment are expected to have an average response of 0.375, compared to the 0 response of the control group. In scenario 2, the treatment group is also expected to have an average response of 0.375, but the control group is expected to have a response of 0.25, resulting in a 0.125 average treatment effect. This would substantially reduce the power to detect a significant difference in scenario 2 compared to 1, given both have equal sample size. The results validate this assumption since the average power of the three simulations in the pre-deletion groups of scenario 1 was 86.8%, whereas in scenario 2 the average power was only 17.4%.

As stated in chapter one, a complete case analysis is only reasonable if the missing data mechanism is MCAR. For both scenarios, we found the CCA to perform well in estimating the treatment effect when MCAR, although both experienced a loss of power. We expect the CCA to be biased for both the MAR and NMAR mechanism, which is observed in both scenarios. The results were substantially more biased for NMAR compared to MAR in both scenarios.
As the models used to simulate the data assume there will be a change in the outcome over time for all subjects, with the exception of the control group in scenario 1, it was assumed LOCF would not perform well in this study. The results of this study agree with this assumption, as all of the simulations had treatment effects that were biased by -39.7 to -47.9% when using LOCF. Most of the estimates were similar for the different simulations, with the exception of the MAR-Scenario 2 simulation, which performed worse.

We would expect the regular multiple imputation using regression method to work well with MCAR data, which was the case for both scenarios. We would expect it to perform reasonably well for the MAR data, since the regression method includes X as a predictor, which is the covariate used to determine the chance of missingness in the MAR simulations. We would expect it to perform the worst in the NMAR mechanism. We found that scenario 1 produced the expected results, but in scenario 2 the MAR data were actually more biased than the NMAR.

The two versions of placebo-imputation performed poorly in all of the simulations. We would expect it to perform slightly better in scenario 2, since half of the treatment subjects had expected values identical to the controls; but the results indicate that the method performed similarly in both scenarios, with scenario 2 actually having results that were slightly more biased. There is very little published work on the uses of placebo-imputation and more investigation are needed to determine if it is a reasonable method to use.
Conclusion

Although easy to perform, the imputation method of last observation carried forward produces consistently biased results when used in a longitudinal clinical trial setting. While the results in this study and others have indicated it tends to underestimate the treatment effect, there are some studies where it over-estimates the effect, as demonstrated by the work of Lane (2009). Thus, it is not a proper justification to state that LOCF is reasonable to use because it diminishes the treatment effect, since this is not guaranteed to be the situation. With the easy availability of computer programs that can perform more sophisticated methods of handling missing data, there is no reason to use LOCF.
References


Appendix A: SAS Code

/********************************************
/* MACRO to simulate and analyze */
/* a randomized trial with two */
/* time points, missing data and */
/* two arms (treatment / control) */
/* Author: Tara Carmack */
/* Last Modified: 2012/5/16 */
/********************************************

/******* VARYING PARAMETERS
 NREPS = number of replicates
 MISS=Missingness pattern (mcar, mar, nmar)
 SEED = starting seed value
 N=number of observations in each group
 PROBX=probability that X=1
 Y1MEAN=mean of baseline values
 Y1VAR=variance of baseline values
 DELTAC=mean change in control group
 DELTANR=mean change in trt group when X=0 (nonresponders)
 DELTAR=mean change in trt group when X=1 (responders)
 DELTAVAR=variance of change (assuming same for all three values)
 MCARPCT=Probability of missingness for MCAR
 ALPHA0=Intercept coefficient used to determine NMAR probability of
 missingness
 ALPHA1=Slope coefficient used to determine NMAR probability of
 missingness
 BETA0=Intercept coefficient used to determine MAR probability of
 missingness
 BETA1=Slope coefficient used to determine MAR probability of missingness
 */

%macro sim(NREPS,MISS, SEED, N, PROBX, Y1MEAN, Y1VAR, DELTAC, DELTANR, DELTAR, DELTAVAR, MCARPCT, BETA0, BETA1, ALPHA0, ALPHA1);
ods listing close;
proc datasets library=work;
delete results_simple results_MI;
run;
data results_simple;
  stop;
runt;
data results_MI;
  stop;
*/
run;

%syscall streaminit(seed); /* set seed first pass through */
%do rep=1 %to &NREPS;

/*Create baseline value, treatment indicator, and X */
data create (drop=trt obs);
  do trt = 0 to 1;
    do obs = 1 to &N;
      treatment=trt;
      X = rand('Bernoulli',&PROBX);
      Y1 = rand('Normal',&Y1MEAN,sqrt(&Y1VAR));
      output;
    end;
  end;
run;

/* create outcome data */
data create;
  set create;
  if treatment=0 then Y2=Y1 -
    rand('Normal',&DELTAC,sqrt(&DELTAVAR));
  else if X=1 then Y2=Y1 -
    rand('Normal',&DELTAR,sqrt(&DELTAVAR));
  else Y2=Y1 - rand('Normal',&DELTANR,sqrt(&DELTAVAR));
run;

/* create missingness indicators */
data create;
  set create;
    M_mcar = rand('Bernoulli',&MCARPCT);
    plogit_mar=&BETA0+&BETA1*X;
    prob_mar=exp(plogit_mar)/(1+exp(plogit_mar));
    M_mar=rand('Bernoulli',prob_mar);
    plogit_nmar=&ALPHA0+&ALPHA1*Y2;
    prob_nmar=exp(plogit_nmar)/(1+exp(plogit_nmar));
    M_nmar=rand('Bernoulli',prob_nmar);
    if M_mcar=1 then Y2_obs_mcar=.;
    else Y2_obs_mcar=Y2;
    if M_mar=1 then Y2_obs_mar=.;
    else Y2_obs_mar=Y2;
    if M_nmar=1 then Y2_obs_nmar=.;
    else Y2_obs_nmar=Y2;
run;

/********** impute missing values **********/

/* LOCF */
data create;
set create;
if M_&&MISS=1 then Y2_locf_&&MISS = Y1;
   else Y2_locf_&&MISS = Y2;
run;

/*Proc MI*/

/* Method A: missing values imputed based on X, treatment, and Y1 */
proc MI data=create out=mioutA NIMPUTE=10;
   monotone method=reg;
   var treatment X Y1 Y2_obs_&&MISS;
run;

/* Method B: recode missing treatments as controls. estimates based on X, Y1, and recoded treatment */
data create;
   set create;
   if treatment=0 then treatmentB=0;
      else if M_&&MISS=1 then treatmentB=0;
      else treatmentB=1;
run;
proc MI data=create out=mioutB NIMPUTE=10;
   monotone method=reg;
   var treatmentB X Y1 Y2_obs_&&MISS;
run;

/* Method C: recode Y2_obs as missing if treatment estimates new Y2_obs by X and Y1 recode Y2_obs as original value if original value is observed*/
data create;
   set create;
   if treatment=0 then Y2_obs_C_&&MISS = Y2_obs_&&MISS;
      else Y2_obs_C_&&MISS=.;
run;
proc MI data=create out=mioutC NIMPUTE=10;
   monotone method=reg;
   var X Y1 Y2_obs_C_&&MISS;
run;
data mioutC;
   set mioutC;
   if M_&&MISS=1 then Y2_obs_&&MISS=Y2_obs_C_&&MISS;
run;

/********** Analyze data***********/

/* Complete Case Analysis */
proc reg data=create ;
model Y2_obs_&&MISS=treatment;
ods output ParameterEstimates=temp;
run;

data temp;
   set temp;
   length method $ 12;
   repetition=&rep;
   method="CCA";
run;
data results_simple;
   set results_simple temp;
run;

/* LOCF */
proc reg data=create;
   model Y2_locf_&&MISS=treatment;
   ods output ParameterEstimates=temp;
run;

data temp;
   set temp;
   repetition=&rep;
   method="LOCF";
run;
data results_simple;
   set results_simple temp;
run;

/* Pre-deletion (no missing data)*/
proc reg data=create;
   model Y2=treatment;
   ods output ParameterEstimates=temp;
run;

data temp;
   set temp;
   repetition=&rep;
   method="Pred";
run;
data results_simple;
   set results_simple temp;
run;

/* Multiple Imputation*/
/* Method A */
proc reg data=mioutA noprint outest=regout covout;
   model y2_obs_&&MISS = treatment;
   by _imputation_;
run;
run;

proc mianalyze data=regout;
   modeleffects intercept treatment;
   ods output parameterestimates=param VarianceInfo=var;
run;

data param;
   set param;
   repetition=&rep;
run;
data var;
   set var;
   repetition=&rep;
run;
data combine;
   merge param var;
   method="MI_A";
run;
data results_mi;
   set results_mi combine;
run;

/* Method B */
proc reg data=mioutB noprint outest=regout covout;
   model y2_obs_&&MISS = treatment;
   by _imputation_; 
run;

proc mianalyze data=regout;
   modeleffects intercept treatment;
   ods output parameterestimates=param VarianceInfo=var;
run;

data param;
   set param;
   repetition=&rep;
run;
data var;
   set var;
   repetition=&rep;
run;
data combine;
   merge param var;
   method="MI_B";
run;
data results_mi;
   set results_mi combine;
run;
/* Method C */
proc reg data=mioutC noprint outest=regout covout;
  model y2_obs_&&MISS = treatment;
  by _imputation_;
run;

proc mianalyze data=regout;
 modeleffects intercept treatment;
  ods output parameterestimates=param VarianceInfo=var;
run;

data param;
  set param;
  repetition=&rep;
run;

data var;
  set var;
  repetition=&rep;
run;

data combine;
  merge param var;
  method="MI_C";
run;
data results_mi;
  set results_mi combine;
run;

dm 'clear log';

%end;

/*** calculate relative bias ****/
data results_mi;
  set results_mi;
  where parm="treatment";
  relbias=(estimate-(\&PROBX*\&DELTAR-(1-\&PROBX)*\&DELTANR+\&DELTAC))/(-\&PROBX*\&DELTAR-(1-\&PROBX)*\&DELTANR+\&DELTAC);
run;

data results_simple;
  set results_simple;
  where variable="treatment";
  relbias=(estimate-(\&PROBX*\&DELTAR-(1-\&PROBX)*\&DELTANR+\&DELTAC))/(-\&PROBX*\&DELTAR-(1-\&PROBX)*\&DELTANR+\&DELTAC);
run;

/*** calculate mean relative bias, proportion rejected ***/
ods listing ;
data results_mi;
  set results_mi;
if probt<=0.05 then reject=1;
    else reject=0;
run;
data results_simple;
    set results_simple;
    if probt<=0.05 then reject=1;
    else reject=0;
run;

proc means data=results_simple;
    var relbias reject;
    class method;
    run;
proc means data=results_mi;
    var relbias reject;
    class method;
    run;

%MEND;