Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display n the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

________________________________ ____________________

Tahera Darensburg Date

PRISE STUDY: A missing data problem

Tahera Darensburg

Masters of Science

Biostatistics

Vicki S. Hertzberg, Ph.D.

Advisor

DuBois F. Bowman, Ph.D.

Advisor

Accepted:

Lisa A. Tedesco, Ph.D.

Dean of the Graduate School

Date

PRISE STUDY: A missing data problem

By

Tahera Darensburg

B.S., Xavier University of LA, 2004

Advisor: Vicki S. Hertzberg, Ph.D.

An abstract of

A thesis submitted to the Faculty of the Graduate School of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science

in Biostatistics

Abstract

PRISE STUDY: A missing data problem

By Tahera Darensburg

The PRISE study was an intervention to encourage African-American women of Grady Healthcare to improve their fitness. This study was subject to large drop outs, presenting a particular problem for analysis. This problem with missing data inspired drop-out analysis and a comparative analysis of imputation. We will analyze the drop-out process as an outcome and determine if any factors measured in the study are associated with drop-out. For imputation, the methods we compare are Average (Avg), Previous (Prev), Post (Post), Last and Next (LaN), Last Value Carried Forward (LVCF), and Next Value Carried Backward (NVCB). We will compare the methods given varying amounts of missingness. We give evidence to support that simple imputation methods can be an effective way to deal with MAR data. The best imputation methods as found in this study were LaN and Post. As the amount of missingness increased both LaN and Post maintained relatively low bias, (proportionate variance) PV close to 1, and consistent variance.

PRISE STUDY: A missing data problem

By

Tahera Darensburg

B.S., Xavier University of LA, 2004

Advisor: Vicki S. Hertzberg, Ph.D.

A thesis submitted to the Faculty of the Graduate School of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science

in Biostatistics

2008

Table of Contents

List of Tables

List of Figures

1. Introduction

1.1 Background: PRISE Study

The purpose of the PRISE intervention was to encourage employed African-American women to initiate and sustain long-term physical activity. Our study population was taken from Atlanta's Grady Health System's 3,000 African-American employees aged 18-50 years old. The study was comprised of two segments: an initial qualitative, cross-sectional component with data collected from various questionnaires and a longitudinal intervention based on Dishman's Theory.[5] PRISE is an acronym that summarizes Dishman's Theory, which stands for **P**reps, **R**eps, **I**ncreased **S**teps, and **E**ncouragement. Each component is outlined in detail below.

The **preps** component focuses on preparing the women for success by reducing the effect of personal barriers that interfere with starting a physical activity program and maintaining healthy eating. **Reps** refers to the resistance and weight training portion of the program. **Increased steps** includes setting and obtaining modest goals for physical activity over the 6-month intervention period. Lastly, the **encouragement** element of the PRISE intervention is meant to improve the likelihood of success in the program with social support and personalized feedback.

Given the overarching health disparities partially attributable to lack of physical activity and obesity, the results of the PRISE study hold considerable value. The purpose of the PRISE study was to show the effectiveness of a work-based intervention and to extend our scope to the health benefits of starting a physical fitness regimen. However, one of the hurdles of this study was the same as previous studies with

African-American women: high drop-out rates, lack of participation, and modest results.[9,13,25,26] These particular issues result in unbalanced data and biased results, presenting considerable complications to completing analysis. The PRISE study served as a motivation to explore reliable techniques for analyzing problems with missing data, and analyze drop-out as an outcome. These results could be applied to similar health studies with high amounts of dropout that focus on African-American women.

1.2 Missing Data

As illustrated by data from the PRISE study, missing data is a common complication to completing analysis. The first step to dealing with missing data is to first determine the nature of the missingness in the data. Three main classifications of missing data are missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR).[6,7]

Utilizing our data we consider Y_i as our dependent variable that represents average daily steps for the women at time i=1,2,3,4 and for the purposes of illustration use X as an arbitrary corresponding independent variable. For data that is MCAR the probability that observation Y_i is missing is unrelated to Y_i observed or missing, however, Y_i may depend on a particular X.[6,7] The PRISE step log data would be MCAR if say a patient had forgotten their pedometer on a particular day and was unable to report. Analysis tends to be unbiased with MCAR data.[6,7]

With MAR data missingness depends on observed Y_i values, but not missing Y_i values. For example, participants with lower ADS may be less likely to report. Particular characteristics of participants may be related to lower ADS. With MAR data we may find that missingness does not depend on Y_i after controlling for another variable. In the case of MAR data we find that bias is introduced. Continuing with our above example if younger participants had lower ADS and thus were less likely to report data, summary statistics and models will be biased since not containing much data from younger participants. Younger participants might be more in shape and have higher average daily steps (ADS); without these measurements our summary statistics might be underestimated. [6,7]

The last classification of data that we will consider will be MNAR. If data is not MCAR or MAR it is MNAR. MNAR data occurs when missingness depends on both observed and missing variables. With MNAR bias is introduced as well.[6,7] Given the type of missing data that a problem has there are various methods that Statisticians have proposed to deal with this issue. The methods that we will look at involve manipulating the data to result in a new dataset that is able to be utilized for analysis.

One method is listwise deletion, where cases are omitted with missing data points. One drawback to this method is this method will result in a much lower sample size. It will also still result in biased results with MAR and MNAR. With the PRISE study, there is so much missing data that listwise deletion is not a wise method, because it will significantly decrease our sample size.

Some statistical software packages use a method called pairwise deletion. With this method subjects are only included in calculations where they have available data. Specifically, when intercorrelation matrices are estimated using all available data. If age and weight parameters of interest only subjects with both measurements available will be included, all others will be deleted. This method still has negative aspects; the parameters of the model will be based on different sets of data with different sample sizes and standard errors. Additionally, with this method it is possible to have an intercorrelation matrix that is not positive definite.

A third collection of methods that are commonly utilized are imputation methods. Imputation is a process that involves replacing missing data in a systematic and realistic way in order to result in a dataset that can be utilized for analysis. There are various imputation methods, some simple and some more sophisticated. There have been several studies that compare different imputation methods and the respective effectiveness of each.[2,12,24] With our research we propose to show that simple imputation methods can be a very effective way of dealing with missing data. These methods are less complicated and perform very well, making them an excellent choice for imputation.

The imputation methods that we will compare are average (Avg), average of previous (Prev), average of future (Post), last and next (LaN), last value carried forward (LVCF), and next value carried backward (NVCB). With the Avg method we replace missing data points with the average for each respective subject.[6] With Prev we replace the missing data with the average of all the preceding values and with Post it is replaced with the average of all of the subsequent values.[6] The next method is the LaN method, where we average the last value and the next value and use that value to replace a missing value.[6] The LVCF method takes the last observed data point and carries it forward to replace a subsequent missing value. Similarly, the NVCB method takes the next value and uses it to replace preceding missing values.[6]

All imputation methods can not be utilized with all instances of missing data. Only the Avg method results in all missing values being replaced. With Prev an initial instance of missing data can not be replaced, and likewise with Post a terminal instance of missing data can not be replaced. LaN, LVCF, and NVCB can not be achieved when there is no existing 'last' or 'next' value. We consider these drawbacks to be minimal when looking at our data. With our particular dataset we have ADS from 31 time points for a 6-month period. All methods would still result in new datasets with a large number of missing data filled in.

With the PRISE study we find intermittent missing data and dropouts. With the amount of missing data that we have it would be beneficial to attempt to fill in some of the missing data. These imputation methods will be applied to our dataset and compared as we simulate varying amounts of missing data in an effort to show that simple imputation methods can be an effective way to deal with MAR data.

2. Methods

2.1 Data Collection

At baseline we will perform an assessment including: a survey instrument to collect demographic data, data from lab results, 2-week self-reported daily steps log, and a treadmill stress test. These results are utilized for the inclusion/exclusion criteria.

To collect psychological data the women will complete the following questionnaires: Exercise Benefits/Barriers Scale (EBBS), Jackson, Hogue, Phillips Contextualized Stress Measure (JHP), Strate, Trait and Anger Expression Inventory (STAXI II), Self-

Efficacy for Exercise Behaviors Scale, Brief symptom survey (BSI), DR Index, and Life Changes Scale. These questionnaires assess perceived barriers and benefits to fitness, measure perceived stress, anger, self-efficacy, and life changing events.[1,3,20,21]

Throughout the study the participants will turn in step logs of their ADS as measured by pedometer. At the end of the study the women will complete EBBS, Self-Efficacy for Exercise Behaviors Scale, BSI, Life Changes Scale, again for follow-up psychological data. They will also undergo a final treadmill stress test.

2.2 Analysis

2.2.1 Overall Analysis

We will calculate descriptive statistics including, but not limited to, age, cholesterol, resting BP, fasting glucose, BMI, ADS, cardiovascular fitness at baseline overall and then stratified by whether the participant had completed the study. Then we complete comparative analysis for participants who completed the program versus those who did not by constructing multiple linear regression models with interaction terms. We constructed models looking at demographic, physiological, and psychological measures separately.

2.2.2 Drop-out analysis

Motivated by previously performed drop-out analysis, we treat drop-out as an outcome and complete a simple analysis of drop-out.[4,14,22,23] We investigate whether the event of dropping out of the study is significantly associated with demographic characteristics (age, annual income, marital status, household size, etc.), physiological characteristics (BMI, diabetes, cardiovascular fitness etc.), or psychological characteristics (EBBS, JHP, and other scales). Our primary means of analysis will be a logistic regression of dropout on the independent variables comprising the measures listed above. We will additionally calculate descriptive statistics on important measures by drop-out.

2.2.3 Imputation

We simulate data using information from our actual data. For all subjects N=250 we simulate data for Y_1, Y_2, Y_3, and Y_4 to represent ADS from week one through week four. Given the characteristics of our data we find that our data fits MAR. We will utilize information from our drop-out analysis and select one factor related to drop-out. We will extend our assumption, for the purposes of our simulation, that this variable is also related to not reporting ADS. Then we will vary the amount of missingness utilizing this fact.

Consider marital status as the variable that we will utilize. In addition to Y_i values each subject will also have marital status. If married women are more likely to not report ADS we will give them a higher probability of missing data than the other participants. We will accomplish this by randomly generating a number from the uniform distribution from (0, 1) for each observation. If we would like 10% missing observations getting a value from 0 to 0.10 will be deleted.

We will vary the amount of missingness (10%, 15%, 20%, and 25%) compare the effectiveness of different imputation methods Avg, Prev, Post, LaN, LVCF, and NVCB utilizing bias and proportionate variance.

$$bias = \frac{\sum (y - \hat{y})}{m}$$

$$PV = \frac{\text{var}(\hat{y})}{\text{var}(y)}$$

We define $\hat{y}$ as the imputed value, y is is the actual value, and m is the number of missing values. A positive bias indicates underestimated imputed values. PV is the proportionate variance. It is a ratio of the observed variance to the true various used to assess under dispersion. When the PV=1 the variance of the imputed values is equal to the true values. A PV value of less than one indicates underestimation. We compute these values for each participant and find the mean across all participants.

2.2.4 Analysis on Complete Data

We complete the analysis on the $N = 83$ participants who completed the program. Although these results are biased and are not accurate study results, they serve as a comparison for our participants who did not complete the program. We will calculate descriptive statistics including, but not limited to, age, cholesterol, resting BP, fasting glucose, BMI, ADS, cardiovascular fitness at baseline and EOS. We will attempt to determine if a significant proportion of the variance in ADS will be accounted for by measures of perceived barriers, perceived benefits, perceived stress, environmental stressors, stress coping factors, personality factors, chronological age, and internalized.

Second we will investigate whether a significant proportion of the variance in improvement in ADS will be accounted for by change in perceived barriers, perceived stress, and self-efficacy compared to base ADS. We will evaluate this hypothesis using several different analytical strategies. Lastly, we will investigate if a significant proportion of the variance in improvement in fitness will be accounted for by change in ADS compared to baseline cardiovascular fitness.

3. Results

3.1 Study population

Our study population is composed of primarily African-American women (88.57%) that work at Grady Memorial Hospital whose average age is 38. The average size of household is 3 and most annual household incomes range from $21,000-80,000 for 59.96% of the women. The majority of the women are non-smokers (88.81%) and 81.46% are not heavy drinkers (only drink occasionally or drink less than once a week). The average weight is 198.34 lbs and average height 64.74 in (almost 5'5"). The 93.93% of the women work full time (>36 hours a week). Table 1 displays demographic characteristics stratified by whether participants completed the program. We find these results to be similar to the average for the entire study population. We find differences in age, weight, and household size.

Table 1. Demographic Characteristics Stratified by Completion

Variable	Non-Completers Mean (±SD) *N=209*	Completers Mean (±SD) *N=270*
Age	35.92 (8.57)	40.27 (7.99)
Weight (lbs)	195.09 (48.91)	199.63 (52.31)
Height (in)	64.64 (3.06)	64.75 (2.72)
Ideal Weight (lbs)	152.14 (24.91)	151.19 (21.80)
Household Size	2.88 (1.76)	3.21 (2.13)
Number of Children	1.04 (1.16)	1.06 (1.19)

*These measurements taken at intake were self reported

There's a low percentage of women in the study that have been diagnosed with heart disease or stroke (<1%), however 6.91% have had a father or brother with MI and 7.29% have had a mother with MI. An overwhelming 5.71% have diabetes and 5.67% of them treat with medicine, 23.89% have hypertension and 22.67 treat with medicine, 13.82% diagnosed with high cholesterol of 4.88% of them treat with medicine, 18.85% are medically considered as obese.

Most of the women are unhappy with their current weight (92.21%) and have tried to lose weight previously 91.53%. The majority (70.59%) have regained some or all of the weight they lost, once or more times-The average amount of weight lost was 28.28 lbs and average amount regained was 13 lbs. The most common methods for weight loss were dieting alone (49.80%), exercising alone (49.00%), diet and exercise (76.89%), pills (39.84%) and frequently skipping meals (38.25%). Atkins (26.69%) and SlimFast (35.86%) were the most popular diets that participants had tried. Weight watchers (27.89%) and Jenny Craig (8.76%) were the most popular formal weight loss programs that participants had enrolled in before.

Only 27.46% of participants currently exercised regularly. 37.76% walk 1-2 days a week, 47.20% do aerobics 1-2 days a week, and 38.40% use exercise machines 1-2 days a week. Most common reasons for exercising were to lose weight (44.80%) and stay healthy (41.20%). The reasons stated for not exercising: not having enough time (39.60%), too tired (40%), and have to take care of family (20.40%).

3.2 Physiological Measures

Several physiological measures were taken at baseline. Table 3 displays some of the physiological variables measured stratified by whether participants had completed the program. Participants who dropped out or were lost to follow-up on average had higher average daily steps (ADS) 7281 versus 6156. BMI tended to be less for non-completers (32.79) than completers (33.32). Total cholesterol was slightly higher in completers (179.04) than non-completers (176.32). Glucose is very similar in both groups. METs are similar as well with completers being slightly higher (7.61) than non-completers (7.53).

Table 2. Physiological Measures Stratified by Completion

Variable	**Non-Completers Mean (±SD)** ***N=209***	**Completers Mean (±SD)** ***N=270***
ADS	7281.91 (3391.75)	6155.88 (2196.35)
BMI	32.79 (7.54)	33.32 (8.24)
Total Cholesterol	176.32 (32.60)	179.04 (33.62)
Glucose	95.83 (24.89)	94.07 (22.05)
METS	7.53 (1.54)	7.61 (1.68)

3.3 Psychological measures

Several psychological measures taken on the participants were similar in average regardless of whether participants had completed the program. A couple of noted differences were in life changes. On average non-completers had a higher score for the life changes scale, 1.99 versus 1.62. Also for DR Index non-completers averaged a score of 3.04 while completers averaged a score of 2.06. For the JHP sub-scale common elements there was a difference in scores. Non-completers score 5.9 on average while completers scored higher at 6.16.

Table 3. Psychological Measures Stratified by Completion

Variable	**Non-Completers Mean (±SD)** ***N=206***	**Completers Mean (±SD)** ***N=137***
EBBS		
Benefits to Exercise	52.00 (12.27)	52.42 (15.08)
Barriers to Exercise	27.35 (7.07)	27.56 (6.27)
Exercise Confidence	3.77 (0.77)	3.84 (0.71)
Brief Symptom Inventory		
Somatization Sub-Scale	7.31 (1.98)	7.14 (1.75)
Depression Sub-Scale	7.93 (2.70)	7.85 (3.15)
Anxiety Sub-Scale	7.99 (2.88)	7.53 (2.31)
JHP		
Workplace	3.66 (0.63)	3.68 (0.63)
Common Elements	5.90 (1.05)	6.16 (1.04)
DR Index	3.04 (0.27)	2.06 (0.26)
Life Changes	1.99 (1.77)	1.62 (1.73)
Weight Efficacy	3.60 (0.65)	3.46 (0.71)

3.4 Comparative Analysis

Several models were constructed to compare women who completed the program to women who did not at baseline. We attempted to construct models looking at demographic, physiological, and psychological measures. The first two models are concerning demographic information. In the first model we look at whether a participant completed the program and the hours worked. Overall, $F=4.88$ $p=0.003$, $r^2=0.06$. We find that based on the model, on average, ADS is lower for women who did not complete the program and this value is even lower as the number of hours that the participant works increases.

Table 4. Model Results: ADS and Hours Worked with Interaction

Variable	Parameter Est	SE	t-value	p-value
Intercept	7770.06	1537.71	5.05	<0.001
Remove	-3503.09	2363.56	-1.48	0.140
Hours Worked	-364.59	341.33	-1.07	0.287
Hours Worked * Remove	1053.20	528.86	1.99	0.048

We also look at age and whether a woman completed the program. Overall we find, $F=6.85$, $p<0.001$, $r^2=0.08$. Older women tended to have slightly higher ADS. However, of women that did not complete the program, the older a woman was would result in a lower average ADS.

Table 5. Model Results: ADS and Age with Interaction

Variable	Parameter Est	SE	t-value	p-value
Intercept	5893.54	1142.98	5.16	<0.001
Remove	4880.92	1671.69	2.92	0.004
Age	4.19	28.06	0.15	0.881
Age * Remove	-100.53	42.95	-2.34	0.020

We did not find any significant models for psychological measures. For physiological measures we construct two models. In the first model with weight as one of the parameters of interest, we find that participants who weigh more have slightly lower average ADS, looking at the interaction term we find that of the women that did not complete the program this is further decreased. Overall we found for this model, $F=10.50$, $p<0.001$, $r^2=0.13$.

Table 6. Model Results: ADS and Weight with Interaction

Variable	Parameter Est	SE	t-value	p-value
Intercept	6765.20	820.34	8.25	<0.001
Remove	5951.00	1752.16	3.40	0.001
Weight	-3.13	4.02	-0.78	0.437
Weight* Remove	-21.54	8.77	-2.46	0.015

With BMI we find similar results. Overall, $F=8.64$, $p<0.001$, $r^2=0.10$. The model reflects that ADS is lower on average for women with lower BMI.

Table 7. Model results: ADS and BMI with Interaction

Variable	Parameter Est	SE	t-value	p-value
Intercept	6640.72	904.80	7.34	<0.001
Remove	5602.74	1549.00	3.62	<0.001
BMI	-13.69	26.47	-0.52	0.605
BMI* Remove	-138.28	45.75	-3.02	0.003

3.5 Complete Analysis

Our complete analysis is the equivalent of analysis with listwise deletion and is a way to just view the results when considering women who did complete the program. This will serve as a way to compare completers to non-completers. This analysis was completed on the N=83 participants who completed the program (i.e. were not removed). First, we construct an OLS model to determine if a portion of the variance in baseline ADS could be accounted for by psychological measures. We find that baseline ADS was statistically significantly associated with Benefit to Exercise Score, Anger Expression Score, and JHP Common Score. The model had r^2=0.07, which is extremely low.

Table 8. Complete Model Results for ADS, N=83

Variable	Parameter Est	SE	t-value	p-value
Intercept	7679.18	2034.84	3.77	<0.001
Benefit to Exercise	-24.47	13.97	-1.75	0.083
Anger Expression	-78.80	50.55	-1.56	0.123
JHP Common	266.55	234.61	1.14	0.259

Next, we construct a GLM to determine if the proportion of variance in improvement in ADS could be accounted for by baseline ADS and psychological measures. We find that it is statistically significantly associated with Baseline ADS, Anger Expression, Weight Efficiency, Life Changes, BSI: Somatization Scale and Exercise Confidence. Overall r^2=0.23, F=30.60, and p<0.0001.

Table 9. Complete Model Results for Improvement in ADS, N=83

Variable	F-value	P-value
Baseline ADS	148.22	<0.001
Anger Expression	14.86	<0.001
Weight Efficiency	7.38	0.007
Life Changes	5.50	0.019
BSI: Somatization Scale	4.33	0.038
Exercise Confidence	3.28	0.071

We construct another GLM adding in change in psychological measures from baseline to end of study. We find that a proportion of variance in ADS is accounted for by baseline ADS, Change in Barriers to Exercise and Change in Total EBBS. Overall, r^2=0.20, F=35.83, and p<0.0001.

Table 10. Complete Model Results for Improvement in ADS with Psychological Measures, N=83

Variable	F-value	P-value
Baseline ADS	94.16	<0.001
Change in Barriers to Exercise	6.77	0.010
Change in Total EBBS	6.54	0.011

We construct an ANOVA to determine if variation in change in cardiovascular fitness can be accounted for by change in ADS. The results were that a difference in cardiovascular fitness is significantly associated with change in ADS, F=11.6, p<0.0001.

3.6 Drop-Out Analysis

We investigate whether the act of dropping out of the study was associated with demographic characteristics or psychological characteristics by performing logistic regression with dropping out as the event of interest. The first model looks at demographic characteristics. We find that marital status, household size, annual income, age, diabetes, and current weight satisfaction are statistically significantly associated with drop-out. A couple of interesting results to point out from the resulting Odds Ratios (OR) listed below are those of Marital Status with an OR of 1.42 and having diabetes with a corresponding OR of 3.53.

Table 11. Dropout Analysis, Demographic Characteristics

Variable	OR	95%CI	p-value
Marital Status	1.42	(1.19,1.69)	< 0.01
Household Size	0.88	(0.79,0.97)	0.01
Annual Income	0.86	(0.76,0.98)	0.02
Age	0.92	(0.89,0.95)	< 0.01
Diabetes	3.53	(1.50,8.30)	< 0.01
Current Weight Satisfaction	0.23	(0.10,0.54)	< 0.01

Next, we investigate whether dropping out was associated with psychological characteristics. We find that drop out was statistically significantly associated with EBBS Total, JHP Total, Anger Expression (hold in), and Anger Expression (lash out). The corresponding OR are listed in the table below.

Table 12. Dropout analysis, Psychological Characteristics

Variable	OR	95%CI	p-value
EBBS Total	0.99	(0.90,1.00)	0.004
JHP Total	0.33	(0.21,0.52)	< 0.001
Anger Expression (Hold in)	0.91	(0.87,0.96)	< 0.001
Anger Expression (Lash out)	1.08	(1.03,1.15)	< 0.001

3.7 Imputation

We compare the following simple imputation methods: Avg, Prev, Post, LVCF, NVCB, and LaN. To revisit the composition, we simulate data for N=250 subjects, allowing 30% to represent married women, with the following assumptions as denoted in the table below:

Time Point	Mean	SE
Y_1	7987.68	213.86
Y_2	8325.02	208.04
Y_3	8444.55	260.46
Y_4	8076.51	217.25

We simulate missing data using three scenarios, being influenced by the results from the drop-out analysis:

1. Married Women experiencing 20% missing data, Other participants 10%

2. Married Women experiencing 30% missing data, Other participants 15%

3. Married Women experiencing 40% missing data, Other participants 20%

Table 13 displays the results for the first scenario. Given the large values that we start with we find that the bias for all methods is not extremely high. Most values underestimate Y_1-Y_4. The NVCB imputation method experiences the largest underestimation and overestimation. LaN slightly underestimates values, however, it experiences the smallest bias. We experience PV close to 1 for all methods for selected measures, which means the variance of our imputed values are close to the variance of our original data.

Table 13. Bias and PV by Imputation Method (Married Women 20% Missing, Other Participants 10%)*

	Avg		Prev		Post		LVCF		NVCB		LaN	
Time	**Bias**	**PV**	**Bias**	**PV**	**Bias**	**PV**	**Bias**	**PV**	**Bias**	**PV**	**Bias**	**PV**
Y_1	-289.12	0.43	-	-	-283.68	0.44	-	-	-352.42	0.50	-	-
Y_2	163.39	0.73	309.05	0.96	61.21	0.81	309.05	0.99	-89.76	1.00	89.86	0.84
Y_3	317.84	0.37	290.44	0.39	318.47	0.42	163.84	0.55	318.47	0.43	194.06	0.43
Y_4	-160.44	0.71	-156.07	0.73	-	-	-329.27	0.92	-	-	-	-

The results from the second scenario are detailed in Table 14. As the proportion of missingness increases we notice comparable results. The bias for all methods is still relatively small. Now we experience the highest and lowest bias for LVCF and NVCB. We

* Values not included could not be computed (i.e. no previous values for Y_1)

notice high PV in all methods except Avg.

Table 14. Bias and PV by Imputation Method (Married Women 30% Missing, Other Participants 15%)[‡]

	Avg		Prev		Post		LVCF		NVCB		LaN	
Time	Bias	PV	Bias	PV	Bias	PV	Bias	PV	Bias	PV	Bias	PV
Y_1	-289.34	0.36	-	-	-283.86	0.38	-	-	-336.41	0.44	-	-
Y_2	147.78	0.64	245.39	0.95	77.69	0.72	245.39	0.98	-62.24	0.98	67.92	0.86
Y_3	296.23	0.36	250.57	0.40	300.86	0.52	128.48	0.55	300.86	0.54	173.99	0.53
Y_4	-186.17	0.60	-183.09	0.62	-	-	-346.89	0.80	-	-	-	-

Table 15 displays the results from the third scenario. As missingness is further increased we still see relatively comparable results. The bias for all methods tends to be small taking into account our starting values. All methods excluding Avg and LaN experience relatively high PV values. NVCB and LVCF contains highest bias values.

Table15. Bias and PV by Imputation Method (Married Women 40% Missing, Other Participants 20%)[§]

	Avg		Prev		Post		LVCF		NVCB		LaN	
Time	Bias	PV	Bias	PV	Bias	PV	Bias	PV	Bias	PV	Bias	PV
Y_1	-265.77	0.25	-	-	-259.28	0.27	-	-	-309.17	0.33	-	-
Y_2	124.17	0.50	213.69	0.70	48.21	0.58	213.69	0.48	-51.79	0.77	56.16	0.64
Y_3	279.61	0.27	227.98	0.31	240.98	0.42	139.07	0.74	240.98	0.45	123.40	0.44
Y_4	-167.95	0.44	-162.43	0.47	-	-0.79	-289.67	0.42	-	-	-	-

We additionally construct graphs displaying 95% CI of the bias. We only utilize one time point, Y_2, for simplicity. We utilize Y_2 since it is one value that can be compared across all methods. We could have also utilized Y_3 for this reason. We Y_2 chose arbitrarily. We

[‡] Values not included could not be computed (i.e. no previous values for Y_1)

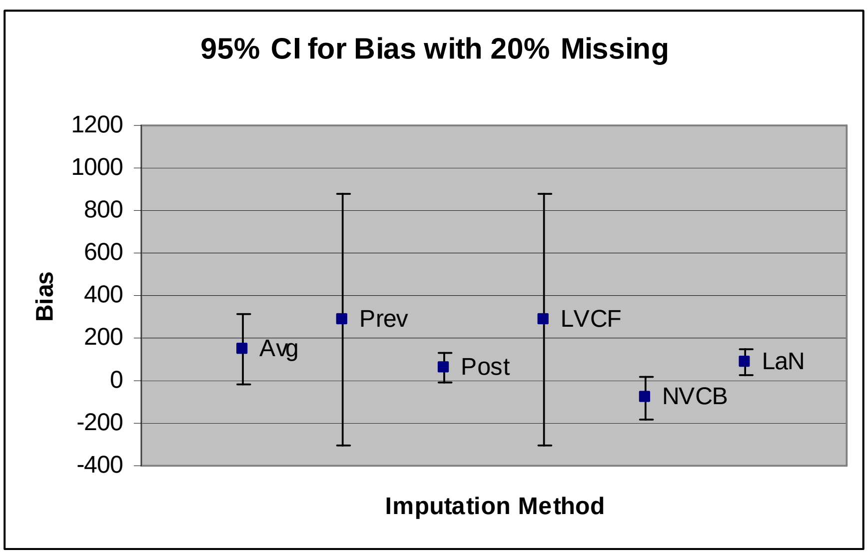
[§] Values not included could not be computed (i.e. no previous values for Y_1)

assume that results are comparable for other time points given table results, and merely wish to compare visually the possible range of the bias.

Figure 1 displays the 95% CI intervals for the bias given the first scenario with 20% missing for married women and 10% missing for other participants. Figure 2 displays the 95% CI intervals for the bias given the first scenario with 30% missing for married women and 15% missing for other participants. Lastly, Figure 3 displays the 95% CI intervals for the bias given the first scenario with 40% missing for married women and 20% missing for other participants.

By visually analyzing Figure 1 we see that LaN, Post, and NVCB appear to be the best methods. For each method we see that the variance is not high. For Post and LaN we see that these two methods slightly underestimate the Y_2. For NVCB, this method slightly overestimates Y_2.

Figure 1. 95% CI for Bias of Y_2 with Different Imputation Methods



With Figure 2 we see that LaN clearly outperforms all other methods. We note that Prev and LVCF seem to be the least reliable methods with the large amount of variance in the imputed values.

Figure 2. 95% CI for Bias of Y_2 with Different Imputation Methods

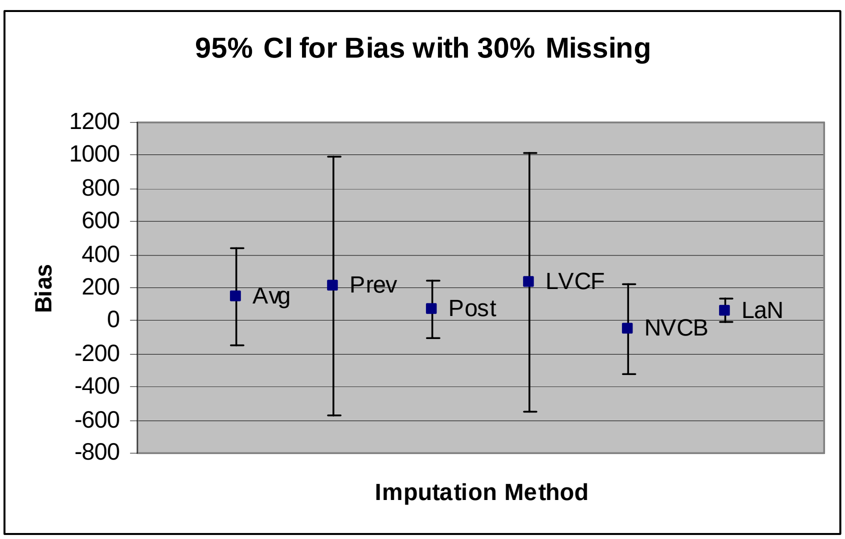
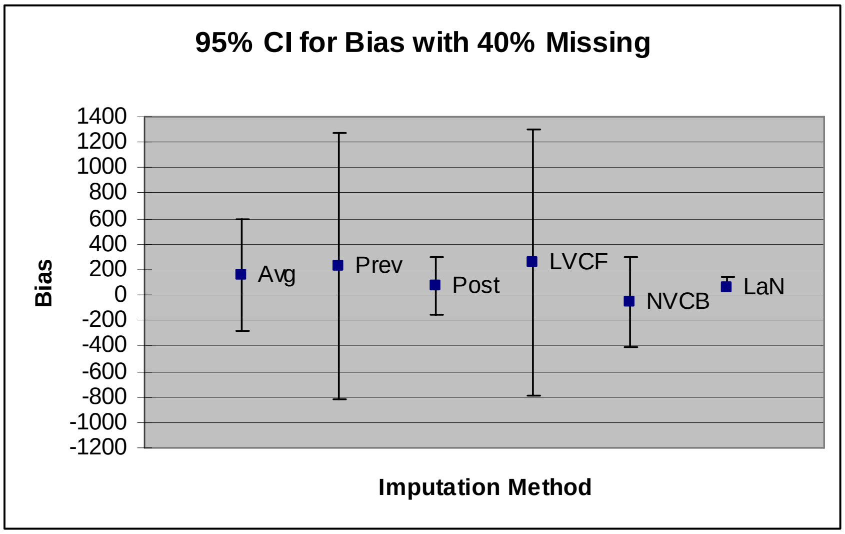


Figure 3 further supports LaN as the preferred method. It clearly outperforms all other methods. As the amount of missingness increases we see that the other methods Avg, Prev, Post, LVCF, and NVCB perform worse. Figure 3 displays the worst performance for Avg, Post, and NVCB. Prev and LVCF do not perform well for any of the scenarios.

Figure 3. 95% CI for Bias of Y_2 with Different Imputation Methods



4. Discussion

From our analysis we are able to capture a large amount of information regarding the differences at baseline for participants who completed the program and those who did not. Additionally, by completing dropout analysis we gain more meaningful information regarding which subsets of the study population that were more likely to dropout of the study.

Hours worked, age, and weight were significantly associated with baseline ADS as well as the interaction between these measures and whether a person completed the program. We find that women who worked longer hours had lower ADS, younger women had lower ADS, and women who weighed less had lower ADS. This could be indicative of older women who had time being more concerned about their health and thus participating more. The average age for completers was 40 verses non-completers at 36. The average weight for completers was 200 and for non-completers 195.

From drop-out analysis we find that one factor that was highly associated with drop-out was marital status. Thus, the different lifestyles and characteristics of our study participants were found to be associated with their respective initial activity levels and associated with whether they completed the program.

By utilizing this information we were able to simulate the data in a fairly realistic manner and apply missing data using an MAR format. We chose to use marital status as a variable that we would control for with missingness. The results support the effectiveness of simple imputation methods to deal with missing data of a MAR format. We find that the most effective imputation method was LaN, followed by Post. It should be revisited, however, that as seen with the tables produced, neither LaN nor Post can replace every missing value. This is not considered a major drawback since both methods will still result in a dataset that is more complete. It would be unrealistic to have a dataset with absolutely no missing data.

Additionally, we find that the other methods with the exclusion of Prev and LVCF performed fairly well as missingness increased. All the methods result in bias measurements of 100-200 on average. Given the fact that Y_1-Y_4 measurements range from 7,000-8000 these are relatively small bias values. Prev and LVCF also fall in this range. However, Prev and LVCF are less reliable due to the very large CI. With looking at PV there were several PV values of 0.60 or greater, indicating very similar variances of the imputed data to the *true* values. Ultimately, we find that certain simple imputation methods were very reliable when dealing with missing data of MAR format. Data imputed by utilizing the best methods would lead to more accurate results.

References

1. Beck A, Steer R, et al. *Manual for the Beck Depression Inventory: BDI-II Manual*. San Antonio, TX: The Psychological Corporation; 1988.

2. Brock GN, Shaffer JR, et. al. Which missing value imputation method to use in expression profiles: a comparative study and two selection schemes. *BMC Bioinformatics*. Vol 9. No 12. pp 1-12.

3. Cohen S. Kamarck T, et al. A Global Easure of Perceived Stress. *Journal of Health and Social Behavior*. Vol 24. pp 385-396.

4. Diggle P, Keward MG. Informative Drop-Out in Longitudinal Data Analysis. *Applied Statistics*. Vol 43, No. 1. pp. 49-93.

5. Dishman RK, Oldenburg B, et. al. Worksite physical activity interventions. *American Journal of Preventive Medicine.* Vol 15, No. 4. pp. 344-361.

6. Fielding S, Fayers PM, et al. Simple imputation methods were inadequate for missing not at random quality of life data. Vol 6. No 57. pp. 1-9.

7. Fielding S, Maclennan G, Cook JA, et al. A review of RCTs in four medical journals to assess the use of imputation to overcome missing data in quality of life outcomes. *Trials*. Volume 9, No 51. pp 1 – 6.

8. Fletcher GF, Balady G, et al. Statement on exercise: benefits and recommendations for physical activity programs for all Americans, AHA. *Circulation*. Volume 94, No. 4. pp 857-862.

9. Gaillard TR, Sherman WM. Importance of Aerobic Fitness in Cardiovascular Risks in Sedentary Overweight and Obese African-American Women. *Nursing Research*. Vol 56, No 6. 407-415.

10. Georgia Division of Public Health. BRFSS GA. Available at http://www.ph.dhr.state.ga.us/epi/brfss/index.shtml. Accessed 4/12/2003.

11. Gibbons RJ, Balady GJ, et al. ACC/AHA 2002 guideline update for exercise testing: A report of the American College of Cardiology/AHA Task Force on Practice Guidelines (Committee on Exercise Testing). Available at www.acc.org/clinical/guidelines/exercise/dirIndex.htm. Accessed 4/12/2003.

12. Horton NJ, Kleinman KP. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *Am Stat*. Vol. 61. No 1. pp. 79-90.

13. Lee IM, Skerrett PJ. Physical activity and all-cause mortality: what is the dose-response relation? *Medicine & Science ins Sports & Exercise*. Volume 33. pp 459-471 and 493-454.

14. Little RJ. Modeling the Drop-Out Mechanism in Repeated Measures Studies. *Journal of ASA*, Vol. 90, No. 431. pp. 1112-1121.

15. Overweight, obesity, and health risk. National Task Force on the Prevention and Treatment of Obesity. *Archives of Internal Medicine*. Volume 160, No 7. pp. 898-904.

16. Pender N. *Health promotion in nursing practice*. 2nd Edition. Norwalk, CT: Appleton-Century-Crofts; 1987.

17. Sallis JF, Pinski RB, et al. The development of self-efficacy scales for health-related diet and exercise behaviors. *Health Education Research*. Vol 3, No 3. pp 283-292.

18. Shrive FM, Stuart H, et al. Dealing with missing data in a multi-question depression scale: a comparison of imputation methods. *BMC Research Methodology*. Vol 6, No 57. pp 1-10.

19. Simpson ME. Testing gender differences in a model for exercise adherence in US Army Reservists [Doctoral]. Columbia, Missouri: Graduate School, University of Missouri, Columbia; 1998.

20. Spielberger CD, Reheiser EC, Sydeman SJ. Measuring the experience, expression, and control of anger. *Issues in Comprehensive Pediatric Nursing*. Vol 18, No 3. pp 207-232.

21. Taylor J, Grundy C. Measuring Black internalization of White stereotypes about African Americans: The Nadanolitization Scale. *Handbook of Tests and Measurements for Black Populations*. Vol 2. Hampton, VA: Cobb & Henry Publishers; 1996: 217-226.

22. Toulomi G, Pocock SJ, et al. Estimation and comparison of rates of change in longitudinal studies with informative drop-outs. *Statistics in Medicine*. Vol 18. pp 1215-1233.

23. Toulomi G, Pocock SJ, et al. Impact of Missing data Due to Selective Dropouts in Cohort Studies and Clinical Trials. *Epidemiology*. Vol. 13, No. 3. pp 347-355.

24. Tuikkala J, Elo L, et al. Missing value imputation improves clustering and interpretation of gene expression microarray data. *BMC Bioinformatics*. Vol 9. No 202. pp 1-14.

25. Yancey AK, McCarthy W, et al. Challenges in Improving Fitness: Results of a Community Based, Randomized, Controlled Lifestyle Change Intervention. *Journal of Women's Health.* Vol. 15, No. 4. pp 412-430.

26. Young DR, Stewart KJ. A Church-based Physical Activity Intervention for African American Women. *Family Community Health*. Vol. 29, No. 2. pp 103-117.