5-2016

Predicting Nonadherence Behavior Towards Mammography Screening Guidelines

Brian L. Trussell
University of Arkansas

Predicting Nonadherence Behavior Towards Mammography Screening Guidelines

An Undergraduate Honors College Thesis

in the

Department of Industrial Engineering

College of Engineering

University of Arkansas

Fayetteville, AR

by

Brian L. Trussell

Thesis Advisor: Dr. Shengfan Zhang

Thesis Reader: Dr. Ashlea Bennett Milburn

Signature Page

Predicting Nonadherence Behavior
Towards Mammography
Screening Guidelines

A thesis submitted in partial fulfillment
of the requirements for the degree of
Bachelors of Science in Industrial Engineering with Honors

by

Brian Trussell

May 2016
University of Arkansas

Dr. Shengfan Zhang
Thesis Director

Dr. Ashlea Bennett Milburn
Committee Member

Acknowledgements

I take this opportunity to thank all of the University of Arkansas Department of Industrial Engineering staff for all of the hard work they put into developing their students. Everybody has the students' best interests at heart, making the past 4 years extremely arduous but even more gratifying. A special thanks to Dr. Tish Pohl and Dr. Ashlea Bennett Milburn for their consistently cheerful attitudes. I also thank my parents for the continuous support and unconditional love they have provided throughout my life.

A special thanks to Fan Wang for taking time out of his schedule to extract BRFSS data into an Excel file for me.

Finally, I would like to sincerely thank Dr. Shengfan Zhang for all of her guidance and support during the honors thesis process. She is a very hardworking and patient professor that cares deeply about her students, not only in academia, but also as young adults going through an important time in their lives.

Abstract

The goal of this research is to examine factors associated with nonadherence behavior toward mammography screening among U.S. women. The 2014 Behavioral Risk Factor Surveillance System (BRFSS) survey data was used for this study, allowing the model to represent a robust sample. A logistic regression model was developed to gain an understanding of influencing factors, including demographic, health-related and behavioral characteristics. Further analysis with logistic regression models stratified by age were conducted to control for the effect of age. The results show that demographic and health related information such as income, number of children, and BMI category can help intervention programs recognize women who are less likely to adhere to mammography screening guidelines. Behavioral factors are the strongest predictor for screening behaviors. It is crucial for women to have a personal physician or health professional that they can routinely see every year. Tracking frequency of doctor visits and routine medical procedures can give great insight into mammography nonadherence, which could ultimately help reduce breast cancer mortality in the U.S.

Table of Contents

Formatted: Font: 12 pt

1. Introduction

Except for non-melanoma skin cancer, breast cancer occurs more than any other type~~s~~ of cancer in American women. According to the Center for Disease Control and Prevention (CDC), it estimated that 246,660 American women will be diagnosed and 40,450 women are predicted to die from breast cancer in the year 2016 [1]. Eating a healthy diet, exercising, and avoiding alcohol can reduce the risk of getting cancer, but there is no guaranteed way to prevent breast cancer. Mammography has long been considered to be the most effective technology for population-based breast cancer screening so women are recommended to receive regular mammogram screenings [2], which has proven to reduce breast cancer mortality by about 28% according to an earlier study [3].

Mammography is a diagnostic and screening tool that uses X-ray imaging to detect breast cancer and diseases. Mammography ~~has the best chance of~~is the best method for early detection of breast cancer which is crucial for minimizing the harm of the disease. Although there is little debate over the benefits of mammograms, multiple organizations have released conflicting guidelines that detail the timing of when a woman should get a screening. The American Cancer Society (ACS) has recently updated their guidelines and now recommends women from the age of 40-44 should have the option to get mammograms, women 45-54 years old need annual screenings, and women 55 and older should switch to biennial screenings [4]. This is similar to the American College of Radiology (ACR) and the Society of Breast Imaging (SBI) except these groups recommend annual screening to start at the age of 40 [5]. The U.S. Preventive Services Task Force, on the other hand, recommends biennial mammogram

screenings from the age of 50-74. Women outside of this age group are encouraged to make a personal decision to get screened since there is insufficient evidence to assess these age groups [6].

Although some recent studies suggest that mammograms are ineffective and lead to emotional distress due to over-diagnosis, a larger amount of literature supports screenings, stating that mammogram screenings reduce breast cancer mortality by up to 48% [7, 8, 9]. The CDC shows that mammogram screening percentages for women in 1987 were 31.9%, 31.7%, and 22.8% for the respective age groups of 40-49, 50 -64, and 65 years and older. After a push for the importance of mammography screenings this percentage increased and plateaued to around 63%, 75% and 67% for the respective age groups [10]. Another study shows that the 5-year survival rates over a similar time frame (1987-1989 to 2001-2007) increased by 6% over all races [11].

The aim of this study is to examine a wide range of factors (behavioral, demographic, and health-related) using logistic regression that can predict if a woman is likely not to adhere to U.S. mammogram guidelines. We consider a woman nonadherent when she had her last mammogram test more than 2 years ago. After preliminary results are found the regression is stratified by age group to further investigate the effect that age has on the non-adherence to mammogram screening. The CDC's 2014 Behavioral Risk Factor Surveillance System data is used in order to have a current representation of a large sample of the U.S. population. Stratified analyses by age group are conducted to further investigate the effect that age has on the nonadherence behavior to mammogram screening. This thesis first summarizes literature related to mammography nonadherence. Next the methodology is detailed, followed by the

results of the regressions. Lastly, the results are discussed along with limitations and future improvements for this study.

2. Literature Review

There have been various studies focusing on analyzing mammogram screening adherence and nonadherence in recent years. Schueler et al. [12] conducted a systematic review on the utilization of mammogram screenings. They included literature that was written in English and analyzed women in the United States and their adherence to mammogram screenings. If the authors found at least three papers that had homogeneous variable definitions and quantitative data the authors included the studies in their analysis. 195 studies between the years of 1988 and 2004 ended up in the paper's analysis. With mammography adherence as the response variable the results showed that women who lacked health coverage had an adjusted odds ratio (OR, discussed in Section 3.5) of 0.47; women who lacked breast cancer screening knowledge resulted in an adjusted OR of 0.46; and women who smoked cigarettes showed an adjusted OR of 0.69. The results indicate that these factors have a significant negative relationship with women's utilization of mammogram screenings. The strongest correlations came from physicians not recommending a mammogram screening, not visiting a physician in the past year, and having had a recent breast exam, resulting in adjusted ORs of 0.16, 0.34, and 9.15 respectively. Based on the results of the review the authors recommend increasing access to physicians and having these physicians encourage Pap testing, mammogram screenings and clinical breast examinations with the knowledge that a woman is much more likely to get another mammogram.

Formatted: Font: 12 pt

Madadi et al. [13] used logistic regression to analyze predicting factors (socio-demographic, health-related, behavioral, and knowledge of breast cancer/mammography characteristics) associated with women's behaviors toward mammography screening. They first focused on all women over the age of 40, splitting the analyses into an age group above and an age group below the age of 65. The second stage of the analysis focuseds on women with poor mammography screening. The 2003 Health Information National Trends Survey (HINTS) data with sample size of 6,369 was used in their analysis. An unmarried marital status and lower income was found to be associated with lower mammogram adherence. They also found that for the two age groups, women with health insurance, a large number of visits to health providers, being advised to have a mammogram, and trust in cancer information predict strong mammography adherence. Based on their findings they recommend sending reminders to women and give suggestions to programs aimed to improve screening rates.

Calvocorresi et al. [14] studied the psychosocial factors that influence the non-adherence of women receiving regular follow-up mammography screenings over time with a specific focus on age and race. The study also used a tree analysis to predict if a woman was at risk of not adhering to guidelines based on a combination of the psychosocial predictors. Using a survey, data was only collected for white and African-American women at 5 Connecticut hospitals. Based on the individual variable logistic regression, women who perceived that they were very likely to develop breast cancer did not adhere to screening guidelines more than other levels of perceived development. Similarly, younger women (age 40-49) that did not believe they were at risk or believed were at low risk of getting breast cancer resulted in an OR above 3. Women who did not receive a recommendation from a health professional (particularly younger

women) or did not receive a reminder to undergo another mammogram screening were far less likely to adhere to the screening guidelines when compared to those who did. The tree analysis showed that the lowest non-adherent women over the age of 50 believed that~~:~~ mammograms were extremely useful; they were moderately susceptible to getting breast cancer in their life time; they were not embarrassed during their mammogram; and that they had a recommendation/reminder to get a mammogram from a healthcare professional. This study shows the potential to apply broad intervention techniques as well as specific intervention techniques based on various demographic characteristics.

Jensen et al. [15] investigated groups of women who did not attend their free biennial breast cancer screening in the Central Denmark Region. The study included women that were invited to participate in a mammogram appointment between the ages of 50-69. Based on the socio-demographic factors pulled from a regional database, women with a lower social status were less likely to attend their screening appointment. More specifically women who made lower income, were unmarried, did not own their own home, and were unemployed were notably less likely to participate in the mammogram program.

Khaliq et al. [16] examine~~d~~s specific~~ally at~~ factors that contribute to hospitalized women's nonadherence to mammography screenings. Data was collected on 250 women over the age of 52 using a bedside survey. The study defined non-adherence to mammography screening as not having had a screening within the 2 years before the survey was taken. The study used a logistic regression to find odds ratios of risk factors. The most significant results came from women who made less than $20,000 per year, smoked tobacco at some point in their lives, or had diabetes. The odds ratios were 3.56, 1.99, and 0.49 with baselines, respectively, of income

greater than $20,000, non-smokers, and women without diabetes. The study suggests health professionals in hospitals should target these groups to help educate and test non-adhering groups of women while they are hospitalized.

This thesis research simultaneously considers a broader range of behavioral, demographic and health-related factors that can predict U.S. women's nonadherence to current mammography guidelines. The large data set also allows for precise results when separating models into five different decadal age groups. Our analysis will also reinforce results found in previous studies using a current, large set of data.

3. Methodology

3.1 Data Source

We use the 2014 Behavioral Risk Factor Surveillance System (BRFSS) data for this study. Every year the CDC's Population Health Surveillance Branch works with the U.S. state health departments and territories to form the BRFSS. The goal of this organization is to collect uniform behavioral and demographic information on Americans in all fifty states as well as other territories. Data is collected via landline and cell phone surveys. The phone numbers are selected at random and the resulting sample must meet a certain criteria established by the BRFSS to ensure the sample is a fair representation. Every area participating in the survey met the criteria in 2014 [17]. The resulting data represents a large, diverse, and up-to-date set of the American population, which is why this data was selected.

Formatted: Font: 12 pt

3.2 Data Processing

Figure 1 shows the study design as well as the sample sizes used during each step of the data processing. First, the raw data file, extracted from [18], was loaded into a Microsoft Excel file in order to filter and obtain desired information. The target population of the study was~~is~~ women ages 40 and above (i.e., the earliest age a woman is recommended for a mammogram), so all males as well as females under the age of 40 were eliminated from the data set. Next the predicting variables were selected which is discussed in more detail in the Section 3.3.

The list-wise deletion method was implemented to remove missing data. Since the regression predicts mammography screening behavior based on individual characteristics, using other methods such as the nearest neighbor technique may produce inaccurate results. We did not identify patterns of missing data, so list-wise deletion sufficed. The survey responses that warranted removal were “Refuse”, and “Don’t Know/Not Sure”. A majority of the time the response of “Not asked or Missing” was removed unless the response gave insight about the question at hand. For example, when asked “During the past 30 days, for about how many days did poor physical or mental health keep you from doing your usual activities, such as self-care, work, or recreation?”, the response of “Not asked or Missing” meant the respondent did not have any poor physical or mental health problems in the last 30 days based on two previous survey questions. In this case the responses were treated as non-missing data and were kept in the data set. The one instance of keeping missing data occurred in the “Health Coverage” variable which was one of the main variables of interest. “Not asked or Missing” represents uncertainty in the insurance status, and this category accounts for over 30% of study population. Thus, the response was kept as a category of its own.

Formatted: Font: 12 pt

3.3 Predictor Selection

The full data set had roughly 270 factors, so a large number of variables had to be eliminated in order to reduce the dependency between the explanatory variables and to create more meaningful, concise results. One objective was to use as much raw data as possible as opposed to using the imputed data. The only circumstances where this was unavoidable was with the "Age Group" and "Race" variables since they had a large amount of missing data. Several questions in BRFSS's survey were secondary, or follow-up questions. These variables were removed if they represented a very small percentage of the overall sample. We categorized three groups of predictors that are most clearly related to mammogram behavior: demographic, health-related, and behavioral characteristic. Predictors were selected if they helped achieve the goal of the analysis: that is if they fell into one of the groups of interest (demographic, health-related, and behavioral). The final variables, as well as the level of each variable, can be found in Table 1. A few predicting variables were combined to reduce the size of the regression and to consolidate similar survey questions. The *Chronic Condition* variable, for instance, equals one if a woman states that she has one of the several major chronic conditions, including coronary heart disease, COPD, emphysema, chronic bronchitis, kidney disease and diabetes. Other variables needed to have combined or modified categories. For example, the *Average Sleep Time* variable combined sleep times to form three categories: less than seven hours, seven to nine hours, and greater than nine hours. A more detailed description of combined variables and corresponding survey questions can be found in Appendix A.

Formatted: Font: 12 pt

Chi-squared tests of independence were first performed on all of the independent variables to determine if there was any relationship between the response and explanatory variables. The results (all $p < 0.01$; results not shown) indicate that all of the variables are associated with a woman's mammogram screening behavior, and are kept in the model. Note that this may be due to the extremely large sample size.

3.4 Data Splitting

Once the appropriate missing data were~~as~~ removed the resulting data were~~as~~ split into two sets, one of which was used to run the logistic regression (training) and the other (testing) to validate the regression. 80% of the data was randomly placed in a training file and the remaining 20% was out into a testing file. After the validation process the two sets were combined back together. The full set was then separated by age group so the model could be stratified by age. That is, a regression was created by only considering one age group at a time.

3.5 Logistic Regression

A multiple logistic regression is a classification model that tries to predict the outcome of a binomial dependent (indicator) variable with multiple independent variables [19]. The logic function that the regression is based on can be found in the equation below.

$$\log(\pi(x)/(1-\pi(x)) = \alpha + \beta * \mathrm{DM} + \gamma * HR + \delta * BH,$$

where $\pi(x)$ is the probability the response variable equals 1, indicating nonadherence (i.e., a woman's last mammogram was more than 2 years ago); x is a vector containing all predictors;

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

DM, HR and BH are vectors of demographic, health-related and behavioral categories, and α, β, γ, and δ are the coefficient vectors of parameter estimates.

The equation above reports coefficient estimates (log-of odds ratios) as the coefficients of the independent categories, but often times odds ratios are reported in medical research [20]. Odds ratios are the exponentiation of the log-of odds ratios. In this analysis the odds ratios are interpreted as the multiplicative relationship between the baseline category of a variable and another category of the same variable. As Figure 1 shows, the same regression is performed for each individual age group as well (removing age as an independent variable). All analyses were conducted with R Version 3.2.5. All code used can be found in Appendix B.

Formatted: Font: 12 pt

3.6 Model Checking

In order to check for the extent of multicollinearity between the independent variables, variable inflation factor (VIF) was used. The most stringent literature considers a VIF above 4 or 5 is an indicator that there is a problem of multicollinearity [21]. The VIFs for all variables range from 1.02 to 2.09 with the exception of the *Yearly Household Income, Employment Status,* and the *Age Group* categories. The resulting VIFs are 2.75, 3.77, and 4.17, respectively, which implies that there might be an issue of multicollinearity. Since the VIF for *Age Group* is greater than 4, it gave us more motivation to stratify by age group so as to eliminate the multicollinearity problem while examining the effect of age.

Formatted: Font: 12 pt

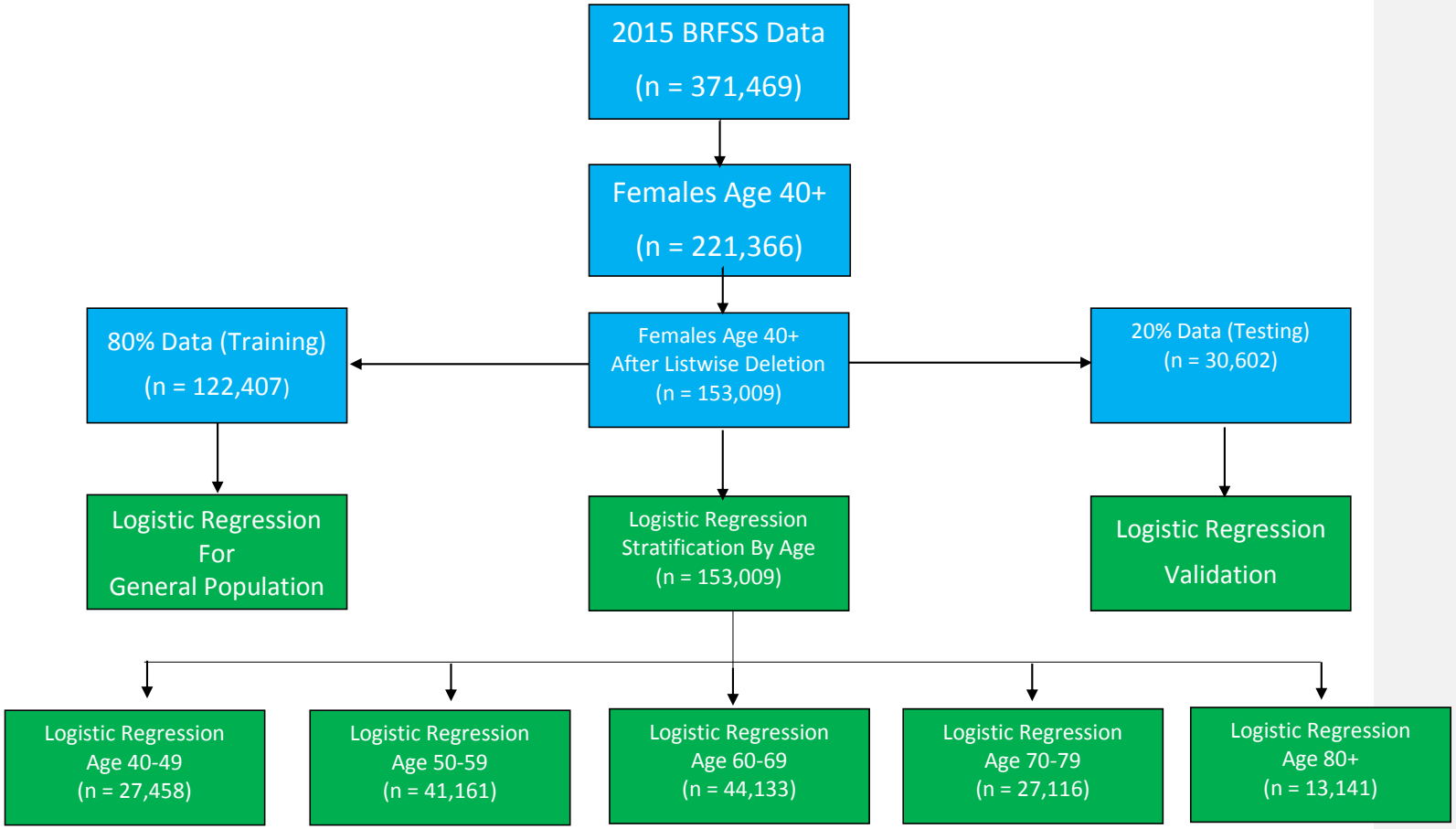

Figure 1: Study Design and Sample Sizes

4. Results

In this section, descriptive statistics for the study population are first summarized using frequency tables and visual representations of the data. Next the association between the predictors and the nonadherence to mammogram screening examined using a logistic regression model is shown. Odds ratios and 95 percent confidence intervals are reported for every category of each variable. After validation, the regression results with age stratification are presented.

4.1 Descriptive Statistics for the Study Population

The count and percentage for each category of the independent variables are presented in Table 1 (column 3 and 4, respectively). Madadi et al. [13] concluded that higher income is correlated to greater mammogram screening adherence. Their study also revealed that women with health insurance are far more likely to adhere. The graphs below helped gather a general understanding of our data when compared to the previous study's results as well as the effect age has on mammogram practices. This initial analysis attempts to compare the most distinct groups by showing the percentage of women who have never had a mammogram next to the percentage of women who had a mammogram within a year of the survey.

Figure 2 shows the resulting percentages based on the women's age groups. For women between the ages of 40-49, the percentage of women who have never had a mammogram is significantly higher than those who had one within the last year. An inverse relationship occurs for women between the ages of 50-79. Figure 3 shows the frequency graphs based on women's household income category. As the income increases, the percent of women that have never

had a mammogram slightly increases. Similarly, as the income group increases the percent of women who had a mammogram within the last year greatly increases. The percent of women who had a mammogram within the last year increases with income level at a far greater rate than the women who have never been screened. It is interesting to observe from Figure 4 that of the women who have Medicaid as their main source of health coverage, a much larger percent of women fell into this category that had a mammogram in the last year when compared to those who have never had one. Employer-paid coverage has the opposite results.

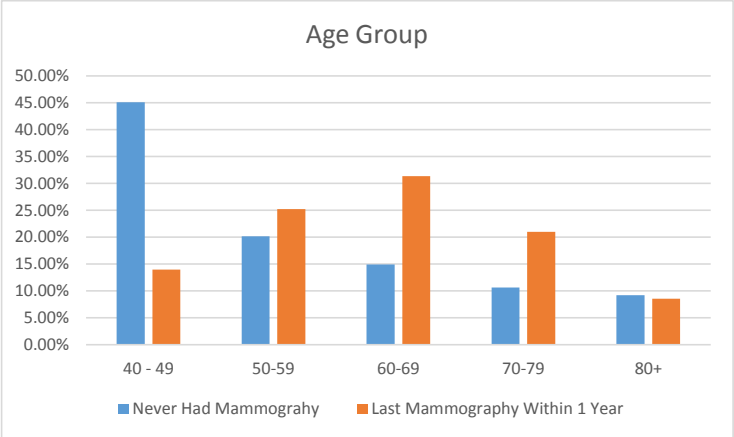

Figure 2: Age Group Mammography Percentages

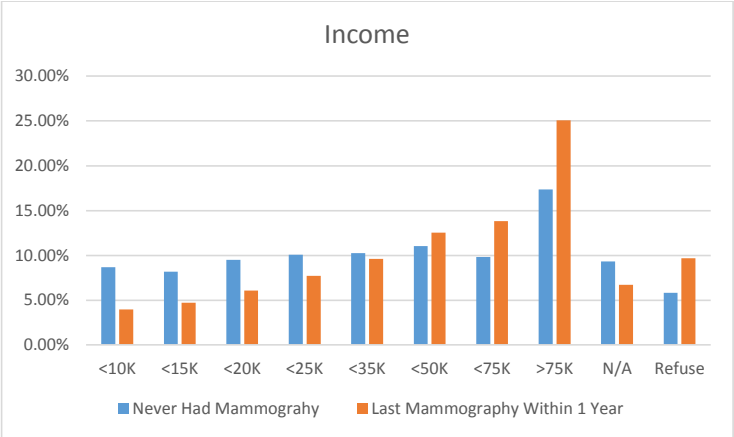

Figure 3: Income Mammography Percentages

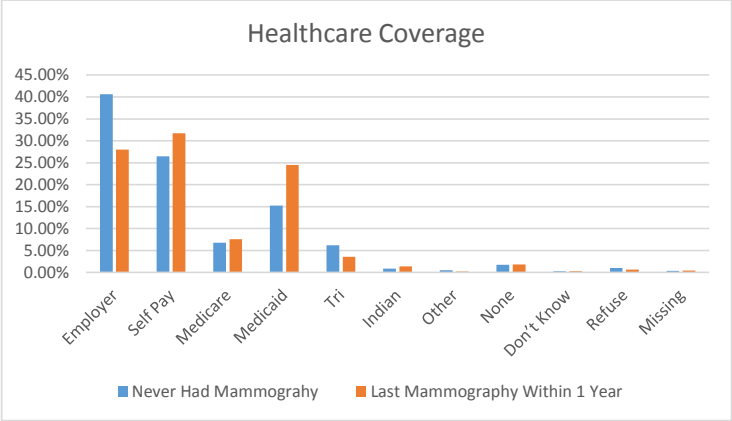

Figure 4: Healthcare Coverage Mammography Percentages

4.2 Logistic Regression Results for the General Population

Table 1 shows all of the odds ratios (ORs) and 95 percent confident intervals (Cis) as well as the reference categories for each variable. Numbers are bolded to highlight significant results discussed later in this section. Reference categories were first selected if a category represented an ideal condition of the population (e.g. women that fell in the "Optimal" category of the *BMI* variable were set as reference category). If a variable did not warrant an ideal condition the category with the largest sample size was chosen for the reference category (e.g. white women represented the largest category, therefore the category was selected as the reference). If there was not a definitively large category, ~~then~~ the category that ~~has an~~ made for the easiest~~y~~ interpretation was chosen.

Table 1: Study Population and Logistic Regression Results

Variables	Categories	Count	% of Total	OR	95% CI
Demographic Factors					
Marital Status	Married	80404	52.5%	--	--
	Divorced	26566	17.4%	1.10	(1.05,1.16)
	Widowed	29298	19.1%	1.10	(1.04,1.15)
	Separated	3049	2.0%	1.09	(0.98,1.22)
	Never Married	11328	7.4%	1.10	(1.03,1.17)
	Unmarried Couple	2364	1.5%	1.13	(1,1.28)
Number of Children in Household	None	123796	80.9%	--	--
	1 Child	14058	9.2%	1.16	(1.1,1.23)
	2 Children	10063	6.6%	**1.41**	(1.32,1.51)
	> 2 Children	5092	3.3%	**1.70**	(1.56,1.86)
Highest Education	Never Attended	110	0.1%	1.02	(0.56,1.86)
	Grades 1 – 8	2808	1.8%	0.97	(0.86,1.09)
	Grades 9 – 11	6784	4.4%	1.02	(0.94,1.11)
	Grade 12 or GED	42490	27.8%	0.93	(0.89,0.97)
	1 - Years College	44011	28.7%	1.02	(0.97,1.06)
	4 or More Years College	56806	37.1%	--	--

Employment Status	Employed for Wages	58246	38.0%	--	--
	Self-Employed	10120	6.6%	1.16	(1.09,1.24)
	Out of Work > 1 Year	3226	2.1%	1.12	(1,1.24)
	Out of Work < 1 Year	2243	1.5%	1.11	(0.98,1.26)
	Homemaker	13076	8.5%	0.97	(0.91,1.03)
	Student	593	0.4%	1.16	(0.92,1.46)
	Retired	52793	34.5%	0.81	(0.77,0.85)
	Unable to Work	12712	8.3%	0.89	(0.83,0.96)
Yearly Household Income Level	< $10,000	7874	5.1%	**1.20**	(1.1,1.32)
	$10,000 - $14,999	9625	6.3%	**1.19**	(1.1,1.29)
	$15,000 - $19,999	12327	8.1%	**1.25**	(1.16,1.34)
	$20,000 - $24,999	15170	9.9%	1.20	(1.12,1.28)
	$25,000 - $34,999	18068	11.8%	1.15	(1.08,1.22)
	$35,000 - $49,999	22919	15.0%	1.09	(1.03,1.15)
	$50,000 - $74,999	24111	15.7%	1.05	(1,1.11)
	> $75,000	42915	28.0%	--	--
Health Coverage Type	Missing	47440	31.0%	1.09	(1.04,1.13)
	Employer Paid	48866	31.9%	--	--
	Family Self-Paid	11555	7.5%	1.10	(1.03,1.17)
	Medicare	34015	22.2%	0.98	(0.93,1.03)
	Medicaid	5867	3.8%	**0.89**	(0.81,0.97)
	Other	4952	3.2%	**0.90**	(0.82,0.99)
	None	314	0.2%	1.02	(0.73,1.43)
Age Group	40-49	27458	17.9%	--	--
	50-59	41161	26.9%	**0.55**	(0.52,0.58)
	60-69	44133	28.8%	**0.48**	(0.45,0.51)
	70-79	27116	17.7%	**0.45**	(0.42,0.48)
	80+	13141	8.6%	**0.82**	(0.75,0.89)
Ethnicity	White	125093	81.7%	--	--
	Black	11904	7.8%	**0.71**	(0.67,0.76)
	Asian	1962	1.3%	**0.81**	(0.7,0.93)
	American Indian/Alaskan Native	2104	1.4%	0.98	(0.87,1.12)
	Hispanic	8598	5.6%	**0.73**	(0.68,0.79)
	Other	3348	2.2%	1.04	(0.94,1.15)
Health-Related Factors					
Poor Health	0 Days	45723	29.9%	--	--
	0 Days with Reported Physical or Mental Health	70805	46.2%	1.02	(0.99,1.06)
	1 - 10 Days	21231	13.9%	1.05	(1,1.1)
	11 - 20 Days	6362	4.2%	0.98	(0.91,1.07)
	21 - 30 Days	8888	5.8%	0.97	(0.9,1.05)
Chronic Condition	Absent	128621	84.0%	--	--
	Present	12356	8.1%	1.04	(1.01,1.08)
Number of Personal Doctors	1 Doctor	12032	7.9%	--	--
	> 1 Doctor	89776	58.6%	0.98	(0.93,1.04)
	None	63233	41.3%	**1.57**	(1.48,1.66)
BMI Category	< 18 (Underweight)	2608	1.7%	**1.43**	(1.27,1.6)
	18 - 24.9 (Optimal)	43790	28.6%	--	--
	25 - 29.9 (Overweight)	48562	31.7%	0.92	(0.89,0.96)
	30 - 39.9 (Obese)	43878	28.7%	0.90	(0.86,0.94)

	40 + (Extremely Obese)	14171	9.3%	0.93	(0.87,0.98)
Difficulty Doing Things Alone	Yes	13762	9.0%	1.29	(1.21,1.37)
	No	139247	91.0%	--	--
Behavioral Factors					
Last Routine Checkup	< 1 Year	122052	79.7%	--	--
	1 - 2 Years	16000	10.5%	1.10	(1.05,1.16)
	2 - 5 Years	7457	4.9%	**2.74**	(2.56,2.92)
	> 5 Years	6734	4.4%	**3.43**	(3.17,3.7)
	Never	766	0.5%	2.22	(1.83,2.7)
Any Exercise in Last Month	> 0 Days	113658	74.2%	--	--
	Never	39351	25.7%	1.04	(1,1.08)
Average Sleep Time	7-9 Hours	101080	66.0%	--	--
	> 9 Hours	46910	30.6%	1.02	(0.99,1.06)
	< 7 Hours	5019	3.3%	1.11	(1.02,1.21)
Last Dentist Visit	< 1 Year	109990	71.8%	--	--
	1 - 2 Years	14349	9.4%	1.38	(1.31,1.45)
	2 - 5 Years	12712	8.3%	**1.68**	(1.59,1.77)
	> 5 Years/Never	15958	10.4%	**1.69**	(1.61,1.78)
Smoke Tobacco	Daily	14982	9.8%	**1.38**	(1.31,1.45)
	Sometimes	5849	3.8%	1.24	(1.15,1.34)
	Never	132178	86.3%	--	--
Use Chewing Tobacco or Snuff	Daily	409	0.3%	**1.38**	(1.05,1.82)
	Sometimes	694	0.5%	0.89	(0.71,1.12)
	Never	151906	99.2%	--	--
Drinking Level	Does Not Drink/Not at Risk	82410	53.8%	1.06	(1.03,1.1)
	Drink Problem	57832	37.8%	1.04	(0.97,1.1)
	At Risk	12767	8.3%	--	--
Last Flu Shot	< 1 Year	79680	52.0%	--	--
	> 1 Year	73329	47.9%	**1.53**	(1.48,1.58)
Last Breast Exam	< 1 Year	90018	58.8%	--	--
	Never	9973	6.5%	**4.27**	(4.02,4.54)
	> 1 Year	53018	34.6%	**4.60**	(4.44,4.77)
Last Pap Test	< 3 Years	98722	64.5%	--	--
	> 3 Years	49568	32.4%	**3.28**	(3.16,3.41)
	Never	4719	3.1%	**3.83**	(3.53,4.17)
Ever Had Hysterectomy	Yes	52773	34.5%	--	--
	No	100236	65.5%	**1.84**	(1.77,1.91)

4.2.1 Demographic Predictors

When compared to women who have no children in their house, women with one, two, or more than two children reported ORs of 1.16 (CI: 1.1–1.23), 1.41 (CI: 1.32-1.51) , and 1.7 (CI: 1.56-1.86) respectively. Categories below $25,000 reported odds ratios between 1.19-1.25 and

categories between $25,000 and $74,999 reported odds ratios between 1.05-1.15. The *Age Group* variable produced the lowest ORs of all the demographic factors. All categories reported odds ratios less that one, with 50-59 resulting in an OR of 0.55 (CI: 0.52-0.58), 60-69 resulting in an OR of 0.48 (CI: 0.45-0.51), 70-79 with the lowest OR of 0.45 (CI: 0.42-0.48) and 80 and above spiking to 0.82 (CI: 0.75-0.89). Surprising results occurred in the *Ethnicity* variable. With *Whites* as the reference category, African-American women reported an OR of 0.71 (CI: 0.67-0.76). Similarly, Hispanics were less likely not to have received a mammogram in the last 2 years with an OR of 0.73 (CI: 0.68-.079).

Women that are self-employed, out of work or are students less likely to participate in a mammogram screening, while women who are retired or are unable to work are more likely to undergo a screening. Based on the overall analysis, the type of health insurance also has a slight effect on mammogram non-adherence. Government funded programs such as Medicare, Medicaid and others (TRICARE, VA, Military, Alaska Native, Indian Health Service, and Tribal Health Services) all reflect less non-adherence with ORs of 0.98 (CI: 0.93-1.03), 0.89 (CI: 0.81-0.97), and 0.90 (CI: 0.82-0.99) respectively. When considering all women above the age of 40 a woman's marital status and highest level of education seems to have a minimal effect when predicting mammogram non-adherence.

4.2.2 Health-Related Predictors

In general, health-related factors seem to have less predicting powers than the other factors. The presence of a chronic condition and self-prescribed poor health do not have an effect on non-adherence to mammogram screening. However, women who reported having trouble doing activities on their own are less likely to adhere than those who reported having no

difficulties. Underweight (BMI < 18) women also had a high OR of 1.43 (CI: 1.27-1.6) when compare to women at an optimal weight. A BMI over 25 also indicates that non-adherence is much less likely with ORs around 0.90 for the three categories. The most significant results came from the *Number of Personal Doctors* variable. If a woman does not have a personal doctor or healthcare provider she is 1.57 times more likely not to adhere to screening guidelines than those having one doctor.

4.2.3 Behavioral Predictors

Several of the behavioral variables produced the strongest results in the logistic regression. The largest ORs in the model showed up in the *Last Breast Exam* variable. Women who have never attended a breast examination were 4.27 (CI: 4.02-4.54) times more likely not to have gotten a mammogram in the last two years of the survey. Women that had their last breast exam more than a year past the survey had the highest OR of 4.6 (OR: 4.44-4.77). Similar results occurred in the *Last Pap Test* variable. Woman that have never had a Pap test reported an OR of 3.83 (CI: 3.53-4.17) and those who had their last Pap test more than three years since the survey reported an OR of 3.28 (CI: 3.16-3.41). Women that have never had a hysterectomy were 1.84 times less likely to have had a mammogram within two years before the survey. Flu shot behavior had a less significant, but notable relationship. An OR of 1.53 (OR: 1.48-1.58) resulted for women who have never had a flu shot, or did not have one within a year of the survey.

Attending routine medical activities seems to be related to mammogram screening practices as well. The *Last Routine Checkup* and *Last Dentist Visit* produces comparable results to each other. With less than one year as the reference categories, the longer the amount of time since

the last appointment the larger the OR became. However, *Last Routine Checkup* is a stronger predictor for mammogram non-adherence with several odds ratios greater than two.

Tobacco use variables also reveal a relationship to screening practices. Daily smokers and daily chewing tobacco both report ORs of 1.38 with CIs of 1.31-1.45 and 1.05-1.82 respectively. Women who sleep less than seven hours per night tend not to adhere with an OR of 1.11 (CI: 1.02-1.21) for the category. The remaining variables have results close to the reference category's odds ratio of one. Drinking habits seem to have very little predicting power at any level, though it does appear that drinkers are slightly more likely not to adhere. Self-reported exercise habits and average sleep time per night also have minimal effects.

4.3 Model Validation

Using the test data, a Receiver-Operating Characteristic (ROC) curve was created to plot the sensitivity against one minus specificity for all points. In this study the sensitivity is the probability of correctly identifying a woman receiving a mammogram in the last 2 years, and the specificity refers to the probability of correctly identifying a woman that did not receive a mammogram in the last 2 years. Figure 5 shows a graph of the resulting ROC curve. When calculating the area under the curve (AUC), a value of 1 represents the model has a perfect predictive power and a value of 0.5 represents the model having no predictive power. The area under an ROC curve combines the effects of sensitivity and one minus the specificity to obtain the validity of a test [22]. The AUC in our validation came out to be 0.837, which indicates that the logistic regression is a good predictor [23] for mammogram non-adherence.

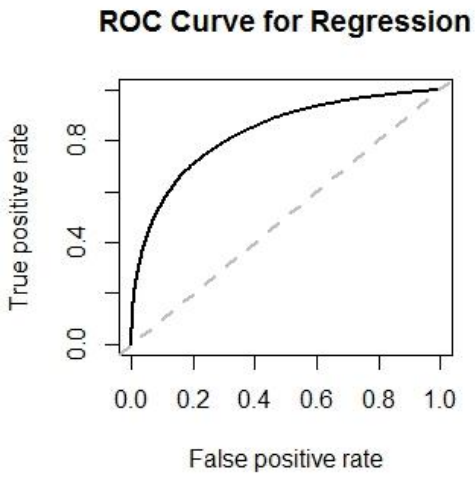

Figure 5: ROC Curve of General Population Logistic Regression

4.4 Stratification Results

The variable inflation factors indicate that there may be an issue with multicollinearity, particularly in the age group. This fact in correspondence with surprisingly low ORs in the age group category led us to perform a stratified logistic regression on age. Table 2 below shows the resulting ORs and confidence intervals. Again, all bolded numbers highlight significant results discussed in this section.

Table 2: Stratified Logistic Regression Results

		Age 40 -49		Age 50-59		Age 60-69		Age 70-79		Age 80+	
Variables	**Categories**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Demographic Factors											
Marital Status	Married	--	--	--	--	--	--	--	--	--	--
	Divorced	1.01	(0.92,1.11)	**1.10**	(1.01,1.2)	**1.11**	(1.02,1.2)	**1.16**	(1.04,1.29)	1.08	(0.9,1.29)
	Widowed	0.97	(0.78,1.21)	1.08	(0.95,1.22)	1.07	(0.98,1.16)	1.20	(1.1,1.3)	1.09	(0.98,1.22)
	Separated	**1.24**	(1.05,1.46)	1.13	(0.95,1.34)	0.97	(0.78,1.22)	1.17	(0.8,1.73)	1.16	(0.56,2.4)
	Never Married	**1.14**	(1.02,1.27)	**1.12**	(1.01,1.25)	**1.14**	(1.01,1.29)	**1.19**	(1,1.41)	0.99	(0.76,1.28)
	Unmarried Couple	**1.25**	(1.04,1.49)	1.01	(0.83,1.23)	1.18	(0.92,1.51)	1.23	(0.79,1.93)	0.53	(0.2,1.45)
Number of Children in Household	None	--	--	--	--	--	--	--	--	--	--
	1 Child	**1.28**	(1.18,1.39)	1.08	(0.99,1.18)	1.14	(0.98,1.33)	1.09	(0.86,1.38)	1.06	(0.68,1.64)
	2 Children	**1.49**	(1.37,1.61)	**1.17**	(1.03,1.33)	1.17	(0.93,1.48)	1.04	(0.7,1.55)	1.24	(0.63,2.47)
	> 2 Children	**1.89**	(1.72,2.09)	**1.42**	(1.18,1.72)	1.06	(0.79,1.43)	1.52	(0.88,2.6)	1.22	(0.52,2.84)
Highest Education	Never Attended	1.15	(0.44,2.96)	2.75	(0.92,8.17)	0.57	(0.16,1.98)	1.03	(0.34,3.1)	0.81	(0.22,3.07)
	Grades 1 - 8	0.96	(0.74,1.24)	0.90	(0.7,1.15)	0.85	(0.68,1.07)	0.82	(0.65,1.03)	1.11	(0.87,1.41)
	Grades 9 - 11	1.00	(0.85,1.18)	1.10	(0.95,1.28)	0.90	(0.77,1.05)	0.94	(0.8,1.1)	1.03	(0.86,1.25)
	Grade 12 or GED	0.99	(0.91,1.09)	1.00	(0.92,1.08)	0.85	(0.78,0.92)	0.85	(0.77,0.93)	1.00	(0.89,1.13)
	1 – 2 Years College	1.12	(1.04,1.21)	1.07	(1,1.16)	0.94	(0.87,1.02)	0.85	(0.77,0.94)	1.03	(0.91,1.17)
	4 or More Years College	--	--	--	--	--	--	--	--	--	--
Employment Status	Employed for Wages	--	--	--	--	--	--	--	--	--	--
	Self-Employed	**1.13**	(1.02,1.26)	**1.23**	(1.11,1.37)	**1.27**	(1.13,1.44)	1.17	(0.94,1.44)	1.00	(0.64,1.56)
	Out of Work > 1 Year	**1.13**	(0.94,1.36)	**1.17**	(1,1.36)	1.13	(0.92,1.38)	1.31	(0.89,1.92)	1.30	(0.64,2.62)
	Out of Work < 1 Year	**1.19**	(0.99,1.43)	1.07	(0.89,1.29)	1.11	(0.85,1.45)	1.08	(0.63,1.84)	0.76	(0.28,2.04)
	Homemaker	**1.11**	(1.01,1.22)	1.10	(0.99,1.24)	1.02	(0.9,1.16)	0.96	(0.81,1.13)	0.90	(0.67,1.21)
	Student	**1.32**	(1.02,1.71)	1.31	(0.87,1.96)	0.73	(0.34,1.54)	1.20	(0.42,3.44)	0.82	(0.13,5.3)
	Retired	0.61	(0.41,0.92)	0.78	(0.68,0.89)	0.84	(0.78,0.91)	0.89	(0.78,1.01)	0.90	(0.69,1.19)
	Unable to Work	1.03	(0.88,1.2)	0.93	(0.82,1.05)	0.86	(0.76,0.97)	1.21	(0.98,1.49)	1.13	(0.79,1.61)
Yearly Household Income Level	< $10,000	1.12	(0.94,1.34)	1.09	(0.93,1.28)	1.31	(1.11,1.55)	1.14	(0.93,1.4)	**1.36**	(1.04,1.78)
	$10,000 - $14,999	1.15	(0.96,1.38)	1.00	(0.85,1.17)	1.33	(1.15,1.55)	1.08	(0.9,1.28)	**1.33**	(1.06,1.67)
	$15,000 - $19,999	**1.29**	(1.1,1.51)	**1.29**	(1.12,1.49)	**1.34**	(1.17,1.53)	1.12	(0.95,1.31)	1.21	(0.97,1.49)
	$20,000 - $24,999	1.31	(1.14,1.5)	1.20	(1.05,1.36)	**1.30**	(1.16,1.47)	1.00	(0.86,1.16)	1.22	(0.99,1.49)
	$25,000 - $34,999	1.22	(1.07,1.38)	1.19	(1.06,1.34)	1.15	(1.03,1.28)	0.97	(0.84,1.12)	1.14	(0.93,1.39)
	$35,000 - $49,999	1.16	(1.05,1.29)	1.17	(1.06,1.29)	1.06	(0.96,1.17)	0.95	(0.83,1.09)	1.15	(0.94,1.4)

	$50,000 - $74,999	1.11	(1.01,1.21)	1.08	(0.98,1.18)	1.04	(0.95,1.15)	0.85	(0.73,0.98)	1.01	(0.8,1.26)
	> $75,000	--	--	--	--	--	--	--	--	--	--
Health Coverage Type	Missing	1.16	(1.08,1.24)	1.17	(1.09,1.26)	1.06	(0.97,1.14)	1.03	(0.9,1.18)	0.89	(0.75,1.06)
	Employer Paid	1.00	--	1.00	--	1.00	--	1.00	--	1.00	--
	Family Self-Paid	1.10	(0.96,1.26)	1.13	(1.01,1.27)	1.01	(0.9,1.14)	1.04	(0.88,1.24)	0.99	(0.81,1.21)
	Medicare	0.85	(0.71,1.03)	0.91	(0.79,1.06)	0.87	(0.8,0.95)	1.00	(0.88,1.14)	0.91	(0.77,1.07)
	Medicaid	0.90	(0.78,1.04)	0.84	(0.72,0.97)	0.91	(0.77,1.08)	1.11	(0.84,1.46)	1.10	(0.77,1.58)
	Other	0.90	(0.75,1.09)	0.96	(0.81,1.14)	0.98	(0.83,1.15)	0.81	(0.64,1.03)	0.80	(0.6,1.07)
	None	0.90	(0.32,2.53)	0.87	(0.44,1.71)	0.86	(0.49,1.53)	0.89	(0.48,1.66)	1.61	(0.75,3.44)
Ethnicity	White	--	--	--	--	--	--	--	--	--	--
	Black	**0.84**	(0.75,0.94)	**0.68**	(0.6,0.76)	**0.63**	(0.56,0.71)	**0.62**	(0.53,0.73)	**0.78**	(0.63,0.96)
	Asian	0.97	(0.79,1.19)	**0.65**	(0.5,0.86)	**0.60**	(0.44,0.81)	0.78	(0.54,1.12)	1.12	(0.76,1.66)
	American Indian/Alaskan Native	1.10	(0.89,1.36)	0.85	(0.69,1.05)	0.91	(0.72,1.16)	0.95	(0.69,1.32)	1.37	(0.84,2.26)
	Hispanic	**0.77**	(0.68,0.86)	**0.76**	(0.67,0.86)	**0.63**	(0.54,0.73)	0.93	(0.78,1.11)	**0.63**	(0.49,0.82)
	Other	0.90	(0.76,1.06)	0.86	(0.71,1.03)	1.08	(0.89,1.3)	1.07	(0.85,1.36)	1.38	(0.99,1.93)
Health-Related Factors											
Poor Health	0 Days	--	--	--	--	--	--	--	--	--	--
	0 Days with Reported Physical or Mental Health	1.03	(0.96,1.1)	1.05	(0.98,1.13)	1.01	(0.94,1.08)	1.06	(0.98,1.15)	1.00	(0.91,1.1)
	1 - 10 Days	1.03	(0.95,1.13)	1.12	(1.02,1.22)	1.01	(0.92,1.11)	1.02	(0.91,1.16)	1.07	(0.92,1.24)
	11 - 20 Days	1.02	(0.87,1.2)	1.03	(0.89,1.19)	0.92	(0.8,1.07)	1.31	(1.1,1.57)	0.86	(0.68,1.08)
	21 - 30 Days	1.01	(0.86,1.2)	0.98	(0.85,1.12)	1.09	(0.95,1.24)	0.94	(0.8,1.1)	0.90	(0.75,1.09)
Chronic Condition	Absent	--	--	--	--	--	--	--	--	--	--
	Present	1.04	(0.97,1.12)	1.05	(0.98,1.12)	1.02	(0.96,1.09)	1.08	(1,1.16)	1.05	(0.96,1.14)
Number of Personal Doctors	1 Doctor	--	--	--	--	--	--	--	--	--	--
	> 1 Doctor	0.95	(0.85,1.06)	0.98	(0.88,1.09)	1.05	(0.94,1.17)	0.97	(0.86,1.09)	0.97	(0.85,1.12)
	None	**1.49**	(1.36,1.63)	**1.58**	(1.44,1.73)	**1.56**	(1.4,1.74)	**1.63**	(1.38,1.92)	**1.41**	(1.12,1.76)
BMI Category	< 18 (Underweight)	1.15	(0.89,1.48)	**1.39**	(1.11,1.74)	**1.38**	(1.11,1.72)	**1.52**	(1.21,1.9)	**1.52**	(1.21,1.91)
	18 - 24.9 (Optimal)	--	--	--	--	--	--	--	--	--	--
	25 - 29.9 (Overweight)	1.01	(0.94,1.09)	0.91	(0.85,0.99)	0.93	(0.87,1.01)	0.85	(0.77,0.92)	0.90	(0.81,0.99)
	30 - 39.9 (Obese)	1.01	(0.93,1.09)	0.90	(0.83,0.97)	0.89	(0.82,0.96)	0.82	(0.75,0.9)	0.86	(0.76,0.96)
	40 + (Extremely Obese)	1.06	(0.96,1.18)	0.89	(0.8,0.98)	0.94	(0.84,1.04)	0.84	(0.72,0.97)	0.86	(0.65,1.13)
Difficulty Doing Things Alone	Yes	**1.17**	(1.02,1.36)	**1.24**	(1.1,1.39)	**1.18**	(1.06,1.32)	**1.31**	(1.16,1.48)	**1.48**	(1.32,1.66)
	No	--	--	--	--	--	--	--	--	--	--

Behavioral Factors											
Last Routine Checkup	< 1 Year	--	--	--	--	--	--	--	--	--	--
	1 - 2 Years	1.14	(1.05,1.24)	1.05	(0.97, 1.13)	1.13	(1.04,1.23)	1.11	(0.99,1.25)	1.10	(0.93,1.28)
	2 - 5 Years	**2.56**	(2.28,2.87)	**2.97**	(2.67, 3.3)	**3.11**	(2.76,3.49)	**2.42**	(2.03,2.88)	**1.74**	(1.36,2.22)
	> 5 Years	**2.59**	(2.25,2.99)	**3.27**	(2.88, 3.71)	**4.21**	(3.67,4.84)	**4.31**	(3.54,5.25)	**2.22**	(1.7,2.92)
	Never	**1.50**	(1.05,2.16)	**1.80**	(1.27, 2.56)	**2.71**	(1.93,3.8)	**2.73**	(1.78,4.17)	**2.12**	(1.18,3.8)
Any Exercise in Last Month	> 0 Days	--	--	--	--	--	--	--	--	--	--
	Never	0.94	(0.87,1.01)	1.00	(0.93, 1.07)	1.08	(1.01,1.15)	1.18	(1.09,1.27)	1.07	(0.98,1.16)
Average Sleep Time	7-9 Hours	--	--	--	--	--	--	--	--	--	--
	> 9 Hours	1.04	(0.98,1.11)	1.05	(0.98, 1.12)	1.02	(0.96,1.09)	1.07	(0.99,1.15)	0.88	(0.8,0.97)
	< 7 Hours	1.17	(0.96,1.43)	1.10	(0.92, 1.31)	1.12	(0.96,1.31)	1.06	(0.9,1.25)	1.01	(0.86,1.2)
Last Dentist Visit	< 1 Year	--	--	--	--	--	--	--	--	--	--
	1 - 2 Years	**1.23**	(1.12,1.36)	**1.41**	(1.29, 1.55)	**1.40**	(1.28,1.54)	**1.43**	(1.27,1.6)	**1.38**	(1.2,1.59)
	2 - 5 Years	**1.39**	(1.25,1.54)	**1.83**	(1.67, 2.02)	**1.74**	(1.58,1.92)	**1.69**	(1.5,1.9)	**1.64**	(1.41,1.91)
	> 5 Years/Never	**1.51**	(1.34,1.7)	**1.77**	(1.6, 1.96)	**1.78**	(1.62,1.95)	**1.71**	(1.55,1.89)	**1.57**	(1.39,1.77)
Do You Smoke?	Daily	1.33	(1.21,1.46)	1.45	(1.33, 1.57)	1.37	(1.25,1.51)	1.52	(1.33,1.73)	1.25	(0.96,1.62)
	Sometimes	1.34	(1.17,1.55)	1.20	(1.06, 1.36)	1.33	(1.16,1.52)	1.23	(1.01,1.5)	1.58	(1.08,2.32)
	Never	--	--	--	--	--	--	--	--	--	--
Use Chewing Tobacco or Snuff	Daily	1.11	(0.7,1.75)	1.13	(0.69, 1.84)	1.09	(0.65,1.83)	2.15	(1.15,4.02)	2.05	(0.79,5.32)
	Sometimes	0.91	(0.6,1.39)	1.01	(0.68, 1.51)	0.71	(0.47,1.06)	0.65	(0.37,1.14)	1.55	(0.82,2.92)
	Never	--	--	--	--	--	--	--	--	--	--
Drinking Level	Does Not Drink	1.06	(1,1.14)	1.03	(0.96, 1.1)	1.09	(1.02,1.17)	1.10	(1.02,1.2)	1.01	(0.92,1.12)
	Drink Problem	1.08	(0.97,1.2)	1.01	(0.91, 1.13)	0.97	(0.87,1.09)	1.09	(0.94,1.26)	0.90	(0.74,1.09)
	At Risk	1.00	--	1.00	--	1.00	--	1.00	--	1.00	--
Last Flu Shot	< 1 Year	--	--	--	--	--	--	--	--	--	--
	> 1 Year	**1.29**	(1.21,1.37)	**1.48**	(1.39, 1.57)	**1.64**	(1.54,1.74)	**1.74**	(1.63,1.87)	**1.51**	(1.39,1.65)
Last Breast Exam	< 1 Year	--	--	--	--	--	--	--	--	--	--
	Never	**3.50**	(3.05,4.02)	**4.35**	(3.82, 4.94)	**4.44**	(3.95,4.98)	**4.80**	(4.28,5.39)	**4.44**	(3.92,5.04)
	> 1 Year	**3.38**	(3.16,3.62)	**4.94**	(4.62, 5.27)	**5.10**	(4.78,5.45)	**5.11**	(4.72,5.54)	**4.79**	(4.35,5.28)
Last Pap Test	< 3 Years	--	--	--	--	--	--	--	--	--	--
	> 3 Years	**3.10**	(2.83,3.4)	**3.81**	(3.56, 4.09)	**3.42**	(3.21,3.65)	**2.88**	(2.66,3.12)	**3.64**	(3.26,4.06)
	Never	**2.59**	(2.07,3.23)	**3.76**	(3.14, 4.51)	**3.64**	(3.08,4.31)	**3.88**	(3.32,4.54)	**4.35**	(3.71,5.11)
Ever Had Hysterectomy	Yes	--	--	--	--	--	--	--	--	--	--
	No	**1.79**	(1.64,1.95)	**2.06**	(1.92, 2.21)	**2.01**	(1.88,2.14)	**1.70**	(1.59,1.83)	**1.53**	(1.41,1.66)

4.4.1 Demographic Predictors

Divorced women, particularly between the ages of 50-79, tend to be non-adherent to mammography guidelines when compared to married women. All odds ratios for the three age groups report are above one with the lower bound confidence interval still above one. Women that have never been married, in all age groups except 80 and over, are also non-adherent. 40-49 year old women are far less likely to receive a mammogram every other year if they have children. If fact, the more children a woman age 40-49 has the more likely she is not to adhere. The reported ORs for one, two, and more than two children are 1.28 (CI: 1.18-1.39), 1.49 (CI: 1.37-1.61), and 1.89 (CI: 1.72-2.09) respectively. Women between the age of 50-59 report similar, but less significant, results.

The *Employment Status* variable reveals notable results, particularly in the younger age groups. Women in their 40's are less likely to get a mammogram screening if they are self-employed (OR = 1.13), out of work for less than one year (OR = 1.13) and more than one year (OR = 1.19), a homemaker (OR = 1.11), or a student (OR = 1.32) when compared to women that are employed for wages. Self-employed women between the ages of 50-59 and 60-69 also report odds ratios greater than one. Reported income levels has a very strong relationship with mammography screening nonadherence. With *greater than $75,000* as the baseline all categories below $25,000 report an OR greater than one across all age groups. For ages 40-49 and 50-59 women are likely not to adhere if they make up to $50,000 per year.

Health coverage type as well as a woman's highest level of education seems to have little predicting power for mammography guidelines. However, ethnicity has the strongest predicting power of the demographic predictors. All black women at every age group report an OR and

upper bound confidence interval below one. Similar results occur for Asian between the ages of 50-69. All odds ratios for Hispanic women are also below one.

4.4.2 Health-Related Predictors

The stratified logistic regression reveals the same results for self-reported *Poor Health* and *Chronic Conditions* as the overall regression. Neither variables seems to have much of an effect on screening practices. Women who do not have a personal health professional report an OR above one for every age group with the lowest OR of 1.41 (OR: 1.12-1.76) falling in the above 80 age group. BMI levels have more significant results than the general model. For age groups 50 and up, every underweight category has an OR greater than one, while the overweight, obese, and extremely obese levels have odds ratios less than one. Women that have difficulty doing things alone in each age group reports and odds ratio greater than one as well.

4.4.3 Behavioral Predictors

Women who have never gone or did not go to a routine checkup within two years of the survey are far more likely not to adhere to screening guidelines. This holds true for all age groups with odds ratios ranging from 1.5 all of the way up to 4.3. Other checkups and procedures have strong results as well. The *Last Breast Exam, Pap Test, Flu Shot, Dentist Visit* and *Ever Had Hysterectomy* variables present ORs higher than one for every age group. All confident intervals do not contain one within their range making the results even stronger. Smoking tobacco appears to have an effect on mammogram nonadherence for all age groups as well with all reported ORs greater than 1.2. Other substance-use variables do not have as much of an effect as smoking tobacco does.

5. Discussion

In this study we evaluate the association between various factors (demographic, health-related and behavioral) and women's nonadherence to mammography screening using the 2014 BRFSS data. With the recently updated ACS and SBI guidelines we define mammography nonadherence as a woman (age 40 and up) not having a mammography within two years before the survey. Based on this definition of nonadherence we find several factors that can help predict nonadherence to mammography screenings. Due to smaller sample sizes, the confidence intervals of stratified logistic regressions tend to be larger, so the results were less conclusive. That being said, the stratified regressions do help paint a clearer picture for certain variables.

For the general model, as well as certain age groups, the demographic factors that are the most distinct predictors are *Number of Children in Household*, *Yearly Household Income,* and *Ethnicity* variables. The greater number of children under the age of 18 in a household indicates that a woman is increasingly less likely to receive a mammogram within the last two years. The stratified model reveals that this is particularly relevant for women between the ages of 40 and 49. In general the lower the income level of a woman's household the lower the chances are that the woman had a mammogram within the last two years. The trend holds very well for all ages below 69. These results are consistent with findings in previous studies [13, 15, 16].

The general and stratified models also reveal that ethnicity has a large effect on mammography nonadherence. White women are less likely to follow guidelines when compared to African-American, Hispanic, and Asian women. These results are very surprising since they contradict the previously performed study by Calvocorresi [14]. Although it is possible that minority

mammography nonadherence has decreased over the years, the results pertaining to ethnicity may be biased. The imputed race variable in use changes missing ethnicity responses to match the most common race of the respondents' region [25]. The imputed data, as well as the large amount of removed data, may have skewed the results for ethnicity, so they should be interpreted with caution.

Our analyses show that health-related characteristics have a moderate effect on nonadherence. The general model shows that underweight women are more likely not to adhere and slightly suggests that women with higher BMIs are less likely to have non-adherent behavior. Stratifying by age groups reveals similar, but much stronger, results for women over the age of 50. Both models support the importance of having at least one personal doctor. Women of all age groups are about one and a half time less likely to adhere than women with one personal doctor. Both models also show a higher nonadherence odds ratio for women that have trouble doing activities alone such as visiting a doctor, or shopping. Women over the age of 70 have the highest ORs for this predictor. Neither model places predicting power on the presence or absence of a chronic condition, but this could be caused by combining too many variables.

The strongest results come from several of the behavioral predictors. The highest odds ratio in the stratified logistic regressions appears in the *Last Breast Exam* category with an OR of 5.11 (CI: 4.74-5.54). This OR shows that woman between the ages of 70 and 79 who did not have a breast exam in the last year are over 5 times more likely not to adhere to mammography screening guidelines than women who had one within the last year. Other age groups show less significant results, but are overall extremely strong predictors. Different screening practices also coincide with breast examination practices. Timing of the last Pap test, last flu shot, and

the performance of a hysterectomy are all very good indictors of a woman's mammogram practices. Regardless of a woman's age, the longer it has been since her last routine checkup the more likely she is not to have a mammogram. These results confirm the study performed by Schueler et al. [12] Dental visits are not as strong of an indicator as general checkups, but it can still be used to predict whether or not a woman will adhere to mammography guidelines. Daily smokers, particularly in women under the age of 80 also seem to be more non-adherent than nonsmokers. Schueler et al. [12] found stronger, but similar results.

5.1 Limitations and Future Studies

This study has several limitations. Firstly, survey data is self-reported and may be biased if a respondent gave false information or omitted certain questions. Around 30 percent of women over the age 40 were removed due to missing data, which can also bias the final results. Another limiting factor was the static response variable which only considers recent mammography screening practices. The variable does not consider intentions for future mammography screenings and cannot explain a woman's rationale for not having a mammogram under current guidelines. Lastly, the response is not adjusted by age group based on the new recommendations from the ACS and SBI, but is instead separated by decadal groups.

There are several improvements that can be made in future mammography nonadherence studies using the BRFSS data. More intuitive age groups can be made to reflect screening guidelines. As well as changing the age groups, future studies can adjust the criteria for nonadherence based on certain guidelines (e.g. women 45-55 could have an adherence cutoff of one year as opposed to two years in correspondence to the new ACS guidelines). It could also

be beneficial to conduct several logistic regressions analyses using more years of data. This could provide insight into patterns of mammography nonadherence over time. Nearest neighbor technique and list-wise deletion may bias the results, but there could be a better way to handle our missing data problem. The ethnicity variable also used imputed data, so more research should analyze nonadherence among races using raw data.

5.2 Conclusion

In summary, the analysis supports several past studies using the most recent BRFSS data [12, 13, 14, 16]. Demographic and health related information such as income, number of children, and BMI category can help intervention programs recognize women who are less likely to adhere to mammography screening guidelines. Behavioral factors are the strongest predictor for screening behaviors. It is crucial for women to have a personal physician or health professional that they can routinely see every year. Tracking frequency of doctor visits and routine medical procedures can give great insight into mammography nonadherence, which could help reduce breast cancer mortality for women in the U.S.

References

[1] American Cancer Society. "Cancer Facts and Figures 2016." 2016. Web. 27 Apr. 2016.

[2] American Cancer Society. "Breast Cancer Prevention and Early Detection." (2015): n. pag. 20 Oct. 2015. Web. 01 Apr. 2016.

[3] Weedon-Fekjær, Harald, Pål R. Romundstad, and Lars J. Vatten. "Modern Mammography Screening and Breast Cancer Mortality: Population Study." The BMJ. CC BY-NC 3.0, 17 June 2014. Web. 29 Apr. 2016. <http://www.bmj.com/content/348/bmj.g3701>.

[4] American Cancer Society. "American Cancer Society Recommendations for Early Breast Cancer Detection in Women without Breast Symptoms." American Cancer Society Recommendations for Early Breast Cancer Detection in Women without Breast Symptoms. American Cancer Society, 20 Oct. 2014. Web. 07 Apr. 2016. <http://www.cancer.org/cancer/breastcancer/moreinformation/breastcancerearlydetection/breast-cancer-early-detection-acs-recs>.

[5] "ACR and SBI Continue to Recommend Regular Mammography Starting at Age 40 -." American College of Radiology. N.p., 20 Oct. 2015. Web. 08 Apr. 2016. <http://www.acr.org/About-Us/Media-Center/Press-Releases/2015-Press-Releases/20151020-ACR-SBI-Recommend-Mammography-at-Age-40>.

[6] "Archived Final Recommendation Statement." US Preventive Services Task Force. N.p., 30 Dec. 2013. Web. 08 Apr. 2016. <http://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/breast-cancer-screening>.

[7] Miller, Anthony B., Claus Wall, Cornelia J. Baines, Teresa To, and Steven A. Narod. "Twenty Five Year Follow-up for Breast Cancer Incidence and Mortality of the Canadian National Breast Screening Study: Randomised Screening Trial." The BMJ. N.p., 16 Jan. 2014. Web. 29 Apr. 2016. <http://www.bmj.com/content/348/bmj.g366>

[8] Magai, C, et al. "Common Psychosocial Factors Underlying Breast Cancer Screening And Breast Cancer Treatment Adherence: A Conceptual Review And Synthesis." Journal Of Women's Health (15409996) 16.1 (2007): 11-23 13p. CINAHL Complete. Mon. 11 Apr. 2016.

[9] Hellquist, Barbro Numan, et al. "Overdiagnosis In The Population-Based Service Screening Programme With Mammography For Women Aged 40 To 49 Years In Sweden." Journal Of Medical Screening 19.1 (2012): 14-19. MEDLINE Complete. Web. 29 Apr. 2016.

[10] "Health, United States, 2104 With Special Feature on Adults Age 55-64." National Center for Health Statistics, 2015. Web. 10 Apr. 2016. <http://www.cdc.gov/nchs/data/hus/hus14.pdf#076>.

[11] Siegel, Rebecca, Deepa Naishadham, and Ahmedin Jemal. "Cancer Statistics, 2012." CA: A Cancer Journal For Clinicians 62.1 (2012): 10-29. MEDLINE Complete. Web. 13 Apr. 2016.

[12] Schueler, Kristin M., Philip W. Chu, and Rebecca Smith-Bindman. "Factors Associated with Mammography Utilization: A Systematic Quantitative Review of the Literature." Journal of Women's Health 17.9 (2008): 1477-498. Web.

[13]Madadi, Mahboubeh, Shengfan Zhang, Karen H. Kim Yeary, and Louise M. Henderson. "Analyzing Factors Associated with Women's Attitudes and Behaviors toward Screening Mammography Using Design-based Logistic Regression." Breast Cancer Res Treat Breast Cancer Research and Treatment 144.1 (2014): 193-204. Web. 26 Apr. 2016

[14]Calvocoressi, Lisa. PSYCHOSOCIAL INFLUENCES ON NON-ADHERENCE TO MAMMOGRAPHY SCREENING GUIDELINES IN AFRICANAMERICAN AND WHITE WOMEN AGES 40-79. Diss. Yale U, 2003. Ann Arbor, MI: ProQuest Information and Learning, 2004. Print.

[15] Jensen, Line Flytkjær, Anette Fischer Pedersen, Berit Andersen, and Peter Vedsted. "Identifying Specific Non-attending Groups in Breast Cancer Screening - Population-based Registry Study of Participation and Socio-demography." BMC Cancer 12.1 (2012): 518. Web. 19 Apr. 2016.

[16] Khaliq, W., Aamar, A., & Wright, S. M. (2015). Predictors of non-adherence to breast cancer screening among hospitalized women. PLoS One, 10(12) <doi:http://0-dx.doi.org.library.uark.edu/10.1371/journal.pone.0145492>

[17] BRFSS, CDC -. "2014 BRFSS Data Overview." (n.d.): n. pag. Http://www.cdc.gov/brfss/annual_data/2014/pdf/overview_2014.pdf. Sept. 2015. Web. 10 Apr. 2016. <http://www.cdc.gov/brfss/annual_data/2014/pdf/overview_2014.pdf>.

[18] "2014 BRFSS Survey Data and Documentation." Centers for Disease Control and Prevention. Centers for Disease Control and Prevention, 11 Sept. 2015. Web. 05 Jan. 2016. <http://www.cdc.gov/brfss/annual_data/annual_2014.html>.

[19] Shalizi, Cosma Rohilla. N.p.: n.p., n.d. Advanced Data Analysis from an Elementary Point of View. 18 Apr. 2016. Web. 20 Apr. 2016. <http://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/ADAfaEPoV.pdf

[20] Bland, Martin J., and Douglas G. Altman. "The Odds Ratio." The BMJ. N.p., 27 May 2000. Web. 12 Apr. 2016. <http://www.bmj.com/content/320/7247/1468.1>.

[21] Montgomery, Douglas C. "Chapter 12." Applied Statistics and Probability for Engineers. 5th ed. N.p.: John Wiley and Sons, 2014. 502. Print.

[22] Kumar, Rajeev, and Abhaya Indrayan. "Receiver Operating Characteristic (ROC) Curve for Medical Researchers." Indian Pediatr Indian Pediatrics 48.4 (2011): 277-87. EBCSO. Web. 26 Apr. 2016.

[23] Chapter 4 - Diagnostic Tests - ROC Curves. University of Georgia, n.d. Web. 16 Apr. 2016. <http://ebp.uga.edu/courses/Chapter%204%20-%20Diagnosis%20I/8%20-%20ROC%20curves.html>.

[24] "Drinking Levels Defined." National Institute on Alcohol Abuse and Alcoholism (NIAAA). N.p., n.d. Web. 06 Mar. 2016. <http://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking>.

[25] Brfss, Cdc -. (n.d.): n. pag. Behavioral Risk Factor Surveillance System 2014 Codebook Report Land-Line and Cell-Phone Data. Center for Disease Control, 12 Aug. 2015. Web. 01 Apr. 2016. <http://

Appendix A: Survey Questions and Responses

Variable	BRFSS Survey *Question*/Responses
No Mammogram Within Last 2 Years	*How long has it been since your last mammogram screening?* Within past year Within past 2 years Within past 5 years 5 or more years ago Don't know/Not sure Never Refuse Not asked or Missing (Refused previous question question asking if she has ever had mammogram)
Poor Health	*During the past 30 days, for about how many days did poor physical or mental health keep you from doing your usual activities, such as self-care, work, or recreation?* Number of days (1-30) None Don't know/Not sure Refuse Not asked or Missing(No poor physical or mental health reported on previous health questions)
Number of Personal Doctors	*Do you have one person you think of as your personal doctor or health care provider? (If "No" ask "Is there more than one or is there no person who you think of as your personal doctor or health care provider?".)* Yes, only one More than one No Not Asked or Missing Refuse
Last Routine Checkup	*About how long has it been since you last visited a doctor for a routine checkup?* Within past year Within past 2 years Within past 5 years 5 or more years ago Don't know/Not sure Never Refuse Not Asked or Missing
Any Exercise in Last Month	*During the past month, other than your regular job, did you participate in any physical activities or exercises such as running, calisthenics, golf, gardening, or walking for exercise?* Yes No Don't know/Not sure Refuse Not Asked or Missing
Average Sleep Time	*On average, how many hours of sleep do you get in a 24-hour period?* Number of hours [1-24] Don't know/Not sure Refuse

Chronic Condition	*(Coronary Heart Disease, COPD, emphysema, chronic bronchitis, kidney disease, diabetes)* Yes (Answer Yes to one or more of the above questions) No (Answered No to all questions)
Last Dentist Visit	*How long has it been since you last visited a dentist or a dental clinic for any reason? Include visits to dental specialists, such as orthodontists.* Within past year Within past 2 years Within past 5 years 5 or more years ago Don't know/Not sure Never Refuse
Marital Status	*Are you: (marital status)* Married Divorced Widowed Seperated Never married A member of an unmarried couple Refuse Not asked or Missing
Number of Children in Household	*How many children less than 18 years of age live in your household?* Number of children (1-99) None Refuse Not asked or Missing
Highest Education	*What is the highest grade or year of school you completed?* Never attended school or only kindergarten Grades 1 through 8 Grades 9 through 11 Grades 12 or GED College 1 year to 3 years College 4 years or more Refuse Not asked or Missing
Employment Status	*Are you currently…?* Employed for Wages Self-Employed Out of work for 1 year or more Out of work for less than 1 year Homemaker Student Retired Unable to Work Refuse Not asked or Missing
Yearly Household Income Level	*Is your annual household income from all sources:* < $10,000 $10,000 - $14,999 $15,000 - $19,999

	$20,000 - $24,999 $25,000 - $34,999 $35,000 - $49,999 $50,000 - $74,999 > $75,000 Don't know/Not sure Refuse Not asked or Missing
BMI Category	*How much do you weigh?/How tall are you?* Combined weight and height to calculate BMI
Difficulty Doing Things Alone	*Because of a physical, mental, or emotional condition, do you have difficulty doing errands alone such as visiting a doctor's office or shopping?* Yes No Don't know/Not sure Refuse Not asked or Missing
Do You Smoke?	*Do you now smoke cigarettes every day, some days, or not at all? (Smoked at least 100 cigs in lifetime)* Every Day Some Days Not at all Don't know/Not sure Refuse Not asked or Missing
Use Chewing Tobacco or Snuff	*Do you currently use chewing tobacco, snuff, or snus every day, some days, or not at all?* Every day Some days Not at all Don't know/Not sure Refuse Not asked or Missing
Drinking Level	*During the past 30 days, how many days per week or per month did you have at least one drink of any alcoholic beverage?/During the past 30 days, on the days when you drank, about how many drinks did you drink on the average?* Does Not Drink/Not at Risk Drink Problem* At Risk**
Last Flu Shot	*During the past 12 months, have you had either a flu shot or a flu vaccine that was sprayed in your nose?* Yes No Don't know/Not sure Refuse Not asked or Missing
Last Breast Exam	*How long has it been since your last breast exam?* Within past year Within past 2 years Within past 5 years 5 or more years ago Don't know/Not sure

	Never
	Refuse
	Not asked or Missing (Never had Breast Exam or Refused previous question)
Last Pap Test	*How long has it been since you had your last Pap test?*
	Within past year
	Within past 2 years
	Within past 5 years
	5 or more years ago
	Don't know/Not sure
	Never
	Refuse
	Not asked or Missing (Never had Pap test or Refused previous question)
Ever Had Hysterectomy	*Have you had a hysterectomy?*
	Yes
	No
	Don't know/Not sure
	Refuse
	Not asked or Missing (Woman is currently pregnant)
Health Coverage Type	*What is the primary source of your health care coverage? Is it…*
	A plan purchased through an employer or union
	A plan that you or another family member buys on your own
	Medicare
	Medicaid or other state program
	TRICARE
	Alaska Native, Indian Health Service, Tribal Health Services
	Some other source
	None
	Don't know/Not sure
	Refuse
	Not asked or Missing
Age Group	*Imputed Age value collapsed above 80*
	Imputed ages from 18-99
Ethnicity	*Imputed race/ethnicity value*
	White, Non-Hispanic
	Black, Non-Hispanic
	Asian, Non-Hispanic
	American Indian/Alaskan Native, Non-Hispanic
	Hispanic
	Other race, Non-Hispanic

Note: Bolded variables combined multiple BRFSS questions

* Heavy Drinking - “Heavy drinking is drinking 5 or more drinks on the same occasion on each of 5 or more days in the past 30 days” [24]

**Low Risk for Developing an Alcohol Use Disorder – “Low-risk drinking is no more than 3 drinks on any single day and no more than 7 drinks per week.” [24]

Appendix B: R Code

```
#Read Training Data Into R
FirstLog <- read.csv("~/Honors Thesis/R Code Files/FirstLog.csv")

#Defining Variables as Categorical (Used similar code for stratification models)
FirstLog$GENHLTH <- factor(FirstLog$GENHLTH)
FirstLog$POORHLTH <- factor(FirstLog$POORHLTH)
FirstLog$PERSDOC2 <- factor(FirstLog$PERSDOC2)
FirstLog$CHECKUP1 <- factor(FirstLog$CHECKUP1)
FirstLog$EXERANY2 <- factor(FirstLog$EXERANY2)
FirstLog$CHECKUP1 <- factor(FirstLog$CHECKUP1)
FirstLog$SLEPTIM1 <- factor(FirstLog$SLEPTIM1)
FirstLog$Chronic.Condition <- factor(FirstLog$Chronic.Condition)
FirstLog$LASTDEN3 <- factor(FirstLog$LASTDEN3)
FirstLog$MARITAL <- factor(FirstLog$MARITAL)
FirstLog$CHILDREN <- factor(FirstLog$CHILDREN)
FirstLog$EDUCA <- factor(FirstLog$EDUCA)
FirstLog$EMPLOY1 <- factor(FirstLog$EMPLOY1)
FirstLog$INCOME2 <- factor(FirstLog$INCOME2)
FirstLog$BMI.CATEGORY <- factor(FirstLog$BMI.CATEGORY)
FirstLog$DIFFALON <- factor(FirstLog$DIFFALON)
FirstLog$Do.You.Smoke. <- factor(FirstLog$Do.You.Smoke.)
FirstLog$USENOW3 <- factor(FirstLog$USENOW3)
FirstLog$At.Risk.Drinking <- factor(FirstLog$At.Risk.Drinking)
FirstLog$FLUSHOT6 <- factor(FirstLog$FLUSHOT6)
FirstLog$Breast.Exam.Category <- factor(FirstLog$Breast.Exam.Category)
FirstLog$PAP...3.Years <- factor(FirstLog$PAP...3.Years)
FirstLog$HADHYST2 <- factor(FirstLog$HADHYST2)
FirstLog$Health.Coverage <- factor(FirstLog$Health.Coverage)
FirstLog$Age.Group <- factor(FirstLog$Age.Group)
FirstLog$X_IMPRACE <- factor(FirstLog$X_IMPRACE)

#Establishing Baseline Categories
contrasts(FirstLog$POORHLTH) <- contr.treatment(5,base = 5)
contrasts(FirstLog$CHILDREN) <- contr.treatment(4,base = 4)
contrasts(FirstLog$EDUCA) <- contr.treatment(6,base = 6)
contrasts(FirstLog$INCOME2) <- contr.treatment(8,base = 8)
contrasts(FirstLog$BMI.CATEGORY) <- contr.treatment(5,base = 2)
contrasts(FirstLog$DIFFALON) <- contr.treatment(2,base = 2)
contrasts(FirstLog$USENOW3) <- contr.treatment(3,base = 3)
contrasts(FirstLog$Do.You.Smoke.) <- contr.treatment(3,base = 3)
contrasts(FirstLog$Health.Coverage) <- contr.treatment(7,base = 2)
#Run Logistic Regression
Logit_Model = glm(MAMM...2.YEAR ~ GENHLTH + POORHLTH + PERSDOC2 + CHECKUP1 + EXERANY2 +
SLEPTIM1 + Chronic.Condition + LASTDEN3 + MARITAL + CHILDREN + EDUCA + EMPLOY1 + INCOME2 +
BMI.CATEGORY + DIFFALON + Do.You.Smoke. + USENOW3 + At.Risk.Drinking + FLUSHOT6 +
```

```
Breast.Exam.Category + PAP...3.Years + HADHYST2 + Health.Coverage + Age.Group + X_IMPRACE, data =
FirstLog, family = binomial)

#Output Coefficients
summary(Logit_Model)

#Output Odds Ratios and 95% Confidence Intervals
exp(cbind(OR = coef(Logit_Model), confint.default(Logit_Model)))

#Install pROC Package
install.packages("pROC")

#Establishing Logistic Regression as Model to be Validated
prob = predict(Logit_Model,type = c("response"))
#Establishing Testing Data
FirstLog$prob = prob

#Running/Outputting ROC Curve and AUC
g <- roc(MAMM...2.YEAR ~ prob, data = FirstLog)
plot(g)

#Calculating VIF
library(car)
vif(Logit_Model)
```