EPIDEMIOLOGICAL STUDY ESTIMATING THE EFFECTS OF SOCIOECONOMIC AND GENERAL HEALTH PREDICTORS ON MENTAL HEALTH USING A ZERO INFLATED NEGATIVE BINOMIAL MODEL

by

Brian William Perez

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
In partial fulfillment of the requirements
for the degree of Master of Science in
Economics

Charlotte

2019

Approved by:

Dr. Thomas Mayock

Dr. Carol Stivender

Dr. Mathew Metzger

©2019
Brian William Perez
ALL RIGHTS RESERVED

ABSTRACT

BRIAN WILLIAM PEREZ. Population study of depression and metabolic illnesses.
(Under the Direction of DR. THOMAS MAYOCK & DR. CAROL STIVENDER)

During the previous thirty years, the United States has experienced epidemic levels of chronic metabolic diseases such as diabetes, heart disease, and obesity. In the same time period, an increase in mental illness has occurred with increased rates of depression, suicide, drug abuse, and extremist acts, such as mass gun violence. Population data from the Centers of Disease Control (CDC) was used to determine if there is an underlying link between physical health and mental health, and the magnitude of effect that each unique attribute has on individuals self-reported mental health. Survey data from the Behavioral Risk Factor Surveillance System (BRFSS) was used to analyze the effect of individual's socioeconomic and general health attributes on the individual's self-reported mental health. Research on metabolic and mental health diseases yielded the included variables. In all, the dependent research variable (Mental Health) and 53 predictors are included in the data set that ranges from 1993-2017 with a total of 5,492,290 observations for each predictor. Using a random sub-sample of 10,000 individuals, three (3) independent, zero inflated negative binomial models were tested across multiple sub samples.

This model's results, along with current metabolic research, point to a link between depression and related mental health disorders and poor physical health. The variable categories of Income, Checkup, Smoking Status, Activity Status, BMI Category, and self-reported General Health had the most consistently statistically significant results. The model was tested across randomly generated sub samples, with individual's General

Health, BMI category, Income and time since last routine Checkup having the greatest impact on the expected number of poor mental health days, keeping all other variables constant.

Subsequent research behind the mechanisms of the statistically significant variables is needed to better target resources and health care policy. To better understand the compounding effects of risk factors on mental health, it is necessary to complete further joint significant studies and analysis. One option moving forward is to focus on fixing patients' existing metabolic deficiencies prior to administering pharmacological treatment for depression and other mental illnesses. Additionally, prospective health policy should target resources to high risk individuals and communities, enabling individuals to receive the care they need.

ACKNOWLEDGMENTS

I am thankful for the University of North Carolina at Charlotte, and the professors of the Economics Department who provided me with a great experience and education the last two years. I am especially grateful for Dr. Stivender and Dr. Depken who helped guide me through this thesis. I am grateful to my mother and father, Deianira and William, who have provided me with support throughout my life. I am also grateful to my sister, Jessica, who supported me when I was lacking motivation. A very special gratitude goes out to my cousin, Ermias Bizuwork for pushing me to be the best the person I can be and for being there when I am in need. I owe much of who I am today to you. I would also like to give a special mention to William Shannon and Devin Whalen; it was fantastic to have two great friends to spend many late nights working through problems together. Finally, to all my friends, family, and colleagues: it was great sharing experiences with all of you during the last two years. Thank you for all your encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BCBSHI	Blue Cross Blue Shield Health Index
BMI	Body Mass Index
BRFSS	Behavioral Risk Factor Surveillance System
CDC	Centers of Disease Control and Prevention
MDD	Major Depressive Disorder
PHQ-9	Patient Health Questionnaire
SSRI	Selective Serotonin Reuptake Inhibitors
WHO	World Health Organization

CHAPTER 1: INTRODUCTION

During the previous thirty years, the United States has experienced epidemic levels of chronic diseases such as diabetes, heart disease, and obesity [19]. During the same time period, an increase in mental illness has occurred, with increased rates of depression, suicide and drug abuse [4,11]. Research was conducted to ascertain if these illnesses are correlated and if so, to determine the statistically significant predictors in identifying high risk individuals, activities and populations.

Research was conducted on the symptoms and variables that lead to the diagnosis of mental illness. Specifically, stress, anxiety, and depressive disorders were the primary mental illnesses researched due to the complexities of mental health. The research conducted on Major Depressive Disorders' (MDD) symptoms and risk factors was used to identify the BRFSS survey questions that would give the most insight on the individuals physical and mental health. These model variables were used in an epidemiological study of MDD, which studied populations of humans to determine how, when, and where MDD occurs. [22]. The subsequent results of the study were analyzed to determine the most statistically significant risk factors of MDD when accounting for individuals' socioeconomic characteristics and relevant risk factors.

Identifying these groups enables policy makers to ensure at risk communities receive the care and resources they need. Additionally, research could be targeted on solving the root problems surrounding depression and anxiety disorders, further demystifying the illnesses. These improvements could lead to refined standards of care and decreasing depression, suicide and drug abuse. These changes would positively affect both individuals and the United States as a whole.

CHAPTER 2: LITERATURE REVIEW

2.1 Major Depressive Disorder

Major Depressive Disorder (MDD) is a medical illness that affects an individual's feelings, thoughts and behavior, which range from mild to severe [1]. Due to the complexities of this illness, diagnosis and treatment vary from patient to patient. Diagnosed symptoms can vary from "feeling sad or having a depressed mood" to "loss of interest or pleasure in activities once enjoyed" and "increased fatigue and…slowed movements, and thoughts of death or suicide." [2]

Depression is the leading cause of disability in the United States, affecting between 4.4 to 8.1 percent of Americans per year [1]. This value ranges between sources as the diagnosis rate is subject to an individual's mental state in a given two-week period. According to a report published in 2018 using the Blue Cross and Blue Shield Health Index (BCBSHI), which analyzes the insurance claims of over 41 million Americans, the diagnosis rate of MDD was 4.4 percent, which equates to 9 million Americans [1]. On the other hand, a separate report published by the CDC in 2018 using the Patient Health Questionnaire (PHQ-9), a symptom screening questionnaire, found that from 2013 to 2016, 8.1 percent of Americans aged 20 and over had symptoms of depression in a given two-week period [3]. Even though the exact diagnosed rate of depression is contested, both reports found that depression is on the rise, causing a significant impact on the overall health of Americans and the US economy.

According to a report by Blue Cross and Blue Shield, the overall diagnosis rate of depression has increased by 33 percent since 2013, with the rate of diagnosis in adolescents and millennials increasing 47 percent since 2013 [1]. Additionally,

adolescent and millennial girls have seen a 65 percent increase in diagnosed depression, while boys have seen a 47 percent increase [1]. This discrepancy between age and gender groups can also be seen in suicide rates. In 2015, the suicide rate among girls between 15 and 19 hit a 40-year high, with suicide rates for teenage girls increasing by 50 percent and the suicide rate for teenage boys increasing by 30 percent between 2007 and 2015 [4].

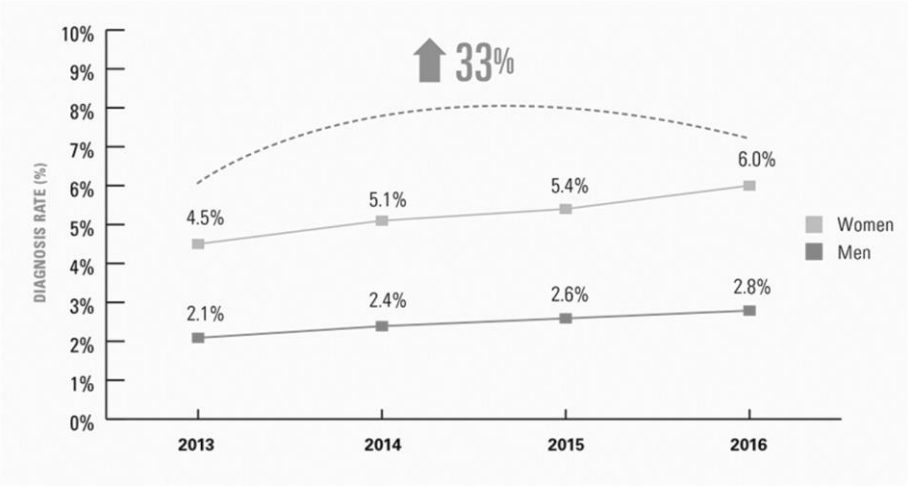

Figure 1:Diagnosis Rate of Major Depression by Sex 2013 -2016 [1].

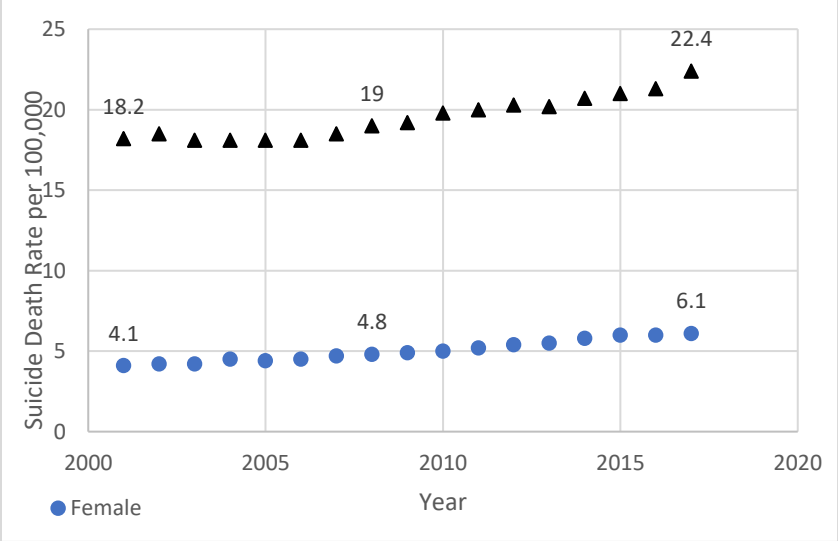

Figure 2: Age Adjusted Suicide Distribution by Sex [4].

These statistics show an increasing prevalence of mental illness in the United States, specifically in adolescents and millennials. Due to the complexity of MDD, the exact causes are not known. However, a variety of factors can increase the chance of becoming depressed, including past trauma, death or loss, certain medications, conflict, genetics, major life events, substance abuse, and serious illness. When selecting model variables, serious illness and substance abuse were the primary focus as data for the other risk factors is difficult to obtain, track and record. Due to MDD's complexity, treatment options available under the standard of care are limited.

Standard of care is defined as "A diagnostic and treatment process that a clinician should follow for a certain type of patient, illness, or clinical circumstance" [5]. The current standard of care for depression starts with a basic evaluation for the most common risk factors. The assessment's primary objective is to determine a potential cause, the depression's severity and history, and the type of treatment plan [7]. The two most common treatment paths are pharmacological management and non-pharmacological management [7]. Pharmacological management relies on choosing an anti-depressant based on a patient's response and history of side effects, the cost, the availability, and the patient's preferences [7]. Non-pharmacological management relies on psychoeducation and psychotherapeutic intervention such as counseling and support groups [7]. The standard of care for many chronic diseases has shifted to primarily using pharmacological solutions. This is also true for anti-depressants, shown in Figure 2 and Figure 5 in Appendix A.

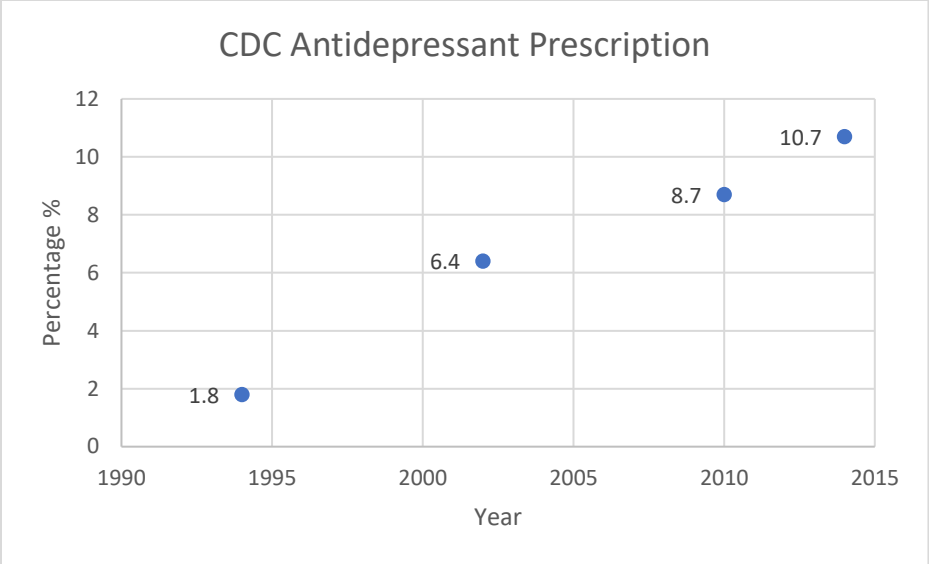

Figure 3: Anti-depressant Prescription Rates [8].

As Figure 3 shows, antidepressant use in the US has increased from 1.8 percent during the early 90's (1988-1994) to 10.8 percent between 2011-2014 [8]. A recent article published by the New York Times found that "some 15.5 million Americans have been taking the medications (antidepressants) for at least five years. The rate has almost doubled since 2010, and more than tripled since 2000" [20].

Antidepressants more specifically selective serotonin reuptake inhibitors (SSRI), are a class of drugs whose primary function is to regulate the hormone serotine [24]. The first SSRI, Prozac, was released to the public in 1988 [24]. This medication quickly became a popular alternative to previous antidepressants [23]. Antidepressants have not improved in the last 30 years as most medications on the market today are classified as SSRIs or SNRIs and act on the regulation of the hormone serotonin and the endocrine system [24].

The endocrine system is the collection of glands that produce hormones that regulate multiple key functions throughout the body [25]. These functions include sleep,

mood, metabolism, growth and development, and sexual reproduction [25]. Issues with the endocrine system occur if hormone levels are too high or too low, or if an individual's body does not respond to hormones in the appropriate ways [26]. The most common endocrine disease in the United States is diabetes. Diabetes is a disease in which the body's ability to produce or respond to the hormone insulin is impaired, resulting in abnormal metabolism of carbohydrates and elevated levels of glucose in the blood and urine [17]. Preclinical research on the gut-brain axis has shown a link between diabetes and depressive disorders.

2.2 Chronic Illness: Gut Brain Axis

The gut-brain axis, also known as the microbiota-gut-brain axis, represents the bidirectional biochemical signaling and communication between the intestines and brain [14]. This system includes the gut microbiome and central nervous system (brain and spinal cord) [14]. In multiple human and rodent studies, it has been established that diet can influence the composition and general health of the microbiome. [14]. The gut microbiota is associated with metabolic disorders such as obesity, diabetes mellitus, and neuropsychiatric disorders such as schizophrenia, autistic disorders, anxiety disorders and major depressive disorders [14]. Recent studies showed that the microbiota could activate the immune and central nervous systems, including commensal and pathogenic microorganisms in the gastrointestinal tract [14]. Gut microorganisms are capable of producing and delivering neuroactive substances, such as serotonin and gamma-aminobutyric acid, which act on the gut-brain axis [14]. Preclinical research in rodents suggested that certain probiotics have antidepressant and anxiolytic properties [14].

Recent research conducted by Blue Cross and Blue Shield shows a link between depression and overall health. The Blue Cross Blue Shield report found that 85 percent of people diagnosed with Major Depression Disorder (MDD) have one or more serious chronic diseases and nearly 30 percent have four or more other chronic health conditions [1]. This equates to "(An individual being) nearly 30 percent less healthy on average than those not diagnosed with major depression. This decrease in overall health translates to nearly 10 years of healthy life lost for both men and women" [1]. These startling statistics point to the high correlation between chronic illness and an individual's mental health.

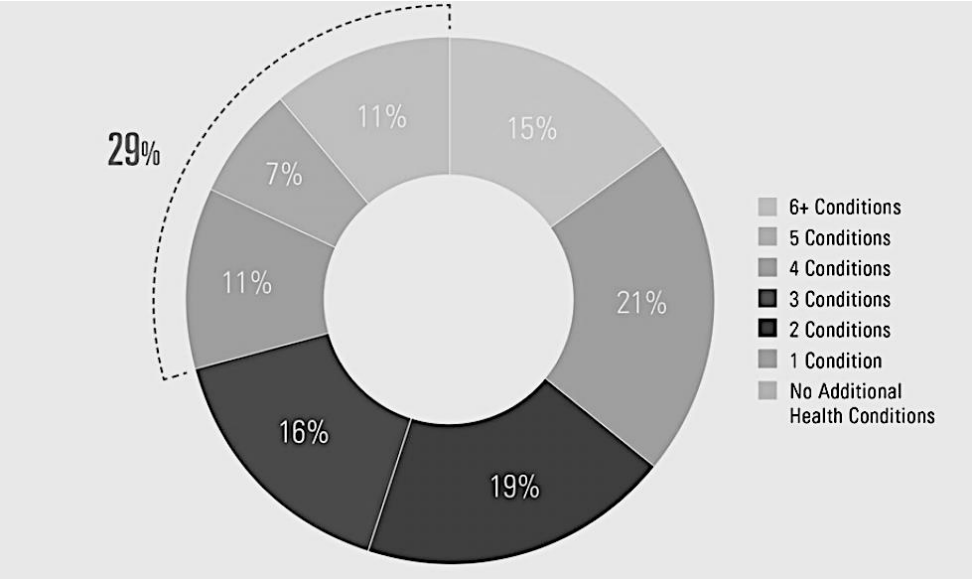

Figure 4: Adults Diagnosed with MDD and Chronic Diseases. [1].

2.3 Chronic Illness: Diabetes and Obesity

Chronic diseases are defined as "conditions that last one year or more and require ongoing medical attention or limit activities of daily living or both" [16]. In the last 30 years, diabetes and obesity have seen the highest increases in the United States. In fact,

these diseases are now considered epidemics by both the Center for Disease Control (CDC) and the World Health Organization (WHO).

In 2017, 84 million US adults were classified as pre-diabetic and an additional 30 million adults were diagnosed with diabetes [12]. In 2016 obesity affected approximately 39.8 percent, or 93.3 million adults in the United States, with 72 percent of American adults considered overweight or obese [18]. According to a report published by the World Health Organization "In 1995, there were an estimated 200 million obese adults worldwide and another 18 million under-five children classified as overweight. As of 2000, the number of obese adults has increased to over 300 million" [19]. In 2016, more than 1.9 billion adults, 18 years and older, were overweight. Of these, over 650 million were obese [19]. These diseases have a large impact not just on the individual's health but also on national economies as costs of health care are continuing to increase. Furthermore, this is an issue of national security as one in three US adults do not meet the height and weight requirements to enlist [27]. Additionally, a study published in 2015 by the National Security Research Division, found that almost 66 percent of service members of the 18,000 that were randomly selected across all service branches, are considered to be overweight or obese [29].

2.4 Chronic Illness: Economic Effects

Obesity and its associated health problems have a significant economic impact on the US health care system. Medical costs associated with individuals who are overweight and obese involve direct and indirect costs. Direct medical costs may include preventive, diagnostic, and treatment services related to obesity. Indirect costs relate to morbidity and mortality costs including productivity measures, which include absenteeism and

presenteeism [13]. The annual nationwide productivity costs of obesity-related absenteeism range between $3.38 billion and $6.38 billion, and the total cost of obesity was estimated at about $147 billion [13].

Individuals diagnosed with diabetes have health expenses that are twice as high than individuals without diabetes. This is due to the increased hospitalization, doctor visits and prescription drug usage. Overall this creates additional strain on the US health care system, with diabetes related medical costs totaling $327 billion yearly.

CHAPTER 3: METHODOLOGY

To further examine the effects of diabetes, obesity and overall general health on an individual's mental health, data was taken from the Behavioral Risk Factor Surveillance System. This phone survey, which started in 1984, is conducted by the CDC and individual states' health departments. The survey consists of an annual standard core, a biannual rotating core, optional modules, and state-added questions. Interview responses from the standard core section of the survey from 1993-2017 were used as model predictors. Two types of model predictors were chosen. Group 1 predictors consisted of control variables of common socioeconomic factors, such as Age, Income, Education, and Sex. Group 2 consisted of general health variables that were selected based on the research conducted. These variables included General Health (self-assessed), Activity Status, BMI Category, Diabetes Status, time since last Checkup, Smoking Status and Drinking Status. The individual survey responses to these questions, with values ranging from 0-9, were converted into individual predictors. Each possible survey response was used as a "dummy" categorical variable. Marital Status was categorized into two possible values, "married" or "not married". Model 3 variable list and descriptions are shown in Table 3 in Appendix C. The 25 individual data sets were combined into one large data set which totaled 53 predictors with 5,492,290 observations for each.

The research variable, labeled "Mental Health" in the data set, was not manipulated and came directly from the following survey question: "Now thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?" [15]. The

value response for this question ranged from 0-30, which represented the number of days in which individuals self-reported having stress, anxiety, depression or emotional issues. A day in which an individual self-reported as having stress, depression, and or problems with emotions will be referred to from this point forward as a "poor mental health day."

Table 1: Frequency Table of Mental Health Research Variable

Mental Health	Freq.	Percent	Mental Health	Freq.	Percent
0	3,711,551	67.58	16	1,787	0.03
1	182,386	3.32	17	1,359	0.02
2	289,528	5.27	18	2,133	0.04
3	165,404	3.01	19	339	0.01
4	82,653	1.5	20	76,074	1.39
5	198,448	3.61	21	5,922	0.11
6	22,332	0.41	22	1,281	0.02
7	77,864	1.42	23	816	0.01
8	15,007	0.27	24	886	0.02
9	2,310	0.04	25	26,136	0.48
10	141,356	2.57	26	1,048	0.02
11	812	0.01	27	1,894	0.03
12	9,409	0.17	28	7,754	0.14
13	1,083	0.02	29	4,704	0.09
14	30,877	0.56	30	298,291	5.43
15	130,846	2.38	Total	5,492,290	100

Due to the high frequency of zeros in the dependent variable (individuals having zero poor mental health days), a zero inflated Poisson model was originally considered, but the test for equidispersion failed. To account for the unequal dispersion of responses, a zero inflated negative binomial regression was used.

The data set contained a high frequency of zero responses, due to this complication, a standard negative binomial model would not distinguish between individuals who responded as having zero mental health days. The zero inflated model

distinguishes individuals who responded as having zero poor mental days with different underlying processes.

The zero inflated negative binomial regression generates two separate models. The first model is a logit model which predicts if an individual can be considered a "certain zero" [21]. A certain zero is an individual who is certain to respond to the research question by claiming zero poor mental health days. A negative binomial model is generated predicting the counts for those individuals who are not certain zeros. The zero inflated negative binomial model then combines the models and STATA output records. The results are divided into two sections: a count section and an inflated "certain zero" section.

CHAPTER 4: RESULTS AND ANALYSIS

Three models were considered using the two groups of parameters. As all variables used in the regression were categorical, except the dependent variable, one predictor in each variable category was omitted in the regression to prevent perfect collinearity. Thus, the results for each predictor were analyzed in relation to the omitted predictor in that variable category. For example, the results of "female" would be compared against the omitted "male" predictor under the Sex variable category. Year was also considered as a possible predictor for Model 3 as data was randomly sampled across 25 years, but it lacked statistical significance in this smaller sample size.

Table 2: Yearly Response Distribution of Random Sub Sample

Year	Freq.	Percent	Cum.
1993	9	0.09	0.09
1994	4	0.04	0.13
1995	2	0.02	0.15
1996	35	0.35	0.50
1997	33	0.33	0.83
1998	32	0.32	1.15
1999	25	0.25	1.40
2000	31	0.31	1.71
2001	62	0.62	2.33
2002	16	0.16	2.49
2003	423	4.23	6.72
2004	545	5.45	12.17
2005	583	5.83	18.00
2006	540	5.40	23.40
2007	708	7.08	30.48
2008	641	6.41	36.89
2009	705	7.05	43.94
2010	730	7.30	51.24
2011	759	7.59	58.83
2012	716	7.16	65.99
2013	732	7.32	73.31
2014	731	7.31	80.62
2015	591	5.91	86.53
2016	690	6.90	93.43
2017	644	6.44	99.87
2018	13	0.13	100.00

4.1 Coefficient Analysis of The Baseline Models 1 & 2

Model 1 only considered the respondents' socioeconomic characteristics as predictors for self-reported poor mental health days. These results can be seen in Figures 22 & 23 in Appendix D. This model's results were very similar to Model 3 with one primary exception; college educated, individuals were found to have .8276 times the self-reported poor mental health days when compared to individuals who had little or no formal education, labeled as "no school". No school individuals were individuals whose highest level of education was either some elementary school, some middle school, or some high school, but no high school diploma. These results were not seen in Model 3, as education was not statistically significant across all categories when compared to the omitted predictor "no school".

Model 2 only considered and analyzed an individual's general health with no consideration of socioeconomic factors. These results can be seen in Figures 24 & 25 in Appendix D. These models proved to be good baselines to compare to Model 3 results. Most predictors that were statistically significant in the baseline models were again statistically significant in Model 3 with their coefficients decreasing in magnitude.

4.2 Coefficient Analysis of Model 3

Model 3 contained thirty-six (36) independent categorical variables from both groups of predictors. Results shown in Table 3 contain the variable category, predictor name and the omitted predictor that the coefficients are compared to. The highlighted values are the statistically significant variables at the 95 percent confidence interval (α = 0.5). These predictors are of most interest, as they indicate individuals who are least

likely to have poor mental health days when compared to the omitted predictors, while keeping all other variables constant.

Table 3:Model 3 Results, Significant Predictors are Highlighted.

Variable Category	Model Variable	Omitted Variable	exp (*B*)	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
Age	Age (25-35)	Age (0-25)	0.845315372	-0.1680455	0.0762699	-2.2	0.028	-0.3175317	-0.0185593
	Age (35-45)		1.101466639	0.0966426	0.0657979	1.47	0.142	-0.0323189	0.225604
	Age (45-55)		1.061466455	0.0596514	0.0609324	0.98	0.328	-0.0597739	0.1790768
	Age (55-65)		1.020212358	0.0200108	0.0667265	0.3	0.764	-0.1107706	0.1507923
	Age (65-100)		0.974734649	-0.02559	0.0694229	-0.37	0.712	-0.1616564	0.1104765
Race	asian	White	0.851944155	-0.1602343	0.2549666	-0.63	0.530	-0.6599596	0.339491
	black		0.799988761	-0.2231576	0.0786005	-2.84	0.005	-0.3772117	-0.0691034
	hispanic		1.099668422	0.0950087	0.0651335	1.46	0.145	-0.0326505	0.2226679
	multiracial		1.099174122	0.0945591	0.1189182	0.8	0.427	-0.1385163	0.3276345
Income	lessthan15k	Less than 10k	0.875452923	-0.1330139	0.0733058	-1.81	0.070	-0.2766906	0.0106628
	lessthan20k		0.970745739	-0.0296907	0.0769919	-0.39	0.700	-0.180592	0.1212107
	lessthan25k		0.981704532	-0.0184649	0.0725729	-0.25	0.799	-0.1607052	0.1237753
	lessthan35k		0.823920858	-0.1936808	0.0753234	-2.57	0.010	-0.3413119	-0.0460497
	lessthan50k		0.787163658	-0.2393191	0.0743608	-3.22	0.001	-0.3850635	-0.0935747
	lessthan75k		0.767858736	-0.2641495	0.0760081	-3.48	0.001	-0.4131227	-0.1151763
	kormore		0.72765365	-0.3179301	0.0781103	-4.07	0.000	-0.4710234	-0.1648368
Education	collegegraduate	No school	0.912648196	-0.0914048	0.067159	-1.36	0.174	-0.223034	0.0402245
	highschoolgraduate		0.926325402	-0.0765297	0.067105	-1.14	0.254	-0.2080531	0.0549938
Employment Status	employed	Unemployed	0.912512952	-0.091553	0.0457341	-2.00	0.045	-0.1811903	-0.0019157
Sex	female	Male	1.065488309	0.0634332	0.0434497	1.46	0.144	-0.0217266	0.148593
Marital Status	married	Not Married	0.940455168	-0.0613913	0.0435354	-1.41	0.158	-0.1467191	0.0239366
Checkup	within1year	5 or More years	0.717852186	-0.3314916	0.1144038	-2.9	0.004	-0.555719	-0.1072642
	within2years		0.685118201	-0.3781639	0.1282135	-2.95	0.003	-0.6294578	-0.12687
	within5years		0.788230988	-0.2379641	0.1221858	-1.95	0.051	-0.477444	0.0015158
Smoking Status	neversmoked	Current Smoker	0.700334361	-0.3561974	0.0535722	-6.65	0.000	-0.4611971	-0.2511978
	formersmoker		0.741385312	-0.2992348	0.0575858	-5.2	0.000	-0.4121009	-0.1863686
	smokessome		0.875276975	-0.1332149	0.080569	-1.65	0.098	-0.2911272	0.0246975
Drinking Status	notheavydrinker	Heavy Drinker	0.965489551	-0.03512	0.0894165	-0.39	0.694	-0.2103731	0.1401331
Activity Status	recentlyactive	Not Recently Active	0.879244731	-0.128692	0.0426316	-3.02	0.003	-0.2122484	-0.0451355
BMI Category	overweight	Obese	0.985903795	-0.0141965	0.0576353	-0.25	0.805	-0.1271596	0.0987666
	normalweight		0.859185871	-0.15177	0.0587046	-2.59	0.010	-0.2668288	-0.0367112
	underweight		0.851130511	-0.1611898	0.0691614	-2.33	0.020	-0.2967436	-0.025636
General Health	excellent	Poor	0.412249265	-0.8861271	0.0902494	-9.82	0.000	-1.063013	-0.7092415
	verygood		0.459712079	-0.7771549	-0.0634198	12.25	0.000	-0.9014554	-0.6528545
	good		0.575663207	-0.5522325	0.0562824	-9.81	0.000	-0.662544	-0.4419211
	fair		0.787822317	-0.2384827	0.0523893	-4.55	0.000	-0.3411639	-0.1358015
	_cons		43.08048025	3.76307	0.1636624	22.99	0.000	3.442297	4.083842

Of the variable categories selected, Income, Checkup, Smoking Status, Activity Status, BMI Category, and General Health had the most consistently significant results when the model was tested across randomly generated sub samples. The model analysis will focus on these categories, and predictors in each category, and how it relates to individuals in a non-theoretical setting.

4.3 Coefficient Analysis of Model 3 - Socioeconomic Predictors

The first variable category, Age, was considered, as the latest research suggests that the highest increase in depression was in adolescents and millennials aged “0-25”. The only consistently significant age group when multiple sub samples were tested was the age group of “25-35”. As the model is a logit regression, the coefficients are interpreted as exp (coefficient). The expected number of self-reported poor mental days in a month for an individual in the Age category “25-35” is exp (-.16804) = .84531 times the expected number of self-reported poor mental health days for an individual in the omitted Age category of “0-25” ,while holding all other variables in the model constant. These results state that individuals in their mid 20’s to early 30’s have 84.5 percent of the expected poor mental health days of individuals in the lower age category. This was the only significant age category with no trend in subsequent Age categories.

The second variable category, Race, had limited results with having only one statistically significant predictor. Individuals that responded as “black (non-Hispanic)” had exp (-.223) = .79998 times the expected number of self-reported poor mental health days as individuals that responded as “white” while holding all other variables in the model constant. This represents individuals that responded as “black”, having 79.9% of the expected poor mental health days of a “white” individual. This was the only Race predictor that was statistically significant across all models.

The Income variable category had statistically significant results with an upward trend. Individuals who responded within higher income categories, self-reported as having less poor mental health days. The expected number of self-reported poor mental health days in a month for an individual who responded as making less than $35,000, is

exp (-.1936) = .8239 times the expected number of self-reported poor mental health days in a month for an individual who makes less than $10,000. The expected number of self-reported poor mental health days in a month for these individuals is exp (-.3179) = .7276 times the expected number of self-reported poor mental health days in a month for an individual who makes less than $10,000. In addition to the trend of decreasing coefficients, as the Income category increases, an increase in statistical significance also occurs with higher Z-Scores.

The next categorical variable, Education, looked at individuals' education levels to determine the role education plays in poor mental health. In Model 1, "college graduate" was statistically significant when compared to "no school" individuals, as shown in Figure 22 in Appendix D. However, this was not seen in Model 3, where the Education categorical predictors were not statistically significant.

Employment status was statistically significant In Model 1, however slightly less statistically significant in the Model 3. The expected number of self-reported poor mental days in a month for an individual that responded as "Employed" is exp(-.09155) = .9125 times the expected number of self-reported poor mental health days as an individual that responded as "Unemployed" while holding all other variables in the model constant.

Marital status was not statistically significant in Model 3, while being statistically significant in Model 1. Sex was not statistically significant in Model 1 or Model 3.

4.4 Coefficient Analysis of Model 3 - General Health Predictors

The category variable "Checkup" represents time since the respondent has seen a doctor for a routine checkup. Both Model 2 and the complete model found that this categorical variable was statistically significant with a positive trend as individuals with

routine checkups were found to have less self-reported days of poor mental health. The expected number of self-reported poor mental days in a month for an individual that responded as “within1year” (having a routine checkup within the year of responding to the survey) is exp(-.33149) = .71785 times the expected number of self-reported poor mental health days as an individual that responded as “5ormoreyears”(having a routine checkup in the last 5 or more years) while holding all other variables in the model constant.

Smoking status of the respondents was statistically significant in both Model 2 and Model 3. As individuals who never smoked or were former smokers had less expected poor mental health days when compared to individuals who were current smokers or smoked some. The expected number of self-reported poor mental health days in a month for an individual that responded as “never smoked” is exp (-.3562) = .7003 times the expected number of self-reported poor mental health days as an individual that responded as “current smoker” while holding all other variables in the model constant. These result show that individuals that never smoked had 70 percent of the expect poor mental health days of current smokers. Drinking status was not statistically significant in both Model 2 and 3. This may be due to respondent survey bias not accurately reporting the amount of drinks per occasion, or the overall prevalence of alcohol in American society.

Many studies have shown the importance of physical activity on physical health and mental health. A primary issue with survey data on exercise is the ability to quantify an individual’s exercise intensity, length and benefit, as individuals may perceive “high level” versus “mild” exercise differently. To approximate exercise, the general survey

question “adults who reported doing physical activity during the past 30 days other than their regular job” was selected. This choice is significant, as it indicates that individuals who engaged in some form of physical exercise were more likely to have less poor mental health days than those who engaged in none. The expected number of self-reported poor mental days in a month for an individual that responded as “recently active” is exp (-.12869) = .8792 times the expected number of self-reported poor mental health days as an individual that responded as “not recently active” while holding all other variables in the model constant.

Body Mass Index (BMI) is a well-known ratio of height to weight used to determine if an individual is obese, overweight, normal weight or underweight. The BMI category was statistically significant in both models as individuals who were categorized as normal weight and underweight could expect less poor mental health days when compared to obese individuals. The expected number of self-reported poor mental days in a month for an individual that was categorized as “normal weight” is exp (-.15177) = .85918 times the expected number of self-reported poor mental health days as an individual that was categorized as “obese” while holding all other variables in the model constant.

Overall, general health was chosen as a predictor to determine the significance of an individual’s self-assessment in relation to the self-reporting of poor mental health days. Individuals who responded as having self-assessed excellent general health were much less likely to have poor mental health days when compared to individuals who responded as having poor general health. These results contained a trend in which the worse the individual assessed his or her general health, it became more likely that

individuals would have poor mental health days. The expected number of self-reported poor mental days in a month for an individual that was categorized as having “excellent” general health is exp(-.88612) = .4122 times the expected number of self-reported poor mental health days as an individual that was categorized as having “poor” general health while holding all other variables in the model constant. These results had the highest impact on the quantity of poor mental health days. Individuals with excellent general health had 41.2 percent of the expected poor mental health days when compared to individuals with poor general health.

CHAPTER 5: CONCLUSION

All three models produced statistically significant results that align with the research conducted. Poor overall health, and the associated risk factors, are consistent predictors in expected poor mental health days. These results, when considered in a non-theoretical sense, points to individuals who are poor, white, not active, and who do not have consistent physical checkups as the individuals most likely to self-report poor mental health days. With these considerations in mind, the medical community needs to develop standards of care that look at the overall health of individuals and solve any underlying chronic health conditions prior to administering pharmalogical solutions. Pharmalogical solutions should be supplemental, and short term, to limit the dependency of these medications. With the latest medical research on gut biome and its effects on mental health, the medical profession can better understand the effects of diet, exercise, and sleep on the health of the brain and endocrine system. Further analysis considering social media, prescription drug use, and the effects of cell phone usage would be beneficial in understanding the recent spikes in mental illness. Subsequent research to determine if there is a connection between mass gun shootings and the effects of physical health, socioeconomics, and demographics would be beneficial.

To stem the epidemic levels of metabolic and mental illnesses, future health policy should target high risk populations and individuals with access to better resources. These resources could include providing high risk populations with life coaches, access to therapists, access to public gyms and food stipends for more nutrient dense food. The underlying risk factors need to be addressed for current and future generations to live productive and healthy lives.

REFERENCES

[1] “Major Depression: The Impact on Overall Health.” *Blue Cross Blue Shield*, 10 May 2018, www.bcbs.com/the-health-of-america/reports/major-depression-the-impact-overall-health.

[2] Parekh M.D, M.P.H, Ranna. “What Is Depression?” *Warning Signs of Mental Illness*, Jan. 2017, www.psychiatry.org/patients-families/depression/what-is-depression.

[3] Brody, D. J., Pratt, L. A., & Hughes, J. P. (2018, February). *Prevalence of Depression Among Adults Aged 20 and Over: United States, 2013–2016*(Rep. No. 303). Retrieved https://www.cdc.gov/nchs/data/databriefs/db303.pdf

[4] QuickStats: Suicide Rates for Teens Aged 15–19 Years, by Sex - United States, 1975–2015 | MMWR. (n.d.). Retrieved from https://www.cdc.gov/mmwr/volumes/66/wr/mm6630a6.htm

[5] Definition of Standard of care. Retrieved from https://www.medicinenet.com/script/main/art.asp?articlekey=33263

[6] Causes of Depression: Genetics, Illness, Abuse, and More. (n.d.). Retrieved from https://www.webmd.com/depression/guide/causes-depression#1

[7] Gautam, S., Jain, A., Gautam, M., Vahia, V. N., & Grover, S. (2017, January). Clinical Practice Guidelines for the management of Depression. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5310101/Jr, W. C. (n.d.).

[8] Selected prescription drug classes used in the past 30 days, by sex and age (Rep.). (2017). Retrieved https://www.cdc.gov/nchs/data/hus/2017/080.pdf

[9] Kane, S. P. (n.d.). Aripiprazole. Retrieved from https://clincalc.com/DrugStats/Drugs/Aripiprazole

[10] The Growing Economic Burden of Depression in the United States. (n.d.). Retrieved from https://www.analysisgroup.com/the-growing-economic-burden-of-depression-in-the-united-states/

[11] The opioid epidemic in America: An update. (n.d.). Retrieved from https://www.bcbs.com/the-health-of-america/reports/the-opioid-epidemic-america-update About Chronic Diseases | CDC. (n.d.). Retrieved from https://www.cdc.gov/chronicdisease/about/index.htm

[12] Centers for Disease Control and Prevention. National Diabetes Statistics Report, 2017. Atlanta, GA: Centers for Disease Control and Prevention, US Department of Health and Human Services; 2017.

[13] The Cost of Diabetes. (n.d.). Retrieved from http://www.diabetes.org/advocacy/news-events/cost-of-diabetes.html

[14] Bădescu, S. V., Tătaru, C., Kobylinska, L., Georgescu, E. L., Zahiu, D. M., Zăgrean, A. M., & Zăgrean, L. (2016). The association between Diabetes mellitus and Depression. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4863499/

[15] CDC - BRFSS. (n.d.). Retrieved from https://www.cdc.gov/brfss/index.html

[16] About Chronic Diseases | CDC. (n.d.). Retrieved from https://www.cdc.gov/chronicdisease/about/index.htm

[17] Prediabetes - Your Chance to Prevent Type 2 Diabetes | CDC. (n.d.). Retrieved from https://www.cdc.gov/diabetes/basics/prediabetes.html

[18] Adult Obesity Facts | Overweight & Obesity | CDC. (n.d.). Retrieved from https://www.cdc.gov/obesity/data/adult.html

[19] Controlling the global obesity epidemic. (2013, December 06). Retrieved from https://www.who.int/nutrition/topics/obesity/en/

[20] Carey, B., & Gebeloff, R. (2018, April 7). Many People Taking Antidepressants Discover They Cannot Quit. Retrieved from https://www.nytimes.com/2018/04/07/health/antidepressants-withdrawal-prozac-cymbalta.html

[21] (n.d.). UCLA. Retrieved from https://stats.idre.ucla.edu/stata/output/zero-inflated-negative-binomial-regression/

[22] (2017, February 7). What Is Epidemiology? Retrieved from https://www.nidcd.nih.gov/health/statistics/what-epidemiology

[23] (n.d.). 25 years after Prozac. Retrieved from https://www.pharmacist.com/25-years-after-prozac

[24] (2018, May 17). Selective serotonin reuptake inhibitors (SSRIs). Retrieved from https://www.mayoclinic.org/diseases-conditions/depression/in-depth/ssris/art-20044825

[25] (n.d.). Endocrine System: Facts, Functions and Diseases. Retrieved from https://www.livescience.com/26496-endocrine-system.html

[26] (n.d.). The Endocrine System. Retrieved from https://www.hormone.org/what-is-endocrinology/the-endocrine-system

[27] Myers, M. (2018, October 15). America's obesity is threatening national security, according to this study. Retrieved from https://www.armytimes.com/news/your-army/2018/10/10/americas-obesity-is-threatening-national-security-according-to-this-study/

[28] Feeney, N. (n.d.). 71% of U.S. Youth Don't Qualify for Military Service, Pentagon Says. Retrieved from https://time.com/2938158/youth-fail-to-qualify-military-service/

[29] Simkins, J. D. (2018, October 3). A staggering number of troops are fat and tired, report says. Retrieved from https://www.militarytimes.com/off-duty/military-culture/2018/10/03/a-staggering-number-of-troops-are-fat-and-tired-report-says/

APPENDIX A: PRESCRIPTION DRUG USE

Table 80. Selected prescription drug classes used in the past 30 days, by sex and age: United States, selected years 1988–1994 through 2011–2014

Excel version (with more data years and standard errors when available): https://www.cdc.gov/nchs/hus/contents2017.htm#080 .

[Data are based on a sample of the civilian noninstitutionalized population]

Age group and Multum Lexicon Plus therapeutic class[1] (common indications for use)	Total			Male			Female		
	1988–1994	1999–2002	2011–2014	1988–1994	1999–2002	2011–2014	1988–1994	1999–2002	2011–2014
All ages	Percent of population with at least one prescription drug in drug class in past 30 days								
Antihyperlipidemic agents (high cholesterol)	1.7	6.5	14.3	1.5	7.1	15.0	1.8	5.8	13.7
Analgesics (pain relief)	7.2	9.4	9.1	5.4	7.3	7.6	9.0	11.3	10.5
Antidepressants (depression and related disorders)	1.8	6.4	10.7	1.2	4.4	7.3	2.3	8.3	13.9
Proton pump inhibitors or H2 antagonists (gastric reflux, ulcers)[2]	2.8	5.3	8.5	2.4	4.7	7.5	3.0	5.9	9.4
Beta-adrenergic blocking agents (high blood pressure, heart disease)	3.1	4.4	7.7	2.7	4.1	7.1	3.5	4.6	8.4
ACE inhibitors (high blood pressure, heart disease)	2.4	4.6	7.3	2.4	4.7	8.1	2.4	4.5	6.6
Antidiabetic agents (diabetes)	2.6	3.7	6.6	2.5	3.7	6.8	2.6	3.8	6.3
Diuretics (high blood pressure, heart disease, kidney disease)[3]	3.4	4.1	5.6	2.3	3.1	4.4	4.4	5.1	6.7
Thyroid hormones (hypothyroidism)	2.3	3.9	5.1	0.8	1.5	1.9	3.7	6.2	8.0
Bronchodilators (asthma, breathing)	2.6	3.5	4.6	2.5	3.1	4.5	2.7	3.8	4.6
Sex hormones (contraceptives, menopause, hot flashes)[4]	...	...	...	...	...	...	9.8	15.2	8.7
Anxiolytics, sedatives, and hypnotics (anxiety, insomnia, and related disorders)	2.8	3.3	5.3	1.9	2.6	4.4	3.6	4.0	6.2
Antihypertensive combinations (high blood pressure)	2.4	2.9	4.1	1.4	1.9	3.3	3.3	3.8	4.8
Anticonvulsants (epilepsy, seizure, and related disorders)	1.4	2.4	4.9	1.2	2.1	4.3	1.6	2.7	5.5
Calcium channel blocking agents (high blood pressure, heart disease)	3.6	4.2	4.7	3.4	3.5	4.6	3.8	4.8	4.7

Figure 5: Prescription Drug Prescribed from 1988-2014 [8].

APPENDIX B: RESEARCH AND MODEL VARIABLES

- All blank and refused values were removed for each question and for each year of data used.

Label: Number of Days Mental Health Not Good
Section Name: Healthy Days — Health Related Quality of Life
Core Section Number: 2
Question Number: 2
Column: 93-94
Type of Variable: Num
SAS Variable Name: MENTHLTH
Question Prologue:
Question: Now thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?

Value	Value Label	Frequency	Percentage	Weighted Percentage
1 - 30	Number of days Notes: _ _ Number of days	142,679	31.71	34.91
88	None	300,134	66.69	63.56
77	Don't know/Not sure	5,026	1.12	1.05
99	Refused	2,176	0.48	0.49
BLANK	Not asked or Missing	1	.	.

Figure 6: Mental Health Statistics taken from 2017 BRFSS Codebook

Label: Imputed age in six groups
Section Name: Calculated Variables
Module Section Number: 8
Question Number: 14
Column: 2033
Type of Variable: Num
SAS Variable Name: _AGE_G
Question Prologue:
Question: Six-level imputed age category

Value	Value Label	Frequency	Percentage	Weighted Percentage
1	Age 18 to 24 Notes: 18 <= _IMPAGE <= 24	26,236	5.83	12.53
2	Age 25 to 34 Notes: 25 <= _IMPAGE <= 34	47,232	10.50	17.35
3	Age 35 to 44 Notes: 35 <= _IMPAGE <= 44	52,100	11.58	16.30
4	Age 45 to 54 Notes: 45 <= _IMPAGE <= 54	69,902	15.53	16.64
5	Age 55 to 64 Notes: 55 <= _IMPAGE <= 64	96,577	21.46	16.68
6	Age 65 or older Notes: _IMPAGE => 65	157,969	35.10	20.50

Figure 7: Age Statistics taken from 2017 BRFSS Codebook

Label: Income Level
Section Name: Demographics
Core Section Number: 8
Question Number: 17
Column: 180-181
Type of Variable: Num
SAS Variable Name: INCOME2
Question Prologue:
Question: Is your annual household income from all sources: (If respondent refuses at any income level, code ´Refused.´)

Value	Value Label	Frequency	Percentage	Weighted Percentage
1	Less than $10,000 Notes: If "no," code 02	18,346	4.11	5.02
2	Less than $15,000 ($10,000 to less than $15,000) Notes: If "no," code 03; if "yes," ask 01	19,334	4.33	4.34
3	Less than $20,000 ($15,000 to less than $20,000) Notes: If "no," code 04; if "yes," ask 02	27,735	6.21	6.45
4	Less than $25,000 ($20,000 to less than $25,000) Notes: If "no," ask 05; if "yes," ask 03	34,222	7.66	7.78
5	Less than $35,000 ($25,000 to less than $35,000) Notes: If "no," ask 06	39,751	8.90	8.71
6	Less than $50,000 ($35,000 to less than $50,000) Notes: If "no," ask 07	53,148	11.90	11.07
7	Less than $75,000 ($50,000 to less than $75,000) Notes: If "no," code 08	59,632	13.35	12.26
8	$75,000 or more	122,763	27.48	28.50
77	Don't know/Not sure	33,328	7.46	8.40
99	Refused	38,426	8.60	7.48
BLANK	Not asked or Missing	3,331	.	.

Figure 8: Income Statistics taken from 2017 BRFSS Codebook

Label: Imputed race/ethnicity value
Section Name: Weighting Variables
Module Section Number: 1
Question Number: 12
Column: 1519-1520
Type of Variable: Num
SAS Variable Name: _IMPRACE
Question Prologue:
Question: Imputed race/ethnicity value (This value is the reported race/ethnicity or an imputed race/ethnicity, if the respondent refused to give a race/ethnicity. The value of the imputed race/ethnicity will be the most common race/ethnicity response for that region of the state)

Value	Value Label	Frequency	Percentage	Weighted Percentage
1	White, Non-Hispanic	344,800	76.62	63.06
2	Black, Non-Hispanic	36,199	8.04	11.66
3	Asian, Non-Hispanic	9,963	2.21	5.37
4	American Indian/Alaskan Native, Non-Hispanic	8,497	1.89	1.01
5	Hispanic	37,233	8.27	16.81
6	Other race, Non-Hispanic	13,324	2.96	2.08

Figure 9: Race Statistics taken from 2017 BRFSS Codebook

Label: Education Level
Section Name: Demographics
Core Section Number: 8
Question Number: 7
Column: 163
Type of Variable: Num
SAS Variable Name: EDUCA
Question Prologue:
Question: What is the highest grade or year of school you completed?

Value	Value Label	Frequency	Percentage	Weighted Percentage
1	Never attended school or only kindergarten	629	0.14	0.31
2	Grades 1 through 8 (Elementary)	10,434	2.32	4.61
3	Grades 9 through 11 (Some high school)	21,624	4.81	8.63
4	Grade 12 or GED (High school graduate)	122,577	27.24	27.90
5	College 1 year to 3 years (Some college or technical school)	124,655	27.70	30.92
6	College 4 years or more (College graduate)	168,390	37.42	27.24
9	Refused	1,701	0.38	0.40
BLANK	Not asked or Missing	6	.	.

Figure 10: Education Statistics taken from 2017 BRFSS Codebook

- Values of 1,2,3 were binned and labeled as “No School”, Values of 4,5 were binned and labeled “High School Graduate”, Values of 6 were labeled “College Graduate”.

Label: Respondents Sex
Section Name: Demographics
Core Section Number: 8
Question Number: 1
Column: 125
Type of Variable: Num
SAS Variable Name: SEX
Question Prologue:
Question: Indicate sex of respondent.

Value	Value Label	Frequency	Percentage	Weighted Percentage
1	Male	198,725	44.16	48.64
2	Female	251,007	55.78	51.31
9	Refused	284	0.06	0.06

Figure 11: Sex Statistics taken from 2017 BRFSS Codebook

Label: Employment Status
Section Name: Demographics
Core Section Number: 8
Question Number: 15
Column: 177
Type of Variable: Num
SAS Variable Name: EMPLOY1
Question Prologue:
Question: Are you currently...?

Value	Value Label	Frequency	Percentage	Weighted Percentage
1	Employed for wages	184,093	40.91	47.26
2	Self-employed	39,592	8.80	9.08
3	Out of work for 1 year or more	10,489	2.33	2.74
4	Out of work for less than 1 year	9,225	2.05	2.87
5	A homemaker	24,376	5.42	6.37
6	A student	12,181	2.71	5.78
7	Retired	133,648	29.70	18.02
8	Unable to work	32,706	7.27	6.89
9	Refused	3,700	0.82	1.00
BLANK	Not asked or Missing	6	.	.

Figure 12: Employment Status Statistics taken from 2017 BRFSS Codebook

- Values of 1,2 were binned and labeled as “Employed” and values of 3-8 were binned and labeled as “Unemployed”. Value 9 and Blank were removed from the dataset.

Label: Marital Status
Section Name: Demographics
Core Section Number: 8
Question Number: 6
Column: 162
Type of Variable: Num
SAS Variable Name: MARITAL
Question Prologue:
Question: Are you: (marital status)

Value	Value Label	Frequency	Percentage	Weighted Percentage
1	Married	232,891	51.75	50.07
2	Divorced	61,437	13.65	10.75
3	Widowed	54,633	12.14	6.88
4	Separated	9,426	2.09	2.56
5	Never married	73,939	16.43	24.15
6	A member of an unmarried couple	14,544	3.23	4.91
9	Refused	3,139	0.70	0.67
BLANK	Not asked or Missing	7	.	.

Figure 13: Marital Statistics taken from 2017 BRFSS Codebook

- Values of 1 were binned and labeled as “Married” and values of 2-6 were binned and labeled as “Not Married”. Value 9 and Blank were removed from the dataset.

Label: General Health
Section Name: Health Status
Core Section Number: 1
Question Number: 1
Column: 90
Type of Variable: Num
SAS Variable Name: GENHLTH
Question Prologue:
Question: Would you say that in general your health is:

Value	Value Label	Frequency	Percentage	Weighted Percentage
1	Excellent	74,356	16.52	17.87
2	Very good	145,479	32.33	31.25
3	Good	142,502	31.67	32.03
4	Fair	62,213	13.82	13.75
5	Poor	24,317	5.40	4.85
7	Don’t know/Not Sure	707	0.16	0.16
9	Refused	439	0.10	0.08
BLANK	Not asked or Missing	3	.	.

Figure 14: General Health Statistics taken from 2017 BRFSS Codebook

- Values of 7 and 9 were removed from the data set.

Label: Computed body mass index categories
Section Name: Calculated Variables
Module Section Number: 8
Question Number: 19
Column: 2049
Type of Variable: Num
SAS Variable Name: _BMI5CAT
Question Prologue:
Question: Four-categories of Body Mass Index (BMI)

Value	Value Label	Frequency	Percentage	Weighted Percentage
1	Underweight Notes: _BMI5 < 1850 (_BMI5 has 2 implied decimal places)	6,925	1.67	1.93
2	Normal Weight Notes: 1850 <= _BMI5 < 2500	128,856	31.16	32.66
3	Overweight Notes: 2500 <= _BMI5 < 3000	149,148	36.06	35.30
4	Obese Notes: 3000 <= _BMI5 < 9999	128,641	31.11	30.10
BLANK	Don't know/Refused/Missing Notes: _BMI5 = 9999	36,446	.	.

Figure 15: BMI Statistics taken from 2017 BRFSS Codebook

Label: Length of time since last routine checkup
Section Name: Health Care Access
Core Section Number: 3
Question Number: 4
Column: 100
Type of Variable: Num
SAS Variable Name: CHECKUP1
Question Prologue:
Question: About how long has it been since you last visited a doctor for a routine checkup? [A routine checkup is a general physical exam, not an exam for a specific injury, illness, or condition.]

Value	Value Label	Frequency	Percentage	Weighted Percentage
1	Within past year (anytime less than 12 months ago)	331,913	73.76	69.74
2	Within past 2 years (1 year but less than 2 years ago)	50,812	11.29	12.78
3	Within past 5 years (2 years but less than 5 years ago)	29,656	6.59	7.98
4	5 or more years ago	27,861	6.19	6.94
7	Don't know/Not sure	5,260	1.17	1.21
8	Never	3,988	0.89	1.22
9	Refused	525	0.12	0.12
BLANK	Not asked or Missing	1	.	.

Figure 16: Check Up Statistics taken from 2017 BRFSS Codebook

- Values of 7,8,9 and Blank were removed from the data set.

Label: (Ever told) you have diabetes
Section Name: Chronic Health Conditions
Core Section Number: 6
Question Number: 12
Column: 117
Type of Variable: Num
SAS Variable Name: DIABETE3
Question Prologue:
Question: (Ever told) you have diabetes (If ´Yes´ and respondent is female, ask ´Was this only when you were pregnant?´. If Respondent says pre-diabetes or borderline diabetes, use response code 4.)

Value	Value Label	Frequency	Percentage	Weighted Percentage
1	Yes	60,440	13.43	10.87
2	Yes, but female told only during pregnancy—Go to Section 07.01 LMTJOIN3	3,283	0.73	0.91
3	No—Go to Section 07.01 LMTJOIN3	377,264	83.83	86.14
4	No, pre-diabetes or borderline diabetes—Go to Section 07.01 LMTJOIN3	8,232	1.83	1.87
7	Don't know/Not Sure—Go to Section 07.01 LMTJOIN3	594	0.13	0.17
9	Refused—Go to Section 07.01 LMTJOIN3	199	0.04	0.04
BLANK	Not asked or Missing	4	.	.

Figure 17: Diabetes Statistics taken from 2017 BRFSS Codebook

- Values of 1,2 were binned and labeled "Positive", Values of 3,4 were binned and Labeled "negative". Values of 7,9, Blank were removed from the dataset.

Label: Leisure Time Physical Activity Calculated Variable
Section Name: Calculated Variables
Module Section Number: 13
Question Number: 1
Column: 2115
Type of Variable: Num
SAS Variable Name: _TOTINDA
Question Prologue:
Question: Adults who reported doing physical activity or exercise during the past 30 days other than their regular job

Value	Value Label	Frequency	Percentage	Weighted Percentage
1	Had physical activity or exercise Notes: EXERANY2 = 1	301,030	66.89	66.15
2	No physical activity or exercise in last 30 days Notes: EXERANY2 = 2	114,783	25.51	24.35
9	Don't know/Refused/Missing Notes: EXERANY2 = 7 or 9 or Missing	34,203	7.60	9.50

Figure 18: Activity Status Statistics taken from 2017 BRFSS Codebook

Label: Computed Smoking Status
Section Name: Calculated Variables
Module Section Number: 9
Question Number: 1
Column: 2054
Type of Variable: Num
SAS Variable Name: _SMOKER3
Question Prologue:
Question: Four-level smoker status: Everyday smoker, Someday smoker, Former smoker, Non-smoker

Value	Value Label	Frequency	Percentage	Weighted Percentage
1	Current smoker - now smokes every day Notes: SMOKE100 = 1 and SMOKEDAY = 1	44,617	9.91	10.44
2	Current smoker - now smokes some days Notes: SMOKE100 = 1 and SMOKEDAY = 2	18,849	4.19	5.06
3	Former smoker Notes: SMOKE100 = 1 and SMOKEDAY = 3	121,722	27.05	22.69
4	Never smoked Notes: SMOKE100 = 2	246,104	54.69	56.75
9	Don't know/Refused/Missing Notes: SMOKE100 = 1 and SMOKEDAY = 9 or SMOKE100 = 7 or 9 or Missing	18,724	4.16	5.05

Figure 19: Smoking Statistics taken from 2017 BRFSS Codebook

Label: Heavy Alcohol Consumption Calculated Variable
Section Name: Calculated Variables
Module Section Number: 11
Question Number: 5
Column: 2068
Type of Variable: Num
SAS Variable Name: _RFDRHV5
Question Prologue:
Question: Heavy drinkers (adult men having more than 14 drinks per week and adult women having more than 7 drinks per week)

Value	Value Label	Frequency	Percentage	Weighted Percentage
1	No Notes: SEX = 1 and _DRNKWEK <= 14 or SEX = 2 and _DRNKWEK <= 7 or ALCDAY5 = 888	395,377	87.86	86.11
2	Yes Notes: SEX = 1 and _DRNKWEK > 14 or SEX = 2 and _DRNKWEK > 7	24,459	5.44	5.68
9	Don't know/Refused/Missing Notes: _DRNKWEK = 99900	30,180	6.71	8.21

Figure 20: Heavy Drinker Statistics taken from 2017 BRFSS Codebook

APPENDIX C: MODEL DUMMY VARIABLES

Table 4: Model Variables Count Frequencies

Variable	0	1	Percent Positive Response (=1)	Total
0-25	6,269	3,731	37 percent	10,000
25-35	9,268	732	7 percent	10,000
35-45	8,905	1,095	11 percent	10,000
45-55	8,693	1,307	13 percent	10,000
5 or More years	9,825	175	2 percent	10,000
55-65	8,695	1,305	13 percent	10,000
65-101	8,170	1,830	18 percent	10,000
75k or More	7,460	2,540	25 percent	10,000
Asian	9,893	107	1 percent	10,000
Black	9,230	770	8 percent	10,000
College Graduate	3,817	6,183	62 percent	10,000
Current Smoker	8,741	1,259	13 percent	10,000
Employed	4,636	5,364	54 percent	10,000
Excellent	8,097	1,903	19 percent	10,000
Fair	8,652	1,348	13 percent	10,000
Female	3,982	6,018	60 percent	10,000
Former Smoker	7,186	2,814	28 percent	10,000
Good	7,064	2,936	29 percent	10,000
Heavy Drinker	9,457	543	5 percent	10,000
High School Graduate	7,081	2,919	29 percent	10,000
Hispanic	9,233	767	8 percent	10,000
Less than 10k	9,474	526	5 percent	10,000
Less than 15k	9,455	545	5 percent	10,000
Less than 20k	9,306	694	7 percent	10,000
Less than 25k	9,063	937	9 percent	10,000
Less than 35k	8,818	1,182	12 percent	10,000
Less than 50k	8,609	1,391	14 percent	10,000
Less than 75k	8,473	1,527	15 percent	10,000
Male	6,018	3,982	40 percent	10,000
Married	4,593	5,407	54 percent	10,000
Multi-Racial	9,841	159	2 percent	10,000
negative	1,123	8,877	89 percent	10,000
Never Smoked	4,582	5,418	54 percent	10,000
No School	9,107	893	9 percent	10,000
Normal Weight	6,623	3,377	34 percent	10,000
Not Heavy Drinker	543	9,457	95 percent	10,000
Not Married	5,407	4,593	46 percent	10,000
Not Recently Active	7,418	2,582	26 percent	10,000
Obese	8,589	1,411	14 percent	10,000
Other	9,954	46	0 percent	10,000
Overweight	6,988	3,012	30 percent	10,000
Poor	9,455	545	5 percent	10,000
Positive	8,877	1,123	11 percent	10,000
Recently Active	2,582	7,418	74 percent	10,000
Smokes Some	9,532	468	5 percent	10,000
Underweight	7,983	2,017	20 percent	10,000
Unemployed	5,364	4,636	46 percent	10,000
Very Good	6,746	3,254	33 percent	10,000
White	2,218	7,782	78 percent	10,000
Within 1 year	2,334	7,666	77 percent	10,000
Within 2 years	8,996	1,004	10 percent	10,000
Within 5 Years	8,845	1,155	12 percent	10,000

Table 5: Research / Model variable descriptions.

Research Independent Variables	Model Variables	Description
Age	v1	0-25
	v2	25-35
	v3	35-45
	v4	45-55
	v6	55-65
	v7	65-101
Income	lessthan10k	Income Less than 10k
	lessthan15k	Income Less than 15k
	lessthan20k	Income Less than 20k
	lessthan25k	Income Less than 25k
	lessthan35k	Income Less than 35k
	lessthan50k	Income Less than 50k
	lessthan75k	Income Less than 75k
	75kormore	Income More than 75k
	Don't know	Don't know their Income
Race	American Indian	Race - Variable Dropped
	Asian	
	Black	
	Hispanic	
	Multiracial	
	Native Hawaiian	Race - Variable Dropped
	Other	Race - Variable Dropped
	White	
Education	College graduate	
	High School graduate	
	No school	
Sex	Female	
	Male	
Employment Status	Employed	
	Unemployed	
Marital Status	Married	
	Not Married	Binned variable -
General Health	Excellent	Self-assessed general health as excellent
	Very Good	Self-assessed general health as very good
	Good	Self-assessed general health as good
	Fair	Self-assessed general health as fair
	Poor	Self-assessed general health as poor
BMI Category	Obese	Based on BMI Score Height/Weight/Age Cat
	Overweight	Based on BMI Score Height/Weight/Age Cat
	Normal Weight	Based on BMI Score Height/Weight/Age Cat
	Underweight	Based on BMI Score Height/Weight/Age Cat
Check UP	Within1year	Time since last check up
	Within2years	Time since last check up
	Within5years	Time since last check up
	5ormoreyears	Time since last check up
Diabetes	Positive	Diagnosed with Diabetes
	Negative	Not Diagnosed with Diabetes
Activity Status	Recently active	
	Not Recently Active	
Smoking	Current smoker	
	Former smoker	
	Smokes some	
	Never smoked	
Heavy Drinking	Not Heavy Drinker	
	Heavy Drinker	
year	1993-2017	

APPENDIX D: MODEL RESULTS

Zero-inflated negative binomial regression — Number of obs = 10,000; Nonzero obs = 3,267; Zero obs = 6,733

Inflation model = logit — Wald chi2(21) = 325.59

Log pseudolikelihood = -16721.83 — Prob > chi2 = 0.0000

mentalhealth	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
mentalhealth						
v2	-.2345211	.0767149	-3.06	0.002	-.3848794	-.0841627
v3	.0884784	.063549	1.39	0.164	-.0360753	.2130321
v4	.1012844	.0577697	1.75	0.080	-.0119422	.214511
v6	.0495003	.0650278	0.76	0.447	-.0779519	.1769526
v7	-.0601077	.0668387	-0.90	0.368	-.1911091	.0708938
asian	-.3045686	.2647667	-1.15	0.250	-.8235018	.2143647
black	-.2042248	.077924	-2.62	0.009	-.356953	-.0514966
hispanic	.1188048	.0659068	1.80	0.071	-.0103701	.2479797
multiracial	.209979	.120287	1.75	0.081	-.0257792	.4457372
lessthan15k	-.0663723	.0703037	-0.94	0.345	-.2041651	.0714205
lessthan20k	-.0136426	.0724047	-0.19	0.851	-.1555532	.128268
lessthan25k	-.0651751	.0687492	-0.95	0.343	-.199921	.0695709
lessthan35k	-.2303483	.0755722	-3.05	0.002	-.3784672	-.0822295
lessthan50k	-.2991177	.0754011	-3.97	0.000	-.4469011	-.1513343
lessthan75k	-.3571852	.0755745	-4.73	0.000	-.5053084	-.2090619
kormore	-.5148937	.0775136	-6.64	0.000	-.6668175	-.3629699
collegegraduate	-.1892302	.0602712	-3.14	0.002	-.3073596	-.0711007
highschoolgraduate	-.1120776	.0615415	-1.82	0.069	-.2326968	.0085416
employed	-.2851156	.0454503	-6.27	0.000	-.3741966	-.1960346
female	.0091071	.0433583	0.21	0.834	-.0758737	.0940878
married	-.1043254	.0438303	-2.38	0.017	-.1902312	-.0184197
_cons	2.733345	.072212	37.85	0.000	2.591812	2.874878

Figure 21: Model 1 - Socioeconomic predictors

inflate						
v2	-.532186	.1056444	-5.04	0.000	-.7392453	-.3251267
v3	-.4131731	.0854973	-4.83	0.000	-.5807447	-.2456016
v4	-.2313067	.0790422	-2.93	0.003	-.3862265	-.0763869
v6	.0596692	.0784581	0.76	0.447	-.0941058	.2134442
v7	.9145796	.0782704	11.68	0.000	.7611725	1.067987
asian	.0840059	.2729549	0.31	0.758	-.4509758	.6189876
black	.3131045	.0965645	3.24	0.001	.1238415	.5023674
hispanic	.2528819	.0961781	2.63	0.009	.0643762	.4413875
multiracial	-.2011787	.1901709	-1.06	0.290	-.5739068	.1715494
lessthan15k	-.1231011	.1219351	-1.01	0.313	-.3620895	.1158873
lessthan20k	.0705739	.1132534	0.62	0.533	-.1513988	.2925465
lessthan25k	.1795205	.1055047	1.70	0.089	-.027265	.3863059
lessthan35k	.2850117	.1029167	2.77	0.006	.0832987	.4867247
lessthan50k	.2788439	.1015861	2.74	0.006	.0797388	.477949
lessthan75k	.2884623	.1042047	2.77	0.006	.0842248	.4926997
kormore	.5462247	.1017068	5.37	0.000	.346883	.7455663
collegegraduate	.1001445	.0944621	1.06	0.289	-.0849978	.2852868
highschoolgraduate	.2017201	.0957902	2.11	0.035	.0139748	.3894655
employed	.0526368	.0589385	0.89	0.372	-.0628805	.1681541
female	-.5140836	.0521499	-9.86	0.000	-.6162956	-.4118716
married	.2278578	.0558814	4.08	0.000	.1183323	.3373834
_cons	.1504237	.1116752	1.35	0.178	-.0684557	.369303
/lnalpha	.3112632	.0400203	7.78	0.000	.2328248	.3897017
alpha	1.365149	.0546337			1.26216	1.47654

Figure 22: Model 1 - Socioeconomic predictors - inflated "certain zeros"

Zero-inflated negative binomial regression — Number of obs = 10,000; Nonzero obs = 3,267; Zero obs = 6,733

Inflation model = logit — Wald chi2(15) = 698.40

Log pseudolikelihood = -16609.92 — Prob > chi2 = 0.0000

mentalhealth	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
mentalhealth						
within1year	-.339328	.1175177	-2.89	0.004	-.5696583	-.1089976
within2years	-.3946668	.1302155	-3.03	0.002	-.6498844	-.1394491
within5years	-.2168635	.1252805	-1.73	0.083	-.4624087	.0286817
neversmoked	-.3920289	.0514209	-7.62	0.000	-.492812	-.2912459
formersmoker	-.3326557	.0560608	-5.93	0.000	-.4425329	-.2227785
smokessome	-.1208138	.0789821	-1.53	0.126	-.2756159	.0339883
notheavydrinker	-.0088147	.0936714	-0.09	0.925	-.1924072	.1747778
recentlyactive	-.1559617	.0420905	-3.71	0.000	-.2384576	-.0734657
overweight	-.0089185	.0555331	-0.16	0.872	-.1177614	.0999244
normalweight	-.1166903	.0567405	-2.06	0.040	-.2278996	-.005481
underweight	-.1124378	.0654681	-1.72	0.086	-.240753	.0158774
excellent	-1.079635	.0851041	-12.69	0.000	-1.246436	-.9128341
verygood	-.9304054	.0566407	-16.43	0.000	-1.041419	-.8193917
good	-.6476442	.0511614	-12.66	0.000	-.7479187	-.5473696
fair	-.2797798	.0486457	-5.75	0.000	-.3751237	-.1844358
_cons	3.586452	.151928	23.61	0.000	3.288679	3.884226

Figure 23: Model 2 - General Health Predictors

inflate						
within1year	-.0889174	.1761832	-0.50	0.614	-.4342301	.2563953
within2years	-.5712781	.1907099	-3.00	0.003	-.9450627	-.1974936
within5years	-.367401	.186969	-1.97	0.049	-.7338535	-.0009486
neversmoked	.301444	.0733645	4.11	0.000	.1576522	.4452358
formersmoker	.4530618	.0791753	5.72	0.000	.297881	.6082425
smokessome	-.1822809	.1246464	-1.46	0.144	-.4265834	.0620217
notheavydrinker	.2557294	.106512	2.40	0.016	.0469696	.4644891
recentlyactive	-.0823869	.0568562	-1.45	0.147	-.1938231	.0290493
overweight	.0429948	.0730139	0.59	0.556	-.1001098	.1860994
normalweight	.0930232	.0733554	1.27	0.205	-.0507508	.2367973
underweight	-.0056246	.0825828	-0.07	0.946	-.1674839	.1562346
excellent	1.71291	.1222226	14.01	0.000	1.473358	1.952462
verygood	1.260188	.111354	11.32	0.000	1.041939	1.478438
good	1.117163	.1097225	10.18	0.000	.9021109	1.332215
fair	.6119377	.1163459	5.26	0.000	.383904	.8399715
_cons	-.9762682	.2294144	-4.26	0.000	-1.425912	-.5266243
/lnalpha	.1819341	.0401267	4.53	0.000	.1032872	.2605811
alpha	1.199535	.0481334			1.10881	1.297684

Figure 24: Model 2 - General Health Predictors - inflated "certain zeros"

inflate						
v2	**-.623947**	**.1083932**	**-5.76**	**0.000**	**-.8363939**	**-.4115002**
v3	**-.3954346**	**.0891049**	**-4.44**	**0.000**	**-.5700769**	**-.2207923**
v4	**-.1561876**	**.0829275**	**-1.88**	**0.060**	**-.3187225**	**.0063472**
v6	**.1311137**	**.0819012**	**1.60**	**0.109**	**-.0294097**	**.2916371**
v7	**.929043**	**.0839711**	**11.06**	**0.000**	**.7644626**	**1.093623**
asian	**.1015021**	**.2751571**	**0.37**	**0.712**	**-.4377959**	**.6408001**
black	**.3163416**	**.0988868**	**3.20**	**0.001**	**.122527**	**.5101563**
hispanic	**.316141**	**.0988023**	**3.20**	**0.001**	**.1224922**	**.5097899**
multiracial	**-.1096332**	**.1933416**	**-0.57**	**0.571**	**-.4885758**	**.2693094**
lessthan15k	**-.0366066**	**.1253862**	**-0.29**	**0.770**	**-.2823589**	**.2091458**
lessthan20k	**.0616117**	**.1152599**	**0.53**	**0.593**	**-.1642936**	**.2875169**
lessthan25k	**.1080256**	**.1077938**	**1.00**	**0.316**	**-.1032465**	**.3192976**
lessthan35k	**.1321693**	**.105285**	**1.26**	**0.209**	**-.0741855**	**.338524**
lessthan50k	**.0904752**	**.1050756**	**0.86**	**0.389**	**-.1154692**	**.2964197**
lessthan75k	**.0596759**	**.1066279**	**0.56**	**0.576**	**-.1493109**	**.2686627**
kormore	**.2331971**	**.1048233**	**2.22**	**0.026**	**.0277472**	**.438647**
collegegraduate	**-.1338665**	**.0978927**	**-1.37**	**0.171**	**-.3257325**	**.0579996**
highschoolgraduate	**.0733703**	**.0982785**	**0.75**	**0.455**	**-.1192521**	**.2659927**
employed	**-.0781742**	**.0612885**	**-1.28**	**0.202**	**-.1982974**	**.041949**
female	**-.5995817**	**.0542918**	**-11.04**	**0.000**	**-.7059917**	**-.4931717**
married	**.1904439**	**.0567424**	**3.36**	**0.001**	**.0792308**	**.3016571**
within1year	**-.1894866**	**.1747648**	**-1.08**	**0.278**	**-.5320193**	**.1530461**
within2years	**-.6161033**	**.190436**	**-3.24**	**0.001**	**-.9893509**	**-.2428556**
within5years	**-.4220705**	**.1857442**	**-2.27**	**0.023**	**-.7861225**	**-.0580185**
neversmoked	**.2254729**	**.0783918**	**2.88**	**0.004**	**.0718278**	**.3791179**
formersmoker	**.2124445**	**.0847275**	**2.51**	**0.012**	**.0463817**	**.3785073**
smokessome	**-.23033**	**.1291132**	**-1.78**	**0.074**	**-.4833872**	**.0227272**
notheavydrinker	**.2051396**	**.108825**	**1.89**	**0.059**	**-.0081535**	**.4184328**
recentlyactive	**-.0517234**	**.0596044**	**-0.87**	**0.386**	**-.1685459**	**.0650991**
overweight	**-.0267886**	**.0765487**	**-0.35**	**0.726**	**-.1768214**	**.1232441**
normalweight	**-.0041531**	**.0784553**	**-0.05**	**0.958**	**-.1579225**	**.1496164**
underweight	**-.0130911**	**.0916204**	**-0.14**	**0.886**	**-.1926637**	**.1664815**
excellent	**2.012389**	**.133819**	**15.04**	**0.000**	**1.750109**	**2.27467**
verygood	**1.487296**	**.1215192**	**12.24**	**0.000**	**1.249123**	**1.725469**
good	**1.270379**	**.1178954**	**10.78**	**0.000**	**1.039308**	**1.501449**
fair	**.6755366**	**.1215812**	**5.56**	**0.000**	**.4372417**	**.9138314**
_cons	**-.7087426**	**.2492974**	**-2.84**	**0.004**	**-1.197357**	**-.2201288**
/lnalpha	**.1364862**	**.0396342**	**3.44**	**0.001**	**.0588046**	**.2141678**
alpha	**1.146239**	**.0454303**			**1.060568**	**1.238831**

Figure 25: Model 3 - Complete model - inflated "certain zeros"