2012

A Comparison of Methods of Analysis to Control for Confounding in a Cohort Study of a Dietary Intervention

Esinhart Hali
Virginia Commonwealth University

A COMPARISON OF METHODS OF ANALYSIS TO CONTROL FOR CONFOUNDING IN A COHORT STUDY OF A DIETARY INTERVENTION

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Biostatistics at Virginia Commonwealth University.

by

HALI SUMMER ESINHART
Bachelor of Science in Biochemistry, University of North Carolina Chapel Hill, 2010

Director: DR DONNA MCCLISH
Director, Graduate Programs in Biostatistics

Virginia Commonwealth University
Richmond, Virginia
August 2012

Acknowledgement

I would like to thank many people for their help and support throughout my schooling career and especially with the completion of my Master's Thesis. First of all, I would like to thank my parents, Jim and Susie Esinhart for supporting and helping me spiritually, mentally, and financially from preschool until now in graduate school. Without you I would have never made it this far. Thank you and I love you!

Next, I would like to thank my sisters, Tiffany, Kellie, and Shelby. You guys are my best friends and have been so helpful and supportive throughout my career. I love you so much and hope the same for you in your studies!

Furthermore I would like to thank my teachers, especially those who helped me on my Master's Thesis; Dr. Donna McClish, Dr. Wen Wan, and Dr. Patricia Carcaise-Edinboro. You all helped me so much and were instrumental on helping me to complete my thesis.

Lastly, I would like to thank all the friends I have made throughout my journey through school who have helped me to learn and grow in my career. All the late night study sessions and trips to Starbucks have finally paid off.

Table of Contents

Page

List of Tables

Page

List of Figures

Page

Abstract

A COMPARISON OF METHODS OF ANALYSIS TO CONTROL FOR CONFOUNDING IN A COHORT STUDY OF A DIETARY INTERVENTION

By Hali Summer Esinhart

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Biostatistics at Virginia Commonwealth University.

Virginia Commonwealth University, 2012

Major Director: Dr. Donna McClish
Director, Graduate Programs in Biostatistics

Comparing samples from different populations can be biased by confounding. There are several statistical methods that can be used to control for confounding. These include; multiple linear regression, propensity score matching, propensity score/logit of propensity score as a single covariate in a linear regression model, stratified analysis using propensity score quintiles, weighted analysis using propensity scores or trimmed scores. The data were from two studies of a dietary intervention (FIBERR and RNP). The outcome variable was change from baseline to one month for eight outcome measures; fat, fiber, and fruits/ vegetables behavior, fat, fiber, and fruits/vegetables intentions, fat and fruits/vegetables self-efficacy. It was found that the propensity score matching and the

quintiles analysis were the two best methods for analyzing this dataset. The weighted analyses were the worst of all the methods compared in analyzing this particular dataset.

CHAPTER 1

Introduction

1.1 Colorectal Cancer

Colorectal cancer is a leading cause of cancer death in the U.S, behind prostate and breast cancer (Terry, 2001). It is also one of the leading cancer deaths in Virginia behind prostate, breast, lungs, and bronchus cancer (Terry, 2001). It has been shown that there is an association between diet and certain cancers (Terry, 2001). All of the major national cancer organizations such as the American Cancer Society and National Cancer Institute currently recommend a healthy diet low in fats, calories, and red meat, and high in whole grains, fruits, and vegetables (Kushi, 2012). It is recommended that each person consumes 2.5 cups of fruits and vegetables daily to decrease their risk of cancer (Kushi, 2012). The American Cancer Society also recommends limiting the amount of alcohol and tobacco use (Kushi, 2012). Being overweight and obesity are directly related to colon cancer, therefore a healthier diet and regular exercise helps reduce the risk (Kushi 2012). One-third of deaths from cancer are directly related to the person's diet and level of physical activity (Kushi, 2012).

In the late 1990's and early 2000's it was recommended to consume a diet low in fats and high in fiber and fruits and vegetables (Auld, 2000). This diet is very similar to what is recommended today to reduce the risk of cancer. Because diet is directly related to colon cancer, consuming a healthier diet could help reduce the risk of colon cancer.

1.2 Rural Nutrition Project (RNP)

The Rural Nutrition Project started in 1999 and was a randomized two-armed dietary intervention trial conducted in Virginia. In this study 754 healthy participants from three different rural physician practices were assessed for dietary and psychosocial behavior at baseline then assigned to receive either tailored feedback and self-help dietary intervention or no interventions. The study's primary outcome article provides more detail on the dietary intervention and primary fat and fiber outcomes (Fries, 2005). The RNP primary outcome was to determine whether participants changed their diet or were willing to change their diet after being educated that a healthier diet can reduce the risk of some cancers (Fries, 2005). Each of the participants was evaluated by phone at baseline, 1 month, 6 months, and 12 months after the intervention was administered. The intervention was a tailored dietary feedback regarding the participant's reported baseline diet, along with a series of booklets that were mailed out to each of the participants in the intervention arm over the course of a month. The booklets gave information on how to improve the participants' diet. The recommended diet was a low fat, high fiber diet in accordance with the current recommendations to lower the chances of colorectal cancer. Data from the participants was collected over the years 1999 to 2003. At one month 224 participants in the intervention group provided outcome data (Fries, 2005). At each of the time points the participant was asked a series of questions to determine their fat, fiber, and fruits and vegetables intake.

1.3 Families in Behavioral Intervention for Risk Reduction (FIBERR) Study

The Families in Behavioral Intervention for Risk Reduction (FIBERR) study was an observational study conducted at the same time as the RNP study using the same intervention as in the RNP study, starting one year later in 2000. For this study 103 participants were recruited from first degree family relatives (FDR) of patients with colorectal cancer. This group of people was chosen because it was thought that first degree relatives of colorectal patients would know the impact of colorectal cancer first hand from the relative and so would be more likely to adhere to the diet intended to lower the risk of colorectal cancer. The patients were first asked if they would consent to their family members being recruited for the study. Family members were contacted and those who satisfied all the inclusion criteria/exclusion criteria and agreed to participant were used in the study. Outcome data for the FIBERR study were collected at 1 and 3 months (Bean, 2008). Baseline and one month (n = 81) follow up data were collected from participants in the FIBERR study were compared to those of the participants in the intervention arm of the RNP study (n = 224). Because both studies measured the same outcome variables and the same baseline measures and intervention, the two studies were deemed comparable.

1.4 The Purpose

The purpose of this thesis is to compare and contrast various statistical methods that can be used to control for confounding in observational studies. This will be done

through the analysis of a data analysis example. The specific example to be used is the comparison of the outcomes for the participants who received the dietary intervention from the RNP study to those of first degree family relatives of patients with colorectal cancer who participated in the FIBERR study. This is a secondary data analysis of the original two studies, and as such is an observational cohort study to determine whether the FIBERR participants are more likely to adhere to the intervention than the RNP participants. Only baseline and 1 month follow-up were compared because these were the only common time points that had been assessed in both studies.

Because the samples in this study were from different populations there are likely to be confounders between the two studies. There are several statistical analysis methods available to control for confounding. The purpose of this study is to compare different methods of analysis to control for potential confounders between the two groups from; the RNP study and the FIBERR study. The methods that will be compared for the FIBERR/RNP study include: directly controlling for the covariates using multiple linear regression, propensity score matching, using propensity score and logit of the propensity score as a covariate in a linear regression model, stratified analysis using quintiles of the propensity score, weighting the propensity score, and trimming the weights of the propensity score. A previous study on the effect of tissue plasminogen activator on death among ischemic stroke patients compared similar methods to control for confounding (Kurth, 2006). The study, though, had an endpoint of death (dichotomous outcome), while the outcome measures for the FIBERR/RNP study are continuous (diet) measures. Thus

the comparison of methods in this thesis extends the results from the Kurth et al (2006) to a different type of outcome measure.

CHAPTER 2

The Data

2.1 Potential Confounding Covariates at Baseline and their Comparisons Between the RNP and FIBERR Study Groups

There were many potential covariates that were in common between the RNP and FIBERR studies. The covariates that were chosen to control for differences were age, gender, ethnicity, marital status, education level, rural or city residence of participant, how many total meals daily the participant ate outside of the home (0-3), how much the participant shopped for, planned and prepared meals, number of months since last doctor's appointment, number of hours weekly the participant spent watching television (a surrogate for physical activity), family social support (a 20 item measure assessing functional social support or family cohesion), and a fat knowledge at baseline score (Bean, 2008). See Table 1 for the distribution of the potential confounders. The vast majority of the participants in both studies were married (66.56%), females (65.57%) in their mid to upper forties. T-tests and chi-squared tests were run on the continuous and categorical confounders respectively to determine which were significantly different between the two studies ($\alpha = 0.10$). Only the covariates that were found significant were used as confounders in the various methods of analysis. The covariates that were significantly different between the studies were ethnicity, education level, whether the participant lived in or out of town, number of meals daily eaten outside of the house, months since last doctor's appointment, and their baseline fat knowledge score.

Table 1: All Potential Confounding Variables

	N	Number (%) in the FIBERR Group	Number (%) in the RNP Group	p-value comparing the RNP and FIBERR groups
Demographic				
Sex				0.8091
Female	200 (65.57)	54 (66.67)	146 (65.18)	
Male	105 (34.43)	27 (33.33)	78 (34.82)	
Ethnicity				**0.0186**
Black	93 (30.49)	16 (19.75)	77 (34.38)	
Non-Black	210 (68.85)	65 (80.25)	147 (80.25)	
Marital Status				0.1771
Married	203 (66.56)	49 (60.49)	154 (68.75)	
Not Married	102 (33.44)	32 (39.51)	70 (31.25)	
Education				**<0.0001**
Some HS or less	37 (12.13)	4 (4.94)	33 (14.80)	
HS	93 (30.49)	16 (19.75)	77 (34.53)	
Some College or Tech School	83 (27.21)	19 (23.46)	64 (28.70)	
College or more	91 (29.84)	42 (51.85)	49 (21.97)	
Town				**<0.0001**
In Town	82 (26.89)	36 (44.44)	46 (20.63)	
Out of Town	222 (72.79)	45 (55.56)	177 (79.37)	
Eating Out				**0.0811**
No Meals Out	161 (52.79)	38 (46.91)	132 (58.93)	
One Meal Out	85 (27.87)	23 (28.40)	62 (27.68)	
Two Meals Out	42 (13.77)	18 (22.22)	25 (11.16)	
All Meals Out	7 (2.30)	2 (2.47)	5 (2.23)	
	N	Mean (SD) in the FIBERR Group	Mean (SD) in the RNP group	p-value comparing the RNP and FIBERR groups
Age	305	46.83 (12.31)	49.08 (13.84)	0.1965
Responsibility for Shopping/Planning/Preparing Meals	299	7.48 (2.25)	7.27 (2.16)	0.4588
Sum of Family Social Support (FSS)	298	12.11 (5.32)	13.02 (4.69)	0.1515
Months Since Last Doctors Visit	294	5.62 (8.44)	3.94 (5.14)	**0.0979**
Number of Hours Watching TV weekly	293	12.18 (9.05)	13.86 (10.78)	0.2163
Fat Knowledge Score	305	5.81 (0.48)	5.53 (1.08)	**0.0017**

Footnote: Significant values were bolded. Age ranged from 19 to 75, Responsibility for Shopping/Planning/Preparing Meals ranged from 3 to 9, Sum of FSS ranged from 5-30, Months Since Last Doctors Visit ranged from 1-48, Hours Watching TV ranged from 0-50, and the Fat Knowledge Score at Baseline ranged from 0-6.

There were 8 outcome variables collected in both studies. These include dietary fat behavior, dietary fiber behavior, fruits and vegetables dietary intake behavior, fat intentions, fiber intentions, fruits and vegetables intentions, fat self-efficacy, and fruits and vegetables self-efficacy.

The behavior measures were based on the Fat and Fiber Behavior questionnaire (FFB). The FFB comprised of 28 questions such as "How often do you trim visible fat from your meat?" or "How often do you eat vegetables at lunch?" Responses were 1 (usually), 2 (sometimes), or 3 (never), giving a score of the overall summary (Fries, 2005). Fat and fiber sub scores were calculated, with lower scores associated with better behavior. Therefore, a decrease over time meant that the participant showed improvement in following the diet, lower fat and higher fiber. Fruit and vegetable behavior was measured as a total number of fruits and vegetables consumed daily. An increase in fruit and vegetable intake over time indicated an improvement in following the diet plan (Carcaise-Edinboro, 2008).

The intention variables measured how much the participant intended to change their eating behavior based on the dietary intervention, and self-efficacy measured how likely the participant thought they would be able to adhere to the dietary intervention. Each variable was measured on a 5 point scale with higher values indicating higher intentions or self-efficacy. An increase over time in these outcome measure indicated an improvement in intentions or self-efficacy.

The outcomes that were compared between the two studies in this thesis were the change from baseline to one month for each of the variables outlined above. Various

methods of controlling for confounders in analysis will be seen in Chapter 3 were implemented in comparing each of the outcome measures.

CHAPTER 3

Methods of Analysis for Control of Confounding

3.1 Literature Review - Controlling for Confounders

There are many statistical analysis methods available to control for confounding between groups. The methods considered in this thesis include controlling for the variables directly in a multiple linear regression model, as well as various ways to control for confounding using propensity scores: propensity score matching, using the propensity score as a covariate in a linear regression model, stratified analysis using quintiles of the propensity score, weighting using the propensity score, and using trimmed weights of the propensity score. Other researchers have compared some or all of these methods using simulation studies or actual data.

D'Agostino (1998) wrote a tutorial on propensity score methods to control for confounding. He pointed out that simple matching and covariate adjustment may not work sufficiently because they are limited to a small number of covariates, while propensity score methods do not have this limitation. An advantage to using the propensity score methods is that a propensity score finds the best probability of treatment groups from the covariates used; therefore over-parameterization should not be a concern. Propensity score methods allow researchers to use observational data which can be less expensive than running a large clinical trial. Using propensity score methods is growing in popularity.

Austin (2009) found that propensity matching and weighting eliminated baseline differences better than stratification or covariate adjustment. He found this true in both real data examples, and simulation experiments. The latter showed matching to be marginally better than weighting. It was also found that doing a propensity score matched analysis created a better balance between the groups than doing a stratified analysis.

Austin et al (2007) looked at the best choice of variables to include in a propensity score model. The found that using only the true confounder in creating the propensity score was ideal. It was found that more matches were created when only using the true confounders. In contrast, Brookhart et al (2006) found that also including variables unrelated to exposure, but related to the outcome variable improved the propensity score model in terms of mean squared error.

In another article Austin (2007) examined the use of propensity score methods to estimate marginal odds ratios. It was found that when using propensity score methods of analysis, marginal treatment effects could be estimated, whereas, when using regression models conditional treatment effects could be estimated.

Schafer and Kang (2008) analyzed a simulated data example of the effect of diet on emotional distress, using regression, and propensity score methods. The paper compared ANCOVA, regression, propensity score matching, weighting, and stratification. In this study, they found that the weighted analysis was not the best, although they mention that typically it is one of the better methods of analysis.

Sturmer et al (2006) reviewed the medical literature to see how propensity scores were used in the literature. They compared results of propensity score methods to that of

usual regression model control of confounding. They concluded that propensity score methods usually did not provide results that were particularly different; specifically they found that in only 13% of studies examined did affect estimates using propensity scores change by more than 20%.

Freedman et al (2008) looked at weighted analysis. They found that, if the propensity score is correctly estimated, then it is recommended to do a weighted analysis rather than a logistic regression model. The only disadvantage found for using a weighted analysis was that it was likely to increase the random error.

Kurth (2006) compared analysis on a dataset of ischemic stroke patients in a German stroke registry using multivariable logistic regression, propensity score matching, regression using propensity score as a covariate, weighted analysis on propensity score (inverse-probability-of-treatment weights and standardized-mortality-ratio weights). The outcome for this study was dichotomous; therefore odds ratios were used to compare the different methods. This study found that the propensity score matching, standardized-mortality ratio weights, and propensity score as a continuous covariate in a regression model had similar odds ratios (less than 1) while the other methods had similar odds ratios greater than 1).

Many papers compared methods of analysis using both real datasets as well as simulation data. In many of the comparisons, dichotomous outcomes were used. In this thesis, continuous outcomes were used. It will be interesting to see which method(s) of analysis best model the data. Based on the literature, all of the methods being considered seemed good for modeling the data.

3.2 Multiple Linear Regression

One commonly used method for addressing confounding variables is to directly control for the covariates in a multiple linear regression model. Multiple linear regression is the most common form of analysis used in practice. It controls for confounding by modeling the dependent variable as a function of the independent variables. They are placed in a linear model. This method may not be appropriate if the data does not fit the linear model well. An advantage to this method, in comparison to the other methods to come, is that it uses the entire sample in the analysis (Kutner, 2004).

Only baseline variables that were significantly different between the groups were to be included in the models. T-tests and chi-squared tests were calculated for continuous and categorical confounders respectively, in order to determine which ones were significantly different between the two studies ($\alpha = 0.10$). The six variables that were found to be significantly different were ethnicity, education, whether or not the participant lived in our out of town, how many meals the participant eats outside of the home, months since last doctor's visit, and fat knowledge score at baseline. After determining which confounders were significantly different between the two groups, they were examined for multicollinearity by looking at the correlation between the six covariates. Multicollinearity is when two or more predictors in the regression model are highly correlated; correlation values greater than 0.5 indicate correlation and potential multicolliniearity. Checking the correlation between the predictors is how to determine if there is multicollinearity using a

Pearson's correlation for the continuous variables and Spearman correlation for the categorical variables.

These potential confounders were used in the linear regression model:

$$y_j = \beta_0 + \gamma group + \beta_1 x_1 + \ldots + \beta_6 x_6 + \varepsilon_j = \gamma group + \mathbf{x_j'\beta} + \boldsymbol{\varepsilon_j}$$

where y is the change from baseline to one month for each of the 8 the outcome measure (this same formula is used for all the outcome measures), $j = 1, \ldots, n$ (n = 305) references the individual subject, *group* refers to the treatment group (FIBERR or RNP), x are the six regressors in the model and ε is the error term. The null hypothesis testing that there is no difference between the two groups in their changes from baseline to 1 month with respect to each of the outcome measures is;

$$H_o : \gamma = 0$$
$$H_a : \gamma \neq 0$$

given that all six regressors are in the model. A Type I error of 0.05 will be used ($\alpha = 0.05$). An ANCOVA test will be run and the least squared means (LSMEANS) for each study as well as the difference in the LSMEANS will be analyzed. The LSMEANS are the means of the outcome variables adjusting for the other variables in the model; the difference in the change from baseline with respect to each study. A difference is then taken of the LSMEANS. P-values as well as a 95% confidence intervals will be calculated on the difference of the means. The 95% confidence intervals were used to determine whether there was a significant difference and the p-values indicated how significant or insignificant the difference is. If the confidence interval contains zero, then it is deemed insignificant, but is significant if it does not contain zero. The adjusted R^2, AICc, RMSE

were also recorded. The results were then compared to the results of various other methods of analysis to determine if there was any difference in the result based on the method of analysis implemented. The model was built using PROC GLM and PROC REG in SAS 9.2.

3.3 Propensity Score Matching

Another method of analysis to control for confounding is to match the two samples by propensity score. A propensity score is a subject's probability of being a part of each respective sample (FIBERR and RNP) conditional on the observed covariates (Rosenbaum, 1983, 1984). An advantage to using propensity score matching is that it analyzes the data by matching like participants from both groups while ignoring participants that are not similar between the two samples. One disadvantage of this is that the sample size is reduced which decreases the power. The confounders used in the propensity score were the same as those in the multiple linear regression model in section 3.2; ethnicity, education level, whether the subject lived in a rural area or not, how often the subject ate outside of their home, months since last doctors visit, and fat knowledge. A logistic regression model was built in order to determine the propensity scores for each of the participants in both studies;

$$propensityscore_j = \frac{e^{(\beta_0+\beta_1 x_{1j}+\ldots+\beta_6 x_{6j})}}{e^{(\beta_0+\beta_1 x_{1j}+\ldots+\beta_6 x_{6j})}+1} = \frac{1}{1+e^{-(\beta_0+\beta_1 x_{1j}+\ldots+\beta_6 x_{6j})}}$$

$$\log it\left(propensityscore_j\right) = \ln\frac{propensityscore_j}{1-propensityscore_j} = \mathbf{x_j'\beta}+\varepsilon_j$$

where the **β** are a vector of the coefficients of the covariates (where are different from the **β** in (1)) and *x* are the covariates in the model.

Caliper matching was performed on the logit of propensity score. Strict caliper matching is one form of nearest neighbor matching that tries to avoid bad matches by placing boundaries as to how far away the matches can be from one another (Todd, 2006). It is a commonly used form of matching in practice. The macro GMATCH was used to perform the matching algorithm. The GMATCH macro matches two groups using a greedy matching algorithm (Kosanke, 2004). FIBERR was the reference sample and 1:1 matching was done. Greedy matching refers to picking the best match for the reference sample without replacement. Specifically, the FIBERR and RNP studies were matched via GMATCH macro on the logit of the propensity score using a caliper of 0.2 of the standard deviation of the logit of the propensity score (Austin, 2009).

After matches were found, paired t-tests and Cochran-Mantel-Haenszel tests were run again with the confounders to verify that none of them were significantly different any more, would be expected. For the outcome variables of interest a mixed models analysis was required, therefore PROC MIXED in SAS 9.2 was used to determine whether there was a difference in the change from baseline between the two studies ($\alpha = 0.05$) ;

$$H_o : \mu_1 - \mu_3 = \mu_2 - \mu_4$$
$$H_a : \mu_1 - \mu_3 \neq \mu_2 - \mu_4$$

μ_1 refers to the mean from the FIBERR study at baseline, μ_2 refers to the mean from the RNP study at baseline, μ_3 refers to the mean from the FIBERR study at 1 month, and μ_4 refers to the mean from the RNP study at 1 month. The results obtained from the

propensity score matched analysis were then compared to the results of all the other analyses performed. The absolute value of the AICc for this method could not be compared to the other methods because the sample size in this analysis was much smaller than that of the other analyses.

3.4 Propensity Score/Logit of Propensity Score as Covariate

Another way to control for the potential confounders is to use the propensity score or the logit of the propensity score as a covariate in a linear regression model;

$$y_j = \beta_0 + \gamma group + \beta x_j + \varepsilon_j$$

where y is the change from baseline for the outcome variable (this same formula is used for each of the 8 outcome measures), $j = 1, \ldots, n$ (n = 305), *group* is the treatment groups (FIBERR and RNP), x refers to the propensity score or the logit of the propensity score, and ε is the error term. As before, the propensity score was calculated using a logistic regression of only the covariates that were significantly different between the two groups as seen above in section 3.2. The propensity score was then placed into the linear regression model as the only covariate in the model. A potential advantage, in comparison to the matching method, is that it uses the entire sample rather than a portion of the sample.

The null hypothesis that was tested was that there is no difference between the two groups in their changes from baseline to 1 month with respect to the outcome variables ($\alpha = 0.05$);

$$H_o : \gamma = 0$$
$$H_a : \gamma \neq 0$$

The results were then compared to the results of various other methods of analysis with respect to adjusted R^2, AICc, and RMSE to determine if there was any difference in the result based on the method of analysis implemented. The model was built using PROC GLM and PROC REG is SAS 9.2.

3.5 Quintiles of Propensity Score

Yet another way to control for the confounders is to do stratified analysis with the strata defined by quintiles of the propensity score. Stratified analysis is useful because it does not assume a linear relationship with the propensity score, as when using the propensity score as a continuous covariate in the linear regression model. An advantage of this method is that is analyzes groups of similar participants in the study and uses the entire sample in the study.

The propensity score was stratified into quintiles, that is the 20th, 40th, 60th, 80th, and 100th percentile of the combined sample of n = 305, which is often referred to as stratification. This technique groups the participants into strata based on baseline characteristics determined by the propensity score (D'Agostino, 1998). Estimates of differences between groups were calculated within each quintile and summed using;

$$\delta = \sum_{k=1}^{K} \frac{n_k}{N} \left[\bar{y}_{0k} - \bar{y}_{1k} \right]$$

$$\text{with } Var\left(\delta\right) = \sum_{k=1}^{K} \left(\frac{n_k}{N} \right)^2 Var\left[\bar{y}_{0k} - \bar{y}_{1k} \right]$$

where *k* indexes the propensity score quintiles, *N* is the total number of participants in the study, n_k is the number of participants in each quintile, $\bar{y}_{0k}$ is the sample mean of the responses in the FIBERR study in the k^{th} quintiles, and $\bar{y}_{1k}$ is the sample mean of the responses in the RNP study in the k^{th} quintiles (Perkins, 2000). The quintles were placed into a class statement in PROC GLM and PROC REG as a categorical covariate using SAS 9.2 to determine the difference between FIBERR and RNP within each outcome measure.

$$y_j = \beta_0 + \gamma group + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon_j$$

where *y* is the change from baseline for the outcome variable (this same formula is used for each of the 8 outcome measures), $j = 1, \ldots, n$ (n = 305), *group* is the treatment groups (FIBERR and RNP), $x_1, \ldots, x_4$ refers to the quintiles of the propensity score (equal to one if in the specified quintile and zero otherwise), and ε is the error term.

The null hypothesis was tested that states there is no difference between the two groups in their changes from baseline to 1 month with respect to fat behavior, fiber behavior, fruits and vegetables behavior, fat intentions, fiber intentions, fruits and vegetables intentions, fat self-efficacy, and fruits and vegetables self-efficacy ($\alpha = 0.05$), respectively;

$$H_o : \gamma = 0$$
$$H_a : \gamma \neq 0$$

The adjusted R^2, AICc, and root mean standard errors on the residuals were then compared to that of all the other methods of analysis.

3.6 Weights of Propensity Score

Weights can be calculated as a function of the propensity score in order to control for the confounders. As before, the propensity score was calculated using a logistic regression of only the six covariates that were significantly different between the two groups.

Weights were then placed on the propensity score as outlined in a previous study (Lee, 2011). The weight of 1 was used for each participant in the FIBERR study and a weight of $\frac{p_j}{1-p_j}$ was used for each participant from the RNP study, where p_j is the j^{th} participant's propensity score. An advantage of using weights is that the RNP group is made to look like the FIBERR group with respect to the propensity score to make it easier to compare (Lee, 2011). Another advantage is that it uses the entire sample in the analysis as compared to the propensity score matching or some forms of trimmed weighted analysis (see next sections). PROC GLM and PROC REG were used to evaluate the weight means. The weighted means were calculated as follows:

$$\frac{\sum_{j=1}^{n} w_j y_j}{\sum_{j=1}^{n} w_j}$$

$$Var\left(\frac{\sum w_j y_j}{\sum w_j}\right) = \frac{1}{\left(\sum w_j\right)^2} \sum w_j^2 \sigma^2$$

The null hypothesis that was tested states that there is no difference between the two groups in their changes from baseline to 1 month with respect to the outcome measures ($\alpha = 0.05$);

3.7 Trimmed Weights of Propensity Score

The weights that are placed on the propensity score can also be trimmed in order to control for the confounders. Trimming puts less weight on the outcomes that have the values of propensity score that were higher than the majority of participants. There are two approaches that can be taken to trim weights. The first method puts less weight on observations with higher propensity scores that may skew results by setting the higher values equal to the cut point. The second method of trimming removes from analysis observations with higher weights that could impact the results (this is sometimes referred to as truncation). An advantage of the first method as compared to matching is that it uses the entire sample in the analysis. A possible disadvantage to the second method (truncation) is that it does not use the entire sample in the analysis. As before, the propensity score was calculated using a logistic regression of only the covariates that were significantly different between the two groups as before in section 3.2.

Trimming was performed at the 90^{th}, 95^{th}, and 99^{th} percentile in order to reduce number of high weights that may have an effect on the analysis. Trimming can be implemented in two different ways. The first way is to set all values of the propensity score greater than the cutpoint (90^{th}, 95^{th}, or 99^{th} percentile) to the value of the cutpoint. The other way is to simply eliminate all values greater than the cutpoint from the analysis

all together. Both methods were implemented and analyzed in a similar fashion to the full weighted model in the previous section (Section 3.6).

3.8 Comparing Across Methods

The different methods of analysis listed will be compared using an adjusted R^2, the AICc, and the root mean squared error (RMSE) to assess which method(s) of analysis is best for analyzing the dataset. AIC (Akaike Information Criterion) measures goodness of fit in a model;

$$AIC = 2k - 2\ln(L)$$

where k is the number of parameters in the model and L is the maximum value of the likelihood function. The AICc is AIC with a greater penalty for extra parameters as follows:

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}$$

where n is the total sample size (Fang 2011). Adjusted R^2 comes from the R^2 given by;

$$R^2 = 1 - \frac{SS_{error}}{SS_{total}}$$

The adjusted R^2 takes into account the varying number of parameters;

$$R^2_{adj} = 1 - \left(1 - R^2\right)\frac{n-1}{n-p-1} = 1 - \frac{SS_{error}}{SS_{total}}\frac{df_t}{df_e}$$

where n is the total sample size, $p(p = k - 1)$ is the total number of regressors in the model, df_t is the degrees of freedom for the total sample (n-*1*), and df_e is the degrees of freedom for the error term (n-p-*1*) (Kutner, 2004).

The root mean squared error is a commonly used value to measuring the accuracy of a model. Adjusted R^2 measure how well the model fits the data which is why it is a good measure to compare amongst the methods of analysis. Values close to 1 are the most desirable.

$$RMSE\left(\theta\right) = \sqrt{MSE\left(\theta\right)} = \sqrt{E\left(\left(\theta - \theta\right)^2\right)}$$

Where θ is the estimator and θ is the parameter being estimated by the model.

Adjusted AIC measures how well the model fits, where smaller values indicate a better fit while adjusting for the number of parameters (k). The smaller values of the root mean squared error also indicates a better fit. R^2 accesses the variation explained by the model. Large values of R^2 indicate a better model.

The methods they compared in the previous study mentioned above were the crude model, multivariate model, propensity score matched, regression of propensity score, deciles of propensity score, and weighted models (Kurth, 2006). Only methods with the same sample size in the analysis can be compared to one another using AICc. All can be compared to one another using the RMSE and adjusted R^2.

CHAPTER 4

Results

4.1 Preliminary Results

All of the covariates that were found to be significantly different between the two groups (p-value < 0.1) and were going to be used in the models as well as introduced in Section 3. Before doing the models, the correlations amongst the variables were calculated and tested. As seen in Tables 2 and 3, the highest correlation coefficient is 0.2 and no multicollinearity is evident.

The first method of analysis that was used was the multiple linear regression model where all the six significant covariates were controlled for directly. Next, the propensity score was found for each of the participants in the study using a logistic regression, where the specific model we used is given as:

$$\log it\left(propensityscore_j\right) = \ln \frac{propensityscore_j}{1 - propensityscore_j} = \mathbf{x_j'\beta} + \varepsilon_j$$

$$\log it\left(propensityscore_j\right) = 4.01 - 0.80x_{1j} - 0.55x_{2j} + 0.84x_{3j} - 0.31x_{4j} - 0.024x_{5j} - 0.20x_{6j}$$

where $j = 1, \ldots, n$ (n = 305), x_1 is ethnicity (black and non-black), x_2 is education level (some high school or less, high school graduate, some college or technical schooling, college graduate or more), x_3 is whether the participant lived in town or not, x_4 number of meals eaten outside of the home (no meals, one meal, two meals, three/all meals), x_5 is months since last doctors visit, and x_6 is the fat knowledge at baseline. Matching was done based on the propensity scores of each of the participants. Caliper matching was

implemented in a 1:1 ratio between the FIBERR study and the RNP study. There were a total of 72 matches created from the GMATCH matching algorithm. The distribution of the propensity score was analyzed as well as the distribution of the logit of the propensity score. It is important that they are normally distributed because if not, then some transformation would need to be made in order to perform the analysis. Also it is important that the distributions for the two studies overlap so that matches can be made and they do, as seen in Figures 1 and 2. The distribution of the propensity score for the FIBERR study is spread out slight more than the RNP study, but there is overlap between the two. The mean of the propensity score of the unmatched set was 0.23 (SE = 0.14) for the RNP study and 0.40 (SE = 0.20) for the FIBERR study. The mean of the propensity score of the matched set was 0.35 (SE = 0.17) for the RNP study and 0.35 (SE = 0.17) for the FIBERR study.

The residuals for the 8 outcomes using, multiple linear regression model were analyzed for normality. As seen in Figure 3-10, the residuals appear to be normally distributed. Had they not been normally distributed, a transformation would have been needed to be applied to the data, such as looking at the log of the data.

Each of the methods used to control for confounding, if successful, should have resulted in producing similar groups at baseline. Thus each of the baseline variable that had been significantly different were tested once again to determine whether they remained significantly different with respect to the different groups (FIBERR/RNP), after controlling using the methods describe in the previous chapter. Table 4 summarizes the results. From the table it appears that the propensity score matching, the linear regression model with the

propensity score or the logit of the propensity score as the single covariate, and stratification were the best ta controlling for confounding as none of the 6 baseline variables remained significantly different. With the multiple linear regression model, 3 of the 6 remained different. None of the weighted analyses adequately controlled for confounding between the two samples.

Table 2: Pearson Correlation Coefficients between Continuous Covariates

	Doctors Visit	Fat Knowledge
Doctors Visit	1.00	0.00559
Fat Knowledge	0.00559	1.00

Table 3: Spearman Correlation Coefficients between Categorical Covariates

	Ethnic	Education	Town	Eat Out
Ethnic	1.00	0.00765	-0.09383	-0.07242
Education	0.00765	1.00	-0.19678	0.13896
Town	-0.09383	-0.19678	1.00	-0.07029
Eat Out	-0.07242	0.13896	-0.07029	1.00

Figure 1: Distribution of Propensity Score

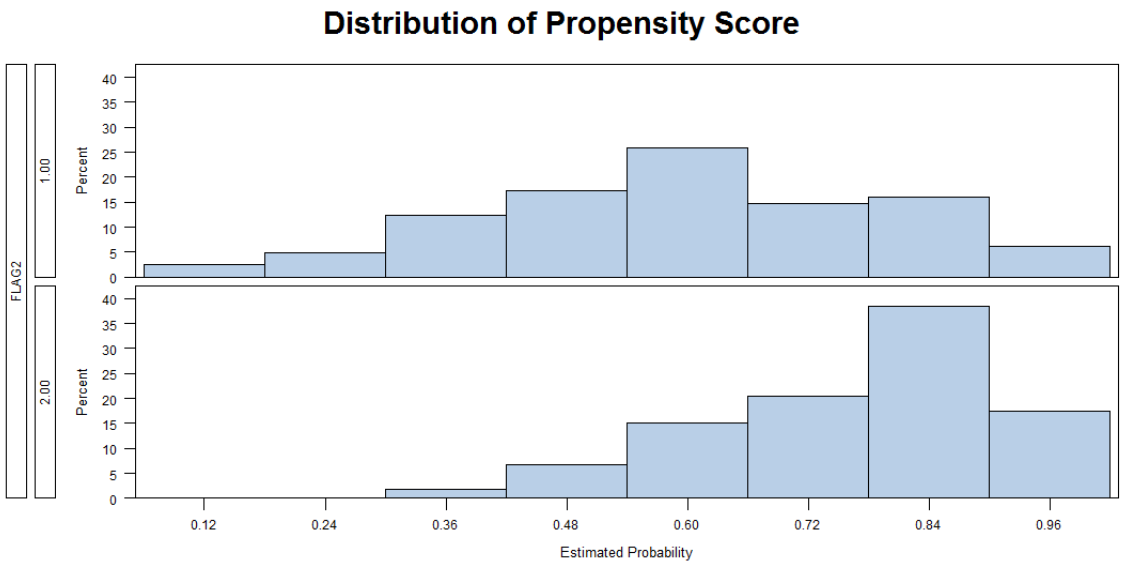


Figure 2: Distribution of Logit of Propensity Score

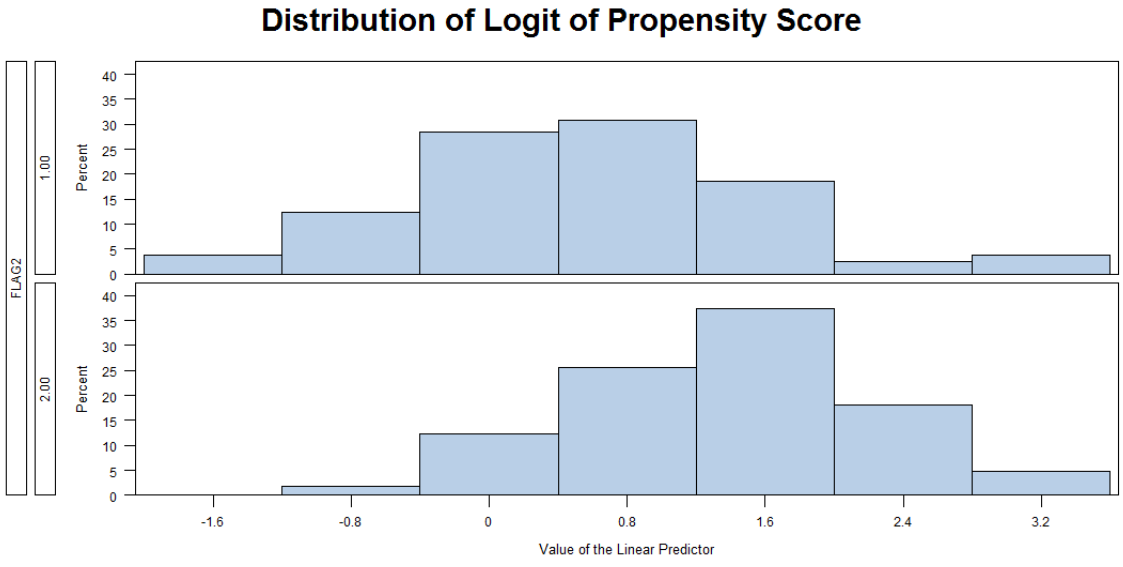


Figure 3: Distribution of Residuals in Fat Behavior from the Multiple Linear Regression Model

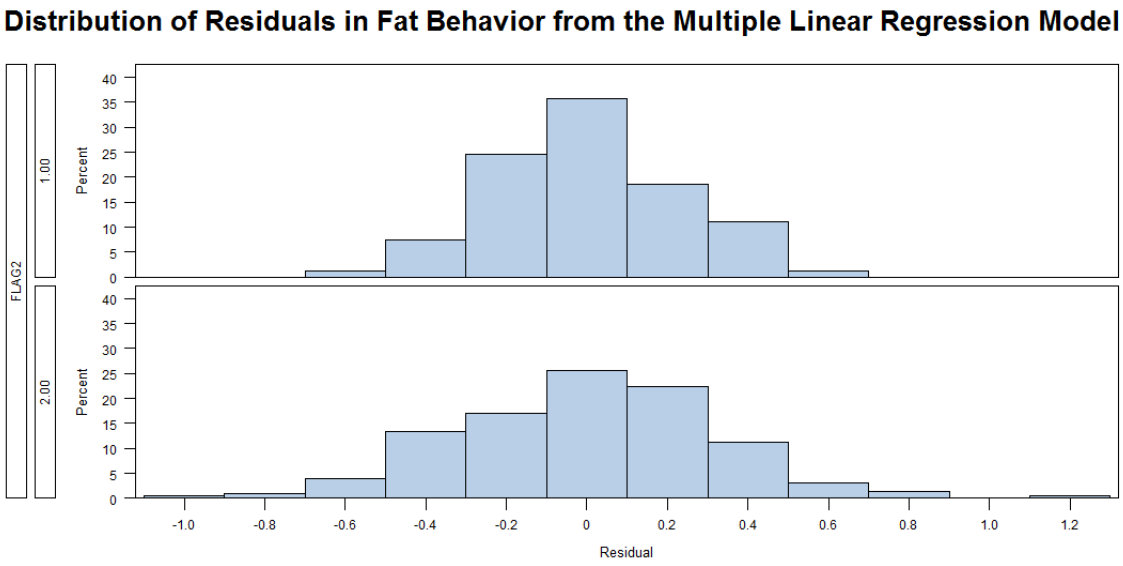


Figure 4: Distribution of Residuals in Fiber Behavior from the Multiple Linear Regression Model

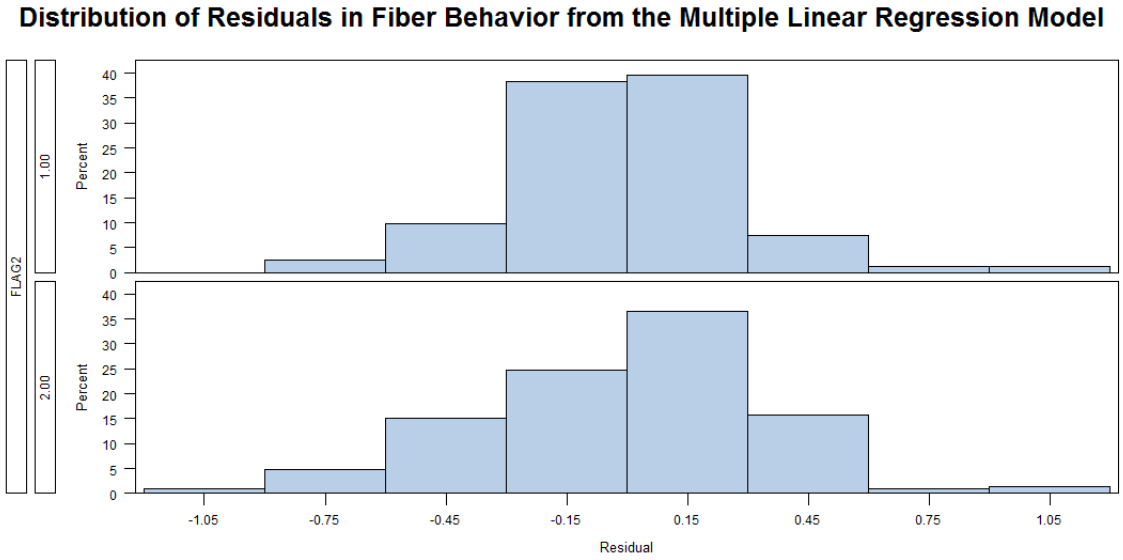


Figure 5: Distribution of Residuals in FV Behavior from the Multiple Linear Regression Model

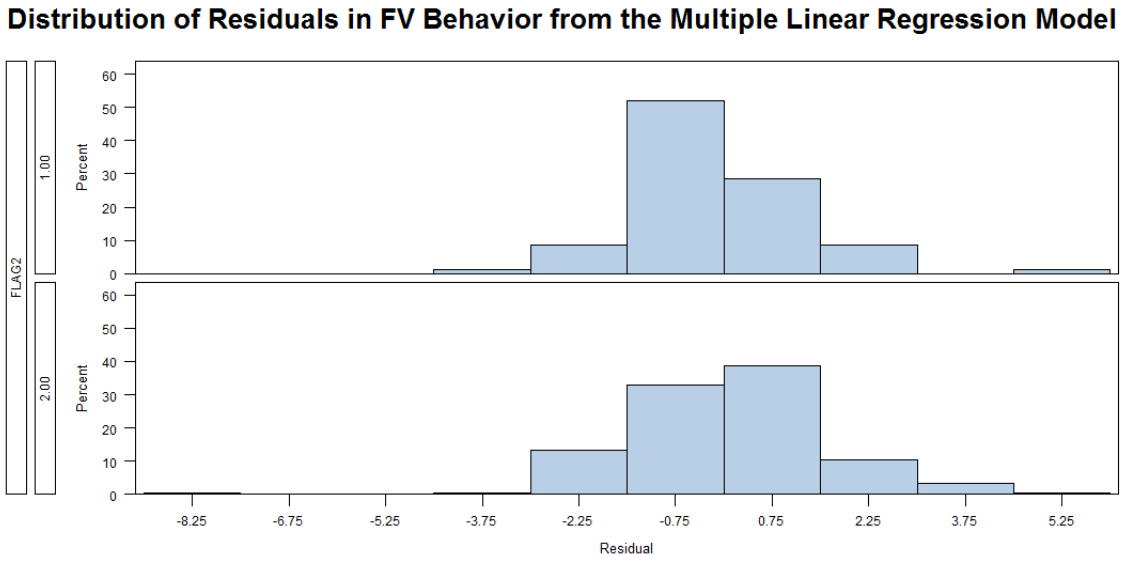


Figure 6: Distribution of Residuals in Fat Intentions from the Multiple Linear Regression Model

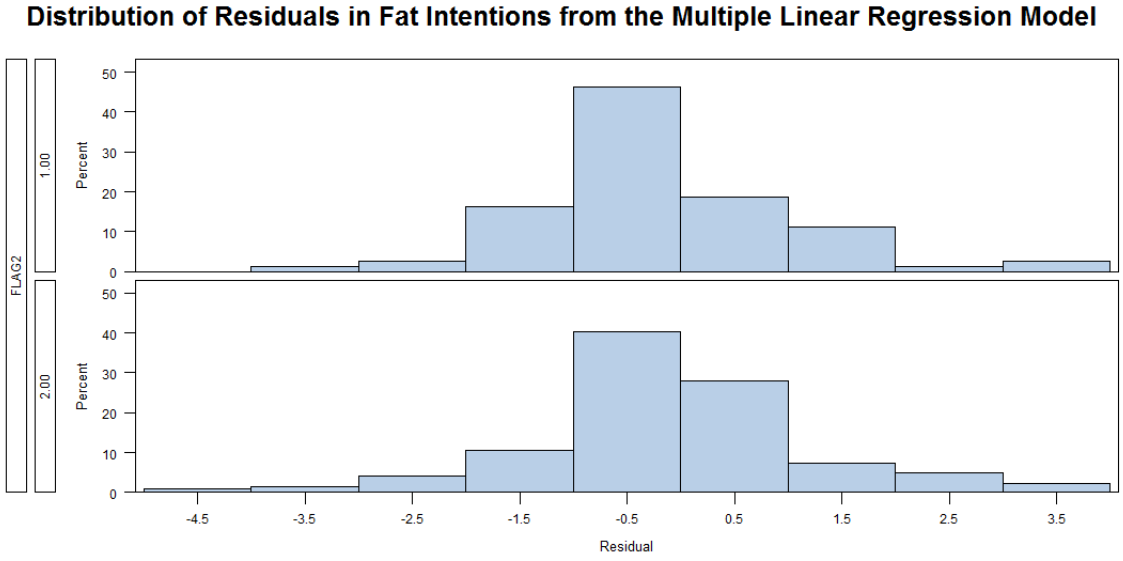


Figure 7: Distribution of Residuals in Fiber Intentions from the Multiple Linear Regression Model

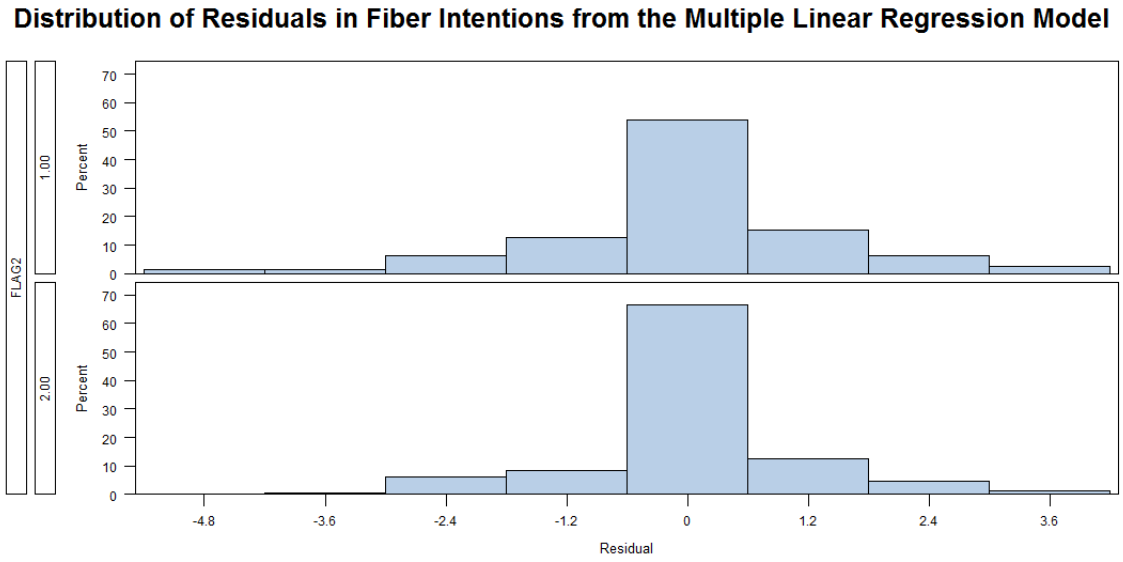


Figure 8: Distribution of Residuals in FV Intentions from the Multiple Linear Regression Model

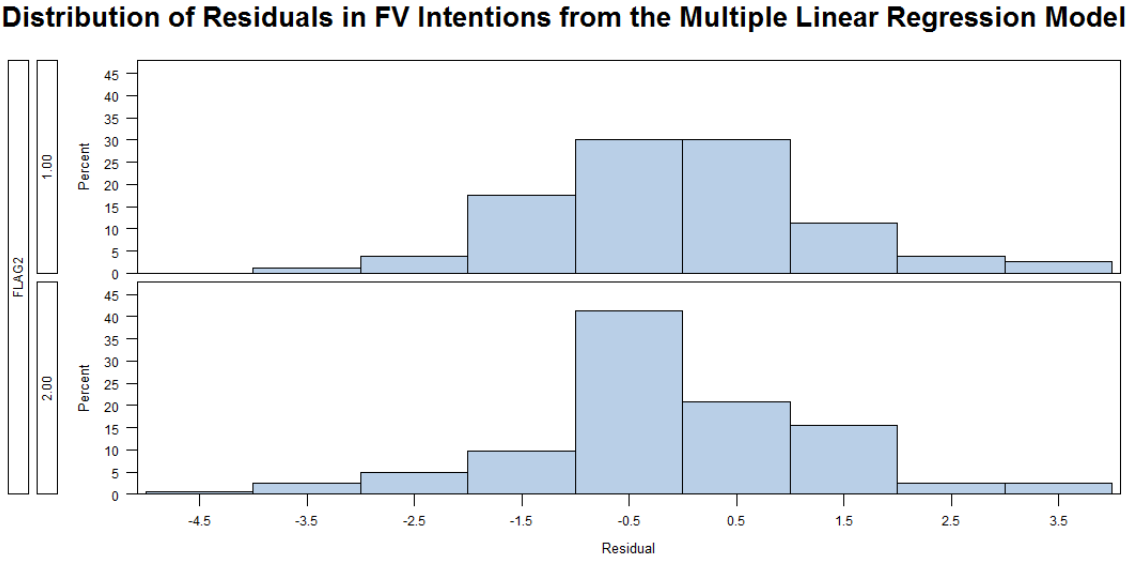


Figure 9: Distribution of Residuals in Fat Self-Efficacy from the Multiple Linear Regression Model

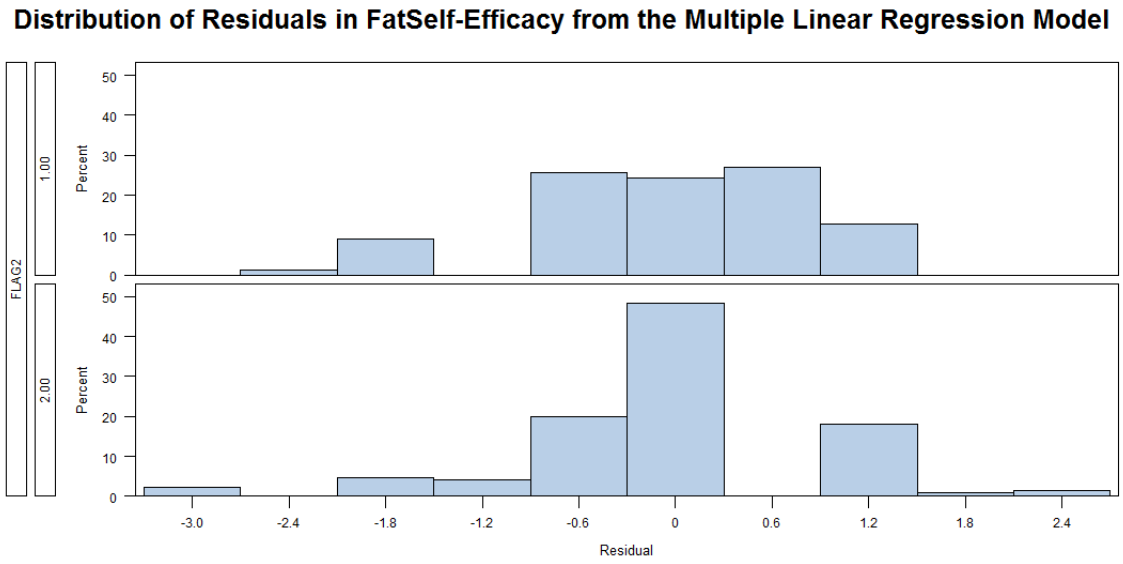
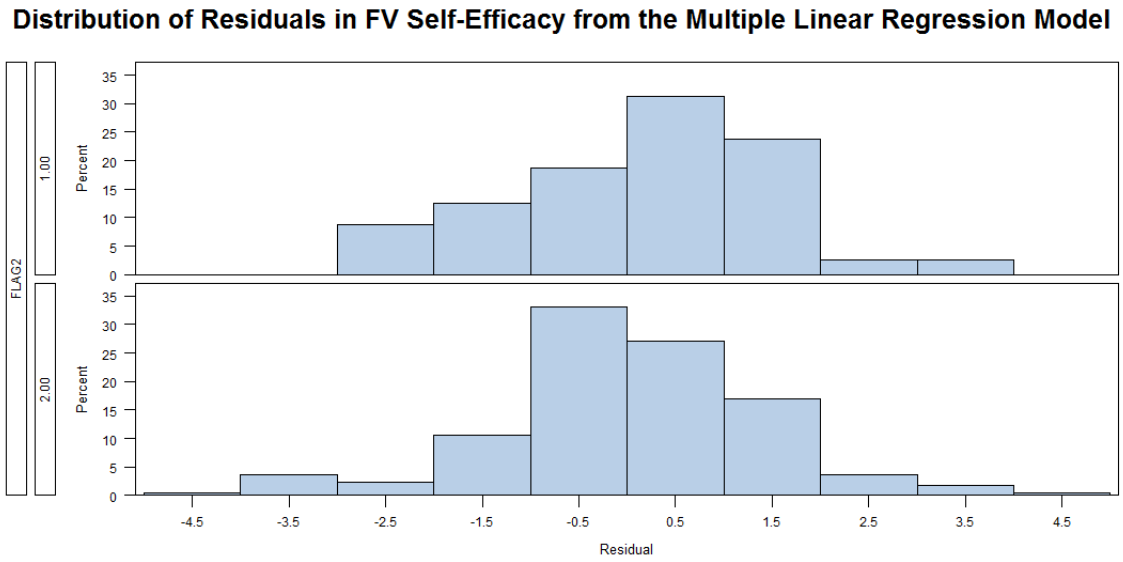


Figure 10: Distribution of Residuals in FV Self-Efficacy from the Multiple Linear Regression Model



Footnote: For Figures 1-10 Flag 1 refers to FIBERR and Flag 2 refers to RNP

Table 4: Re-Examining the Baseline Variables to Verify Whether Confounding was Controlled For by Each Method of Analysis

Method	Education	Ethnicity	Town	Eating Out	Doctor's Visit	Fat Knowledge
Propensity Score Matching						
Multiple Linear Regression	X	X	X			
Propensity Score as Cont. Covariate						
Quintiles of the Propensity Score						
Weight using Propensity Score	X	X	X	X	X	
Trimmed Weights of Propensity Score (if ≥90/95/99 then make the value the propensity at 90/95/99)	X	X	X	X	X	
	X	X	X	X	X	
	X	X	X	X	X	
Trimmed Weights of Propensity Score (if ≥90/95/99 then ignore the observation from analysis)	X	X	X		X	
	X	X	X		X	
	X	X	X	X	X	

Footnote: An 'X' indicates that there was a significant different between the two groups (FIBERR/RNP) with respect to that baseline variable.

4.2 Fat Behavior

The outcome measure, dietary fat behavior, was analyzed using all the various methods of analysis. The non-adjusted (crude) model ($p = 0.0113$), the model where covariates were controlled for directly in a multiple linear regression model ($p = 0.0450$), the propensity score and logit of the propensity score of each participant was used a covariate in the linear regression ($p = 0.0484$ and $p = 0.0405$ respectively), and the quintiles analysis ($p = 0.0147$) showed a significant difference in the differences between the two studies from change from baseline, rejecting the null hypothesis that there was no difference in the change from baseline to one month. The trimmed analysis where all values greater than the 95^{th} percentile were ignored from analysis showed marginal significance ($p = 0.0506$). The propensity score matching and all of the other forms of weighted and trimmed weights analysis showed no significant difference in the changes from baseline between the two studies, shown in Table 5. Because of varying sample sizes in the analysis the other methods that could be compared using AICc were the multiple linear regression model, using propensity score/logit of propensity score as a covariate in a linear regression model, quintiles analysis, the weighted analysis, and the trimmed weighted analysis where all values at the cut point were set equal to the cut point, could be compared. Also, throughout the analyses it appeared that the various analyses grouped into 3 groups;(1) propensity score matching alone,(2) multiple linear regression, linear regression with either propensity score or logit of the propensity score as the single covariate, and stratified analysis (3) the weighted and trimmed weighted analyses. The

linear regression model with propensity score as the covariate had the best (lowest) value of AICc amongst all of the methods with like sample sizes. The linear regression model with propensity score as the single covariate had the greatest R^2 in comparison to all of the different methods of analysis. The propensity score matching had the smallest root mean square error in comparison to all of the other methods of analysis. From group (2), the linear regression model with propensity score as the single covariate appeared to be the best method analysis for the data.

Table 5: Fat Behavior

Method	N	$FIBERR_{adj}$ (SE)	RNP_{adj} (SE)	Difference (CI)	p-value	R^2_{adj}	AIC_{adj}	Root MSE
Unadjusted Comparison	305	-0.090 (0.034)	-0.19 (0.020)	0.10 (0.023, 018)	**0.0113**	0.0177	-729.35	0.3015
Propensity Score Matching	144	-0.092 (0.026)	-0.14 (0.026)	0.044 (-0.031, 0.012)	0.2495	NA	108.1	0.1935
Multiple Linear Regression	292	-0.099 (0.035)	-0.19 (0.021)	0.086 (0.0019, 0.17)	**0.0450**	0.0148	-689.38	0.3032
Propensity Score as Cont. Covariate	292	Propensity: -0.10 (0.036)	-0.19 (0.021)	0.086 (0.00063, 0.17)	**0.0484**	0.0154	-694.18	0.3031
	292	Logit of Propensity: -0.10 (0.036)	-0.19 (0.021)	0.089 (0.0039, 0.17)	**0.0405**	0.0150	-694.05	0.3031
Quintiles of the Propensity Score	292	-0.094 (0.036)	-0.19 (0.021)	0.095 (0.010, 0.18)	**0.0283**	0.0147	-691.06	0.3032
Weight using Propensity Score	292	-0.090 (0.082)	-0.19 (0.022)	0.10 (-0.066, 0.27)	0.2319	0.0015	-173.10	0.7410
Trimmed Weights of Propensity Score (if ≥90/95/99 then make the value the propensity at 90/95/99)	292	90%: -0.090 (0.076)	-0.20 (0.022)	0.11 (-0.042, 0.27)	0.1592	0.0036	-221.16	0.6824
	292	95%: -0.090 (0.080)	-0.20 (0.022)	0.11 (-0.052, 0.27)	0.1821	0.0027	-189.01	0.7211
	292	99%: -0.090 (0.081)	0.19 (0.022)	0.10 (-0.062, 0.27)	0.2188	0.0018	-176.64	0.7365
Trimmed Weights of Propensity Score (if ≥90/95/99 then ignore the observation from analysis)	262	90%: -0.092 (0.070)	-0.19 (0.022)	0.094 (-0.049, 0.24)	0.1974	0.0026	-260.15	0.6064
	277	95%: -0.092 (0.075)	-0.24 (0.022)	0.15 (-0.0039, 0.30)	0.0560	0.0096	-223.88	0.6652
	290	99%: -0.090 (0.079)	-0.21 (0.021)	0.12 (-0.039, 0.28)	0.1367	0.0042	-198.31	0.7080

Footnote: FIBERR and RNP were measured 1 month minus baseline. The 'difference' column was subtracted FIBERR minus RNP. Negative values indicate improvement for dietary fat behavior. Significant values were bolded.

4.3 Fiber Behavior

The difference in the change from baseline between FIBERR and RNP was analyzed using all the methods outlined above. All of the methods of analysis gave the same result that there was no significant difference between the studies in the change from baseline, as can be seen in Table 6. All of the methods failed to reject the null hypothesis that there was no difference in the change from baseline between the groups. When comparing the adjusted R^2 values amongst all the methods of analysis, quintiles analysis had the greatest value. For those with like sample size, the method with the smallest AICc was the quintiles analysis. The propensity score matching had the smallest root mean squared error. Likewise, the weighted analyses appeared to be the worst methods of the data. From the second group, it appeared that multiple linear regression model was the best model for the data with respect to Fiber Behavior.

Table 6: Fiber Behavior

Method	N	$FIBERR_{adj}$ (SE)	RNP_{adj} (SE)	Difference (CI)	p-value	R^2_{adj}	AIC_{adj}	Root MSE
Unadjusted Comparison	305	-0.10 (0.038)	-0.19 (0.023)	0.082 (-0.0041, 0.17)	0.0628	0.0082	-658.22	0.3388
Propensity Score Matching	144	-0.089 (0.039)	-0.15 (0.039)	0.066 (-0.043, 0.17)	0.2326	NA	198.2	0.2335
Multiple Linear Regression	292	-0.12 (0.040)	-0.18 (0.024)	0.061 (-0.035, 0.16)	0.2108	0.0140	-618.42	0.3423
Propensity Score as Cont. Covariate	292	Propensity: -0.11 (0.041)	-0.18 (0.024)	0.070 (-0.027, 0.17)	0.1587	0.0054	-620.50	0.3438
	292	Logit of Propensity: -0.12 (0.041)	-0.18 (0.024)	0.063 (-0.033, 0.16)	0.2003	0.0075	-621.14	0.3435
Quintiles of the Propensity Score	292	-0.11 (0.040)	-0.19 (0.024)	0.074 (-0.021, 0.17)	0.1286	0.0262	-623.79	0.3402
Weight using Propensity Score	292	-0.10 (0.093)	-0.22 (0.025)	0.12 (-0.070, 0.31)	0.2156	0.0019	-101.89	0.8371
Trimmed Weights of Propensity Score (if ≥90/95/99 then make the value the propensity at 90/95/99)	292	90%: -0.10 (0.086)	-0.21 (0.024)	0.11 (-0.067, 0.28)	0.2232	0.0017	-150.33	0.7704
	292	95%: -0.10 (0.091)	-0.22 (0.025)	0.12 (-0.066, 0.30)	0.2069	0.0021	-115.26	0.8181
	292	99%: -0.10 (0.093)	-0.23 (0.025)	0.12 (-0.068, 0.31)	0.2087	0.0020	-103.13	0.8353
Trimmed Weights of Propensity Score (if ≥90/95/99 then ignore the observation from analysis)	262	90%: -0.11 (0.076)	-0.17 (0.024)	0.061 (-0.095, 0.22)	0.4424	-0.0016	-215.54	0.6602
	277	95%: -0.11 (0.086)	-0.23 (0.025)	0.13 (-0.051, 0.30)	0.1622	0.0035	-151.59	0.7579
	290	99%: -0.10 (0.092)	-0.24 (0.025)	0.13 (-0.052, 0.32)	0.1563	0.0035	-109.89	0.8246

Footnote: FIBERR and RNP were measured 1 month minus baseline. The 'difference' column was subtracted FIBERR minus RNP. Negative values indicate improvement for dietary fiber behavior. Significant values were bolded.

4.4 Fruits and Vegetables Dietary Intake Behavior

Fruits and vegetables dietary intake behavior was also analyzed using all the various methods of analysis. For this outcome variable, the propensity score matching and the method using the propensity score a continuous covariate in a linear regression model showed a significant difference between FIBERR and RNP in their changes from baseline ($p = 0.0061$ and $p = 0.0416$ respectively), rejecting the null hypothesis that there was no change from baseline between the groups. Analysis using all covariates, as well as the analysis using the logit of the propensity score had marginal statistical significance ($p = 0.0655$ and $p = 0.0509$ respectively). All of the other methods of analysis showed no significant difference between the studies in the change from baseline as can be seen in Table 7. When analyzing the AICc values amongst the various methods of analysis with the same sample size in the analysis, regression model were propensity score was the only regressor, was the best fit for the data. For adjusted R^2, the multiple linear regression model had the greatest value and the propensity score matching had the smallest mean square error. Only two methods gave significant results and a few others were very close. It could be possible that those eliminated from analysis in the propensity score matching (one of the methods that gave significant results) had a big impact on the analysis. Again, the weighted and trimmed weighted analyses were much worse that the other analyses. The multiple linear regression again, appears to be the best method of analysis amongst group 2.

Table 7: Fruits and Vegetables Behavior

Method	N	$FIBERR_{adj}$ (SE)	RNP_{adj} (SE)	Difference (CI)	p-value	R^2_{adj}	AIC_{adj}	Root MSE
Unadjusted Comparison	296	0.43 (0.17)	0.74 (0.11)	-0.31 (-0.71, 0.092)	0.1311	0.0077	264.35	1.5576
Propensity Score Matching	144	0.36 (0.16)	1.00 (0.16)	-0.64 (-1.09, -0.19)	**0.0061**	NA	1018.4	0.4695
Multiple Linear Regression	284	0.40 (0.18)	0.80 (0.11)	-0.40 (-0.83, 0.026)	0.0655	0.0145	248.42	1.3227
Propensity Score as Cont. Covariate	284	Propensity: 0.36 (0.18)	0.82 (0.11)	-0.45 (-0.88, -0.017)	**0.0416**	0.0079	245.13	1.5316
	284	Logit of Propensity: 0.38 (0.18)	0.81 (0.11)	-0.43 (-0.86, 0.0016)	0.0509	0.0065	245.52	1.5326
Quintiles of the Propensity Score	284	0.39 (0.18)	0.81 (0.11)	-0.42 (-0.85, 0.0093)	0.0551	0.0015	249.95	1.5365
Weight using Propensity Score	284	0.43 (0.39)	0.79 (0.11)	-0.36 (-1.15, 0.43)	0.3709	-0.0007	707.61	3.4757
Trimmed Weights of Propensity Score (if ≥90/95/99 then make the value the propensity at 90/95/99)	284	90%: 0.43 (0.36)	0.79 (0.10)	-0.36 (-1.09, 0.37)	0.3353	-0.0002	665.99	3.2188
	284	95%: 0.43 (0.37)	0.82 (0.10)	-0.38 (-1.15, 0.38)	0.3255	-0.0001	692.69	3.3737
	284	99%: 0.43 (0.38)	0.80 (0.11)	-0.37 (-1.15, 0.42)	0.3577	-0.0005	709.32	3.4615
Trimmed Weights of Propensity Score (if ≥90/95/99 then ignore the observation from analysis)	256	90%: 0.47 (0.34)	0.74 (0.11)	-0.27 (-0.97, 0.43)	0.4517	-0.0017	554.43	2.9416
	270	95%: 0.46 (0.35)	0.80 (0.11)	-0.34 (-1.06, 0.39)	0.3638	-0.0006	617.98	3.1290
	282	99%: 0.43 (0.37)	0.83 (0.10)	-0.40 (-1.17, 0.36)	0.3018	0.0003	686.44	3.3654

Footnote: FIBERR and RNP were measured 1 month minus baseline. The 'difference' column was subtracted FIBERR minus RNP. Positive values indicate improvement for fruits and vegetables behavior. Significant values were bolded.

4.5 Fat Intentions

For the outcome measure, dietary fat intentions, when implementing all various methods of analysis outline above, all of the methods gave the same result. All of the methods of analysis showed that there was no significant difference between the studies in the change from baseline as shown in Table 8. In other words, all of the methods of analysis failed to reject the null hypothesis. With comparing the AICc values amongst the methods with the sample size in the analysis, the quintiles analysis gave the smallest value. Comparing the adjusted R^2 values amongst all the methods of analysis, the unadjusted analysis had the largest value. The propensity score matching analysis gave the smallest root mean square error. Again, the weighted analyses were the worst in comparison to all the others and for group 2, it appears that the quintiles analysis was the best method.

Table 8: Fat Intentions

Method	N	$FIBERR_{adj}$ (SE)	RNP_{adj} (SE)	Difference (CI)	p-value	R^2_{adj}	AIC_{adj}	Root MSE
Unadjusted Comparison	298	0.30 (0.15)	0.40 (0.088)	-0.10 (-0.44, 0.23)	0.5427	0.0013	158.99	1.3013
Propensity Score Matching	142	0.34 (0.14)	0.21 (0.14)	0.13 (-0.26, 0.52)	0.5166	NA	902.5	0.4416
Multiple Linear Regression	285	0.34 (0.16)	0.39 (0.094)	-0.045 (-0.41, 0.33)	0.8122	-0.0159	166.28	1.3196
Propensity Score as Cont. Covariate	285	Propensity: 0.37 (0.16)	0.38 (0.094)	-0.012 (-0.38, 0.36)	0.9508	-0.0006	156.76	1.3097
	285	Logit of Propensity: 0.35 (0.16)	0.39 (0.094)	-0.038 (-0.41, 0.33)	0.8391	-0.0067	165.26	1.3111
Quintiles of the Propensity Score	285	0.37 (0.16)	0.38 (0.094)	-0.016 (-0.39, 0.35)	0.9342	-0.0045	160.87	1.3122
Weight using Propensity Score	285	0.30 (0.35)	0.38 (0.093)	-0.079 (-0.78 0.62)	0.8245	-0.0034	644.40	3.0864
Trimmed Weights of Propensity Score (if ≥90/95/99 then make the value the propensity at 90/95/99)	285	90%: 0.30 (0.31)	0.41 (0.088)	-0.11 (-0.74, 0.51)	0.7216	-0.0031	575.09	2.7330
	285	95%: 0.30 (0.32)	0.41 (0.089)	-0.11 (-0.77, 0.55)	0.7428	-0.0032	608.72	2.8991
	285	99%: 0.30 (0.34)	0.39 (0.092)	-0.094 (-0.79, 0.60)	0.7903	-0.0033	637.04	3.0468
Trimmed Weights of Propensity Score (if ≥90/95/99 then ignore the observation from analysis)	256	90%: 0.29 (0.28)	0.40 (0.090)	-0.11 (-0.69, 0.48)	0.7207	-0.0034	457.75	2.4355
	270	95%: 0.30 (0.29)	0.43 (0.085)	-0.13 (-0.72, 0.46)	0.6714	-0.0031	502.26	2.5253
	283	99%: 0.30 (0.30)	0.54 (0.083)	-0.24 (-0.85, 0.38)	0.4499	-0.0015	564.13	2.6997

Footnote: FIBERR and RNP were measured 1 month minus baseline. The 'difference' column was subtracted FIBERR minus RNP. Positive values indicate improvement for fat intentions. Significant values were bolded and marginally significant values were indicated with at *.

4.6 Fiber Intentions

Likewise, dietary fiber intentions was analyzed using all the outline methods of analysis and gave the same result. There was no significant different between FIBERR and RNP in the change from baseline to one month in the participants' intentions to increase the amount of fiber in their diet, which is evident in Table 9. In other words, all of the methods of analysis failed to reject the null hypothesis that there was no difference in the change from baseline between FIBERR and RNP. When comparing the AICc values between the like analyses, the linear regression model were the propensity and the logit of the propensity score were used as the single covariate had the smallest value. When comparing adjusted R^2, the unadjusted model had the greatest value. The propensity score matching had the smallest root mean square error. Similarly as before, the weighted analyses were similar to each other and much worse than the other methods. The linear regression model where propensity score was the single covariate appears to be the best from group 2.

Table 9: Fiber Intentions

Method	N	$FIBERR_{adj}$	RNP_{adj}	Difference (CI)	p-value	R^2_{adj}	AIC_{adj}	Root MSE
Unadjusted Comparison	294	0.42 (0.15)	0.45 (0.090)	-0.026 (-0.37, 0.32)	0.5427	0.0013	158.99	1.3170
Propensity Score Matching	139	0.40 (0.16)	0.53 (0.16)	-0.13 (-0.59, 0.32)	0.5542	NA	904.3	0.4767
Multiple Linear Regression	281	0.43 (0.16)	0.46 (0.097)	-0.032 (-0.41, 0.35)	0.8676	-0.0228	174.92	1.3454
Propensity Score as Cont. Covariate	281	Propensity: 0.44 (0.16)	0.45 (0.096)	-0.011 (-0.39, 0.37)	0.9568	-0.0067	165.32	1.3347
	281	Logit of Propensity: 0.43 (0.16)	0.45 (0.096)	-0.021 (-0.40, 0.36)	0.9120	-0.0069	165.32	1.3349
Quintiles of the Propensity Score	281	0.42 (0.16)	0.46 (0.096)	-0.034 (-0.42, 0.35)	0.8596	-0.0172	171.18	1.3417
Weight using Propensity Score	281	0.42 (0.30)	0.44 (0.081)	-0.017 (-0.63, 0.59)	0.9555	-0.0036	546.07	2.6329
Trimmed Weights of Propensity Score (if ≥90/95/99 then make the value the propensity at 90/95/99)	281	90%: 0.42 (0.29)	0.43 (0.082)	-0.0056 (-0.59, 0.58)	0.9849	-0.0036	522.29	2.5238
	281	95%: 0.42 (0.30)	0.44 (0.082)	-0.014 (-0.62, 0.59)	0.9643	-0.0036	541.52	2.6117
	281	99%: 0.42 (0.30)	0.44 (0.081)	-0.019 (-0.63, 0.59)	0.9514	-0.0036	545.96	2.6324
Trimmed Weights of Propensity Score (if ≥90/95/99 then ignore the observation from analysis)	255	90%: 0.44 (0.28)	0.43 (0.087)	-0.0096 (-0.56, 0.58)	0.9735	-0.0039	440.66	2.3634
	269	95%: 0.44 (0.29)	0.44 (0.084)	-0.0044 (-0.59, 0.59)	0.9984	-0.0037	493.55	2.4934
	279	99%: 0.42 (0.30)	0.46 (0.083)	-0.038 (-0.65, 0.57)	0.9017	-0.0036	542.78	2.6357

Footnote: FIBERR and RNP were measured 1 month minus baseline. The 'difference' column was subtracted FIBERR minus RNP. Positive values indicate improvement for fiber intentions. Significant values were bolded and marginally significant values were indicated with at *.

4.7 Fruits and Vegetables Dietary Intake Intentions

In a similar fashion, the outcome variable fruits and vegetables dietary intake intention was analyzed and all the various methods of analysis gave the same result. There was no significant difference between the studies in the change from baseline in the participants' intentions to increase the number of fruits and vegetables they eat, as seen in Table 10. In other words, all of the methods failed to reject the null hypothesis. When comparing the like analyses, the linear regression model where the logit of the propensity score was the single covariate used gave the smallest AICc value. Furthermore, when comparing all the analyses, the multiple linear regression model and the propensity score matching gave the best values of adjusted R^2 and root mean square error respectively. The multiple linear regression model appears to be the best from group 2 and again the weighted analyses appear to be the worst of all the various methods.

Table 10: Fruits and Vegetables Intentions

Method	N	$FIBERR_{adj}$	RNP_{adj}	Difference (CI)	p-value	R^2_{adj}	AIC_{adj}	Root MSE
Unadjusted Comparison	298	0.39 (0.15)	0.33 (0.092)	0.072 (-0.29, 0.41)	0.7272	0.0004	182.74	1.3543
Propensity Score Matching	141	0.38 (0.15)	0.30 (0.15)	0.087 (-0.33, 0.50)	0.6761	NA	873.5	0.4549
Multiple Linear Regression	285	0.46 (0.16)	0.32 (0.10)	0.14 (-0.24, 0.52)	0.4644	0.0145	176.14	1.3427
Propensity Score as Cont. Covariate	285	Propensity: 0.46 (0.16)	0.32 (0.097)	0.15 (-0.24, 0.53)	0.4520	-0.0006	175.27	1.3347
	285	Logit of Propensity: 0.46 (0.16)	0.032 (0.097)	0.15 (-0.23, 0.53)	0.4456	0.0004	175.00	1.3349
Quintiles of the Propensity Score	285	0.43 (0.16)	0.32 (0.097)	0.11 (-0.27, 0.49)	0.5576	0.0001	178.06	1.3524
Weight using Propensity Score	285	0.39 (0.31)	0.36 (0.083)	0.024 (-0.60, 0.65)	0.9494	-0.0034	644.40	2.7475
Trimmed Weights of Propensity Score (if ≥90/95/99 then make the value the propensity at 90/95/99)	285	90%: 0.39 (0.29)	0.36 (0.085)	0.025 (-0.58 0.63)	0.9353	-0.0035	554.20	2.6347
	285	95%: 0.39 (0.30)	0.36 (0.083)	0.024 (-0.59, 0.64)	0.9386	-0.0035	567.36	2.6962
	285	99%: 0.39 (0.31)	0.37 (0.083)	0.022 (-0.60, 0.65)	0.9452	-0.0035	577.77	2.7459
Trimmed Weights of Propensity Score (if ≥90/95/99 then ignore the observation from analysis)	256	90%: 0.32 (0.30)	0.39 (0.097)	-0.065 (-0.69, 0.56)	0.8371	-0.0038	495.13	2.6199
	270	95%: 0.32 (0.30)	0.32 (0.088)	0.0062 (-0.60, 0.61)	0.9839	-0.0037	517.70	2.5987
	283	99%: 0.39 (0.30)	0.42 (0.082)	-0.036 (-0.65, 0.57)	0.9083	-0.0035	557.79	2.6697

Footnote: FIBERR and RNP were measured 1 month minus baseline. The 'difference' column was subtracted FIBERR minus RNP. Positive values indicate improvement for fruits and vegetables intentions. Significant values were bolded and marginally significant values were indicated with at *.

4.8 Fat Self-Efficacy

Fat self-efficacy was analyzed by all the methods of analysis. All of the various methods gave the same result that there was no significant difference between the studies in the participants' self-efficacy in fat consumption. The details are outline in Table 11. In other words, all of the methods of analysis failed to reject the null hypothesis that there was no difference in the change from baseline between the groups. When comparing all the methods of analysis, the unadjusted model and the propensity score matching gave the best values of adjusted R^2 and root mean square error respectively. Also, when comparing the like analyses, the linear regression with propensity as the covariate, gave the best value of AICc. The multiple linear regression model also was the best from group 2. Again, the weighted analyses were the worst.

Table 11: Fat Self-Efficacy

Method	N	$FIBERR_{adj}$	RNP_{adj}	Difference (CI)	p-value	R^2_{adj}	AIC_{adj}	Root MSE
Unadjusted Comparison	294	-0.35 (0.10)	-0.18 (0.063)	-0.17 (-0.41, 0.069)	0.1622	0.0033	-47.23	0.9197
Propensity Score Matching	139	-0.34 (0.12)	-0.28 (0.12)	-0.060 (-0.39, 0.27)	0.7179	NA	-666.6	0.3986
Multiple Linear Regression	281	-0.34 (0.11)	-0.19 (0.066)	-0.15 (-0.41, 0.11)	0.2521	-0.0014	-37.13	0.9235
Propensity Score as Cont. Covariate	281	Propensity: -0.34 (0.11)	-0.20 (0.067)	-0.14 (-0.41, 0.12)	0.2963	-0.0013	-41.73	0.9235
	281	Logit of Propensity: -0.34 (0.11)	-0.19 (0.067)	-0.14 (-0.41, 0.12)	0.2813	-0.0014	-41.71	0.9236
Quintiles of the Propensity Score	281	-0.36 (0.10)	-0.19 (0.066)	-0.16 (-0.43, 0.099)	0.2212	-0.0047	-37.88	0.9251
Weight using Propensity Score	281	-0.35 (0.22)	-0.20 (0.057)	-0.15 (-0.59, 0.29)	0.5109	-0.0020	363.24	1.9018
Trimmed Weights of Propensity Score (if ≥90/95/99 then make the value the propensity at 90/95/99)	281	90%: -0.35 (0.21)	-0.19 (0.059)	-0.15 (-0.58, 0.27)	0.4824	-0.0018	342.85	1.8340
	281	95%: -0.35 (0.21)	-0.20 (0.058)	-0.14 (-0.58, 0.29)	0.5192	-0.0021	359.73	1.8899
	281	99%: -0.35 (0.22)	-0.20 (0.057)	-0.15 (-0.58, 0.29)	0.5129	-0.0020	363.20	1.9016
Trimmed Weights of Propensity Score (if ≥90/95/99 then ignore the observation from analysis)	252	90%: -0.36 (0.21)	-0.14 (0.065)	-0.22 (-0.64, 0.20)	0.3078	0.0002	285.71	1.7558
	266	95%: -0.37 (0.21)	-0.21 (0.063)	-0.16 (-0.60, 0.28)	0.4665	-0.0018	331.01	1.8560
	279	99%: -0.35 (0.22)	-0.21 (0.059)	-0.14 (-0.58, 0.30)	0.5393	-0.0022	362.12	1.9067

Footnote: FIBERR and RNP were measured 1 month minus baseline. The 'difference' column was subtracted FIBERR minus RNP. Negative values indicate improvement for fat self-efficacy. Significant values were bolded and marginally significant values were indicated with at *.

4.9 Fruits and Vegetables Self-Efficacy

The outcome variable, fruits and vegetables self-efficacy, of all the methods of analysis only one of them found that there was a significant difference between the two studies in the change from baseline. The only method that found the significant difference was the propensity score matching ($p = 0.0220$), rejecting the null hypothesis in favor of the alternative hypothesis that there was a difference in the change from baseline between FIBERR and RNP. There were a few other methods that had marginally significant results; multiple linear regression ($p = 0.0567$), quintiles analysis s ($p = 0.0565$), and linear regression where propensity score and logit of the propensity score was the only single covariate ($p = 0.0733$ and $p = 0.0662$ respectively). All of the other methods resulted in no significant difference between the studies in the change from baseline to one month, as seen in Table 12. When comparing the like analyses with respect to sample size, the multiple linear regression model gave the lowest AICc value. It also gave the best adjusted R^2 value when compared to all of the methods of analysis. The propensity score matching gave the best root mean squared error. Again, the weighted analyses were similar to each other and much worse than all the other methods of analysis. From group 2, the multiple linear regression model was the best.

Table 12: Fruits and Vegetables Self-Efficacy

Method	N	$FIBERR_{adj}$	RNP_{adj}	Difference (CI)	p-value	R^2_{adj}	AIC_{adj}	Root MSE
Unadjusted Comparison	299	0.29 (0.16)	0.096 (0.095)	0.19 (-0.17, 0.55)	0.2962	0.0037	203.86	1.4015
Propensity Score Matching	142	0.30 (0.15)	-0.19 (0.15)	0.49 (0.072, 0.90)	**0.0220**	NA	869.1	0.4565
Multiple Linear Regression	286	0.42 (0.16)	0.046 (0.098)	0.37 (-0.011, 0.76)	0.0567	0.0429	190.11	1.3744
Propensity Score as Cont. Covariate	286	Propensity: 0.41 (0.17)	0.049 (0.10)	0.36 (-0.034, 0.76)	0.0733	0.0119	194.04	1.3965
	286	Logit of Propensity: 0.41 (0.17)	0.048 (0.099)	0.37 (-0.025, 0.76)	0.0662	0.0151	193.10	1.3942
Quintiles of the Propensity Score	286	0.43 (0.17)	0.046 (0.10)	0.38 (-0.011, 0.78)	0.0565	0.0083	198.08	1.3991
Weight using Propensity Score	286	0.29 (0.34)	0.20 (0.090)	0.088 (-0.60, 0.77)	0.7992	-0.0033	630.27	2.9993
Trimmed Weights of Propensity Score (if ≥90/95/99 then make the value the propensity at 90/95/99)	286	90%: 0.29 (0.32)	0.17 (0.091)	0.12 (-0.53, 0.76)	0.7212	-0.0031	594.82	2.8190
	286	95%: 0.29 (0.33)	0.19 (0.090)	0.10 (-0.57, 0.77)	0.7671	-0.0032	619.28	2.9422
	286	99%: 0.29 (0.34)	0.20 (0.090)	0.087 (-0.50, 0.77)	0.8021	-0.0033	630.06	2.9982
Trimmed Weights of Propensity Score (if ≥90/95/99 then ignore the observation from analysis)	257	90%: 0.16 (0.29)	0.065 (0.094)	0.095 (-0.51, 0.70)	0.7586	-0.0035	481.55	2.5421
	272	95%: 0.18 (0.31)	0.14 (0.091)	0.045 (-0.59, 0.68)	0.8905	-0.0036	549.39	2.7352
	284	99%: 0.29 (0.33)	0.25 (0.090)	0.037 (-0.64, 0.71)	0.9148	-0.0035	615.01	2.9424

Footnote: FIBERR and RNP were measured 1 month minus baseline. The 'difference' column was subtracted FIBERR minus RNP. Positive values indicate improvement for fruits and vegetables self-efficacy. Significant values were bolded and marginally significant values were indicated with at *.

CHAPTER 5

Discussion

5.1 Discussion of Overall Outcomes

Three of the outcome measure showed a different between groups in improvement in diet via at least one method of analysis; fat behavior, fruits and vegetables behavior, and fruits and vegetables self-efficacy. For the fat behavior, all of the methods of analysis except the propensity score matching, weighting, and trimmed weighting did not show significant results. Results of methods which were significant implied that there was a greater improvement in diet with respect to eating less fat in the RNP study than in the FIBERR study. Also, for fruits and vegetables behavior the propensity score matching and treating the propensity score as a continuous covariate in a linear regression model resulted in a significant difference between the two studies in the change from baseline. The participants in the RNP study increased the number of fruits and vegetables they consume more than those in the FIBERR study. Lastly, the fruits and vegetables self-efficacy variable showed a significant difference between the studies in the change from baseline to one month using the propensity score matching method. This method's results imply that when comparing outcomes, FIBERR study participants thought they would be able to increase the number of fruits and vegetables they eat more than the participants in the RNP study. In fact, the RNP study showed a decline is self-efficacy whereas the FIBERR study showed an improvement over the baseline to 1 month time period using the propensity score matching method. This contradicts what actually happened as mentioned above.

The original hypothesis was that the participants in the FIBERR study would be more likely to adhere to the diet intervention because of the knowledge they gained having a family member with colon cancer. For many of the outcome variables, it was that the RNP study participants had a greater improvement than the FIBERR study participants. This could be due to FIBERR participants starting off with better diets than the RN P participants, leaving little room to improve. They could have started eating healthier prior to the study.

5.2 Comparison of Methods of Analysis

There were a number of methods of analysis that were implemented with this particular dataset to control for potential confounders; controlling for covariates directly in a multiple linear regression model, propensity score matching, using the propensity score as a continuous covariate in a linear regression model, analyzing quintiles of the propensity score, weighting the propensity score, and various way of trimming the weights on the propensity score. The ones that could be compared using the AICc were those with the same sample size in the analysis (multiple linear regression model, using propensity score/logit of propensity score as a covariate in a linear regression model, quintiles analysis, the weighted analysis, and the trimmed weighted analysis where all values at the cut point were set equal to the cut point). The RMSE and adjusted R^2 could be compared across all of the methods of analysis. Furthermore the methods grouped themselves into 3 groups; (1) propensity score matching, (2) multiple linear regression, linear regression with propensity score or the logit of propensity score as the single covariate, and the quintiles

analysis (3) the weighted and trimmed weighted analyses. The weighted analyses were the worst of all the other methods with respect to the adjusted R^2, AICc, and root mean square error. In addition, the weighted analyses failed to control for the baseline confounding. It appears that doing a weighted or trimmed weighted analysis was not the best method of analysis for this particular dataset. This was unexpected because much of the literature gave reason to believe that the weighted analysis would have been one of the better methods of analysis. Had the weights been trimmed further at the 85^{th} or 80^{th} percentile, then the weighting analysis may have model the data better.

The propensity score matching method eliminated many participants from analysis, analyzing only144 from the total sample of 305. This could have potentially lowered the power for this particular method. In the fruits and vegetables self-efficacy outcome, propensity score matching was the only method to show significant results and in the fat behavior, it showed insignificant results where others gave significant results. Further analysis was run comparing the unmatched set to the matched set in the propensity score matching analysis to determine if there were any outstanding outliers in the unmatched set that caused the results to differ in the propensity score matching in comparison to the other methods (Tables 13-16). Tables 13 and 14 contain the comparison of the baseline variables considered for confounding and tables 15 and 16 contain the outcome variables analyzed at baseline, 1 month and the change from baseline.

Table 13: Comparison of Matched verses Unmatched Set in RNP of Potential Confounders

RNP	Count (%) in the Matched Group (N = 72)	Count (%) in the Unmatched group (N = 152)	p-value comparing the Matched and Unmatched groups
Covariates			
Ethnicity			**0.0084**
Black	16 (22.22)	61 (40.13)	
Non-Black	56 (77.78)	91 (59.87)	
Education			**<0.0001**
Some HS or Less	3 (4.17)	30 (19.87)	
HS	16 (22.22)	61 (40.40)	
Some College or Tech School	17 (23.61)	47 (31.13)	
College or More	36 (50.00)	13 (8.61)	
Gender			0.9829
Male	25 (34.72)	53 (34.87)	
Female	47 (65.28)	99 (65.13)	
Town			**0.0012**
In Town	24 (33.33)	22 (14.57)	
Out of Town	48 (66.67)	129 (85.43)	
Eating Out			0.5782
No Meals Out	39 (54.17)	93 (61.18)	
One Meal Out	24 (33.33)	38 (25.00)	
Two Meals Out	8 (11.11)	17 (11.18)	
All Meals Out	1 (1.39)	4 (2.63)	
Marital Status			0.8774
Married	50 (69.44)	104 (68.42)	
Not Married	22 (30.56)	48 (31.58)	
	Mean (SD) in the Matched Group (N = 72)	Mean (SD) in the Unmatched group (N = 152)	p-value comparing the Matched and Unmatched groups
Age	46.26 (13.65)	50.42 (1.25)	**0.0355**
Doctor's Visits	4.32 (5.72)	3.75 (4.83)	0.4474
Fat Knowledge	5.86 (0.42)	5.38 (1.25)	**0.0015**
TV Hours	14.39 (11.05)	13.60 (10.67)	0.6135
Sum of FSS	12.99 (4.41)	13.04 (4.84)	0.9355
Meals Shopped/ Planned/Prepared	7.33 (6.83)	7.24 (6.89)	0.7643

Footnote: Significant values were bolded.

Table 14: Comparison of Matched verses Unmatched Set in FIBERR of Potential Confounders

FIBERR	Count (%) in the Matched Group (N = 72)	Count (%) in the Unmatched group (N = 152)	p-value comparing the Matched and Unmatched groups
Covariates			
Ethnicity			0.1144
Black	16 (22.22)	0 (0.00)	
Non-Black	56 (77.78)	9 (100.00)	
Education			**0.0244**
Some HS or Less	4 (5.56)	0 (0.00)	
HS	16 (22.22)	0 (0.00)	
Some College or Tech School	19 (26.39)	0 (0.00)	
College or More	33 (45.83)	9 (100.00)	
Gender			0.1336
Male	22 (30.56)	5 (55.56)	
Female	50 (69.44)	4 (44.44)	
Town			**0.0004**
In Town	27 (32.50)	9 (100.00)	
Out of Town	45 (62.50)	0 (0.00)	
Eating Out			**0.0118**
No Meals Out	38 (52.78)	0 (0.00)	
One Meal Out	19 (26.39)	4 (44.44)	
Two Meals Out	14 (19.44)	4 (44.44)	
All Meals Out	1 (1.39)	1 (11.11)	
Marital Status			0.0771
Married	46 (63.89)	3 (33.33)	
Not Married	26 (36.11)	6 (66.67)	
	Mean (SD) in the Matched Group (N = 72)	Mean (SD) in the Unmatched group (N = 152)	p-value comparing the Matched and Unmatched groups
Age	47.35 (12.28)	42.67 (12.46)	0.2850
Doctor's Visits	4.56 (5.22)	14.11 (19.49)	**0.0010**
Fat Knowledge	5.79 (0.50)	6 (0.00)	0.2191
TV Hours	12.91 (9.29)	6.44 (3.40)	**0.0427**
Sum of FSS	11.93 (5.14)	13.56 (6.75)	0.3908
Meals Shopped/ Planned/Prepared	7.56 (2.21)	6.89 (2.57)	0.4050

Footnote: Significant values were bolded.

Table 15: Comparison of Matched verses Unmatched Set in RNP of Outcome Variables

RNP	Mean (SD) in the Matched Group (N = 72)	Mean (SD) in the Unmatched group (N = 152)	p-value comparing the Matched and Unmatched groups
Baseline			
Fat Behavior	1.92 (0.30)	2.09 (0.36)	**0.0004**
Fiber Behavior	2.22 (0.35)	2.26 (0.37)	0.7232
FV Behavior	3.04 (1.71)	2.51 (1.61)	**0.0260**
Fat Intentions	3.35 (1.32)	3.26 (1.31)	0.6401
Fiber Intentions	2.92 (1.36)	3.09 (1.30)	0.3471
FV Intentions	3.48 (1.16)	3.61 (1.28)	0.4707
Fat Self-Efficacy	2.06 (0.84)	1.91 (0.87)	0.2286
FV Self-Efficacy	4.00 (1.02)	3.62 (1.31)	**0.0337**
One Month			
Fat Behavior	1.78 (0.32)	1.88 (0.34)	**0.0454**
Fiber Behavior	2.07 (0.42)	2.05 (0.38)	0.8544
FV Behavior	4.01 (1.56)	3.14 (1.77)	**0.0005**
Fat Intentions	3.54 (1.12)	3.73 (1.19)	0.2467
Fiber Intentions	3.39 (1.26)	3.50 (1.31)	0.5593
FV Intentions	3.74 (1.22)	3.92 (1.22)	0.3233
Fat Self-Efficacy	1.80 (0.73)	1.73 (0.72)	0.5372
FV Self-Efficacy	3.77 (1.17)	3.85 (1.28)	0.6886
Change from Baseline			
Fat Behavior	-0.14 (0.27)	-0.22 (0.34)	0.0789
Fiber Behavior	-0.15 (0.36)	-0.20 (0.35)	0.3429
FV Behavior	1.00 (1.39)	0.62 (1.69)	0.0789
Fat Intentions	0.21 (1.12)	0.50 (1.42)	0.1394
Fiber Intentions	0.49 (1.07)	0.43 (1.31)	0.7640
FV Intentions	0.30 (1.23)	0.34 (1.42)	0.8485
Fat Self-Efficacy	-0.27 (0.98)	-0.13 (0.92)	0.3014
FV Self-Efficacy	-0.21(1.29)	0.24 (1.46)	**0.0262**

Footnote: Significant values were bolded.

Table 16: Comparison of Matched verses Unmatched Set in FIBERR of Outcome Variables

FIBERR	Mean (SD) in the Matched Group (N = 72)	Mean (SD) in the Unmatched group (N = 9)	p-value comparing the Matched and Unmatched groups
Baseline			
Fat Behavior	1.97 (0.37)	1.95 (0.31)	0.8677
Fiber Behavior	2.24 (0.32)	2.18 (0.29)	0.5976
FV Behavior	3.13 (1.44)	3.11 (1.45)	0.9784
Fat Intentions	3.36 (1.32)	3.00 (1.00)	0.4329
Fiber Intentions	3.24 (1.29)	3.00 (1.12)	0.5857
FV Intentions	3.56 (1.27)	3.44 (1.00)	0.8041
Fat Self-Efficacy	1.94 (0.84)	2.11 (0.78)	0.5976
FV Self-Efficacy	3.63 (1.18)	3.67 (1.22)	0.9210
One Month			
Fat Behavior	1.88 (0.34)	1.87 (0.47)	0.9296
Fiber Behavior	2.15 (0.35)	1.95 (0.25)	0.0932
FV Behavior	3.49 (1.64)	4.11 (1.17)	0.2710
Fat Intentions	3.66 (1.15)	3.11 (1.36)	0.1872
Fiber Intentions	3.58 (1.06)	3.56 (0.88)	0.9530
FV Intentions	3.90 (0.96)	4.00 (1.32)	0.7817
Fat Self-Efficacy	1.60 (0.65)	1.67 (0.71)	0.7738
FV Self-Efficacy	3.89 (0.95)	4.00 (1.12)	0.7431
Change from Baseline			
Fat Behavior	-0.092 (0.23)	-0.081 (0.31)	0.9011
Fiber Behavior	-0.089 (0.30)	-0.23 (0.19)	0.1573
FV Behavior	0.36 (1.37)	1.00 (1.80)	0.2059
Fat Intentions	0.32 (1.25)	0.11 (0.78)	0.6209
Fiber Intentions	0.41 (1.55)	0.56 (1.33)	0.7825
FV Intentions	0.37 (1.30)	0.56 (1.67)	0.6914
Fat Self-Efficacy	-0.33 (0.89)	-0.44 (0.73)	0.7197
FV Self-Efficacy	0.28 (1.42)	0.33 (0.71)	0.9148

Footnote: Significant values were bolded.

The unmatched set in the FIBERR study was very small N = 9 therefore the comparisons may not be as powerful as the comparisons between the matched and unmatched set in the RNP study where the unmatched sample was N = 152 because of the small sample size. As seen in the table, there is a significant difference in the matched unmatched sets in the RNP study between the outcome variables fat behavior at baseline and one month, fruits and vegetables at baseline and one month, and fruits and vegetables self-efficacy at baseline. These are the outcome variables that were found to be significantly different using at least one of the methods of analysis.

For the fat behavior outcome, the method using propensity score matching did not give significant results. It appears that the methods where certain participants are eliminated or weighted differently give different results, perhaps, indicating that the participants in the unused group are much different from those in the used group, which is evident in the comparison of the matched verses unmatched sets.

For the outcome variable fruits and vegetables behavior the only methods of analysis that resulted in significant results were the propensity score matching and the method where the propensity score is treated as a continuous variable in a linear regression model. This can be attributed to the fact that the groups used in the propensity score matching were significantly from those not used in the matching. Had there been a bigger sample size to choose the matches from, the propensity score matching method may have been more accurate.

Likewise, for the fruits and vegetables self-efficacy variables, the matched set was significantly different between the matched and unmatched set at baseline measurement, 1

month, and for the change from baseline. The propensity score matching method was the only method that gave significantly different results in the change from baseline. This could be because it used only a portion of the total sample. Those participants not used in the matched sample could have been the reason for why all other methods resulted in an insignificant difference in the studies in change from baseline to one month.

Analysis using propensity score in a linear regression model assumes that the propensity score is linearly related to the outcome variable. To examine this assumption, quintiles analysis was further analyzed to determine whether the change from baseline was linear across the quintiles for each of the outcome variables. If they are linear than the assumption made that propensity score is linearly related to the outcome variable is reasonable. If the plots do not appear to be linear, then the assumption may be not be valid, indicating that using the propensity score in a linear regression model may not have been the optimal fit. As seen in the Figures 11-26, most of the plots do not appear to be linear, indicating that using the propensity score as a single covariate in a linear regression model may not have been the optimal fit for the data.

Figure 11: Change from Baseline for Fat Behavior for FIBERR

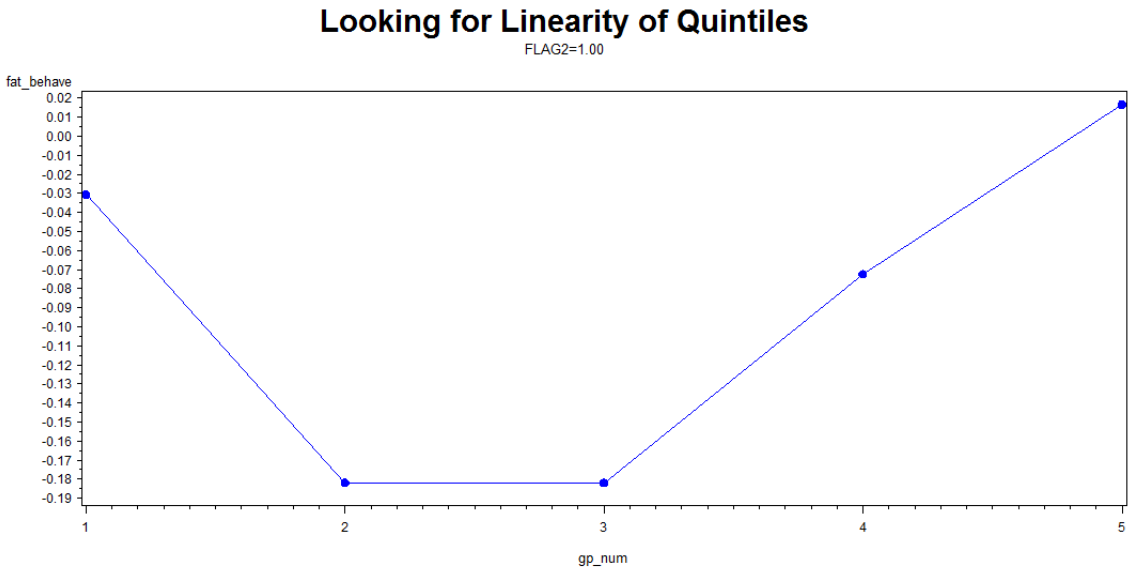


Figure 12: Change from Baseline for Fat Behavior for RNP

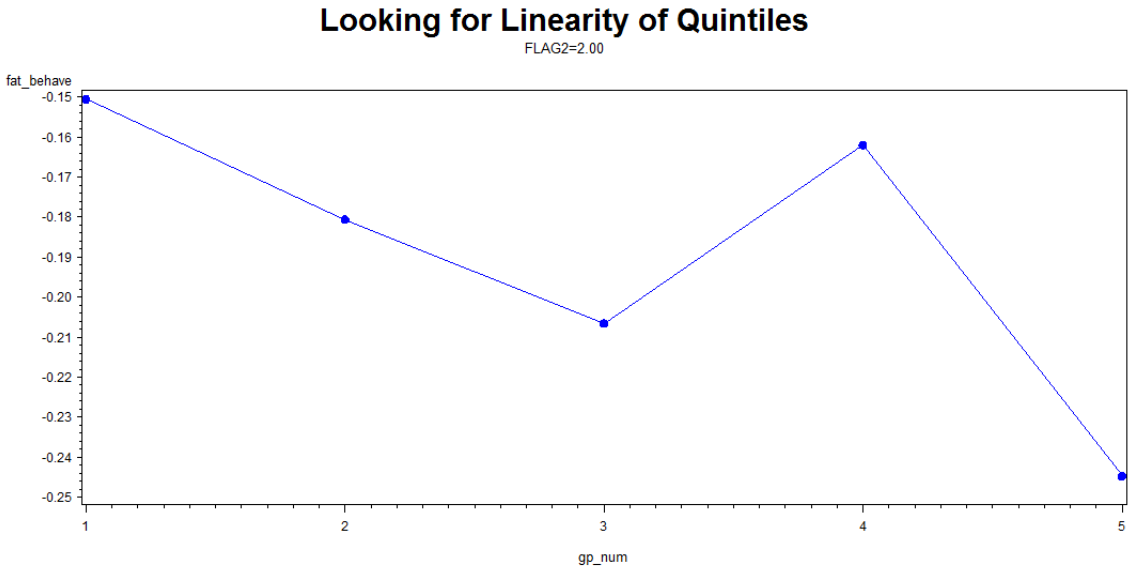


Figure 13: Change from Baseline for Fiber Behavior for FIBERR

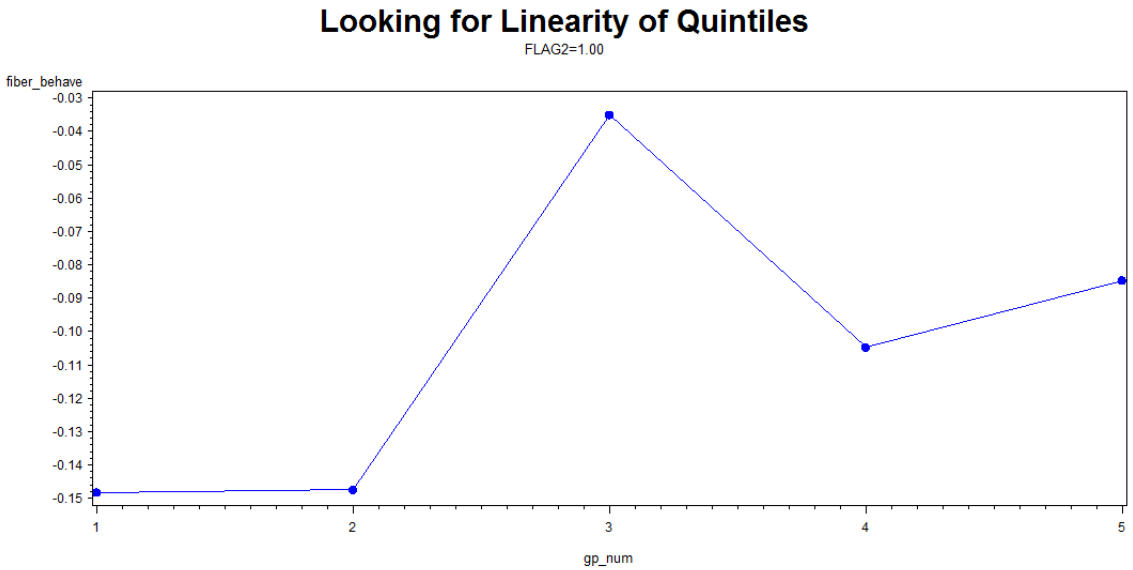


Figure 14: Change from Baseline for Fiber Behavior for RNP

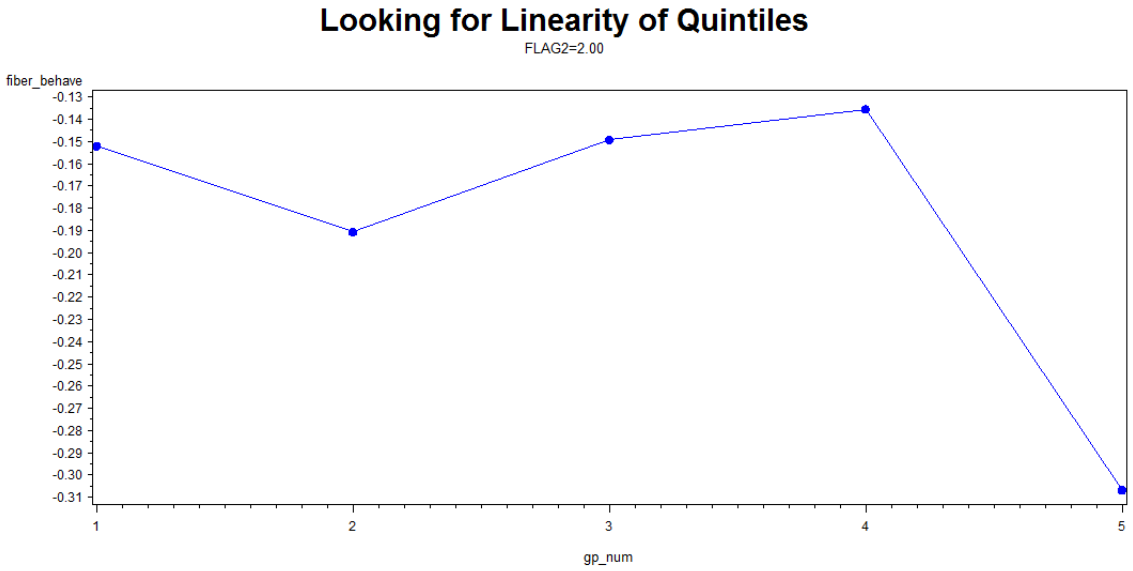


Figure 15: Change from Baseline for FV Behavior for FIBERR

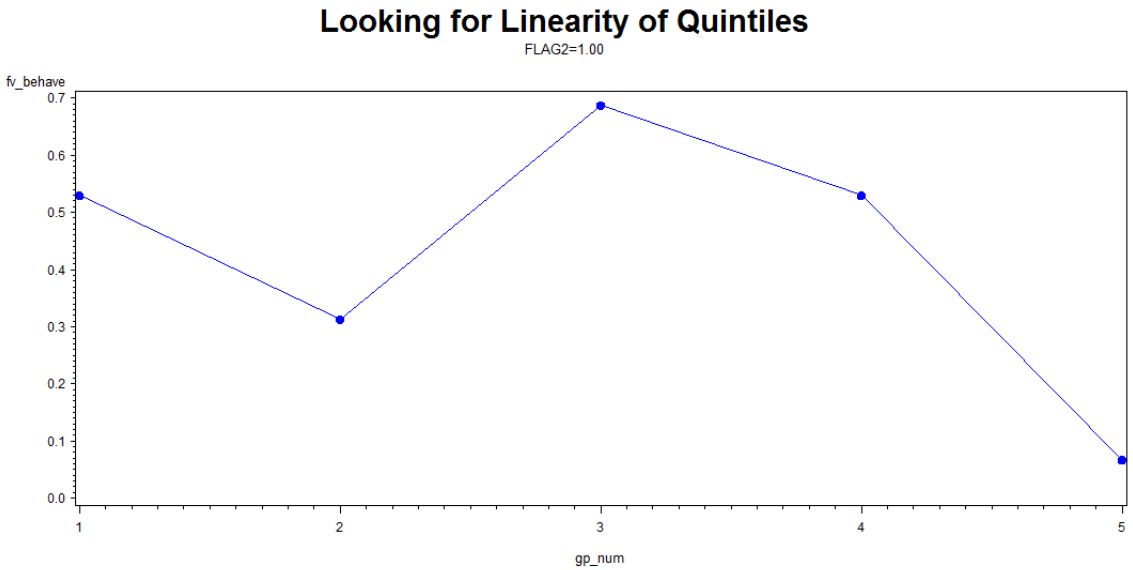


Figure 16: Change from Baseline for FV Behavior for RNP

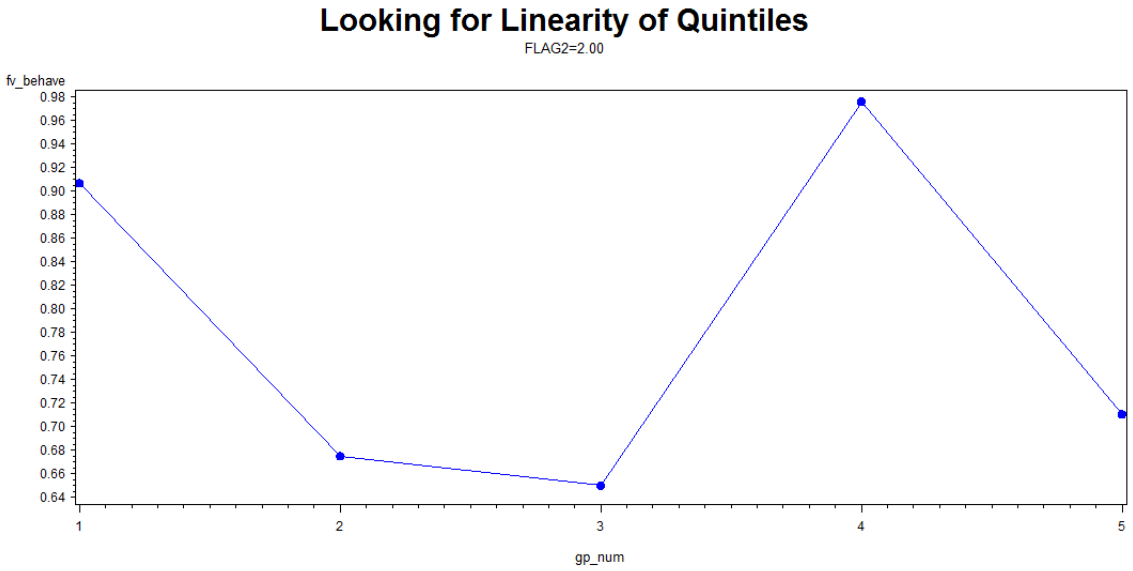


Figure 17: Change from Baseline for Fat Intention for FIBERR

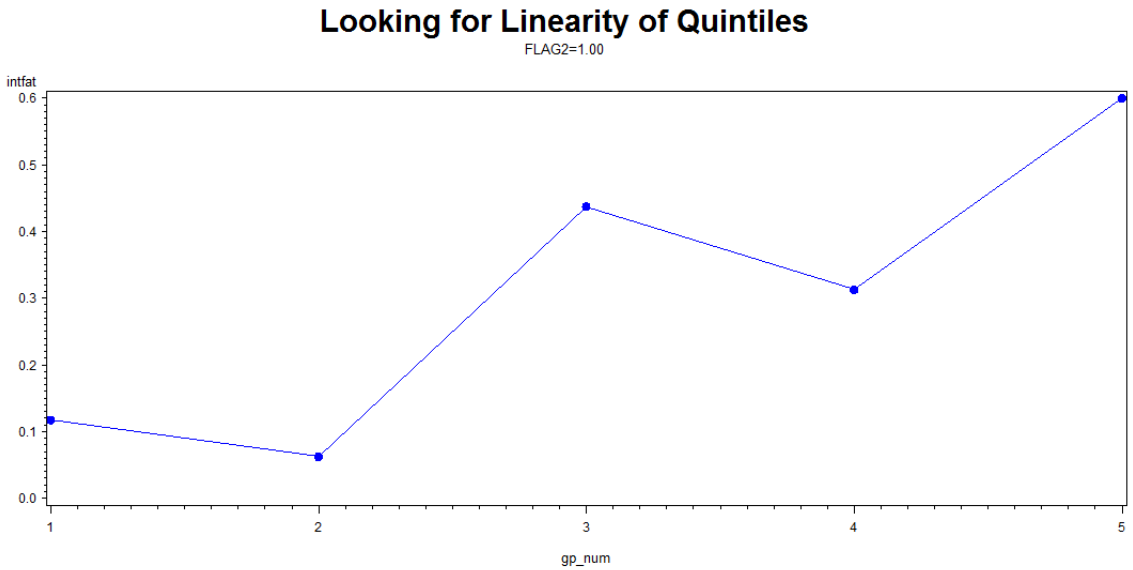


Figure 18: Change from Baseline for Fat Intentions for RNP

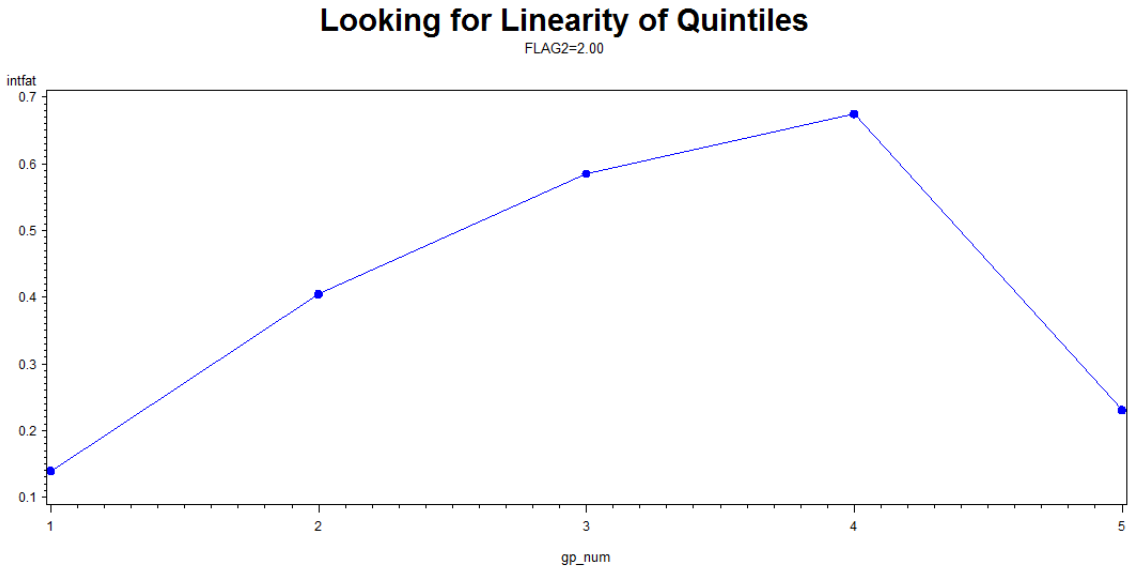


Figure 19: Change from Baseline for Fiber Intentions for FIBERR

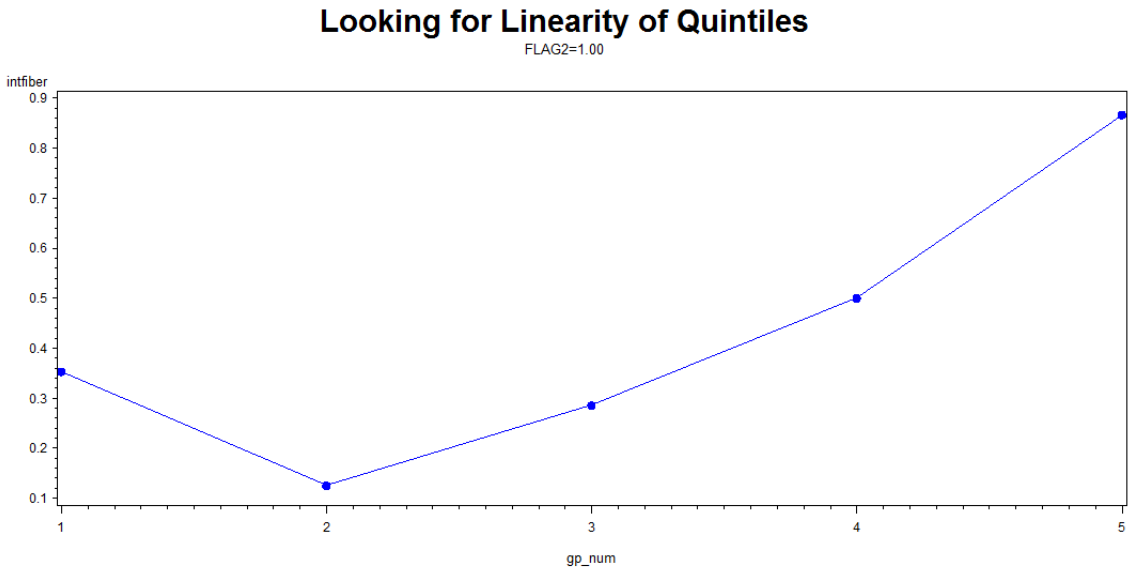


Figure 20: Change from Baseline for Fiber Intentions for RNP

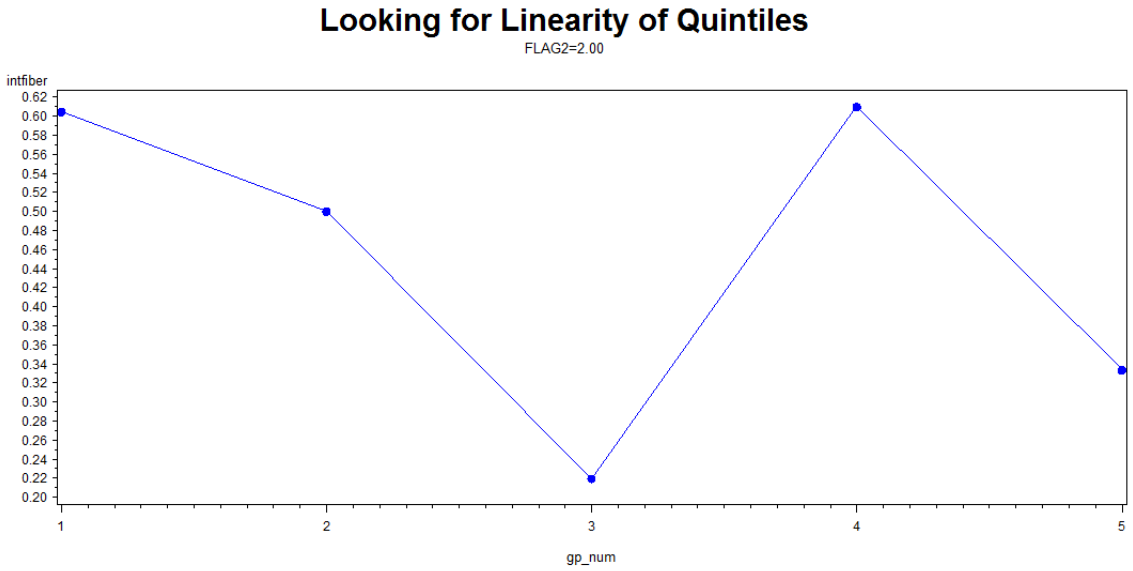


Figure 21: Change from Baseline for FV Intentions for FIBERR

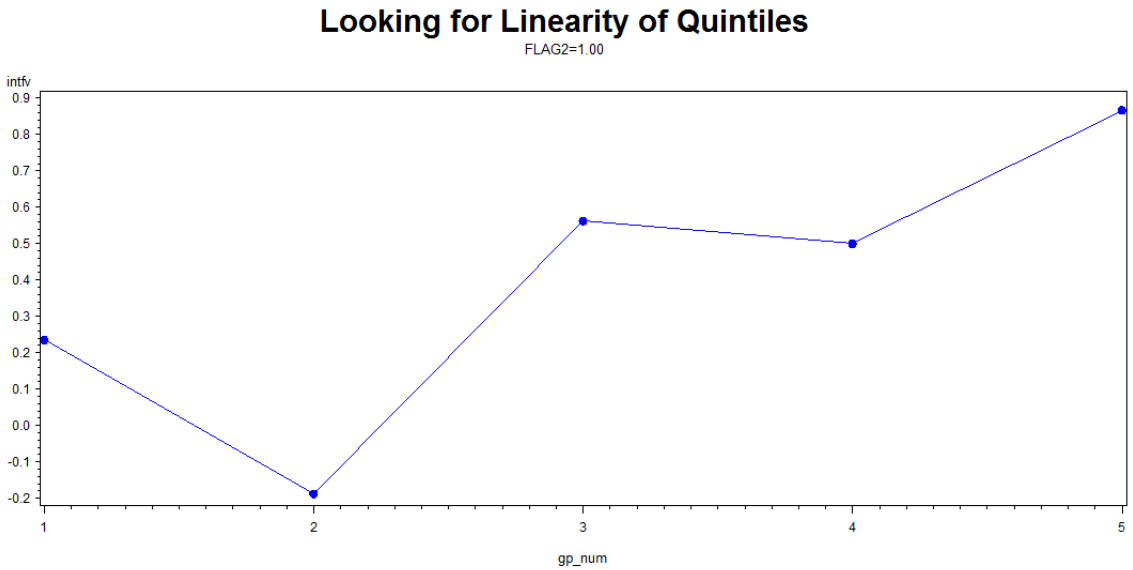


Figure 22: Change from Baseline for FV Intentions for RNP

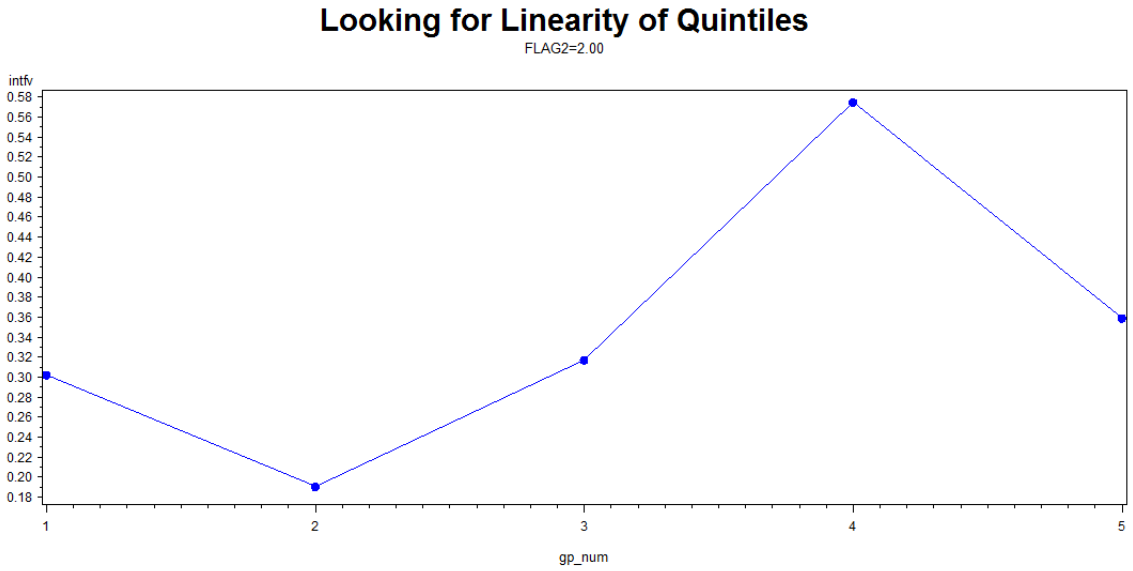


Figure 23: Change from Baseline for Fat Self-Efficacy for FIBERR

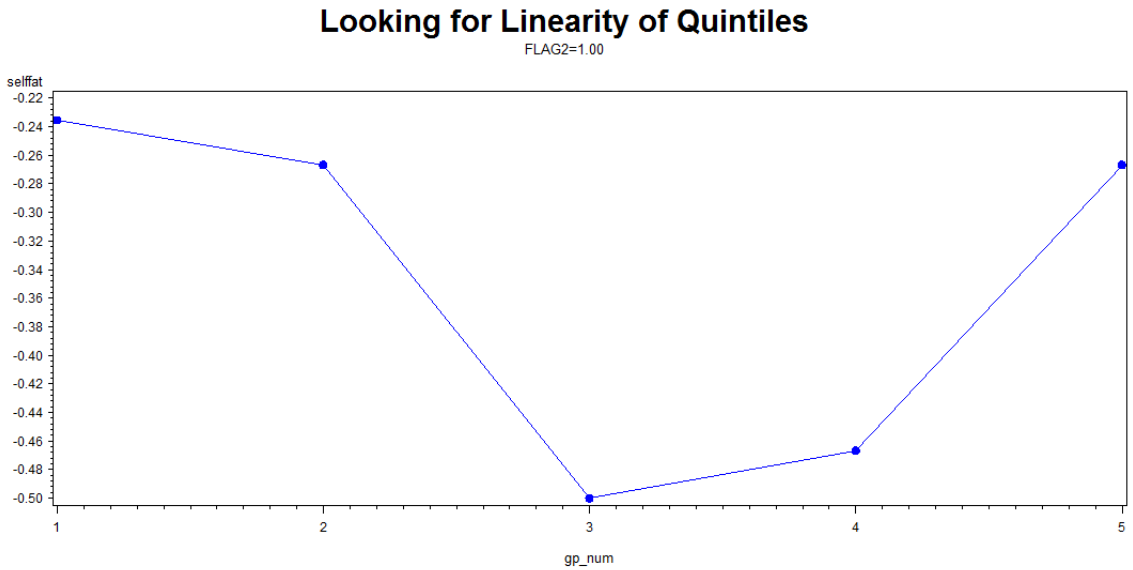


Figure 24: Change from Baseline for Fat Self-Efficacy for RNP

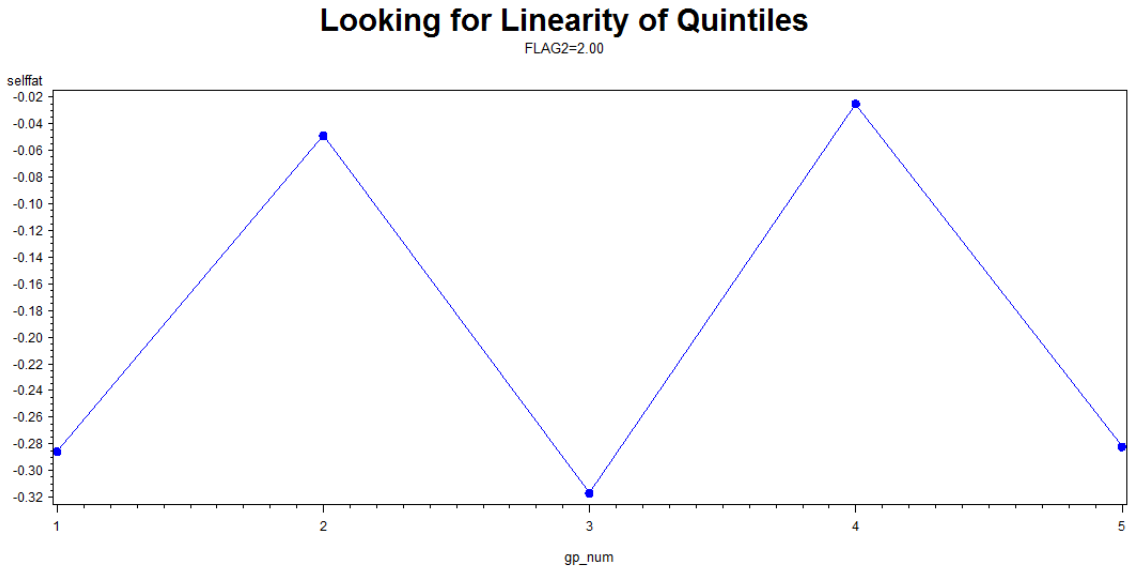


Figure 25: Change from Baseline for FV Self-Efficacy for FIBERR

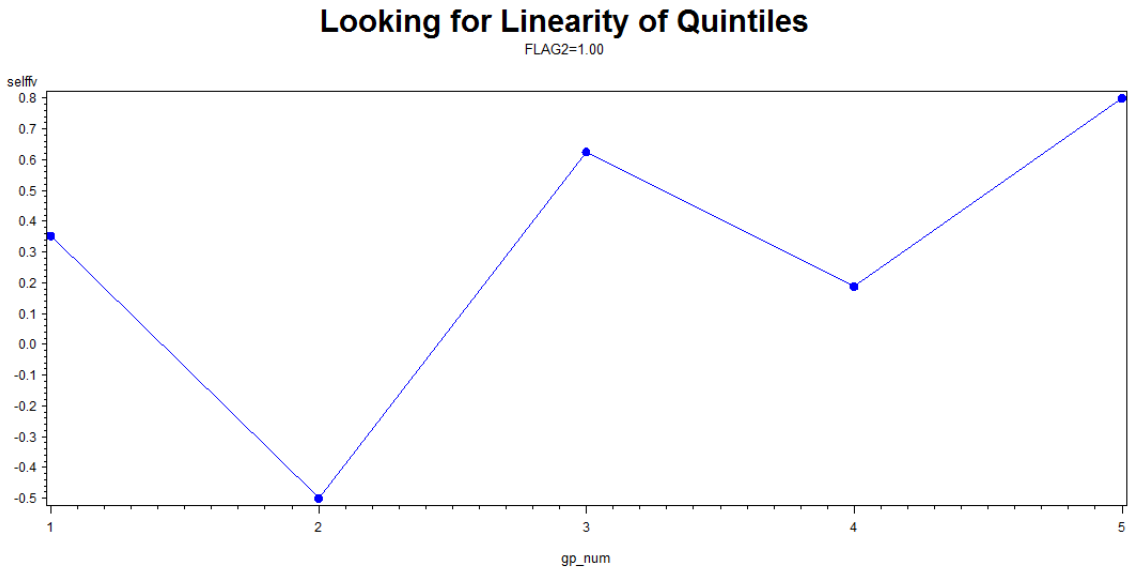
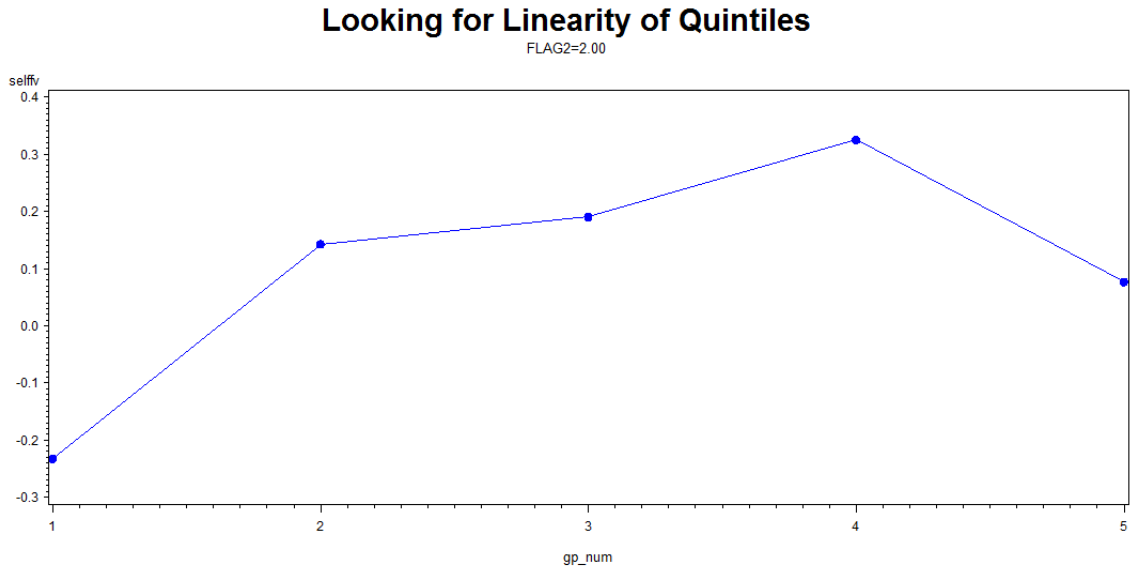


Figure 26: Change from Baseline for FV Self-Efficacy for RNP



5.3 Limitations and Conclusions

There are a few possible limitations to this study. To begin, there were only 12 potential covariates that were analyzed for significant differences. There may have been others that were collected from FIBERR and RNP that may have been good to have used as a potential covariate. Some covariates were collected in one study and not the other or were not collected all together. An example of the former is BMI, which was collected in RNP, but not the FIBERR study. It may have been helpful to have used BMI in the model. Only covariates that were significantly different between the two studies were used to calculate the propensity score. Some covariates that could have clinical significance were left out such as age and gender. It is possible that they should have been left in the model. This could explain why the adjusted R^2 is so small, because they covariates used may not model the outcomes in the best fashion. Also, more complex modeling could have been done in developing the propensity score. Quadratic or interaction terms may have been a better fit for developing the propensity score than what was used.

Different methods of matching may have been better for this dataset. This may have given different results and may have been interesting to compare to the other methods of analysis. Propensity score matching was used as the matching method because the intent was to compare various other ways to use propensity scores to control for confounding. Another limitation to this study is that the results are based on one dataset. It may have been helpful to compare the methods of analysis using various datasets rather than just one. The results could have varied. Different datasets may have given different results especially with respect to the weighted analyses. It could have also been helpful to

simulate data rather than using a real dataset. With a simulation, the true results would have been known, and thus the ability to evaluate the methods improved.

Overall, all the methods of analysis were fairly good in analyzing the data except the weighted analyses. The propensity score matching, multiple linear regression, stratified (quintile) method, and linear regression using propensity score or the logit of the propensity score as the covariate controlled for confounding the best, based on Table 4. These methods of analysis also gave the best adjusted R^2, AICc, and root mean square error when comparing all the methods. Furthermore, analyzing the linearity of the quintiles showed that the stratified analysis may have been a better method of analysis in comparison to using the propensity score as a covariate in a linear regression model. Therefore, the two methods that best analyzed the data while controlling for confounders was the propensity score matching and the stratified quintiles model of all the methods considered in this thesis.

Literature Cited

Literature Cited

Auld, G. W., and Bruhn C. M., McNulty J., et al. (2000). Reported Adoption of Dietary Fat and Fiber Recommendations Among Consumers. J Am Diet Assoc 100: 52-8.

Austin, P. C. (2009). Some Methods of Propensity-Score Matching had Superior Performance to Others: Results of an Empirical Investigation and Monte Carlo Simulations. Biom J 51: 171-184.

Austin, P.C. (2009). The Relative Ability of Different Propensity Score Methods to Balance Measured Covariates Between Treated and Untreated Subjects in Observational Studies. Med Decis Making 29: 661-677.

Austin, P. C., Grootendorst, P., Anderson, G. M. (2007). A Comparison of the Ability if Different Propensity Score Models to Balance Measured Variables Between Treated and Untreated Subjects: A Monte Carlo Study. Stat Med 26; 734-753.

Austin, P. C., Grootendorst, P., Normand, SL. T., Anderson, G. M., (2007). Conditioning on the Propensity Score can Results in Biased Estimation of Common Measures of Treatment Effect: A Monte Carlo Study. Stat Med 26; 754-768.

Austin, P. C., Mamdani, M. M., Stukel, T. A., Anderson, G. M., and Tu, J. V. (2005). The use of the Propensity Score for Estimating Treatment Effects: Administrative verses Clincal Data. Stat Med 24: 1563-1578.

Bean, M. K., Mazzeo, S. E., and Fries, E. (2008). Family Factors as Correlates of Diet in Relatives of Colon Cancer Patients. Am J Health Behav 32(4): 347-355.

Caracise-Edinboro, P., McClish, D., Kracen, A. C., Bowen, D., Fries, E. (2008). Fruit and Vegetable Dietary Behavior in Response to Low-Intensity Dietary Intervention: The Rural Physician Cancer Prevention Project. J Rural Health 24(3): 299-305.

D'Agostino, R. B. (1998). Tutorial in Biostatistics; Propensity Score Methods for Bias Reduction in Comparison of Treatment to Non-Randomized Control Group. Stat Med 17: 2265-2281.

Fang, Y. (2011). Asymptotic Equivalence between Cross-Validations of Akaike Criteria in Mixed-Effects Models. J Data Sci 9: 15-21.

Freedman, D. A., Berk, R. A. (2008). Weighing Regressions by Propensity Score. Evaluation Review 32: 392-409.

Fries, E., Edinboro, P., McClish, D., Manion, L., Bowen, D., Beresford, S. A. A., and Ripley, J. (2005). Randomized Trial of a Low-Intensity Dietary Intervention in Rural Residents The Rural Physician Cancer Prevention Project. Am J Prev Med 28(2).

Kosanke, J. and Bergstralh, E. (2004). GMATCH. Mayo Clinic College of Medicine

Kurth, T., Walker, A. M., Glynn, R. J., Chan, K. A., Gaziano, J. M., Berger, K., and Robins, J. M. (2006). Results of Multivariable Logistic Regression, Propensity Matching, Propensity Adjustment, and Propensity-based Weighting under Conditions of Non-uniform Effect. Am J Epidemiol 163: 262-270.

Kushi, L. H., Doyle, C., McCullough, M., Rock, C. L., Demark-Wahnefried, W., Bandera, E. V., Gapstur, S., Patel, A. V., Andrews, K., Gansler, T. (2012). The American Cancer Society 2010 Nutrition and Physical Activity Guidelines Advisory Committee (2012), American Cancer Society guidelines on nutrition and physical activity for cancer prevention. CA: Cancer J Clin 62: 30–67.

Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2004). Applied Linear Statistical Models. (5^{th} ed.). McGraw-Hill and Irwin.

Lee, B. K., Lessler, J., and Stuart, E. A. (2011). Weight Trimming and Propensity Score Weighting. PLoS One 6 (3):1-6.

Perkins, S. M., Wanzhu, T., Underhill, M. G., Zhou, X. H., and Murray, M. D.(2000). The Use of Propensity Scores in Pharmacoepidemiologic Research. Pharmacoepidemiol Drug Saf 9: 93-101.

Rosenbaum, P. R., and Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika 70: 71-55.

Rosenbaum, P. R., and Rubin, D. B. (1984). Reducing Bias in Observational Studies Using Sub-classification on the Propensity Score. J Am Stat Assoc 79: 516-524.

Schafer, J. L. and Kang, J. (2008). Average Causal Effects From Nonrandomized Studies: A Practical Guide and Simulated Example. Psych Methods 13(4): 279-313.

Stuart, E. A. (2010). Matching Methods for Causal Inference: A Review and Look Forward. Stat Sci 25(1): 1-21.

Stürmer, R., Joshi, M., Glynn, R. J., Avorn, J., Rothman, K. J., Schneeweis, S. (2006). A Review of the Application of Propensity Score Methods Yielded Increasing Use, Advantages in Specific Settings, but not Substantially Different Estimates Compared with Conventional Multivariable Methods. J Clin Epidemiol 59: 437-447.

Terry, P., Giovannucci, E., Michael, K. B., et al. (2001). Fruit, Vegetables, Dietary Fiber, and Risk of Colorectal Cancer. J Natl Cancer Inst. 93: 525-533.

Todd, P. E. (2007). Matching Estimators. in P. Newman, M. Milgate, and J. Eatwell, eds., The New Palgrave—A Dictionary of Economics, Vol. forthcoming, New York: Macmillan

Weitzen, S., Lapane, K. L., Toledano, A. Y., Hume, A. L., Mor, V. (2004). Principles for Modeling Propensity Scores in Medical Research: a Systematic Literature Review. Pharmacoepidemil Drug Saf 13: 841-853.

Westreich, D., Cole, S. R., Funk, M. J., Brookhart, M. A., Stürmer, T. (2011). A Role of the *C*-Statistic in Variable Selection for Propensity Score Model. Pharmacoepidemil Drug Saf 20: 317-320.

APPENDIX A

SAS Code for Crude Model

```
libname library 'C:\Users\Owner\Documents\SSRP_Thesis\Datasets';
run;

/*Combining the Datasets for FIBERR and RNP into one dataset: */

data rnp1;
set library.rnp;
keep id_x1 cond2 fbrmnx1 fbrmnx2 fatmnx1 fatmnx2 sumfatx1 scfat9x1
scfat9x2 scfat8x1 scfat8x2 scfbrx1 scfbrx2 fv1x1 fv1x2 fv7x1 fv7x2 fv6x1
fv6x2 ethnic
educ gender age income marital town tvhrs drvisit epplanx1 epprepx1
epshopx1 fss1x1 fss2x1 fss3x1 fss4x1 fxx5x1 epbrkx1 epdinx1
eplunx1;
if fatmnx2=. then delete;
if cond2=1 then delete;
if ethnic=1 then ethnic=1;
if ethnic=2 then ethnic=2;
if ethnic=3 then ethnic=2;
if ethnic=4 then ethnic=2;
if ethnic=5 then ethnic=2;
if ethnic=6 then ethnic=2;
if ethnic=7 then ethnic=2;
run;
data rnp2;
set rnp1;
if cond2=2 then cond2=2;
run;
data rnp2;
set rnp2(rename=(cond2=cond fbrmnx1=fiber_base fbrmnx2=fiber_1m
fatmnx1=fat_base fatmnx2=fat_1m sumfatx1=fatknow_base
scfat9x1=selffat_base
scfat9x2=selffat_1m scfat8x1=intfat_base scfat8x2=intfat_1m
scfbrx1=intfib_base scfbrx2=intfib_1m fv1x1=fv_base fv1x2=fv_1m
fv7x1=selffv_base
fv7x2=selffv_1m fv6x1=intfv_base fv6x2=intfv_1m epplanx1=epplan
epprepx1=epprep epshopx1=epshop fss1x1=fss1 fss2x1=fss2
fss3x1=fss3 fss4x1=fss4 fxx5x1=fss5 ethnic=ethnic educ=educ
epbrkx1=brkfast epdinx1=dinner eplunx1=lunch id_x1=id));
format row_num ;
id_num=_N_;
__idco=id_num+1000;
if dinner=. then dinner=0;
if dinner=1 then dinner=0;
if dinner=2 then dinner=1;
if dinner=3 then dinner=0;
if lunch=. then lunch=0;
if lunch=1 then lunch=0;
if lunch=2 then lunch=1;
```

```
if lunch=3 then lunch=0;
if brkfast=. then brkfast=0;
if brkfast=1 then brkfast=0;
if brkfast=2 then brkfast=1;
if brkfast=3 then brkfast=0;
meals = epshop + epplan + epprep;
eat_out = brkfast + dinner + lunch;
sum_fss = fss1 + fss2 + fss3 + fss4+ fss5;
if 1=< educ=<3 then educ=1;
if educ=4 then educ=2;
if educ=5 then educ=3;
if educ=6 then educ=3;
if educ>6 then educ=4;
if marital=1 then marital=1;
if marital=2 then marital=2;
if marital=3 then marital=2;
if marital=4 then marital=2;
if marital=5 then marital=2;
drop lunch dinner brkfast fss1 fss2 fss3 fss4 fss5 epplan epshop epprep;
run;

data fiber;
set library.fiberr1m_2011;
keep id flag2 fbrmean ffbfbmx2 fatmean ffbftmx2 fatkntot soc9 soc9x2 soc8
soc8x2 soc10 soc10x2 fv1 fv1x2 fv9 fv9x2 fv8 fv8x2 ethnic educ gender
age income marital town tvhrs docvisit epplan epprep epshop fss1 fss2
fss3 fss4 fss5 epbrfst epdinr eplunch;
fatkntot=fatkntot+1;
if ethnic=1 then ethnic=1;
if ethnic=2 then ethnic=2;
if ethnic=3 then ethnic=2;
if ethnic=4 then ethnic=2;
run;
data fiber;
set fiber(rename=(flag2=cond fbrmean=fiber_base ffbfbmx2=fiber_1m
fatmean=fat_base ffbftmx2=fat_1m fatkntot=fatknow_base soc9=selffat_base
soc9x2=selffat_1m soc8=intfat_base soc8x2=intfat_1m soc10=intfib_base
soc10x2=intfib_1m fv1=fv_base fv1x2=fv_1m fv9=selffv_base
fv9x2=selffv_1m  fv8=intfv_base fv8x2=intfv_1m docvisit=drvisit
ethnic=ethnic educ=educ epbrfst=brkfast epdinr=dinner eplunch=lunch
id=id));
format row_num ;
id_num=_N_;
__idca=id_num+1000;
if dinner=. then dinner=0;
if dinner=1 then dinner=0;
if dinner=2 then dinner=1;
if dinner=3 then dinner=0;
if lunch=. then lunch=0;
if lunch=1 then lunch=0;
if lunch=2 then lunch=1;
if lunch=3 then lunch=0;
if brkfast=. then brkfast=0;
```

```
if brkfast=1 then brkfast=0;
if brkfast=2 then brkfast=1;
if brkfast=3 then brkfast=0;
meals = epshop + epplan + epprep;
eat_out = brkfast + dinner + lunch;
sum_fss = fss1 + fss2 + fss3 + fss4+ fss5;
if 1=< educ=<3 then educ=1;
if educ=4 then educ=2;
if educ=5 then educ=3;
if educ=6 then educ=3;
if educ>6 then educ=4;
if marital=1 then marital=1;
if marital=2 then marital=2;
if marital=3 then marital=2;
if marital=4 then marital=2;
if marital=5 then marital=2;
drop lunch dinner brkfast fss1 fss2 fss3 fss4 fss5 epplan epshop epprep;
run;

data fiber_rnp;                /*Combined Dataset*/
set fiber rnp2;
run;


/* Running t-tests and chi-squared tests to determine which varaibles
should be in the model */

proc freq data=fiber_rnp;
tables cond*ethnic/chisq ;
tables cond*educ/chisq ;
tables cond*gender/chisq ;
tables cond*marital/chisq ;
tables cond*town/chisq ;
tables cond*eat_out/chisq ;
run;

proc ttest data=fiber_rnp;
class cond;
var age tvhrs drvisit fatknow_base meals sum_fss;
run;

/*Checking for Multicollinearity*/

proc corr data=fiber_rnp;
var drvisit fatknow_base  ;
run;

proc corr data=fiber_rnp spearman;
var ethnic educ town eat_out;
run;


data first_analysis;    /*The outcome variables change from baseline */
```

```
set fiber_rnp;
fat_behave=fat_1m - fat_base;
fiber_behave=fiber_1m - fiber_base;
fv_behave=fv_1m - fv_base;
intfv=intfv_1m - intfv_base;
selffat=selffat_1m - selffat_base;
intfat=intfat_1m - intfat_base;
intfiber=intfib_1m - intfib_base;
selffv=selffv_1m - selffv_base;
output first_analysis;
run;

/*Crude Model: not controlling for confoudning*/
proc glm data=first_analysis;
      class cond;
      model fat_behave= cond ;
      lsmeans cond/cl pdiff stderr ;
      title 'fat behavior crude';
      run;
proc reg data=first_analysis ;
      model fat_behave = cond /selection=adjrsq aic bic;
      title 'fat behavior crude';
      run;

proc glm data=first_analysis;
      class cond;
      model fiber_behave= cond /cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fiber behavior crude';
      run;
proc reg data=first_analysis ;
      model fiber_behave = cond  /selection=adjrsq aic bic;
      title 'fiber behavior crude';
      run;

proc glm data=first_analysis;
      class cond;
      model fv_behave= cond /cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fv behavior crude';
      run;
proc reg data=first_analysis ;
      model fv_behave = cond  /selection=adjrsq aic bic;
      title 'fv behavior crude';
      run;

proc glm data=first_analysis;
      class cond;
      model intfat= cond /cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fat intentions crude';
      run;
proc reg data=first_analysis ;
```

```
      model intfat = cond  /selection=adjrsq aic bic;
      title 'fat intentions crude';
      run;

proc glm data=first_analysis;
      class cond;
      model intfiber= cond /cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fiber intentions crude';
      run;
proc reg data=first_analysis ;
      model intfiber = cond  /selection=adjrsq aic bic;
      title 'fiber intentions crude';
      run;

proc glm data=first_analysis;
      class cond;
      model intfv= cond /cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fv intentions crude';
      run;
proc reg data=first_analysis ;
      model intfv = cond  /selection=adjrsq aic bic;
      title 'fv intentions crude';
      run;

proc glm data=first_analysis;
      class cond;
      model selffat= cond /cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fat self efficacy crude';
      run;
proc reg data=first_analysis ;
      model selffat = cond /selection=adjrsq aic bic;
      title 'fat self efficacy crude';
      run;

proc glm data=first_analysis;
      class cond;
      model selffv= cond /cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fv self efficacy crude';
      run;
proc reg data=first_analysis ;
      model selffv = cond  /selection=adjrsq aic bic;
      title 'fv self efficacy crude';
      run;
```

APPENDIX B

SAS Code for Multiple Linear Regression Model

```
libname library 'C:\Users\Owner\Documents\SSRP_Thesis\Datasets';
run;

/*Combining the Datasets for FIBERR and RNP into one dataset: */

data rnp1;
set library.rnp;
keep id_x1 cond2 fbrmnx1 fbrmnx2 fatmnx1 fatmnx2 sumfatx1 scfat9x1
scfat9x2 scfat8x1 scfat8x2 scfbrx1 scfbrx2 fv1x1 fv1x2 fv7x1 fv7x2 fv6x1
fv6x2 ethnic
educ gender age income marital town tvhrs drvisit epplanx1 epprepx1
epshopx1 fss1x1 fss2x1 fss3x1 fss4x1 fxx5x1 epbrkx1 epdinx1
eplunx1;
if fatmnx2=. then delete;
if cond2=1 then delete;
if ethnic=1 then ethnic=1;
if ethnic=2 then ethnic=2;
if ethnic=3 then ethnic=2;
if ethnic=4 then ethnic=2;
if ethnic=5 then ethnic=2;
if ethnic=6 then ethnic=2;
if ethnic=7 then ethnic=2;
run;
data rnp2;
set rnp1;
if cond2=2 then cond2=2;
run;
data rnp2;
set rnp2(rename=(cond2=cond fbrmnx1=fiber_base fbrmnx2=fiber_1m
fatmnx1=fat_base fatmnx2=fat_1m sumfatx1=fatknow_base
scfat9x1=selffat_base
scfat9x2=selffat_1m scfat8x1=intfat_base scfat8x2=intfat_1m
scfbrx1=intfib_base scfbrx2=intfib_1m fv1x1=fv_base fv1x2=fv_1m
fv7x1=selffv_base
fv7x2=selffv_1m fv6x1=intfv_base fv6x2=intfv_1m epplanx1=epplan
epprepx1=epprep epshopx1=epshop fss1x1=fss1 fss2x1=fss2
fss3x1=fss3 fss4x1=fss4 fxx5x1=fss5 ethnic=ethnic educ=educ
epbrkx1=brkfast epdinx1=dinner eplunx1=lunch id_x1=id));
format row_num ;
id_num=_N_;
__idco=id_num+1000;
if dinner=. then dinner=0;
if dinner=1 then dinner=0;
if dinner=2 then dinner=1;
if dinner=3 then dinner=0;
if lunch=. then lunch=0;
if lunch=1 then lunch=0;
if lunch=2 then lunch=1;
```

```
if lunch=3 then lunch=0;
if brkfast=. then brkfast=0;
if brkfast=1 then brkfast=0;
if brkfast=2 then brkfast=1;
if brkfast=3 then brkfast=0;
meals = epshop + epplan + epprep;
eat_out = brkfast + dinner + lunch;
sum_fss = fss1 + fss2 + fss3 + fss4+ fss5;
if 1=< educ=<3 then educ=1;
if educ=4 then educ=2;
if educ=5 then educ=3;
if educ=6 then educ=3;
if educ>6 then educ=4;
if marital=1 then marital=1;
if marital=2 then marital=2;
if marital=3 then marital=2;
if marital=4 then marital=2;
if marital=5 then marital=2;
drop lunch dinner brkfast fss1 fss2 fss3 fss4 fss5 epplan epshop epprep;
run;

data fiber;
set library.fiberr1m_2011;
keep id flag2 fbrmean ffbfbmx2 fatmean ffbftmx2 fatkntot soc9 soc9x2 soc8
soc8x2 soc10 soc10x2 fv1 fv1x2 fv9 fv9x2 fv8 fv8x2 ethnic educ gender
age income marital town tvhrs docvisit epplan epprep epshop fss1 fss2
fss3 fss4 fss5 epbrfst epdinr eplunch;
fatkntot=fatkntot+1;
if ethnic=1 then ethnic=1;
if ethnic=2 then ethnic=2;
if ethnic=3 then ethnic=2;
if ethnic=4 then ethnic=2;
run;
data fiber;
set fiber(rename=(flag2=cond fbrmean=fiber_base ffbfbmx2=fiber_1m
fatmean=fat_base ffbftmx2=fat_1m fatkntot=fatknow_base soc9=selffat_base
soc9x2=selffat_1m soc8=intfat_base soc8x2=intfat_1m soc10=intfib_base
soc10x2=intfib_1m fv1=fv_base fv1x2=fv_1m fv9=selffv_base
fv9x2=selffv_1m  fv8=intfv_base fv8x2=intfv_1m docvisit=drvisit
ethnic=ethnic educ=educ epbrfst=brkfast epdinr=dinner eplunch=lunch
id=id));
format row_num ;
id_num=_N_;
__idca=id_num+1000;
if dinner=. then dinner=0;
if dinner=1 then dinner=0;
if dinner=2 then dinner=1;
if dinner=3 then dinner=0;
if lunch=. then lunch=0;
if lunch=1 then lunch=0;
if lunch=2 then lunch=1;
if lunch=3 then lunch=0;
if brkfast=. then brkfast=0;
```

```
if brkfast=1 then brkfast=0;
if brkfast=2 then brkfast=1;
if brkfast=3 then brkfast=0;
meals = epshop + epplan + epprep;
eat_out = brkfast + dinner + lunch;
sum_fss = fss1 + fss2 + fss3 + fss4+ fss5;
if 1=< educ=<3 then educ=1;
if educ=4 then educ=2;
if educ=5 then educ=3;
if educ=6 then educ=3;
if educ>6 then educ=4;
if marital=1 then marital=1;
if marital=2 then marital=2;
if marital=3 then marital=2;
if marital=4 then marital=2;
if marital=5 then marital=2;
drop lunch dinner brkfast fss1 fss2 fss3 fss4 fss5 epplan epshop epprep;
run;

data fiber_rnp;                /*Combined Dataset*/
set fiber rnp2;
run;


/* Running t-tests and chi-squared tests to determine which varaibles
should be in the model */

proc freq data=fiber_rnp;
tables cond*ethnic/chisq ;
tables cond*educ/chisq ;
tables cond*gender/chisq ;
tables cond*marital/chisq ;
tables cond*town/chisq ;
tables cond*eat_out/chisq ;
run;

proc ttest data=fiber_rnp;
class cond;
var age tvhrs drvisit fatknow_base meals sum_fss;
run;

/*Checking for Multicollinearity*/

proc corr data=fiber_rnp;
var drvisit fatknow_base  ;
run;

proc corr data=fiber_rnp spearman;
var ethnic educ town eat_out;
run;


data first_analysis; /*The outcome variables change from baseline */
```

```
set fiber_rnp;
fat_behave=fat_1m - fat_base;
fiber_behave=fiber_1m - fiber_base;
fv_behave=fv_1m - fv_base;
intfv=intfv_1m - intfv_base;
selffat=selffat_1m - selffat_base;
intfat=intfat_1m - intfat_base;
intfiber=intfib_1m - intfib_base;
selffv=selffv_1m - selffv_base;
output first_analysis;
run;


/*Controlling for covarites in a Multiple Linear Regression Model*/

proc glm data=first_analysis;
      class cond;
      model fat_behave= cond ethnic educ town eat_out drvisit
fatknow_base;
      lsmeans cond/cl pdiff stderr ;
      title 'fat behavior analysis 1';
      run;
proc reg data=first_analysis ;
      model fat_behave = cond ethnic educ town eat_out drvisit
fatknow_base /selection=adjrsq aic bic;
      title 'fat behavior analysis 1';
      output out=first_analysis residual=resid1;
      run;

 proc univariate data=first_analysis;     /*checking to see if the
residuals are normally distributed; they look to be*/
  class cond;
  var resid1;
  histogram resid1;
  title 'Distribution of Residuals in Fat Behavior from the Multiple
Linear Regression Model';
run;

proc glm data=first_analysis;
      class cond;
      model fiber_behave= cond ethnic educ town eat_out drvisit
fatknow_base/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fiber behavior analysis 1';
      run;
proc reg data=first_analysis ;
      model fiber_behave = cond ethnic educ town eat_out drvisit
fatknow_base /selection=adjrsq aic bic;
      title 'fiber behavior analysis 1';
      output out=first_analysis residual=resid2;
      run;
```

```
 proc univariate data=first_analysis;     /*checking to see if the
residuals are normally distributed; they look to be*/
  class cond;
  var resid2;
  histogram resid2;
  title 'Distribution of Residuals in Fiber Behavior from the Multiple
Linear Regression Model';
run;

proc glm data=first_analysis;
      class cond;
      model fv_behave= cond ethnic educ town eat_out drvisit
fatknow_base/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fv behavior analysis 1';
      run;
proc reg data=first_analysis ;
      model fv_behave = cond ethnic educ town eat_out drvisit
fatknow_base /selection=adjrsq aic bic;
      title 'fv behavior analysis 1';
      output out=first_analysis residual=resid3;
      run;

 proc univariate data=first_analysis;     /*checking to see if the
residuals are normally distributed; they look to be*/
  class cond;
  var resid3;
  histogram resid3;
  title 'Distribution of Residuals in FV Behavior from the Multiple
Linear Regression Model';
run;

proc glm data=first_analysis;
      class cond;
      model intfat= cond ethnic educ town eat_out drvisit
fatknow_base/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fat intentions analysis 1';
      run;
proc reg data=first_analysis ;
      model intfat = cond ethnic educ town eat_out drvisit fatknow_base
/selection=adjrsq aic bic;
      title 'fat intentions analysis 1';
      output out=first_analysis residual=resid4;
      run;

 proc univariate data=first_analysis;     /*checking to see if the
residuals are normally distributed; they look to be*/
  class cond;
  var resid4;
  histogram resid4;
  title 'Distribution of Residuals in Fat Intentions from the Multiple
Linear Regression Model';
```

```
run;

proc glm data=first_analysis;
      class cond;
      model intfiber= cond ethnic educ town eat_out drvisit
fatknow_base/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fiber intentions analysis 1';
      run;
proc reg data=first_analysis ;
      model intfiber = cond ethnic educ town eat_out drvisit fatknow_base
/selection=adjrsq aic bic;
      title 'fiber intentions analysis 1';
      output out=first_analysis residual=resid5;
      run;

 proc univariate data=first_analysis;     /*checking to see if the
residuals are normally distributed; they look to be*/
  class cond;
  var resid5;
  histogram resid5;
  title 'Distribution of Residuals in Fiber Intentions from the Multiple
Linear Regression Model';
run;

proc glm data=first_analysis;
      class cond;
      model intfv= cond ethnic educ town eat_out drvisit
fatknow_base/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fv intentions analysis 1';
      run;
proc reg data=first_analysis ;
      model intfv = cond ethnic educ town eat_out drvisit fatknow_base
/selection=adjrsq aic bic;
      title 'fv intentions analysis 1';
      output out=first_analysis residual=resid6;
      run;

 proc univariate data=first_analysis;     /*checking to see if the
residuals are normally distributed; they look to be*/
  class cond;
  var resid6;
  histogram resid6;
  title 'Distribution of Residuals in FV Intentions from the Multiple
Linear Regression Model';
run;

proc glm data=first_analysis;
      class cond;
      model selffat= cond ethnic educ town eat_out drvisit
fatknow_base/cli;
      lsmeans cond/cl pdiff stderr ;
```

```
      title 'fat self efficacy analysis 1';
      run;
proc reg data=first_analysis ;
      model selffat = cond ethnic educ town eat_out drvisit fatknow_base
/selection=adjrsq aic bic;
      title 'fat self efficacy analysis 1';
      output out=first_analysis residual=resid7;
      run;

 proc univariate data=first_analysis;     /*checking to see if the
residuals are normally distributed; they look to be*/
  class cond;
  var resid7;
  histogram resid7;
  title 'Distribution of Residuals in FatSelf-Efficacy from the Multiple
Linear Regression Model';
run;

proc glm data=first_analysis;
      class cond;
      model selffv= cond ethnic educ town eat_out drvisit
fatknow_base/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fv self efficacy analysis 1';
      run;
proc reg data=first_analysis ;
      model selffv = cond ethnic educ town eat_out drvisit fatknow_base
/selection=adjrsq aic bic;
      title 'fv self efficacy analysis 1';
      output out=first_analysis residual=resid8;
      run;

 proc univariate data=first_analysis;     /*checking to see if the
residuals are normally distributed; they look to be*/
  class cond;
  var resid8;
  histogram resid8;
  title 'Distribution of Residuals in FV Self-Efficacy from the Multiple
Linear Regression Model';
run;


/*checking that there is no longer significant differences between
baseline variables*/
proc glm data=first_analysis;
      class cond;
      model ethnic= cond educ town eat_out drvisit fatknow_base/cli;
      run;
proc glm data=first_analysis;
      class cond;
      model educ= cond ethnic town eat_out drvisit fatknow_base/cli;
      run;
proc glm data=first_analysis;
```

```
      class cond;
      model town= cond ethnic educ eat_out drvisit fatknow_base/cli;
      run;
proc glm data=first_analysis;
      class cond;
      model eat_out= cond ethnic educ town drvisit fatknow_base/cli;
      run;
proc glm data=first_analysis;
      class cond;
      model drvisit= cond ethnic educ town eat_out fatknow_base/cli;
      run;
proc glm data=first_analysis;
      class cond;
      model fatknow_base= cond ethnic educ town eat_out drvisit /cli;
      run;
```

APPENDIX C

SAS Code for Propensity Score Matching

```
libname library 'C:\Users\Owner\Documents\SSRP_Thesis\Datasets';
run;

data rnp1;
set library.rnp;
keep id_x1 cond2 fbrmnx1 fbrmnx2 fatmnx1 fatmnx2 sumfatx1 scfat9x1
scfat9x2 scfat8x1 scfat8x2 scfbrx1 scfbrx2 fv1x1 fv1x2 fv7x1 fv7x2 fv6x1
fv6x2 ethnic
educ gender age income marital town tvhrs drvisit epplanx1 epprepx1
epshopx1 fss1x1 fss2x1 fss3x1 fss4x1 fxx5x1 epbrkx1 epdinx1
eplunx1;
if fatmnx2=. then delete;
if cond2=1 then delete;
if ethnic=1 then ethnic=1;
if ethnic=2 then ethnic=2;
if ethnic=3 then ethnic=2;
if ethnic=4 then ethnic=2;
if ethnic=5 then ethnic=2;
if ethnic=6 then ethnic=2;
if ethnic=7 then ethnic=2;
run;
data rnp2;
set rnp1;
if cond2=2 then cond2=0;
run;
data rnp2;
set rnp2(rename=(cond2=cond fbrmnx1=fiber_base fbrmnx2=fiber_1m
fatmnx1=fat_base fatmnx2=fat_1m sumfatx1=fatknow_base
scfat9x1=selffat_base
scfat9x2=selffat_1m scfat8x1=intfat_base scfat8x2=intfat_1m
scfbrx1=intfib_base scfbrx2=intfib_1m fv1x1=fv_base fv1x2=fv_1m
fv7x1=selffv_base
fv7x2=selffv_1m fv6x1=intfv_base fv6x2=intfv_1m epplanx1=epplan
epprepx1=epprep epshopx1=epshop fss1x1=fss1 fss2x1=fss2
fss3x1=fss3 fss4x1=fss4 fxx5x1=fss5 ethnic=ethnic educ=educ
epbrkx1=brkfast epdinx1=dinner eplunx1=lunch id_x1=id));
format row_num ;
id_num=_N_;
__idco=id_num+1000;
if dinner=. then dinner=0;
if dinner=1 then dinner=0;
if dinner=2 then dinner=1;
if dinner=3 then dinner=0;
if lunch=. then lunch=0;
if lunch=1 then lunch=0;
if lunch=2 then lunch=1;
if lunch=3 then lunch=0;
if brkfast=. then brkfast=0;
```

```
if brkfast=1 then brkfast=0;
if brkfast=2 then brkfast=1;
if brkfast=3 then brkfast=0;
meals = epshop + epplan + epprep;
eat_out = brkfast + dinner + lunch;
sum_fss = fss1 + fss2 + fss3 + fss4+ fss5;
if 1=< educ=<3 then educ=1;
if educ=4 then educ=2;
if educ=5 then educ=3;
if educ=6 then educ=3;
if educ>6 then educ=4;
if marital=1 then marital=1;
if marital=2 then marital=2;
if marital=3 then marital=2;
if marital=4 then marital=2;
if marital=5 then marital=2;
drop lunch dinner brkfast fss1 fss2 fss3 fss4 fss5 epplan epshop epprep;
run;

data fiber;
set library.fiberr1m_2011;
keep id flag2 fbrmean ffbfbmx2 fatmean ffbftmx2 fatkntot soc9 soc9x2 soc8
soc8x2 soc10 soc10x2 fv1 fv1x2 fv9 fv9x2 fv8 fv8x2 ethnic educ gender
age income marital town tvhrs docvisit epplan epprep epshop fss1 fss2
fss3 fss4 fss5 epbrfst epdinr eplunch;
fatkntot=fatkntot+1;
if ethnic=1 then ethnic=1;
if ethnic=2 then ethnic=2;
if ethnic=3 then ethnic=2;
if ethnic=4 then ethnic=2;
run;
data fiber;
set fiber(rename=(flag2=cond fbrmean=fiber_base ffbfbmx2=fiber_1m
fatmean=fat_base ffbftmx2=fat_1m fatkntot=fatknow_base soc9=selffat_base
soc9x2=selffat_1m soc8=intfat_base soc8x2=intfat_1m soc10=intfib_base
soc10x2=intfib_1m fv1=fv_base fv1x2=fv_1m fv9=selffv_base
fv9x2=selffv_1m  fv8=intfv_base fv8x2=intfv_1m docvisit=drvisit
ethnic=ethnic educ=educ epbrfst=brkfast epdinr=dinner eplunch=lunch
id=id));
format row_num ;
id_num=_N_;
__idca=id_num+1000;
if dinner=. then dinner=0;
if dinner=1 then dinner=0;
if dinner=2 then dinner=1;
if dinner=3 then dinner=0;
if lunch=. then lunch=0;
if lunch=1 then lunch=0;
if lunch=2 then lunch=1;
if lunch=3 then lunch=0;
if brkfast=. then brkfast=0;
if brkfast=1 then brkfast=0;
if brkfast=2 then brkfast=1;
```

```
if brkfast=3 then brkfast=0;
meals = epshop + epplan + epprep;
eat_out = brkfast + dinner + lunch;
sum_fss = fss1 + fss2 + fss3 + fss4+ fss5;
if 1=< educ=<3 then educ=1;
if educ=4 then educ=2;
if educ=5 then educ=3;
if educ=6 then educ=3;
if educ>6 then educ=4;
if marital=1 then marital=1;
if marital=2 then marital=2;
if marital=3 then marital=2;
if marital=4 then marital=2;
if marital=5 then marital=2;
drop lunch dinner brkfast fss1 fss2 fss3 fss4 fss5 epplan epshop epprep;
run;

data fiber_rnp;
set fiber rnp2;
run;


/* t-tests and chi-squared tests to find significance of possible
variables for the model */

proc freq data=fiber_rnp;
tables cond*ethnic/chisq ;
tables cond*educ/chisq ;
tables cond*gender/chisq ;
tables cond*marital/chisq ;
tables cond*town/chisq ;
tables cond*eat_out/chisq ;
run;

proc ttest data=fiber_rnp;
class cond;
var age tvhrs drvisit fatknow_base meals sum_fss;
run;

proc means data=fiber_rnp;
var age tvhrs drvisit fatknow_base meals sum_fss;
run;

/*Creating a propensity score*/

proc logistic data=fiber_rnp ;
  class cond educ/param=ref ref=first;
  model cond =  ethnic educ town eat_out drvisit fatknow_base /* age
gender*/;
  output out=propen pred=propensity xbeta=predlogit;
      run;

proc means data=fiber_rnp n;
```

```
var ethnic educ town eat_out drvisit fatknow_base age gender;
run;

DATA PROPEN;
      SET PROPEN;
      LOGIT = LOG(PROPENSITY/(1-PROPENSITY));
      PROC PRINT;
      RUN;

      proc means data=fiber_rnp;
      class cond;
      var age gender ethnic educ town eat_out drvisit fatknow_base;
      run;

proc means data = propen;
var logit  propensity predlogit;
run;


proc freq data = propen;
tables propensity;
run;

/*Distribution of Propensity Score*/
proc univariate data=propen;
  class cond;
  var propensity;
  histogram propensity;
  title 'Distribution Propensity Score';
run;

/*Distribution of Logit of the Propensity Score*/
Proc univariate data=propen;
  class cond;
  var predlogit;
  histogram predlogit;
  title 'Distribution of Logit of Propensity Score';
  run;


data propen1;
SET PROPEN;
idnum = id_num + 1000;
drop id_num;
run;

/*Finding the mean of the propensity score*/
proc sort data=propen; by cond;
proc means data=propen;
by cond;
var propensity;
run;
```

```
/*creating propensity score matches; note that I tried creating the
matches with and without the two nonsignificant
  variables that I threw into the model becaues of clinical significance
(age and gender) to see which would gives more
  matches  0.18067614*/

 %include 'C:\Users\Owner\Documents\SSRP_Thesis\Datasets\gmatch.sas';
%gmatch(data=propen1, group=cond, id=idnum, mvars=predlogit, wts=1,
dmaxk=0.1965  , dist=1,
       ncontls=1,seedca=234098,seedco=0489);
         run;


/*Merging Datasets to create a Matched Dataset*/
data out;
set __out;
format row_num ;
match_num=_N_;
*drop __dij __cont_n __cotime __catime __dtime __wt1 __ca1 __co1 __absd1
__d1;
run;

Data Out_RNP;
Set Out;
Drop  __idca;
run;
Proc Sort data=Out_RNP; by __idco;
run;
Data Out_RNP1;
Merge rnp2 out_rnp;
by __idco;
if __Co1=. then delete;
idd=__idco;
run;

Data Out_Fib;
Set Out;
Drop __idco;
run;
Proc Sort data=Out_fib; by __idca;
Data Out_Fib1;
Merge fiber Out_Fib;
by __idca;
if __ca1=. then delete;
idd=__idca;
run;

Proc Sort data=Out_Rnp1; by match_num;
run;
data matched_set;
set Out_Fib1 Out_RNP1;
by match_num;
m_num=match_num;
```

```
idnum=idd;
run;

/*Finding the mean of the propensity score for the matched set*/
proc sort data=matched_set;by cond idnum ;
proc sort data=propen1;by cond idnum ;
data propen2;
merge matched_set propen1;
by cond idnum ;
if m_num=. then delete;
run;
proc means data=propen2;
by cond;
var propensity;
run;


/*Creating a Unmatched Dataset to Compare the Matched and Unmatched
Datasets*/
Data RNP_NOT;
set Out;
Drop  __idca;
run;
Proc Sort data=RNP_NOT; by __idco;
run;
Data RNP_NOT1;
Merge rnp2 out_rnp;
by __idco;
if __Co1=. then output;
idd=__idco;
run;

Data FIB_NOT;
Set Out;
Drop __idco;
run;
Proc Sort data=FIB_NOT; by __idca;
Data FIB_NOT1;
Merge fiber Out_Fib;
by __idca;
if __ca1=. then output;
idd=__idca;
run;

Proc Sort data=RNP_NOT1; by match_num;
run;
data not_matched;
set RNP_NOT1 FIB_NOT1;
by match_num;
m_num=match_num;
*drop match_num;
run;
```

```
proc freq data=not_matched;
tables cond*ethnic/chisq ;
tables cond*educ/chisq ;
tables cond*gender/chisq ;
tables cond*town/chisq;
tables cond*eat_out/chisq ;
tables cond*marital/chisq;
run;

proc ttest data=not_matched;
class cond;
var age drvisit fatknow_base tvhrs meals sum_fss;
run;

data not_matched1;
set not_matched;
if cond=1 then compare=1;
if cond=0 then compare=1;
run;
data matched_set1;
set matched_set;
if cond=1 then compare=2;
if cond=0 then compare=2;
run;
data compare;
set not_matched1 matched_set1;
run;
proc freq data=compare;
tables compare*ethnic/chisq ;
tables compare*educ/chisq ;
tables compare*gender/chisq ;
tables compare*town/chisq;
tables compare*eat_out/chisq ;
tables compare*marital/chisq;
run;

proc ttest data=compare;
class compare;
var age drvisit fatknow_base tvhrs meals sum_fss fat_base fiber_base
intfv_base
selffat_base intfat_base intfib_base fv_base selffv_base fat_1m fiber_1m
intfv_1m
selffat_1m intfat_1m intfib_1m fv_1m selffv_1m;
run;


/*end of comparison*/



/*Testing variables to check that they are no longer signficantly
different between FIBERR and RNP*/
```

```
proc freq data=matched_set;
tables m_num*cond*ethnic/cmh ;
tables m_num*cond*educ/cmh ;
tables m_num*cond*gender/cmh ;
tables m_num*cond*town/cmh ;
tables m_num*cond*eat_out/cmh ;
tables m_num*cond*marital/cmh;
run;

proc means data=matched_set;
var age drvisit fatknow_base tvhrs meals sum_fss;
run;

proc mixed data=matched_set;
class m_num cond;
model age=cond ;
repeated/subject = m_num type=un;
estimate 'age' cond 1 -1 ;
contrast 'age' cond 1 -1 ;
run;
proc mixed data=matched_set;
class m_num cond;
model drvisit=cond ;
repeated/subject = m_num type=un;
estimate 'dr visit' cond 1 -1 ;
contrast 'dr visit' cond 1 -1 ;
run;
proc mixed data=matched_set;
class m_num cond;
model fatknow_base=cond ;
repeated/subject = m_num type=un;
estimate 'fat knowlege' cond 1 -1 ;
contrast 'fat knowlege' cond 1 -1 ;
run;
proc mixed data=matched_set;
class m_num cond;
model tvhrs=cond ;
repeated/subject = m_num type=un;
estimate 'tv hours' cond 1 -1 ;
contrast 'tv hours' cond 1 -1 ;
run;
proc mixed data=matched_set;
class m_num cond;
model meals=cond ;
repeated/subject = m_num type=un;
estimate 'meals' cond 1 -1 ;
contrast 'meals' cond 1 -1 ;
run;
proc mixed data=matched_set;
class m_num cond;
model sum_fss=cond ;
repeated/subject = m_num type=un;
```

```
estimate 'fss score' cond 1 -1 ;
contrast 'fss score' cond 1 -1 ;
run;


/*Preparing Outcome Variables for Analysis*/

Data Out_Fib2;
Set Out_fib1;
drop fat_base fiber_base intfv_base selffat_base intfat_base intfib_base
fv_base selffv_base  town
fatknow_base ethnic educ drvisit age gender marital income tvhrs;
gptm=cond;
if gptm=1 then gptm=3;
m_num=match_num;
fat_behave=fat_1m;
fiber_behave=fiber_1m;
fv_behave=fv_1m;
intfv=intfv_1m;
selffat=selffat_1m;
intfat=intfat_1m;
intfiber=intfib_1m;
selffv=selffv_1m;
drop fat_1m fiber_1m fv_1m intfv_1m selffat_1m intfat_1m intfib_1m
selffv_1m;
run;

Data Out_Fib3;
Set Out_fib1;
drop fat_1m fiber_1m intfv_1m selffat_1m intfat_1m intfib_1m fv_1m
selffv_1m town
fatknow_base  ethnic educ  drvisit age gender marital income tvhrs;
gptm=cond;
if gptm=1 then gptm=1;
m_num=match_num;
fat_behave=fat_base;
fiber_behave=fiber_base;
fv_behave=fv_base;
intfv=intfv_base;
selffat=selffat_base;
intfat=intfat_base;
intfiber=intfib_base;
selffv=selffv_base;
drop fat_base fiber_base fv_base intfv_base selffat_base intfat_base
intfib_base selffv_base;
run;

Data Out_rnp2;
Set Out_rnp1;
drop fat_base fiber_base intfv_base selffat_base intfat_base intfib_base
fv_base selffv_base town
fatknow_base  ethnic educ  drvisit age gender marital income tvhrs;
gptm=cond;
```

```
if gptm=0 then gptm=4;
m_num=match_num;
fat_behave=fat_1m;
fiber_behave=fiber_1m;
fv_behave=fv_1m;
intfv=intfv_1m;
selffat=selffat_1m;
intfat=intfat_1m;
intfiber=intfib_1m;
selffv=selffv_1m;
drop fat_1m fiber_1m fv_1m intfv_1m selffat_1m intfat_1m intfib_1m
selffv_1m;
run;

Data Out_rnp3;
Set Out_rnp1;
drop fat_1m fiber_1m intfv_1m selffat_1m intfat_1m intfib_1m fv_1m
selffv_1m town
fatknow_base  ethnic educ drvisit age gender marital income tvhrs;
gptm=cond;
if gptm=0 then gptm=2;
m_num=match_num;
fat_behave=fat_base;
fiber_behave=fiber_base;
fv_behave=fv_base;
intfv=intfv_base;
selffat=selffat_base;
intfat=intfat_base;
intfiber=intfib_base;
selffv=selffv_base;
drop fat_base fiber_base fv_base intfv_base selffat_base intfat_base
intfib_base selffv_base;
run;

data for_glm;
set matched_set;
fat_behave=fat_1m - fat_base;
fiber_behave=fiber_1m - fiber_base;
fv_behave=fv_1m - fv_base;
intfv=intfv_1m - intfv_base;
selffat=selffat_1m - selffat_base;
intfat=intfat_1m - intfat_base;
intfiber=intfib_1m - intfib_base;
selffv=selffv_1m - selffv_base;
run;

data for_analysis;
set Out_fib3 out_rnp3 Out_fib2 Out_RNP2 ;
drop match_num;
run;

/*Analysis of Propensity Score Matching*/
```

```
proc means data=matched_set;
class cond;
var fat_base fiber_base intfv_base selffat_base intfat_base intfib_base
fv_base selffv_base
fat_1m fiber_1m intfv_1m selffat_1m intfat_1m intfib_1m  fv_1m selffv_1m;
run;


proc mixed data=for_analysis;
class m_num gptm ;
model fat_behave=gptm ;
repeated/subject =m_num type=un;
estimate 'group effect_fatmean' gptm -1 1 1 -1/cl;
contrast 'group effect_fatmean' gptm -1 1 1 -1;
lsmeans gptm/pdiff cl;
run;

proc mixed data=for_analysis;
class m_num gptm;
model fiber_behave=gptm ;
repeated/subject = m_num type=un;
estimate 'group effect_fibermean' gptm -1 1 1 -1/cl;
contrast 'group effect_fibermean' gptm -1 1 1 -1;
lsmeans gptm/pdiff cl;
run;

proc mixed data=for_analysis;
class m_num gptm;
model fv_behave=gptm ;
repeated/subject = m_num type=un;
estimate 'group effect_fv mean' gptm -1 1 1 -1/cl;
contrast 'group effect_fv mean' gptm -1 1 1 -1;
lsmeans gptm/pdiff cl;
run;

proc mixed data=for_analysis;
class m_num gptm;
model intfv=gptm;
repeated/subject = m_num type=un;
estimate 'group effect_fv intention' gptm -1 1 1 -1/cl;
contrast 'group effect_fv intention' gptm -1 1 1 -1;
lsmeans gptm/pdiff cl;
run;

proc mixed data=for_analysis;
class m_num gptm;
model selffat=gptm;
repeated/subject = m_num type=un;
estimate 'group effect_fat self efficacy' gptm -1 1 1 -1/cl;
contrast 'group effect_fat self efficacy' gptm -1 1 1 -1;
lsmeans gptm/pdiff cl;
run;
```

```
proc mixed data=for_analysis;
class m_num gptm;
model intfat=gptm;
repeated/subject = m_num type=un;
estimate 'group effect_fat intention' gptm -1 1 1 -1/cl;
contrast 'group effect_fat intention' gptm -1 1 1 -1;
lsmeans gptm/pdiff cl;
run;

proc mixed data=for_analysis;
class m_num gptm;
model intfiber=gptm;
repeated/subject = m_num type=un;
estimate 'group effect_fiber intention' gptm -1 1 1 -1/cl;
contrast 'group effect_fiber intention' gptm -1 1 1 -1;
lsmeans gptm/pdiff cl;
run;

proc mixed data=for_analysis;
class m_num gptm;
model selffv=gptm;
repeated/subject = m_num  type=un;
estimate 'group effect_fv self efficacy' gptm -1 1 1 -1/cl;
contrast 'group effect_fv self efficacy' gptm -1 1 1 -1;
lsmeans gptm/pdiff cl;
run;
```

APPENDIX D

SAS Code for Propensity Score as a Covariate

```
libname library 'C:\Users\Owner\Documents\SSRP_Thesis\Datasets';
run;

data rnp1;
set library.rnp;
keep id_x1 cond2 fbrmnx1 fbrmnx2 fatmnx1 fatmnx2 sumfatx1 scfat9x1
scfat9x2 scfat8x1 scfat8x2 scfbrx1 scfbrx2 fv1x1 fv1x2 fv7x1 fv7x2 fv6x1
fv6x2 ethnic
educ gender age income marital town tvhrs drvisit epplanx1 epprepx1
epshopx1 fss1x1 fss2x1 fss3x1 fss4x1 fxx5x1 epbrkx1 epdinx1
eplunx1;
if fatmnx2=. then delete;
if cond2=1 then delete;
if ethnic=1 then ethnic=1;
if ethnic=2 then ethnic=2;
if ethnic=3 then ethnic=2;
if ethnic=4 then ethnic=2;
if ethnic=5 then ethnic=2;
if ethnic=6 then ethnic=2;
if ethnic=7 then ethnic=2;
run;
data rnp2;
set rnp1;
if cond2=2 then cond2=2;
run;
data rnp2;
set rnp2(rename=(cond2=cond fbrmnx1=fiber_base fbrmnx2=fiber_1m
fatmnx1=fat_base fatmnx2=fat_1m sumfatx1=fatknow_base
scfat9x1=selffat_base
scfat9x2=selffat_1m scfat8x1=intfat_base scfat8x2=intfat_1m
scfbrx1=intfib_base scfbrx2=intfib_1m fv1x1=fv_base fv1x2=fv_1m
fv7x1=selffv_base
fv7x2=selffv_1m fv6x1=intfv_base fv6x2=intfv_1m epplanx1=epplan
epprepx1=epprep epshopx1=epshop fss1x1=fss1 fss2x1=fss2
fss3x1=fss3 fss4x1=fss4 fxx5x1=fss5 ethnic=ethnic educ=educ
epbrkx1=brkfast epdinx1=dinner eplunx1=lunch id_x1=id));
format row_num ;
id_num=_N_;
__idco=id_num+1000;
if dinner=. then dinner=0;
if dinner=1 then dinner=0;
if dinner=2 then dinner=1;
if dinner=3 then dinner=0;
if lunch=. then lunch=0;
if lunch=1 then lunch=0;
if lunch=2 then lunch=1;
if lunch=3 then lunch=0;
if brkfast=. then brkfast=0;
```

```
if brkfast=1 then brkfast=0;
if brkfast=2 then brkfast=1;
if brkfast=3 then brkfast=0;
meals = epshop + epplan + epprep;
eat_out = brkfast + dinner + lunch;
sum_fss = fss1 + fss2 + fss3 + fss4+ fss5;
if 1=< educ=<3 then educ=1;
if educ=4 then educ=2;
if educ=5 then educ=3;
if educ=6 then educ=3;
if educ>6 then educ=4;
if marital=1 then marital=1;
if marital=2 then marital=2;
if marital=3 then marital=2;
if marital=4 then marital=2;
if marital=5 then marital=2;
drop lunch dinner brkfast fss1 fss2 fss3 fss4 fss5 epplan epshop epprep;
run;

data fiber;
set library.fiberr1m_2011;
keep id flag2 fbrmean ffbfbmx2 fatmean ffbftmx2 fatkntot soc9 soc9x2 soc8
soc8x2 soc10 soc10x2 fv1 fv1x2 fv9 fv9x2 fv8 fv8x2 ethnic educ gender
age income marital town tvhrs docvisit epplan epprep epshop fss1 fss2
fss3 fss4 fss5 epbrfst epdinr eplunch;
fatkntot=fatkntot+1;
if ethnic=1 then ethnic=1;
if ethnic=2 then ethnic=2;
if ethnic=3 then ethnic=2;
if ethnic=4 then ethnic=2;
run;
data fiber;
set fiber(rename=(flag2=cond fbrmean=fiber_base ffbfbmx2=fiber_1m
fatmean=fat_base ffbftmx2=fat_1m fatkntot=fatknow_base soc9=selffat_base
soc9x2=selffat_1m soc8=intfat_base soc8x2=intfat_1m soc10=intfib_base
soc10x2=intfib_1m fv1=fv_base fv1x2=fv_1m fv9=selffv_base
fv9x2=selffv_1m  fv8=intfv_base fv8x2=intfv_1m docvisit=drvisit
ethnic=ethnic educ=educ epbrfst=brkfast epdinr=dinner eplunch=lunch
id=id));
format row_num ;
id_num=_N_;
__idca=id_num+1000;
if dinner=. then dinner=0;
if dinner=1 then dinner=0;
if dinner=2 then dinner=1;
if dinner=3 then dinner=0;
if lunch=. then lunch=0;
if lunch=1 then lunch=0;
if lunch=2 then lunch=1;
if lunch=3 then lunch=0;
if brkfast=. then brkfast=0;
if brkfast=1 then brkfast=0;
if brkfast=2 then brkfast=1;
```

```
if brkfast=3 then brkfast=0;
meals = epshop + epplan + epprep;
eat_out = brkfast + dinner + lunch;
sum_fss = fss1 + fss2 + fss3 + fss4+ fss5;
if 1=< educ=<3 then educ=1;
if educ=4 then educ=2;
if educ=5 then educ=3;
if educ=6 then educ=3;
if educ>6 then educ=4;
if marital=1 then marital=1;
if marital=2 then marital=2;
if marital=3 then marital=2;
if marital=4 then marital=2;
if marital=5 then marital=2;
drop lunch dinner brkfast fss1 fss2 fss3 fss4 fss5 epplan epshop epprep;
run;

data fiber_rnp;
set fiber rnp2;
run;


/* t-test and chi-squared tests to find significance of possible
variables for the model */

proc freq data=fiber_rnp;
tables cond*ethnic/chisq ;
tables cond*educ/chisq ;
tables cond*gender/chisq ;
tables cond*marital/chisq ;
tables cond*town/chisq ;
tables cond*eat_out/chisq ;
run;

proc ttest data=fiber_rnp;
class cond;
var age tvhrs drvisit fatknow_base meals sum_fss;
run;

/*creating a propensity score*/

proc logistic data=fiber_rnp ;
  class cond educ/param=ref ref=first;
  model cond =  ethnic educ town eat_out drvisit fatknow_base /* age
gender*/;
  output out=propen pred=propensity xbeta=predlogit;
      run;


proc means data=fiber_rnp n;
var ethnic educ town eat_out drvisit fatknow_base age gender;
run;
```

```
DATA PROPEN;
      SET PROPEN;
      LOGIT = LOG(PROPENSITY/(1-PROPENSITY));
      PROC PRINT;
      RUN;

      proc means data=fiber_rnp;
      class cond;
      var age gender ethnic educ town eat_out drvisit fatknow_base;
      run;

proc means data = propen;
var logit  propensity predlogit;
run;


proc freq data = propen;
tables propensity;
run;

proc univariate data=propen;
  class cond;
  var propensity;
  histogram propensity;
run;

data propen1;
SET PROPEN;
idnum = id_num + 1000;
drop id_num;
run;

Proc Logistic data= propen1;
class cond educ/param=ref ref=first;
  model cond =  ethnic educ town eat_out drvisit fatknow_base /* age
gender*/;
  run;

data analysis;
set propen1;
fat_behave=fat_1m - fat_base;
fiber_behave=fiber_1m - fiber_base;
fv_behave=fv_1m - fv_base;
intfv=intfv_1m - intfv_base;
selffat=selffat_1m - selffat_base;
intfat=intfat_1m - intfat_base;
intfiber=intfib_1m - intfib_base;
selffv=selffv_1m - selffv_base;
run;

/*Analysis of Method; analyzed both the propensity score and logit of the
propensity score*/
```

```
proc sort data=analysis;
by cond;
run;

proc means data= analysis;
by cond;
var fat_behave fiber_behave fv_behave intfv selffat intfat intfiber
selffv
fat_1m fat_base fiber_1m fiber_base fv_1m fv_base intfv_1m intfv_base
selffat_1m selffat_base intfat_1m
intfat_base intfib_1m intfib_base selffv_1m selffv_base;;
run;

proc glm data=analysis;
      class cond;
      model fat_behave= cond  propensity/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fat behavior analysis 3 prop';
      run;
proc reg data=analysis;
      model fat_behave= cond  propensity/selection=adjrsq aic bic;
      title 'fat behavior analysis 3 prop';
      output out=analysis residual=resid;
      run;
proc univariate data=analysis;      /*checking to see if the residuals
are normally distributed; they look to be*/
      class cond;
      var resid;
      histogram resid;
      run;


proc glm data=analysis;
      class cond;
      model fat_behave= cond  predlogit/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fat behavior analysis 3 logit';
      run;
proc reg data=analysis;
      model fat_behave= cond  predlogit/selection=adjrsq aic bic;
      title 'fat behavior analysis 3 logit';
      run;

proc glm data=analysis;
      class cond;
      model fiber_behave= cond  propensity/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fiber behavior analysis 3 prop';
      run;
proc reg data=analysis;
      model fiber_behave= cond  propensity/selection=adjrsq aic bic;
      title 'fiber behavior analysis 3 prop';
      run;
```

```
proc glm data=analysis;
      class cond;
      model fiber_behave= cond  predlogit/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fiber behavior analysis 3 logit';
      run;
proc reg data=analysis;
      model fiber_behave= cond  predlogit/selection=adjrsq aic bic;
      title 'fiber behavior analysis 3 logit';
      run;

proc glm data=analysis;
      class cond;
      model fv_behave= cond  propensity/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fv behavior analysis 3 prop';
      run;
proc reg data=analysis;
      model fv_behave= cond  propensity/selection=adjrsq aic bic;
      title 'fv behavior analysis 3 prop';
      run;
proc glm data=analysis;
      class cond;
      model fv_behave= cond  predlogit/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fv behavior analysis 3 logit';
      run;
proc reg data=analysis;
      model fv_behave= cond  predlogit/selection=adjrsq aic bic;
      title 'fv behavior analysis 3 logit';
      run;

proc glm data=analysis;
      class cond;
      model intfat= cond  propensity/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fat intentions analysis 3 prop';
      run;
proc reg data=analysis;
      model intfat= cond  propensity/selection=adjrsq aic bic;
      title 'fat intentions analysis 3 prop';
      run;
proc glm data=analysis;
      class cond;
      model intfat= cond  predlogit/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fat intentions analysis 3 logit';
      run;
proc reg data=analysis;
      model intfat= cond  predlogit/selection=adjrsq aic bic;
      title 'fat intentions analysis 3 logit';
      run;
```

```
proc glm data=analysis;
      class cond;
      model intfiber= cond  propensity/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fiber intentions analysis 3 prop';
      run;
proc reg data=analysis;
      model intfiber= cond  propensity/selection=adjrsq aic bic;
      title 'fiber intentions analysis 3 prop';
      run;
proc glm data=analysis;
      class cond;
      model intfiber= cond  predlogit/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fiber intentions analysis 3 logit';
      run;
proc reg data=analysis;
      model intfiber= cond  predlogit/selection=adjrsq aic bic;
      title 'fiber intentions analysis 3 logit';
      run;

proc glm data=analysis;
      class cond;
      model intfv= cond  propensity/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fv intentions analysis 3 prop';
      run;
proc reg data=analysis;
      model intfv= cond  propensity/selection=adjrsq aic bic;
      title 'fv intentions analysis 3 prop';
      run;
proc glm data=analysis;
      class cond;
      model intfv= cond  predlogit/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fv intentions analysis 3 logit';
      run;
proc reg data=analysis;
      model intfv= cond  predlogit/selection=adjrsq aic bic;
      title 'fv intentions analysis 3 logit';
      run;

proc glm data=analysis;
      class cond;
      model selffat= cond  propensity/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fat self effiacy analysis 3 prop';
      run;
proc reg data=analysis;
      model selffat= cond  propensity/selection=adjrsq aic bic;
      title 'fat self effiacy analysis 3 prop';
      run;
proc glm data=analysis;
```

```
      class cond;
      model selffat= cond  predlogit/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fat self effiacy analysis 3 logit';
      run;
proc reg data=analysis;
      model selffat= cond  predlogit/selection=adjrsq aic bic;
      title 'fat self effiacy analysis 3 logit';
      run;

proc glm data=analysis;
      class cond;
      model selffv= cond  propensity/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fv self effiacy analysis 3 prop';
      run;
proc reg data=analysis;
      model selffv= cond  propensity/selection=adjrsq aic bic;
      title 'fv self effiacy analysis 3 prop';
      run;
proc glm data=analysis;
      class cond;
      model selffv= cond  predlogit/cli;
      lsmeans cond/cl pdiff stderr ;
      title 'fv self effiacy analysis 3 logit';
      run;
proc reg data=analysis;
      model selffv= cond  predlogit/selection=adjrsq aic bic;
      title 'fv self effiacy analysis 3 logit';
      run;


/*checking that there is no longer significant differences between
baseline variables*/
proc glm data=analysis;
      class cond;
      model ethnic= cond  propensity;
      run;
proc glm data=analysis;
      class cond;
      model ethnic= cond  predlogit;
      run;
proc glm data=analysis;
      class cond;
      model educ= cond  propensity;
      run;
proc glm data=analysis;
      class cond;
      model educ= cond  predlogit;
      run;
proc glm data=analysis;
      class cond;
```

```
      model town= cond  propensity;
      run;
proc glm data=analysis;
      class cond;
      model town= cond  predlogit;
      run;
proc glm data=analysis;
      class cond;
      model eat_out= cond  propensity;
      run;
proc glm data=analysis;
      class cond;
      model eat_out= cond  predlogit;
      run;
proc glm data=analysis;
      class cond;
      model drvisit= cond  propensity;
      run;
proc glm data=analysis;
      class cond;
      model drvisit= cond  predlogit;
      run;
proc glm data=analysis;
      class cond;
      model fatknow_base= cond  propensity;
      run;
proc glm data=analysis;
      class cond;
      model fatknow_base= cond  predlogit;
      run;
```

APPENDIX E

SAS Code for Quintiles Analysis

```
libname library 'C:\Users\Owner\Documents\SSRP_Thesis\Datasets';
run;

data rnp1;
set library.rnp;
keep id_x1 cond2 fbrmnx1 fbrmnx2 fatmnx1 fatmnx2 sumfatx1 scfat9x1
scfat9x2 scfat8x1 scfat8x2 scfbrx1 scfbrx2 fv1x1 fv1x2 fv7x1 fv7x2 fv6x1
fv6x2 ethnic
educ gender age income marital town tvhrs drvisit epplanx1 epprepx1
epshopx1 fss1x1 fss2x1 fss3x1 fss4x1 fxx5x1 epbrkx1 epdinx1
eplunx1;
if fatmnx2=. then delete;
if cond2=1 then delete;
if ethnic=1 then ethnic=1;
if ethnic=2 then ethnic=2;
if ethnic=3 then ethnic=2;
if ethnic=4 then ethnic=2;
if ethnic=5 then ethnic=2;
if ethnic=6 then ethnic=2;
if ethnic=7 then ethnic=2;
run;
data rnp2;
set rnp1;
if cond2=2 then cond2=2;
run;
data rnp2;
set rnp2(rename=(cond2=cond fbrmnx1=fiber_base fbrmnx2=fiber_1m
fatmnx1=fat_base fatmnx2=fat_1m sumfatx1=fatknow_base
scfat9x1=selffat_base
scfat9x2=selffat_1m scfat8x1=intfat_base scfat8x2=intfat_1m
scfbrx1=intfib_base scfbrx2=intfib_1m fv1x1=fv_base fv1x2=fv_1m
fv7x1=selffv_base
fv7x2=selffv_1m fv6x1=intfv_base fv6x2=intfv_1m epplanx1=epplan
epprepx1=epprep epshopx1=epshop fss1x1=fss1 fss2x1=fss2
fss3x1=fss3 fss4x1=fss4 fxx5x1=fss5 ethnic=ethnic educ=educ
epbrkx1=brkfast epdinx1=dinner eplunx1=lunch id_x1=id));
format row_num ;
id_num=_N_;
__idco=id_num+1000;
if dinner=. then dinner=0;
if dinner=1 then dinner=0;
if dinner=2 then dinner=1;
if dinner=3 then dinner=0;
if lunch=. then lunch=0;
if lunch=1 then lunch=0;
if lunch=2 then lunch=1;
if lunch=3 then lunch=0;
if brkfast=. then brkfast=0;
```

```
if brkfast=1 then brkfast=0;
if brkfast=2 then brkfast=1;
if brkfast=3 then brkfast=0;
meals = epshop + epplan + epprep;
eat_out = brkfast + dinner + lunch;
sum_fss = fss1 + fss2 + fss3 + fss4+ fss5;
if 1=< educ=<3 then educ=1;
if educ=4 then educ=2;
if educ=5 then educ=3;
if educ=6 then educ=3;
if educ>6 then educ=4;
if marital=1 then marital=1;
if marital=2 then marital=2;
if marital=3 then marital=2;
if marital=4 then marital=2;
if marital=5 then marital=2;
drop lunch dinner brkfast fss1 fss2 fss3 fss4 fss5 epplan epshop epprep;
run;

data fiber;
set library.fiberr1m_2011;
keep id flag2 fbrmean ffbfbmx2 fatmean ffbftmx2 fatkntot soc9 soc9x2 soc8
soc8x2 soc10 soc10x2 fv1 fv1x2 fv9 fv9x2 fv8 fv8x2 ethnic educ gender
age income marital town tvhrs docvisit epplan epprep epshop fss1 fss2
fss3 fss4 fss5 epbrfst epdinr eplunch;
fatkntot=fatkntot+1;
if ethnic=1 then ethnic=1;
if ethnic=2 then ethnic=2;
if ethnic=3 then ethnic=2;
if ethnic=4 then ethnic=2;
run;
data fiber;
set fiber(rename=(flag2=cond fbrmean=fiber_base ffbfbmx2=fiber_1m
fatmean=fat_base ffbftmx2=fat_1m fatkntot=fatknow_base soc9=selffat_base
soc9x2=selffat_1m soc8=intfat_base soc8x2=intfat_1m soc10=intfib_base
soc10x2=intfib_1m fv1=fv_base fv1x2=fv_1m fv9=selffv_base
fv9x2=selffv_1m  fv8=intfv_base fv8x2=intfv_1m docvisit=drvisit
ethnic=ethnic educ=educ epbrfst=brkfast epdinr=dinner eplunch=lunch
id=id));
format row_num ;
id_num=_N_;
__idca=id_num+1000;
if dinner=. then dinner=0;
if dinner=1 then dinner=0;
if dinner=2 then dinner=1;
if dinner=3 then dinner=0;
if lunch=. then lunch=0;
if lunch=1 then lunch=0;
if lunch=2 then lunch=1;
if lunch=3 then lunch=0;
if brkfast=. then brkfast=0;
if brkfast=1 then brkfast=0;
if brkfast=2 then brkfast=1;
```

```
if brkfast=3 then brkfast=0;
meals = epshop + epplan + epprep;
eat_out = brkfast + dinner + lunch;
sum_fss = fss1 + fss2 + fss3 + fss4+ fss5;
if 1=< educ=<3 then educ=1;
if educ=4 then educ=2;
if educ=5 then educ=3;
if educ=6 then educ=3;
if educ>6 then educ=4;
if marital=1 then marital=1;
if marital=2 then marital=2;
if marital=3 then marital=2;
if marital=4 then marital=2;
if marital=5 then marital=2;
drop lunch dinner brkfast fss1 fss2 fss3 fss4 fss5 epplan epshop epprep;
run;

data fiber_rnp;
set fiber rnp2;
run;


/* t-tests and chi-squared tests to find significance of possible
variables for the model */

proc freq data=fiber_rnp;
tables cond*ethnic/chisq ;
tables cond*educ/chisq ;
tables cond*gender/chisq ;
tables cond*marital/chisq ;
tables cond*town/chisq ;
tables cond*eat_out/chisq ;
run;

proc ttest data=fiber_rnp;
class cond;
var age tvhrs drvisit fatknow_base meals sum_fss;
run;

/*creating a propensity score*/

proc logistic data=fiber_rnp ;
  class cond educ/param=ref ref=first;
  model cond =  ethnic educ town eat_out drvisit fatknow_base /* age
gender*/;
  output out=propen pred=propensity xbeta=predlogit;
      run;

proc means data=fiber_rnp n;
var ethnic educ town eat_out drvisit fatknow_base age gender;
run;

DATA PROPEN;
```

```
      SET PROPEN;
      LOGIT = LOG(PROPENSITY/(1-PROPENSITY));
      PROC PRINT;
      RUN;

      proc means data=fiber_rnp;
      class cond;
      var age gender ethnic educ town eat_out drvisit fatknow_base;
      run;

proc means data = propen;
var logit  propensity predlogit;
run;


proc freq data = propen;
tables propensity;
run;

proc univariate data=propen;
  var propensity;
  title 'propensity';
  output out=quint pctlpre=p_  pctlpts=20 40 60 80 100;
run;

/*Creating Quintiles of the propensity score*/

data quintile;
merge quint propen;
if p_20=. then p_20=0.5702474318;
if p_40=. then p_40=0.7324023404;
if p_60=. then p_60=0.8026081979;
if p_80=. then p_80=0.8704126427;
if p_100=. then p_100=0.9625928874;
if propensity=. then delete;
if 0<=propensity then gp=p_20;
if propensity =< p_20 then gp=p_20;
if p_20 < propensity =<p_40 then gp=p_40;
if p_40 < propensity =<p_60 then gp=p_60;
if p_60 < propensity =<p_80 then gp=p_80;
if p_80 < propensity =<p_100 then gp=p_100;
if gp=p_20 then gp_num=1;
if gp=p_40 then gp_num=2;
if gp=p_60 then gp_num=3;
if gp=p_80 then gp_num=4;
if gp=p_100 then gp_num=5;
run;

proc sort data= quintile; by gp_num;run;

/*Checking the distribution of the Quintiles for equality across them*/

proc freq data=quintile;
```

```
tables gp_num*ethnic/chisq ;
tables gp_num*educ/chisq ;
tables gp_num*gender/chisq ;
tables gp_num*marital/chisq ;
tables gp_num*town/chisq ;
tables gp_num*eat_out/chisq ;
tables gp_num*age/chisq;
tables gp_num*tvhrs/chisq;
tables gp_num*drvisit/chisq;
tables gp_num*fatknow_base/chisq;
tables gp_num*meals/chisq;
tables gp_num*sum_fss/chisq;
run;

proc freq data=quintile;
tables gp*cond;
run;

/*Creating outcome variables*/

data quintile_analysis;
set quintile;
fat_behave=fat_1m - fat_base;
fiber_behave=fiber_1m - fiber_base;
fv_behave=fv_1m - fv_base;
intfv=intfv_1m - intfv_base;
selffat=selffat_1m - selffat_base;
intfat=intfat_1m - intfat_base;
intfiber=intfib_1m - intfib_base;
selffv=selffv_1m - selffv_base;
if gp_num=1 then d1=1; else d1=0;
if gp_num=2 then d2=1; else d2=0;
if gp_num=3 then d3=1; else d3=0;
if gp_num=4 then d4=1; else d4=0;
if gp_num=5 then d5=1; else d5=0;
run;

proc sort data=quintile_analysis; by gp_num cond;run;
proc means data=quintile_analysis;
by gp_num cond;
var fat_behave fiber_behave fv_behave intfv selffat intfat intfiber
selffv ;
output out=quint_means ;
run;

/*checking the linearity of the quintiles in change from baseline*/

data quint_mean;
set quint_means;
if _stat_='N' then delete;
if _stat_='MIN' then delete;
if _stat_='MAX' then delete;
if _stat_='STD' then delete;
```

```
run;

/*plotting the quintiles*/

proc sort data=quint_mean; by cond gp_num;
symbol1 interpol=join value=dot color=blue ;
proc gplot data=quint_mean;
by cond;
plot fat_behave*gp_num fiber_behave*gp_num fv_behave*gp_num intfat*gp_num
intfiber*gp_num intfv*gp_num
selffat*gp_num selffv*gp_num;
title 'Looking for Linearity of Quintiles';
run;

/*Quintiles Anlysis*/

proc glm data=quintile_analysis;
      class cond gp ;
      model fat_behave= cond gp ;
      lsmeans cond/cl pdiff stderr ;
      title 'fat behavior analysis 4';
      run;
Proc reg data=quintile_analysis;
      model fat_behave= d1 d2 d3 d4 d5 cond /selection=adjrsq aic bic;
      title 'fat behavior analysis 4';
      run;

proc glm data=quintile_analysis;
      class cond gp_num;
      model fiber_behave= cond gp_num ;
      lsmeans cond/cl pdiff stderr ;
      title 'fiber behavior analysis 4';
      run;
Proc reg data=quintile_analysis;
      model fiber_behave= d1 d2 d3 d4 d5 cond /selection=adjrsq aic bic;
      title 'fiber behavior analysis 4';
      run;

proc glm data=quintile_analysis;
      class cond gp_num ;
      model fv_behave= cond gp_num;
      lsmeans cond/cl pdiff stderr ;
      title 'fv behavior analysis 4';
      run;
Proc reg data=quintile_analysis;
      model fv_behave=d1 d2 d3 d4 d5 cond /selection=adjrsq aic bic;
      title 'fv behavior analysis 4';
      run;

proc glm data=quintile_analysis;
      class cond gp_num;
      model intfat= cond gp_num ;
      lsmeans cond/cl pdiff stderr ;
```

```
      title 'fat intentions analysis 4';
      run;
Proc reg data=quintile_analysis;
      model intfat= d1 d2 d3 d4 d5 cond /selection=adjrsq aic bic;
      title 'fat intentions analysis 4';
      run;

proc glm data=quintile_analysis;
      class cond gp_num ;
      model intfiber= cond gp_num;
      lsmeans cond/cl pdiff stderr ;
      title 'fiber intentions analysis 4';
      run;
Proc reg data=quintile_analysis;
      model intfiber= d1 d2 d3 d4 d5 cond /selection=adjrsq aic bic;
      title 'fiber intentions analysis 4';
      run;

proc glm data=quintile_analysis;
      class cond gp_num;
      model intfv= cond gp_num ;
      lsmeans cond/cl pdiff stderr ;
      title 'fv intentions analysis 4';
      run;
Proc reg data=quintile_analysis;
      model intfv= d1 d2 d3 d4 d5 cond /selection=adjrsq aic bic;
      title 'fv intentions analysis 4';
      run;

proc glm data=quintile_analysis;
      class cond gp_num;
      model selffat= cond gp_num ;
      lsmeans cond/cl pdiff stderr ;
      title 'fat self-effiacy analysis 4';
      run;
Proc reg data=quintile_analysis;
      model selffat= d1 d2 d3 d4 d5 cond /selection=adjrsq aic bic;
      title 'fat self-effiacy analysis 4';
      run;

proc glm data=quintile_analysis;
      class cond gp_num;
      model selffv= cond gp_num;
      lsmeans cond/cl pdiff stderr ;
      title 'fv self-effiacy analysis 4';
      run;
Proc reg data=quintile_analysis;
      model selffv= d1 d2 d3 d4 d5 cond /selection=adjrsq aic bic;
      title 'fv self-effiacy analysis 4';
      run;

/*checking that there is no longer significant differences between
baseline variables*/
```

```
proc glm data=quintile_analysis;
      class cond ;
      model ethnic= cond gp;
      run;
proc glm data=quintile_analysis;
      class cond ;
      model educ= cond gp;
      run;
proc glm data=quintile_analysis;
      class cond ;
      model town= cond gp;
      run;
proc glm data=quintile_analysis;
      class cond ;
      model eat_out= cond gp;
      run;
proc glm data=quintile_analysis;
      class cond;
      model drvisit= cond gp;
      run;
proc glm data=quintile_analysis;
      class cond ;
      model fatknow_base= cond gp;
      run;
```

APPENDIX F

SAS Code for Weighted Analysis

```
      libname library 'C:\Users\Owner\Documents\SSRP_Thesis\Datasets';
run;

data rnp1;
set library.rnp;
keep id_x1 cond2 fbrmnx1 fbrmnx2 fatmnx1 fatmnx2 sumfatx1 scfat9x1
scfat9x2 scfat8x1 scfat8x2 scfbrx1 scfbrx2 fv1x1 fv1x2 fv7x1 fv7x2 fv6x1
fv6x2 ethnic
educ gender age income marital town tvhrs drvisit epplanx1 epprepx1
epshopx1 fss1x1 fss2x1 fss3x1 fss4x1 fxx5x1 epbrkx1 epdinx1
eplunx1;
if fatmnx2=. then delete;
if cond2=1 then delete;
if ethnic=1 then ethnic=1;
if ethnic=2 then ethnic=2;
if ethnic=3 then ethnic=2;
if ethnic=4 then ethnic=2;
if ethnic=5 then ethnic=2;
if ethnic=6 then ethnic=2;
if ethnic=7 then ethnic=2;
run;
data rnp2;
set rnp1;
if cond2=2 then cond2=2;
run;
data rnp2;
set rnp2(rename=(cond2=cond fbrmnx1=fiber_base fbrmnx2=fiber_1m
fatmnx1=fat_base fatmnx2=fat_1m sumfatx1=fatknow_base
scfat9x1=selffat_base
scfat9x2=selffat_1m scfat8x1=intfat_base scfat8x2=intfat_1m
scfbrx1=intfib_base scfbrx2=intfib_1m fv1x1=fv_base fv1x2=fv_1m
fv7x1=selffv_base
fv7x2=selffv_1m fv6x1=intfv_base fv6x2=intfv_1m epplanx1=epplan
epprepx1=epprep epshopx1=epshop fss1x1=fss1 fss2x1=fss2
fss3x1=fss3 fss4x1=fss4 fxx5x1=fss5 ethnic=ethnic educ=educ
epbrkx1=brkfast epdinx1=dinner eplunx1=lunch id_x1=id));
format row_num ;
id_num=_N_;
__idco=id_num+1000;
if dinner=. then dinner=0;
if dinner=1 then dinner=0;
if dinner=2 then dinner=1;
if dinner=3 then dinner=0;
if lunch=. then lunch=0;
if lunch=1 then lunch=0;
if lunch=2 then lunch=1;
if lunch=3 then lunch=0;
if brkfast=. then brkfast=0;
```

```
if brkfast=1 then brkfast=0;
if brkfast=2 then brkfast=1;
if brkfast=3 then brkfast=0;
meals = epshop + epplan + epprep;
eat_out = brkfast + dinner + lunch;
sum_fss = fss1 + fss2 + fss3 + fss4+ fss5;
if 1=< educ=<3 then educ=1;
if educ=4 then educ=2;
if educ=5 then educ=3;
if educ=6 then educ=3;
if educ>6 then educ=4;
if marital=1 then marital=1;
if marital=2 then marital=2;
if marital=3 then marital=2;
if marital=4 then marital=2;
if marital=5 then marital=2;
drop lunch dinner brkfast fss1 fss2 fss3 fss4 fss5 epplan epshop epprep;
run;

data fiber;
set library.fiberr1m_2011;
keep id flag2 fbrmean ffbfbmx2 fatmean ffbftmx2 fatkntot soc9 soc9x2 soc8
soc8x2 soc10 soc10x2 fv1 fv1x2 fv9 fv9x2 fv8 fv8x2 ethnic educ gender
age income marital town tvhrs docvisit epplan epprep epshop fss1 fss2
fss3 fss4 fss5 epbrfst epdinr eplunch;
fatkntot=fatkntot+1;
if ethnic=1 then ethnic=1;
if ethnic=2 then ethnic=2;
if ethnic=3 then ethnic=2;
if ethnic=4 then ethnic=2;
run;
data fiber;
set fiber(rename=(flag2=cond fbrmean=fiber_base ffbfbmx2=fiber_1m
fatmean=fat_base ffbftmx2=fat_1m fatkntot=fatknow_base soc9=selffat_base
soc9x2=selffat_1m soc8=intfat_base soc8x2=intfat_1m soc10=intfib_base
soc10x2=intfib_1m fv1=fv_base fv1x2=fv_1m fv9=selffv_base
fv9x2=selffv_1m  fv8=intfv_base fv8x2=intfv_1m docvisit=drvisit
ethnic=ethnic educ=educ epbrfst=brkfast epdinr=dinner eplunch=lunch
id=id));
format row_num ;
id_num=_N_;
__idca=id_num+1000;
if dinner=. then dinner=0;
if dinner=1 then dinner=0;
if dinner=2 then dinner=1;
if dinner=3 then dinner=0;
if lunch=. then lunch=0;
if lunch=1 then lunch=0;
if lunch=2 then lunch=1;
if lunch=3 then lunch=0;
if brkfast=. then brkfast=0;
if brkfast=1 then brkfast=0;
if brkfast=2 then brkfast=1;
```

```
if brkfast=3 then brkfast=0;
meals = epshop + epplan + epprep;
eat_out = brkfast + dinner + lunch;
sum_fss = fss1 + fss2 + fss3 + fss4+ fss5;
if 1=< educ=<3 then educ=1;
if educ=4 then educ=2;
if educ=5 then educ=3;
if educ=6 then educ=3;
if educ>6 then educ=4;
if marital=1 then marital=1;
if marital=2 then marital=2;
if marital=3 then marital=2;
if marital=4 then marital=2;
if marital=5 then marital=2;
drop lunch dinner brkfast fss1 fss2 fss3 fss4 fss5 epplan epshop epprep;
run;

data fiber_rnp;
set fiber rnp2;
run;


/* t-tests and chi-squared tests to find significance of possible
variables for the model */

proc freq data=fiber_rnp;
tables cond*ethnic/chisq ;
tables cond*educ/chisq ;
tables cond*gender/chisq ;
tables cond*marital/chisq ;
tables cond*town/chisq ;
tables cond*eat_out/chisq ;
run;

proc ttest data=fiber_rnp;
class cond;
var age tvhrs drvisit fatknow_base meals sum_fss;
run;

/*creating a propensity score*/

proc logistic data=fiber_rnp ;
  class cond educ/param=ref ref=first;
  model cond =  ethnic educ town eat_out drvisit fatknow_base /* age
gender*/;
  output out=propen pred=propensity xbeta=predlogit;
      run;


proc means data=fiber_rnp n;
var ethnic educ town eat_out drvisit fatknow_base age gender;
run;
```

```
DATA PROPEN;
      SET PROPEN;
      LOGIT = LOG(PROPENSITY/(1-PROPENSITY));
      PROC PRINT;
      RUN;

      proc means data=fiber_rnp;
      class cond;
      var age gender ethnic educ town eat_out drvisit fatknow_base;
      run;

proc means data = propen;
var logit  propensity predlogit;
run;


proc freq data = propen;
tables propensity;
run;


data propen1;
SET PROPEN;
idnum = id_num + 1000;
drop id_num;
run;

/*Creating outcome variables*/

data analysis;
set propen1;
fat_behave=fat_1m - fat_base;
fiber_behave=fiber_1m - fiber_base;
fv_behave=fv_1m - fv_base;
intfv=intfv_1m - intfv_base;
selffat=selffat_1m - selffat_base;
intfat=intfat_1m - intfat_base;
intfiber=intfib_1m - intfib_base;
selffv=selffv_1m - selffv_base;
run;

/*Adding weights to the propensity score*/

data weight_analysis;
set analysis;
weight=propensity;
do weight=weight/(1-weight);
end;
if cond=1 then weight=1;
if cond=2 then weight=weight;
run;

proc univariate data=weight_analysis;
```

```
by cond;
var weight;
histogram weight;
run;
/*Weights Analysis*/


proc glm data=weight_analysis;
      class cond;
      model fat_behave= cond ;
      weight weight;
      lsmeans cond/cl pdiff stderr ;
      title 'fat behavior analysis 5';
      run;
proc reg data=weight_analysis ;
      model fat_behave = cond /selection=adjrsq aic bic;
      weight weight;
      title 'fat behavior analysis 5';
      run;

proc glm data=weight_analysis;
      class cond;
      model fiber_behave= cond ;
      weight weight;
      lsmeans cond/cl pdiff stderr ;
      title 'fiber behavior analysis 5';
      run;
proc reg data=weight_analysis ;
      model fiber_behave = cond /selection=adjrsq aic bic;
      weight weight;
      title 'fiber behavior analysis 5';
      run;

proc glm data=weight_analysis;
      class cond;
      model fv_behave= cond ;
      weight weight;
      lsmeans cond/cl pdiff stderr ;
      title 'fv behavior analysis 5';
      run;
proc reg data=weight_analysis ;
      model fv_behave = cond /selection=adjrsq aic bic;
      weight weight;
      title 'fv behavior analysis 5';
      run;

proc glm data=weight_analysis;
      class cond;
      model intfat= cond ;
      weight weight;
      lsmeans cond/cl pdiff stderr ;
      title 'fat intentions analysis 5';
      run;
```

```
proc reg data=weight_analysis ;
      model intfat = cond /selection=adjrsq aic bic;
      weight weight;
      title 'fat intentions analysis 5';
      run;

proc glm data=weight_analysis;
      class cond;
      model intfiber= cond ;
      weight weight;
      lsmeans cond/cl pdiff stderr ;
      title 'fiber intentions analysis 5';
      run;
proc reg data=weight_analysis ;
      model intfiber = cond /selection=adjrsq aic bic;
      weight weight;
      title 'fiber intentions analysis 5';
      run;

proc glm data=weight_analysis;
      class cond;
      model intfv= cond ;
      weight weight;
      lsmeans cond/cl pdiff stderr ;
      title 'fv intentions analysis 5';
      run;
proc reg data=weight_analysis ;
      model intfat = cond /selection=adjrsq aic bic;
      weight weight;
      title 'fv intentions analysis 5';
      run;

proc glm data=weight_analysis;
      class cond;
      model selffat= cond ;
      weight weight;
      lsmeans cond/cl pdiff stderr ;
      title 'fat self-efficacy analysis 5';
      run;
proc reg data=weight_analysis ;
      model selffat = cond /selection=adjrsq aic bic;
      weight weight;
      title 'fat self-efficacy analysis 5';
      run;

proc glm data=weight_analysis;
      class cond;
      model selffv= cond ;
      weight weight;
      lsmeans cond/cl pdiff stderr  ;
      title 'fv self-efficacy analysis 5';
      run;
proc reg data=weight_analysis ;
```

```
      model selffv = cond /selection=adjrsq aic bic;
      weight weight;
      title 'fv self-efficacy analysis 5';
      run;


/*checking that there is no longer significant differences between
baseline variables*/
proc glm data=weight_analysis;
      class cond;
      model ethnic= cond ;
      weight weight;
      run;
proc glm data=weight_analysis;
      class cond;
      model educ= cond ;
      weight weight;
      run;
proc glm data=weight_analysis;
      class cond;
      model town= cond ;
      weight weight;
      run;
proc glm data=weight_analysis;
      class cond;
      model eat_out= cond ;
      weight weight;
      run;
proc glm data=weight_analysis;
      class cond;
      model drvisit= cond ;
      weight weight;
      run;
proc glm data=weight_analysis;
      class cond;
      model fatknow_base= cond ;
      weight weight;
      run;
```

APPENDIX G

SAS Code for Trimmed Weights Analysis

```
libname library 'C:\Users\Owner\Documents\SSRP_Thesis\Datasets';
run;

data rnp1;
set library.rnp;
keep id_x1 cond2 fbrmnx1 fbrmnx2 fatmnx1 fatmnx2 sumfatx1 scfat9x1
scfat9x2 scfat8x1 scfat8x2 scfbrx1 scfbrx2 fv1x1 fv1x2 fv7x1 fv7x2 fv6x1
fv6x2 ethnic
educ gender age income marital town tvhrs drvisit epplanx1 epprepx1
epshopx1 fss1x1 fss2x1 fss3x1 fss4x1 fxx5x1 epbrkx1 epdinx1
eplunx1;
if fatmnx2=. then delete;
if cond2=1 then delete;
if ethnic=1 then ethnic=1;
if ethnic=2 then ethnic=2;
if ethnic=3 then ethnic=2;
if ethnic=4 then ethnic=2;
if ethnic=5 then ethnic=2;
if ethnic=6 then ethnic=2;
if ethnic=7 then ethnic=2;
run;
data rnp2;
set rnp1;
if cond2=2 then cond2=2;
run;
data rnp2;
set rnp2(rename=(cond2=cond fbrmnx1=fiber_base fbrmnx2=fiber_1m
fatmnx1=fat_base fatmnx2=fat_1m sumfatx1=fatknow_base
scfat9x1=selffat_base
scfat9x2=selffat_1m scfat8x1=intfat_base scfat8x2=intfat_1m
scfbrx1=intfib_base scfbrx2=intfib_1m fv1x1=fv_base fv1x2=fv_1m
fv7x1=selffv_base
fv7x2=selffv_1m fv6x1=intfv_base fv6x2=intfv_1m epplanx1=epplan
epprepx1=epprep epshopx1=epshop fss1x1=fss1 fss2x1=fss2
fss3x1=fss3 fss4x1=fss4 fxx5x1=fss5 ethnic=ethnic educ=educ
epbrkx1=brkfast epdinx1=dinner eplunx1=lunch id_x1=id));
format row_num ;
id_num=_N_;
__idco=id_num+1000;
if dinner=. then dinner=0;
if dinner=1 then dinner=0;
if dinner=2 then dinner=1;
if dinner=3 then dinner=0;
if lunch=. then lunch=0;
if lunch=1 then lunch=0;
if lunch=2 then lunch=1;
if lunch=3 then lunch=0;
if brkfast=. then brkfast=0;
```

```
if brkfast=1 then brkfast=0;
if brkfast=2 then brkfast=1;
if brkfast=3 then brkfast=0;
meals = epshop + epplan + epprep;
eat_out = brkfast + dinner + lunch;
sum_fss = fss1 + fss2 + fss3 + fss4+ fss5;
if 1=< educ=<3 then educ=1;
if educ=4 then educ=2;
if educ=5 then educ=3;
if educ=6 then educ=3;
if educ>6 then educ=4;
if marital=1 then marital=1;
if marital=2 then marital=2;
if marital=3 then marital=2;
if marital=4 then marital=2;
if marital=5 then marital=2;
drop lunch dinner brkfast fss1 fss2 fss3 fss4 fss5 epplan epshop epprep;
run;

data fiber;
set library.fiberr1m_2011;
keep id flag2 fbrmean ffbfbmx2 fatmean ffbftmx2 fatkntot soc9 soc9x2 soc8
soc8x2 soc10 soc10x2 fv1 fv1x2 fv9 fv9x2 fv8 fv8x2 ethnic educ gender
age income marital town tvhrs docvisit epplan epprep epshop fss1 fss2
fss3 fss4 fss5 epbrfst epdinr eplunch;
fatkntot=fatkntot+1;
if ethnic=1 then ethnic=1;
if ethnic=2 then ethnic=2;
if ethnic=3 then ethnic=2;
if ethnic=4 then ethnic=2;
run;
data fiber;
set fiber(rename=(flag2=cond fbrmean=fiber_base ffbfbmx2=fiber_1m
fatmean=fat_base ffbftmx2=fat_1m fatkntot=fatknow_base soc9=selffat_base
soc9x2=selffat_1m soc8=intfat_base soc8x2=intfat_1m soc10=intfib_base
soc10x2=intfib_1m fv1=fv_base fv1x2=fv_1m fv9=selffv_base
fv9x2=selffv_1m  fv8=intfv_base fv8x2=intfv_1m docvisit=drvisit
ethnic=ethnic educ=educ epbrfst=brkfast epdinr=dinner eplunch=lunch
id=id));
format row_num ;
id_num=_N_;
__idca=id_num+1000;
if dinner=. then dinner=0;
if dinner=1 then dinner=0;
if dinner=2 then dinner=1;
if dinner=3 then dinner=0;
if lunch=. then lunch=0;
if lunch=1 then lunch=0;
if lunch=2 then lunch=1;
if lunch=3 then lunch=0;
if brkfast=. then brkfast=0;
if brkfast=1 then brkfast=0;
if brkfast=2 then brkfast=1;
```

```
if brkfast=3 then brkfast=0;
meals = epshop + epplan + epprep;
eat_out = brkfast + dinner + lunch;
sum_fss = fss1 + fss2 + fss3 + fss4+ fss5;
if 1=< educ=<3 then educ=1;
if educ=4 then educ=2;
if educ=5 then educ=3;
if educ=6 then educ=3;
if educ>6 then educ=4;
if marital=1 then marital=1;
if marital=2 then marital=2;
if marital=3 then marital=2;
if marital=4 then marital=2;
if marital=5 then marital=2;
drop lunch dinner brkfast fss1 fss2 fss3 fss4 fss5 epplan epshop epprep;
run;

data fiber_rnp;
set fiber rnp2;
run;


/* t-tests and chi-squared tests to find significance of possible
variables for the model */

proc freq data=fiber_rnp;
tables cond*ethnic/chisq ;
tables cond*educ/chisq ;
tables cond*gender/chisq ;
tables cond*marital/chisq ;
tables cond*town/chisq ;
tables cond*eat_out/chisq ;
run;

proc ttest data=fiber_rnp;
class cond;
var age tvhrs drvisit fatknow_base meals sum_fss;
run;

/*creating a propensity score*/

proc logistic data=fiber_rnp ;
  class cond educ/param=ref ref=first;
  model cond =  ethnic educ town eat_out drvisit fatknow_base /* age
gender*/;
  output out=propen pred=propensity xbeta=predlogit;
      run;


proc means data=fiber_rnp n;
var ethnic educ town eat_out drvisit fatknow_base age gender;
run;
```

```
DATA PROPEN;
      SET PROPEN;
      LOGIT = LOG(PROPENSITY/(1-PROPENSITY));
      PROC PRINT;
      RUN;

      proc means data=fiber_rnp;
      class cond;
      var age gender ethnic educ town eat_out drvisit fatknow_base;
      run;

proc means data = propen;
var logit  propensity predlogit;
run;


proc freq data = propen;
tables propensity;
run;

proc univariate data=propen;
  *class cond;
  var propensity;
  output out=weight pctlpre=p_  pctlpts=90 95 99;
run;

/*Creating outcome variables*/

data analysis;
set propen;
fat_behave=fat_1m - fat_base;
fiber_behave=fiber_1m - fiber_base;
fv_behave=fv_1m - fv_base;
intfv=intfv_1m - intfv_base;
selffat=selffat_1m - selffat_base;
intfat=intfat_1m - intfat_base;
intfiber=intfib_1m - intfib_base;
selffv=selffv_1m - selffv_base;
run;

/*Trimming the Weights*/

data trim;
merge weight analysis;
if p_90=. then p_90=0.9071551248;
if p_95=. then p_95=0.9372892647;
if p_99=. then p_99=0.9570148604;
if propensity=>p_90 then propensity=p_90;     /*setting everying 90% and
greater equal to 90%*/
*if propensity=>p_95 then propensity=p_95;         /*setting everying 95%
and greater equal to 95%*/
*if propensity=>p_99 then propensity=p_99;         /*setting everying 99%
and greater equal to 99%*/
```

```
*if propensity>p_90 then delete;                    /*ignoring everying
greater than 90%*/
*if propensity>p_95 then delete;                          /*ignoring
everying greater than 95%*/
*if propensity>p_99 then delete;                    /*ignoring everying
greater than 99%*/
run;

data trim_analysis;
set trim;
trim=propensity;
do trim=trim/(1-trim);
end;
if cond=1 then trim=1;
if cond=2 then trim=trim;
run;

/*Trimmed Weights Analysis*/

proc glm data=trim_analysis;
      class cond;
      model fat_behave=cond ;
      weight trim;
      lsmeans cond/cl pdiff stderr ;
      title 'fat behavior analysis 6';
      run;
proc reg data=trim_analysis ;
      model fat_behave = cond /selection=adjrsq aic bic;
      weight trim;
      title 'fat behavior analysis 6';
      run;

proc glm data=trim_analysis;
      class cond;
      model fiber_behave=cond ;
      weight trim;
      lsmeans cond/cl pdiff stderr ;
      title 'fiber behavior analysis 6';
      run;
proc reg data=trim_analysis ;
      model fiber_behave = cond /selection=adjrsq aic bic;
      weight trim;
      title 'fiber behavior analysis 6';
      run;

proc glm data=trim_analysis;
      class cond;
      model fv_behave=cond ;
      weight trim;
      lsmeans cond/cl pdiff stderr ;
      title 'fv behavior analysis 6';
      run;
proc reg data=trim_analysis ;
```

```
        model fv_behave = cond /selection=adjrsq aic bic;
        weight trim;
        title 'fv behavior analysis 6';
        run;

proc glm data=trim_analysis;
        class cond;
        model intfat=cond ;
        weight trim;
        lsmeans cond/cl pdiff stderr ;
        title 'fat intentions analysis 6';
        run;
proc reg data=trim_analysis ;
        model intfat = cond /selection=adjrsq aic bic;
        weight trim;
        title 'fat intentions analysis 6';
        run;

proc glm data=trim_analysis;
        class cond;
        model intfiber=cond ;
        weight trim;
        lsmeans cond/cl pdiff stderr ;
        title 'fiber intentions analysis 6';
        run;
proc reg data=trim_analysis ;
        model intfiber = cond /selection=adjrsq aic bic;
        weight trim;
        title 'fiber intentions analysis 6';
        run;

proc glm data=trim_analysis;
        class cond;
        model intfv=cond ;
        weight trim;
        lsmeans cond/cl pdiff stderr ;
        title 'fv intentions analysis 6';
        run;
proc reg data=trim_analysis ;
        model intfv = cond /selection=adjrsq aic bic;
        weight trim;
        title 'fv intentions analysis 6';
        run;

proc glm data=trim_analysis;
        class cond;
        model selffat=cond ;
        weight trim;
        lsmeans cond/cl pdiff stderr ;
        title 'fat self-efficacy analysis 6';
        run;
proc reg data=trim_analysis ;
        model selffat = cond /selection=adjrsq aic bic;
```

```
      weight trim;
      title 'fat self-efficacy analysis 6';
      run;

proc glm data=trim_analysis;
      class cond;
      model selffv=cond ;
      weight trim;
      lsmeans cond/cl pdiff stderr ;
      title 'fv self-efficacy analysis 6';
      run;
proc reg data=trim_analysis ;
      model selffv = cond /selection=adjrsq aic bic;
      weight trim;
      title 'fv self-efficacy analysis 6';
      run;

/*checking that there is no longer significant differences between
baseline variables*/
proc glm data=trim_analysis;
      class cond;
      model ethnic=cond ;
      weight trim;
      run;
proc glm data=trim_analysis;
      class cond;
      model educ=cond ;
      weight trim;
      run;
proc glm data=trim_analysis;
      class cond;
      model town=cond ;
      weight trim;
      run;
proc glm data=trim_analysis;
      class cond;
      model eat_out=cond ;
      weight trim;
      run;
proc glm data=trim_analysis;
      class cond;
      model drvisit=cond ;
      weight trim;
      run;
proc glm data=trim_analysis;
      class cond;
      model fatknow_base=cond ;
      weight trim;
      run;
```

VITA

Hali Summer Esinhart was born June 5th 1988 in Richmond, VA to James Douglas Esinhart and Susan Chandler Esinhart. She grew up at Wrightsville Beach, NC where she attended John T. Hoggard High School where she received the Athletic Booster Club's scholarship for showing outstanding achievements both academically and athletically.

She attended the University of North Carolina Chapel Hill where she received a Bachelor of Science in Biochemistry. While attending UNC, she did research under Dr. Malcolm D. E. Forbes in the topic of Electronic Paramagnetic Resonance. The research led to a publication in the Langmuir of the American Chemical Society. It can be found as follows:

Caregnato P., Jarocha L. E., Esinhart H.S., Lebedeva N. V., Tarasov V. F., Forbes M. D. E. "Electrostatic Control of Spin Exchange Between Mobile Spin-Correlated Radical Pairs Created in Micellar Solutions" *Langmuir.* 2011 May 3; 27(9):5304-9.Epub 2011 April 8.

She then attended Virginia Commonwealth University, where she studied Biostatistics. While attending VCU, she was a teaching assistant for one semester for Dr. Al Best and Dr. Tina Cunningham in an introductory Biostatistics course. She also received the Best New Presenter Award at the Biostatistics Student Research Symposium in her first year. A paper is currently being completed on her topic of research, which was extended to her thesis.