Spring 2019

Effect of Severe Economic Recession on the Psychological Distress: Evidence of Modifying Effect of Risky Behaviors and Insurance Status

Lumi Bakos

EFFECT OF SEVERE ECONOMIC RECESSION ON THE PSYCHOLOGICAL DISTRESS: EVIDENCE OF MODIFYING EFFECT OF RISKY BEHAVIORS AND INSURANCE STATUS

by

Lumi Bakos

Bachelor of Science
Youngstown State University, 2001

Master of Arts
Youngstown State University, 2002

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Health Services Policy and Management

The Norman J. Arnold School of Public Health

University of South Carolina

2019

Accepted by:

Ronnie Horner, Major Professor

Christine DiStefano, Committee Member

Elizabeth Crouch, Committee Member

Ryan Carlson, Committee Member

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

DEDICATION

I dedicate this work to my husband Dr. Jason Bakos for all the patience, guidance and love he has given me over the years. He has been my biggest supporter. I also want to thank my children, Jade, Justin, and Julian for showing me that life is more than just a dissertation. I could have not done it without you.

ACKNOWLEDGEMENTS

I would like to thank Dr. Ronnie Horner, Dr. Christine DiStefano, Dr. Elizabeth Crouch, and Dr. Ryan Carlson for helping me see the light at the end of the tunnel. You have graciously have stepped in and helped me finish my dissertation for that I am forever in your debt.

A special THANK YOU to Dr. Christine DiStefano for her patience and encouragement throughout this process. I would have never done it without your support.

ABSTRACT

The investigation identified the effect of the 2007-2009 economic recession on the level of psychological distress (PD) as modified by engaging in risky behaviors and insurance status. The objective was to determine if engaging in risky behaviors and insurance status had an impact upon the level of psychological distress experienced during a recession; the timeframe for the study is 2007 representing pre-recession to beginning of recession year, to 2009 representing the recession year, to 2012-2013 representing post-recession years. The study involved the use of a nationwide non-institutionalized adult population, the Mental Health and Stigma optional module found in the BRFSS (Behavioral Risk Factor Surveillance System) survey available in years 2007, 2009, 2012, and 2013. Because the datasets are cross-sectional, this research used the Integrative Data Analysis (IDA) framework. IDA allows for analysis of multiple independent data sets that are pooled into one. Here, pooling was conducted by matching demographic characteristics of participants (sex, race, age, income, and education level) across waves to provide a view of target variables at the population level. Our research found that income had a small effect on PD, pre-, during, and post-recession with no significant effect on the association between PD over time. Increased level of uninsured (percentage of individuals without insurance) had a significant positive effect on PD. Alcohol consumption had a small negative effect on PD with no disruption in the association between psychological distress over time. Smoking had a larger positive effect on PD with a decrease of percentage of individuals smoking from 2009 to 2012

which could have played a part in the disruption of the relationship of PD between the years. In addition, we found that low psychological distress decreased smoking while moderate and severe psychological distress increased smoking. This is an important finding for smoking cessation researchers, policy makers, and medical personnel to review their policies and programs on smoking cessation.

Policymakers along with mental health providers and insurance companies should investigate the possibility of recognizing PD as a mental illness and include moderate to severe psychological distress as a coverage option.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

K	Interval of telephone numbers.
N	Population count of telephone numbers in the frame.
n	Desired Sample Size.
K_i	Number of non-missing observed variables in row i of the data set.
$\sum_i$	Model-implied covariance matrix.
x_i	Individual's score vector.
μ_i	Population mean vector.
x	Column vector of the q observed variables.
Λ_x	q x n matrix of fixed and estimated loadings of the n factors.
ζ	Column vector of latent variables.
δ	Column of vector uniqueness (errors).
λ	Estimated factor loadings.

LIST OF ABBREVIATIONS

AUC ....Area Under Receiver Operating Characteristic Curve

BMI .... Body Mass Index

BRFSS..Behavioral Risk Factor Surveillance System

CDC .... Centers for Disease Control

CFA..Confirmatory Factor Analysis

CFI .... Comparative Fit Index

CHIS .... California Health Interview Survey

CI...Confidence Interval

CIDI-SF...................................Composite International Diagnostic Interview Short Form

DF .... Degrees of Freedom

DSM-5...Diagnostic and Statistical Manual 5th Edition

DSS .... Disproportional Stratified Sample

DUI .... Driving under the Influence

DWI.. Driving While Intoxicated

EFA .... Exploratory Factor Analysis

FIML ....Full Information Maximum Likelihood

GAD .... Generalized Anxiety Disorder

GAF.. Global Assessment of Functioning Scale

GFI .... Goodness of Fit

HC .... Healthcare

IDA ....Integrative Data Analysis

IRT .... Item Response Theory

K6/K10.. Kessler 6/10 Psychological Distress Scale

LGM...Latent Growth Modeling

MAR .... Missing at Random

MCAR..Missing Completely at Random

MEPS ....Medical Expenditure Panel Survey

MHPAEA..Mental Health Parity and Addiction Equity Act

MI..Measurement Invariance

ML.. Maximum Likelihood

MLM ....Maximum Likelihood Estimation with Robust Estimates

NHIS .... National Health Interview Survey

NHANES ....National Health and Nutrition Examination Survey

NFI ....Normed Fit Index

NNFI ....Non-Normed Fit Index

NSDUH.. National Survey of Drug Use and Health

PAF .... Principal Axis Factor

PC...Principal Components

PD .... Psychological Distress

ROC ....Receiver Operating Characteristic

RMSEA... Root Mean Square Error of Approximation

SCID ....Structured Clinical Interview for DSM-IV

SEM ....Structural Equation Modeling

SES... Socio Economic Status

SMI ....Severe Mental Illness

SPD .... Serious Psychological Distress

SRMR ....Standardized Root Mean Square Residual

TLI ....Tucker-Lewis Index

WHO ....World Health Organization

WHO-DAS............................ World Health Organization Disability Assessment Schedule

YPLL...Years of Potential Life Lost

CHAPTER 1

STUDY BACKGROUND AND SIGNIFICANCE

1.1. INTRODUCTION

In mid-2007, the economic collapse triggered a recession that lasted in the U.S. through mid-2009. The recession was marked by a sharp downturn of US stocks, high foreclosure rates, and increasing unemployment rates (US Labor Statistics, 2016). During the period between December 1, 2007 and June 1, 2009, the U.S. unemployment rate rose from 5% to 9.5%, a higher increase than any in recent decades (US Labor Statistics, 2016). Income losses coupled with housing foreclosures increased the risk for mental health issues, including long term depression and anxiety/stress, and led to increased suicide mortality risk (Barr, 2012; McLaughlin, 2012). The recession was also associated with lower health care utilization (Mortensen & Chen, 2013; Travers, et al., 2017).

Prior work has found associations between economic or financial stress and tobacco and alcohol use (Davalos et al., 2012; Dee, 2001; Compton et al., 2014). Unemployed people are more likely to report heavy alcohol use, tobacco use, and alcohol abuse or dependence, even when adjusting analyses for major depression an1d socio-demographic variables (Compton et al., 2014). During the 2007-2009 recession, the combination of increased unemployment and the burst of the housing bubble caused an increase in the incidence of mental problems among the adult population in the United States. Previous research examined the association between mental distress and smoking and alcohol consumption, and other work examined the effect of recession on mental wellbeing (McLaughlin et al., 2012; Alley et al., 2011; Pollack et al., 2011; Lasser et al., Le Cook et al, 2014), but to our knowledge none of these previous efforts have researched the effect of risky behaviors (represented by alcohol consumption and cigarette smoking) and insurance status on the level of psychological stress during economic recession. Previous research examined the association between risky behaviors and clinical mental illness (Lasser et al., 2000; Miles et al., 2003; Sanchez-Villegas et al., 2008), but not the association between risky behaviors, insurance status and psychological distress, especially in the context of the 2007-2009 economic recession. By researching these associations one can understand how psychological distress brought on by economic distress might be modified (increase or decrease) by risky behaviors and insurance status in the general population.

1.2. STUDY BACKGROUND

The purpose of this study was to investigate the extent to which the relationship between risky behaviors (represented by smoking/drinking alcohol) affected the level of psychological distress among adult population in times of high economic stress (recession) versus low economic stress (non-recession) and whether psychological distress is further mediated by insurance status that may affect access to mental health services.

To complete the study, a multiyear data set from a large-scale national survey was used, the Behavioral Risk Factor Surveillance System (BRFSS); more specifically, the Kessler-6 survey questions as part of the Mental Health and Stigma Module were the key data elements. BRFSS is a self-reported survey that collects behavioral risk data at the state and local level of more than 400,000 adults each year. BRFSS surveys are used to collect behavioral health risk data at the state and local level and it has been a powerful tool in decision making and building health promotion activities. As a result, multiple studies examined the methods, reliability and validity of the data and concluded that the BRFSS data are reliable, and the methods are valid (Adams, Park, & Irwin, 2015; Burger & Reither, 2014; Dwyer-Lindgre, et al., 2013; Elliott & West, 2015). Regarding mental health status, Andersen et al., 2003 and Kapp et al., 2009 found moderate to excellent reliability across quality of life measures and measures of mental distress.

1.3. CONTRIBUTION TO EXISTING LITERATURE

The scales used, and outcomes identified by studies of risky behaviors/insurance status on mental health vary. Many of these studies have looked at the association between risky behaviors such as alcohol consumption and mental health assessed with the Short Form 36 mental health component score (Bell & Britton, 2014); tobacco use and depression symptoms as assessed by the National Health and Nutrition Examination Survey (NHANES) symptom-based questionnaire (Pratt & Brody, 2010); smoking and tobacco cessation and mental illness as defined by the Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised (Lasser et al., 2000); or recession and depressive symptoms as assessed by the Center of Epidemiologic Studies Depression scale (McInerney et al., 2013; Riumallo-Herl et al., 2014). Some have even looked at barriers to mental illness treatment (Rowen et al., 2013, Walker et al., 2015; van Beljouw et al., 2010). They have found that individuals with mental health issues were more likely to have public insurance and gaps in mental health coverage created cost barriers to mental health care among public and private insurance policyholders.

No study has examined the effect of risky behaviors and insurance status on psychological distress before, during, and after the U.S. recession of 2007. This is an important research topic because if left untreated, serious psychological distress (psychological discomfort) may progress into serious mental illness (diagnosable mental illness) and, if economic recession heightens the likelihood of psychological distress, a large number of citizens will be put at risk. Understanding the association between psychological distress and risky behavior can be useful in addressing government programs designed to deal with recession impacts on the workforce that can either be newly implemented or modified to address issues related to coverage of mental health services and risky behaviors.

In addition, most of the existing literature used regression analysis to examine the hypothesized associations. A unique contribution of this study is the use of structural equation modeling (SEM) tools in conjunction with the integrative data analysis (IDA) framework to examine associations over time between latent/unobserved variables (psychological distress, depression symptoms, and anxiety symptoms). SEM is a general, powerful multivariate technique that allows for examination of the relationship between a latent variable (variable that cannot be directly measured) and an observed variable. In addition, SEM allows for analysis of complex and dynamic relationships between observed and unobserved variables. The technique allows for multiple, cross-sectional datasets to be pooled into a single dataset to examine relations between different participant groups, different scales, and/or across time. (Curran & Hussong, 2009).

The unique features of this study include:

1. Use of nationwide, multistate, datasets from multiple years in the analysis.

2. Inclusion of 4 periods in time namely 2007, 2009, 2012, and 2013 (before, during, and after U.S. recession).
3. An examination of psychological distress trends during the economic recession.
4. Use of the integrative data framework to model trends at the population level by using multiple years of cross-sectional data.
5. Deepening of the existing literature by providing further evidence of the validity of the K6 scale.
6. Examination of the relationship between alcohol and tobacco use and psychological distress in the general population during economic hardship.

1.4. LIMITATIONS OF THE STUDY

There are several limitations of the study which include survey design limitations and methodological limitations. One limitation of the BRFSS survey is the use of a cross-sectional design. The BRFSS provides a general, cross-sectional view of the population's health; however, the same individuals are not assessed at different time points. Here, using the IDA method to link the multiple cross-sectional datasets allows for generalization of results at population level. In addition, the BRFSS survey is a multi-part survey asking questions on general conditions and not specific conditions. As a result, additional longitudinal survey analysis that looks at specific mental illness will be needed to find population-specific or cause-specific reasons behind associations found by this research.

Another limitation of the BRFSS survey is recall bias. The survey asks raters to report information over the past 30 days, over the past years, or questions pertaining to certain health problems. The responder might have a hard time remembering an event or may be hesitant to report illnesses or behaviors that have negative connotations, thus causing bias in the dataset.

The Mental Illness and Stigma survey is an optional module in the BRFSS survey measuring psychological distress via depression and anxiety symptoms questions. It does not measure any specific mental illness. Further, the available sample is restricted as not all the states elected to use it. Hence the geographical differences and the gap in survey years used with these analyses.

The current research examines the association between psychological distress and risky behaviors (smoking and alcohol use) during economic recession as represented by the Mental Illness and Stigma survey years and does not account for the severity of the economic loss on the population (i.e., unemployment level, salary suppression, housing foreclosure level, etc.). Also, this study is using a novel methodological design, integrative data analysis, to identify general population trends. Finally, the purpose of this research is not to examine serious and persistent mental illness as defined by the Diagnostic and Statistical Manual of Mental Disorder 5th Edition (American Psychiatric Association, 2013).

1.5. CONCLUSION

Mental health of the population is an important public health topic in the United States, with an estimated 18.1 percent or 43.6 million adults 18 years or older reporting any mental illness in the past year in 2014, 4.1 percent with serious mental illness (Center for Behavioral Health Statistics and Quality, 2015), and 14% with minor mental problems. Research has found that depression is a risk factor for diabetes and heart disease (Druss & Walker, 2011; Edge et al., 2002), and obesity (Scott et al., 2008; Clum et al., 2014). In addition, mental illness has a significant impact on quality of life for the individual and his or her family (Horwath et al, 1992). Prior literature found that 15-32 years of life are lost to mental health problems (Lutterman et al, 2003; Colton et al, 2006; Miller et al, 2006). In general, people with mental health issues who have co-morbidities incur higher health care costs compared to people who don't have mental health issues with the same co-morbidities (Gameroff & Olfson, 2006; Kim & Lee, 2006). The recession 2007-2009 created a life of increased stress for many Americans with feelings of shame, loss, and regret (Libman et al., 2012), and even suicide (Houle & Light; 2014).

Research suggests that depression and anxiety are each a risk factor for smoking and alcohol consumption. In 2014, nearly 17 of every 100 adults (16.8 percent) smoked cigarettes (CDC, 2015), and 1 in 6 adults (17.1 percent) were binge drinking in 2010 (CDC, 2012). Both smoking and alcohol consumption are modifiable behaviors in the general population, but when they are coupled with mental health issues these behaviors are not as easily modifiable (Pratt & Brody, 2010; Lasser et al, 2000). Prior research tried to find the link between these behaviors and mental health and questioned its temporality (Bell & Britton, 2014). The question of temporality still remains: Do people with mental health issues consume more alcohol and smoke cigarettes to alleviate mental health issues or does engaging in unhealthy behaviors (which are associated with other diseases such as heart disease) cause mental health issues (i.e. stress, depression, anxiety)? Our research is trying to identify such associations and if found it could have a significant impact on workplace and healthcare policies.

CHAPTER 2

LITERATURE REVIEW

This chapter begins with a brief review of mental health terminology, a review of literature that examines the current state of mental health in the United States population, and a review of literature that examines the consequences of mental illness. Next, it describes the background of the 2007-2009 economic recession, specifically with regard to the housing bubble, its effect on the US economy and families, and literature that examines the mental health effects and coping mechanisms of the population during similar events.

2.1. MENTAL HEALTH AND MENTAL ILLNESS

Mental health is "a state of well-being in which the individual realizes his or her own abilities, can cope with the normal stress of life, can work productively and fruitfully, and is able to make a contribution to his or her community" (Centers for Disease Control and Prevention, 2013; World Health Organization, 2001). Mental illness is defined as "collectively all diagnosable mental disorders" or "health conditions that are characterized by alterations in thinking, mood, behavior (or some combination thereof) associated with distress and/or impaired functioning" (US Department of Health and Human Services, 1999; Centers for Disease Control and Prevention, 2013). There are five mental illness categories recognized by professional psychiatry experts characterized by mood, thought, and/or behavior, and diagnosed based on the Diagnostic and Statistical

Manual 5th edition revised (DSM-5) (American Psychiatric Association, 2013): depression, anxiety, psychotic disorders, bipolar disorder, and dementia/Alzheimer's Disease. Depression and anxiety are the most common type of mental illness. It is common for a person to experience occasional periods of sadness without causing substantial impact on the individual's life, while diagnosable depression interferes with daily life and causes anguish to the individual and their families (National Institute of Mental Health, 2016). In order to differentiate these types of depression, the American Psychiatric Association developed minimal criteria to classify a person as depressed: a depressed person must have at least five of the following symptoms for a continuous period of at least two weeks: depressed or sad mood, diminished interest in activities which used to be pleasurable, weight gain or loss, psychomotor agitation or retardation, fatigue, inappropriate guilt, difficulties concentrating, and/or recurrent thoughts of death (American Psychiatric Association, 2013; Centers for Disease Control and Prevention, 2013). The Mental Illness and Stigma survey within the BRFSS contains 6 questions representing depression and anxiety symptoms in the past 30 days. However, there are other clinical criteria used by psychiatrists based on the DSM-5 with the recent revisions. The BRFSS questions serve as a screening tool to identify persons with psychological distress who are at higher risk to qualify for a clinical diagnosis and treatment of depression and anxiety disorder (Kessler et al., 2002; Kessler et al., 2003; Kessler et al., 2010).

2.1.1. Depression

Depression is common, but most people with depression and anxiety either do not have access or do not seek treatment (Centers for Disease Control and Prevention, 2016).

Depression can manifest in varying severities, from minor to severe. This study explores the association between tobacco use, alcohol abuse, and uninsurance on psychological distress (short time frame, within the last 30 days, and symptoms less severe).

Depression is caused by genetic, biological, environmental, or/and psychological factors. DSM-5 diagnostic criteria (American Psychiatric Association, 2013) for major depressive disorder (MDD) includes five or more of the following symptoms present for more than 2 weeks and a change from the person's baseline; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.

1. Depressed mood or irritability most of the day, nearly every day as indicated by either subjective report (e.g., feels sad or empty) or observation made by others (e.g., appears tearful).
2. Decreased interest or pleasure in all, or almost all activities most of each day.
3. Significant weight change (more than 5% of body weight in a month) or change in appetite.
4. Change in sleep: insomnia or hypersomnia nearly every day.
5. Change in activity: psychomotor agitation or retardation nearly every day.
6. Fatigue or loss of energy nearly every day.
7. Feelings of worthlessness or excessive or inappropriate guilt nearly every day.
8. Diminished ability to think or concentrate, or more indecisiveness.
9. Thoughts of death or suicide or has suicide plan.

The Mental Illness and Stigma survey includes questions that assess the level of depression symptoms. The questions are part of a psychological distress scale that was developed specifically to screen for serious mental illness (SMI) in the population

(Kessler et al., 2002). Detailed description of the scale is listed in the Psychological Distress section below. Four of the questions ask respondents how frequently they experienced symptoms of depression in the past 30 days. These depression symptoms are: sadness, hopelessness, worthlessness, and fatigue.

2.1.2. Anxiety

Most people have occasional anxiety caused by an upcoming test, making an important decision, before a presentation, before and during an interview, etc. (National Institute of Mental Health, 2016). This type of anxiety is temporary and normal. On the other hand, the symptoms of more pervasive types of anxiety disorder (for example generalized anxiety disorder or GAD) do not pass after the task is performed. As with depression, these feelings last for a substantial amount of time and interfere with daily life. The DSM-5 criteria (American Psychiatric Association, 2013) for GAD include:

1. Excessive anxiety and worry (apprehensive expectation), occurring most days for at least 6 months, about a number of events or activities (such as work or school performance).
2. The individual finds it difficult to control the worry.
3. The anxiety and worry are associated with three (or more) of the following six symptoms:
    a. Restlessness or feeling keyed up or on edge.
    b. Being easily fatigued.
    c. Difficulty concentrating or mind going blank.
    d. Irritability.
    e. Muscle tension.

f. Sleep disturbance (difficulty falling asleep or staying asleep, or restless unsatisfying sleep).

4. The anxiety, worry, or physical symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
5. The disturbance is not due to the direct of physiological effects of a substance (e.g., a drug abuse, a medication) or a general medical condition (e.g., hyperthyroidism) and does not occur exclusively during a mood disorder, a psychotic disorder, or pervasive developmental disorder.

The Mental Illness and Stigma survey includes two questions that assess the level of anxiety symptoms, intended to serve as a screen for anxiety disorder in the population (Kessler et al., 2002). Detailed description of the scale is listed in the Psychological Distress section below. The questions ask respondents how frequently they experienced symptoms of anxiety in the past 30 days. These anxiety symptoms are: nervousness and feeling restless/fidgety.

2.1.3. Psychological Distress

Psychological distress is a broad term that describes psychological discomfort. At one time or another every person experiences psychological distress of different levels and can be described as feelings of sadness, anxiety, hopelessness. In more extreme cases psychological distress indicates a more severe mental distress. R.C. Kessler developed the Kessler 10 and Kessler 6 (the shorter version) surveys that measure the level of psychological distress in adults. Research has shown that the survey is a good instrument for identifying persons with possible serious mental illness. (Kessler et al., 2002; Kessler et al., 2003; Kessler et al., 2010).

The K6 survey can be found on the BRFSS questionnaire, more specifically the first six questions in the Mental Illness and Stigma Questionnaire. It assesses the level of psychological distress (Kessler et al., 2002) with four questions describing the depression symptoms (sadness, hopelessness, worthlessness, and fatigue), and two questions describing anxiety symptoms (fidgety and nervousness).

Kessler-6 Survey: The Kessler-10 and the subset Kessler-6 surveys were developed by R.C. Kessler in 2002. The K6 scale is included in the BRFSS survey, more specifically the Mental Illness and Stigma module, which will be used as a tool in this research.

The goal for Kessler and his team was to create a short screening scale that would tie the non-specific psychological distress symptoms to the clinical definitions of SMI, more specifically the US Substance Abuse and Mental Health Service Administration definition of a 12-month SMI (serious mental illness) to be used in a redesigned US National Health Interview Survey (NHIS). This definition includes 12-month DSM-IV disorder along with a GAF (global assessment functioning score; Endicott et al., 1976) of less than 60 for the worst month in the 12-month period (Kessler et al., 2002). GAF was used to measure the social, occupational, and psychological functioning of an individual. The GAF scale is measured from 1 to 100 with lower values indicating increasing severity. A GAF less than 60 includes moderate symptoms to severe symptoms. GAF scale can be found on page 34 in the DSM-IV manual (American Psychiatric Association, 2000). GAF was removed with the new and updated DSM-5. Like BRFSS, NHIS is a large civilian non-institutionalized population survey, but the data is collected via personal household interviews and not telephone survey.

Methodology: The creation and analysis of the survey was done in multiple steps using multiple sets of data: (1) the mail pilot survey: nationally representative sample of the continental US that oversampled minorities for a total of 1,403 adult (age 18+) respondents, for a 54.8% response rate. The survey consisted of 45 questions (out of initial 612 questions) using a 4-level response scale (most of the time, some of the time, a little of the time, and never) plus socio-demographic questions; (2) the telephone pilot survey: a revised set of scale questions based on the results of the mail pilot survey was administered to a national representative sample of the continental US for a total of 1,574 adult respondents (age 25+), for a response rate of 40.4%. The survey consisted of 32 questions, 28 questions from the original mail pilot survey plus one question split into two separate questions (feeling ashamed or guilty into feeling ashamed and feeling guilty), and three additional questions meant to increase the precision in the depressed mood (feeling hopeless) and vigilance (feeling angry, feeling resentful) domains. In addition, another scale response was added "all of the time" meant to increase the prevision at the upper end of the distribution; (3) the clinical reappraisal survey: the K10 and K6 scales were developed at this stage based on the results of mail and telephone surveys using the Item Response Theory (IRT) methods. With this survey the authors administered the surveys in a two-stage convenience sample. In the first stage was a telephone screening survey of 1,000 adult respondents (age 18+) from Boston Metropolitan Area. In the second stage, 155 of the first stage respondents had face-to-face interviews, oversampling respondents with perceived mental health problems. In this interview the K10 was administered followed by the 12-month version of the Structured Clinical Interview for DSM-IV (SCID). In this step, the authors determined if the K10 is a useful screen for SMI. In addition, the data from second-stage interview was weighted to match the distribution from the 1997 NIHS of age, sex, education, and a four-category classification of the score on the 6-question (K6) scale.

The K6 was subsequently introduced in the 1997 (N= 36,116) and 1998 (N=32,440) NIHS surveys. The responses from the both NIHS national surveys were analyzed to cross-validate the IRT results from the clinical reappraisal survey results. IRT is a statistical approach that is used to evaluate the quality of measures, in this case it was used to evaluate the contribution of each item to the sensitivity of the total scale in the severity range of the distribution. For the items included in the final scale, the authors chose the items that had standardized severity parameters greater than 0.8 for the mail pilot survey, and 0.9-0.99 for the telephone pilot survey. A minimum standardized sensitivity parameter of higher than 1.0 was imposed on the items (the 1.0 level is the minimum required sensitivity level found in the IRT literature). In addition, the items selected were required to have consistent severity across demographic subsamples (the scale scores have the same meaning across different demographic samples). The comparison between the K10/K6 scales to the DSM-IV/SCID diagnoses in the clinical reappraisal survey was done via the Receiver Operating Characteristic (ROC) curve analysis (Hanley & McNeil, 1082). The ROC analyzes the relationship between the sensitivity of each value of the screening scale in predicting the following two outcomes: (1) a 12-month DSM-IV diagnosis with a GAF score in the range of 0-70; and (2) a12-month DSM-IV diagnosis with a GAF score in the range of 0-50. The area under ROC curve was used to evaluate the accuracy of the scale and is the probability that a randomly chosen case and a randomly chosen non-case will be correctly discriminated based on their screening scale score.

Findings: Kessler and his team found that both scales, K10 and K6 are sensitive in the upper 90-99th percentile range of the standardized distribution, with Cronbach's Alpha (internal consistency reliability) for K10=0.93 and K6=0.89 (telephone pilot survey). Kessler et al. also found that both scales (K10 and K6) have very good discrimination between cases and non-cases of DSM-IV disorders, with a ROC of 0.876 for K10 and 0.879 for K6. A better result was found for severe cases where the area under the curve for K10 was 0.955 and for K6 was 0.950. Severity parameters were found to be similar across the demographic subsamples, with Pearson correlations between 0.76 – 0.99 for K10 scale and 0.98 – 0.99 for the K6 scale (Kessler et al., 2002).

As a follow up to the 2002 development of the K10 and K6 scales, Kessler et al., tested further the two scales as a screening tool for SMI in the general population (Kessler et al., 2003). The study used the K10 and K6 scales to measure psychological distress, the World Health Organization (WHO) Composite International Diagnostic Interview Short-Form (CIDI-SF) scales to screen for several DSM-IV anxiety and mood disorders, and the WHO Disability Assessment Schedule (WHO-DAS; Rehm et al., 1999) to screen for activity limitations associated with the DSM disorders. The same two-step sample and measurement methodology was used in this project as was used in the 2002 project, with a final sample of 150 respondents. The order of the scales administered to the respondents were as follows: CIDI-SF, K10/K6, and WHO-DAS. The team defined SMI as meeting the criteria for at least one of the DSM-IV/SCID diagnoses and having a GAF score of less than 60. The internal consistency results were replicated by Kessler in 2003 with a Cronbach Alpha of 0.93 for the K10 and 0.89 for the K6 scales. Both scales K10 and K6 are statistically significant predictors of SMI, with K6 to be a better predictor in terms of area under ROC curve (AUC) at 0.86 vs. 0.85. In addition, Kessler et al., found the optimal cut-off point the K6 scale for the serious mental illness to be higher than 13 (items coded 0-4 and summing all the responses). The sensitivity (proportion of individuals with SMI who are identified by the tool) of the K6 scale for this cut-off point was 0.36, while the specificity (the proportion of individuals that are correctly identified as not having SMI) was 0.96 with a total classification accuracy of 0.92.

Prochaska et al., 2012 used the K6 as part of the 2007 California Health Interview Survey (CHIS) to build on Kessler et al., 2003 results and identified a lower threshold score indicative of moderate psychological distress. CHIS is a telephone survey conducted bi-annually of the adult non-institutionalized population in California. The total sample used in the analysis was 50,880. The authors used the ROC analysis to calculate the optimal threshold to maximize the sum of the specificity and sensitivity values. For the K6 scale a score of 5 (items coded 0-4 and summing all the responses) was found to be the optimal threshold for moderate psychological distress with a sensitivity of 0.76 and a specificity of 0.75, with a total classification accuracy of 0.74 (little variance by ethnic/racial group). Based on these studies, the current investigation studies psychological distress as measured by the K6 tool and assesses the association of psychological distress with tobacco and alcohol use, particularly in association with the economic downturn of 2008-09.

2.2. STATUS OF MENTAL HEALTH IN UNITED STATES AND THE CONSEQUENCES OF MENTAL HEALTH ILLNESS

2.2.1. Status of Mental Health in United States

According to CDC (CDC, 2011) the number of individuals with mental illness increased between 2006-2009. Figures 2.1 to 2.3 show the increasing trend in mental health illness in the US population from 2006, to 2007 and through 2009. In these figures, depression is defined from the Patient Health Questionnaire-8 based on a severity score of 10 or higher. Psychological distress is calculated using the aggregate Kessler-6 score of 13 or higher. Mentally unhealthy period is based on the answers to the survey question "Now thinking about your mental health, which includes stress, depression, and problem with emotions, for how many days during the last 30 days was your mental health not good?" The quartiles are based on point estimates. Although the maps do not measure the same mental health issue, they show a trend in most states of a higher prevalence of mental health issues. The darker shades of blue indicate higher prevalence of mental health issue.

Table 2.1 shows that from 2006 to 2014 the frequency of Americans with serious psychological distress (SPD) has been increasing from 2.8 percent in 2006-2007 to 3.4 percent in 2013-2014. Adults aged 45-64 had the highest SPD rate ranging from 3.7 percent to 4.5 percent during the same time period.

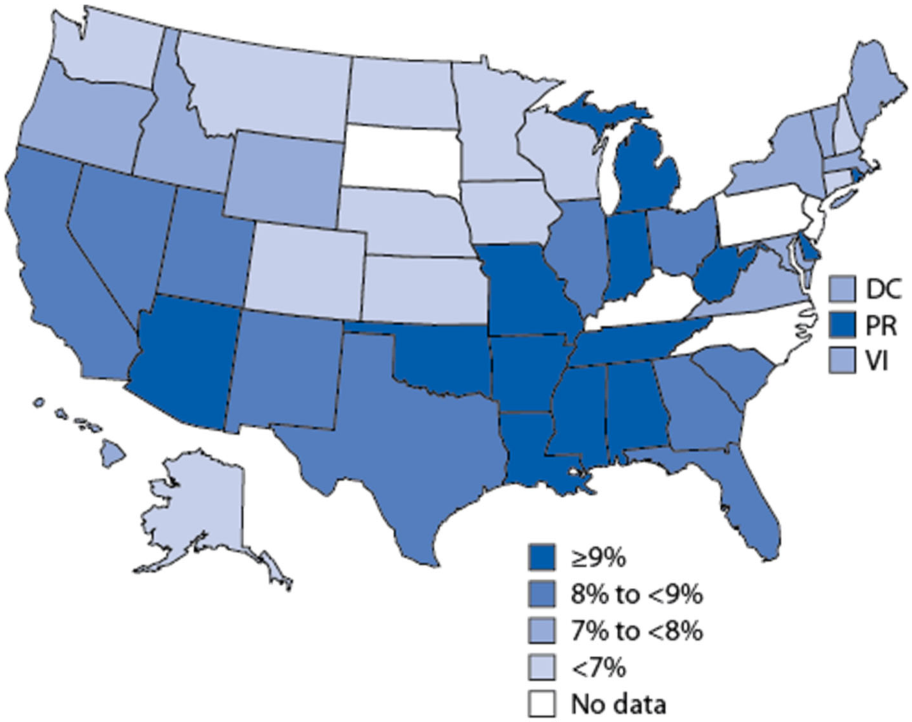

Figure 2.1. Prevalence of Current Depression Among Adults Aged ≥18 years, by State Quartile — Behavioral Risk Factor Surveillance System, United States, 2006

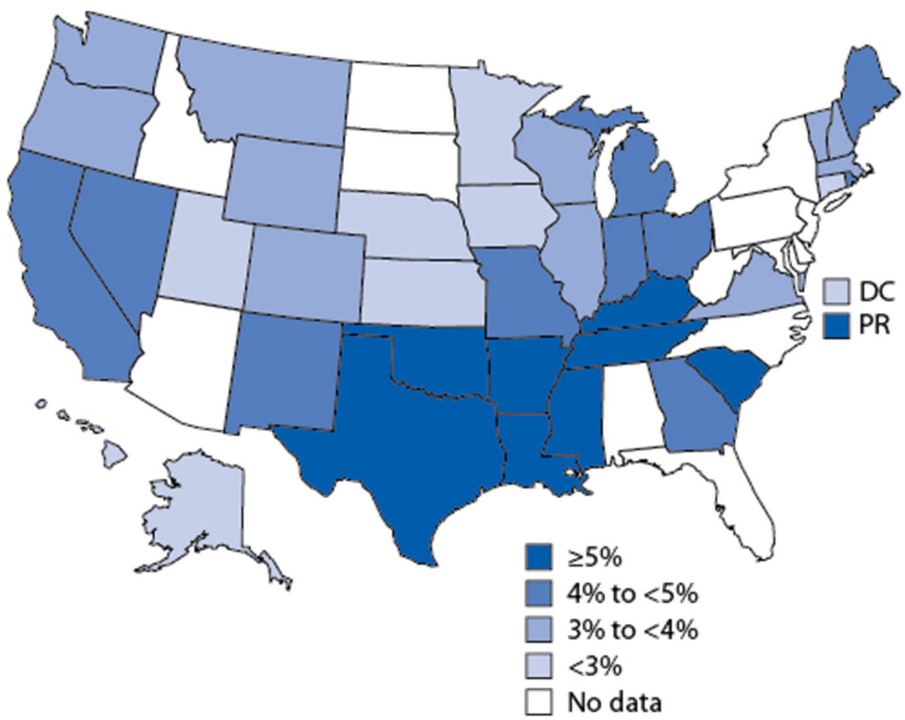

Figure 2.2. Prevalence of Serious Psychological Distress Among Adults aged ≥18 Years, by State Quartile — Behavioral Risk Factor Surveillance System, United States, 2007

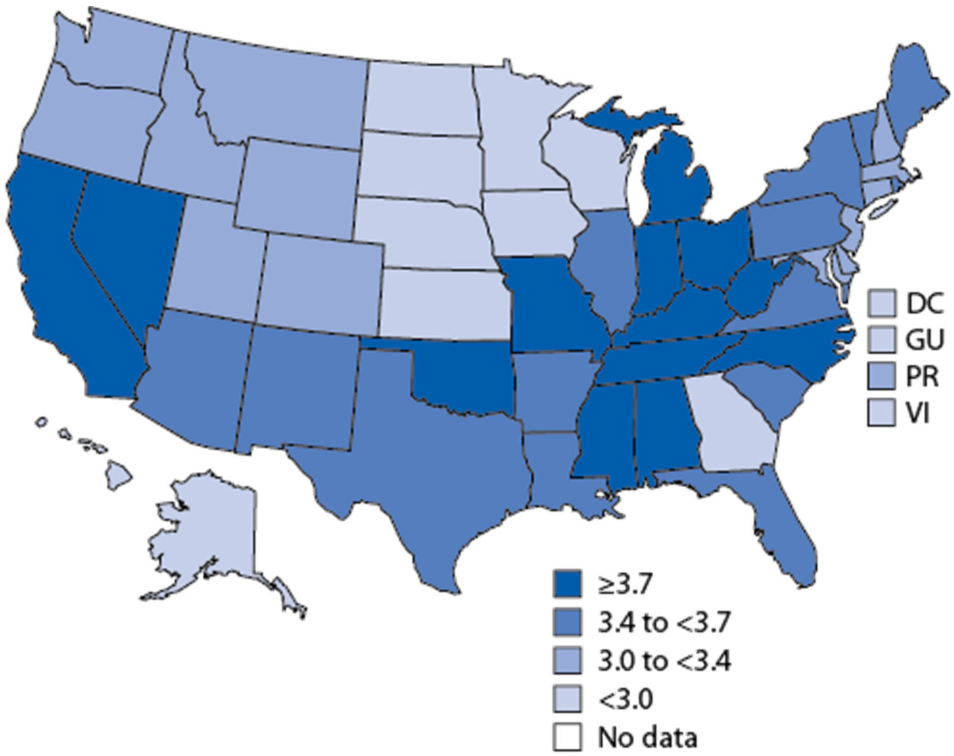

Figure 2.3. Mean Number of Mentally Unhealthy Days During Past 30 Days Among Adults aged ≥18 Years, by State Quartile — Behavioral Risk Factor Surveillance System, United States, 2009

2.2.2. Mental Health/Mental Illness and Comorbid Health Conditions

Untreated depression is a risk factor for developing major physical health problems such as diabetes and heart disease (Druss & Walker, 2011; Edge et al., 2003). In addition, meta-analyses of longitudinal studies suggest a significant association between depression and obesity (Scott et al., 2008). Both obesity and depression are associated with higher health care costs across inpatient and outpatient services (Simon et al., 2011). Clum et al, 2014 found that increased depressive symptoms were directly associated with increased BMI (body mass index) in women and indirectly with increased healthy eating self-efficacy, increased emotional eating, and decreased exercise self-

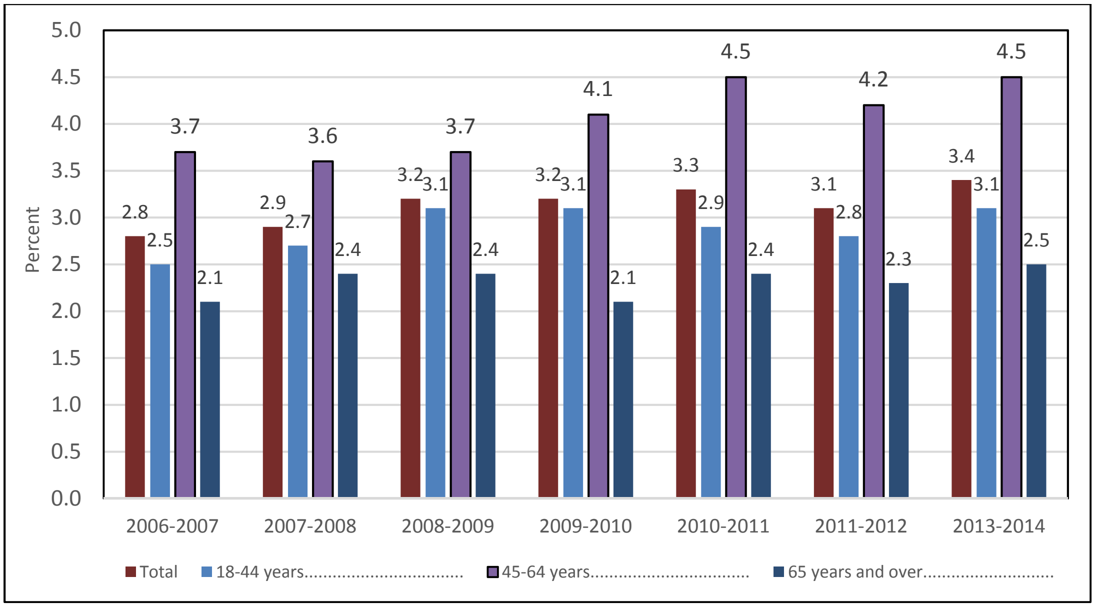

Table 2.1. Percentage of Persons Aged 18 and Over with Serious Psychological Distress in the Past 30 Days by Age: United States, 2006-2014
Source: CDC, National Health Interview Survey, 2006-2014

efficacy (Clum et al., 2014). In general, people with mental health issues and co-morbidities incur higher health care costs compared to people that don't have mental health issues with the same co-morbidities (Gameroff & Olfson, 2006; Kim & Lee, 2006).

2.2.3. <u>Risk for substance use and quality of life</u>

Research suggests that depression and anxiety are a risk factor for smoking, alcohol consumption, physical inactivity, and sleep disturbance. The positive association between mental health and physical health has a direct impact on the cost of medical care (Lutterman et al., 2003). Forman-Hoffman et al., 2014 found that there is a statistically significant and robust association between psychological distress and mortality, even after adjusting for confounding effects such as smoking and chronical health conditions. Multiple studies in multiple states find that adults with mental health illness who were receiving treatment had a mean years of potential life lost (YPLL) of between 13.5 and

32.2 (Lutterman et al., 2003; Colton et al., 2006). This estimate is similar to another study done in Ohio public mental health hospitals between 1998 and 2002 where the mean YPLL was 32.0 (Miller et al., 2006).

According to Schaller and Stevens, investment in health is a function of market-based health care inputs, plus the individual's time devoted to improving health, plus the individual's human capital. All these inputs, along with the direct effects of job loss on employment and earnings, and ultimately changes in health insurance coverage, indicate a potentially strong link between job loss and health (Schaller & Stevens, 2015). Their research on a large uninstitutionalized U.S. population sample found that displacement (involuntary loss of employment) increases the probability of poor or fair self-reported physical health status by 16% of the baseline probability (or 1.37 percentage points). They also found positive and significant increases in the incidence of arthritis and diabetes following job displacement which could be aggravated by loss of health insurance leading to lower health care utilization, less usage or no usage of prescription drugs, and increased use of emergency rooms.

Mental health disorders have a significant impact on the quality of life of the affected individual and their families. Horwath et al., 1992 found that untreated depression increases the risk of developing major depression by 4.4 percent within one year. Forsell 2007 found that minor depression increases the risk of major depression four to five times in a three-year period. As a result, Greenberg et al, 2015 found the economic burden of depression, including direct and indirect costs (workplace costs, suicide costs, etc.) to be estimated at $210.5 billion in 2010.

2.3. THE ECONOMIC DOWNTURN OF 2007-2019

2.3.1. The Economic Downturn of 2007-2009

In December 2007 the United States had fallen into an economic recession. Recession is characterized by a slowdown in the economy, a slowdown of business activity, and a reduction in the amount of goods and services produced and sold. According to the National Bureau of Economic Research (the official arbiter of U.S. recessions) the most recent recession in the United States began in December of 2007 and ended in June 2009, though many statistics that indicate that the US economy has not yet returned to pre-recession values (Bureau of Labor Statistics; 2012).

The most common characteristic of any recession is an increasing unemployment rate caused by slowdown of the economy and business activity. In December 2007, the national employment rate was 5.0 percent and had been at that level or below for the previous 30 months indicating a healthy economy. At the end of recession in June 2009, the unemployment rate went up to 9.5 percent and peaked in October of 2009 to 10.0 percent (Bureau of Labor Statistics; 2012). This sharp increase in unemployment caused an unprecedented increase in home foreclosures caused not only by the higher unemployment rate but also by the subprime mortgage crisis. According to RealtyTrac, more than 2.3 million houses went into foreclosure in 2008 (an increase of 81 percent from the previous year), and more than 2.8 million properties foreclosing in 2010 and 2.8 million in 2011 (RealtyTrac, 2009-2011).

The subprime mortgage crisis was triggered by a large decline in home prices caused by the "housing bubble," leaving homeowners owing more on the mortgage than the house was worth, making it more difficult for homeowners to borrow money and refinance their loans.

In the years leading up to the recession, banks issued adjustable-rate mortgages, allowing customers to purchase homes beyond their means due to the low mortgage rates. After the recession, these low adjustable rates began to increase, causing higher monthly payments and a resulting increase in mortgage delinquencies. These delinquencies forced some homeowners to leave their homes and forced the subprime banking industry to shrink drastically, causing a reduction in the stock market and a loss in retirement savings.

2.3.2. Mental Health Status during the Economic Downturn of 2007-2009

Research has found that the recession had a tremendous impact on mental health, ranging from less severe mental problems to serious mental health problems. According to McLaughlin et al., exposure to foreclosure predicted symptoms for major depression (McLaughlin et al., 2012). In addition, Houle and Light found that an increase in foreclosure rates was associated with an increase in the suicide rates especially among adults aged 46 to 64 years (Houle and Light; 2014). More commonly, unemployment and foreclosure cause stressful life invoking feelings of shame, loss, and regret (Libman et al., 2012) and is associated with mental health issues such as anxiety and depression (Alley et al, 2011; McLaughlin et al., 2012; Pollack et al, 2011; Cannuscio et al., 2012). In addition, unemployment and foreclosure are also associated with physical health problems and increased mortality (Sullivan & Von Wachter, 2009).

2.4. BARRIERS TO MENTAL HEALH CARE UTILIZATION

2.4.1. Unemployment and insurance

As mentioned above, unemployment is a barrier to mental health care utilization, as is decreased earnings for the employed (Jacobson et al., 1993; Couch & Placzek, 2010) reducing healthcare access or coverage. Schaller and Stevens found that displaced workers are more likely to lose their insurance with a job loss; 35.2% of the restricted sample of workers who were likely to face larger changes in insurance coverage compared with 17% of the full sample. They were also more likely to reduce their health care utilization (5.3% decrease in the probability of a doctor's visit after displacement) (Schaller & Stevens, 2015). The high uninsurance rate among people with mental illness contributes to cost as a barrier to treatment (Rowan et al., 2013). Walker et al., 2015 found that 75% of uninsured adults with any mental illness and 56% of uninsured adults with serious/severe mental illness (SMI) did not receive treatment.

2.4.2. Limited insurance coverage of mental health issues

A significant barrier to mental health care is limited healthcare coverage. The Mental Health Parity and Addiction Equity Act (MHPAEA) of 2008 requires insurers to match the financial requirements (co-pays, deductibles, etc.) and treatment limitations (visit limits, etc.) for mental health care with the financial requirements and treatment limitations for physical health care (medical, surgical, etc.). This law put mental healthcare on par with medical/surgical healthcare coverage.

However, there are two problems with the MHPEA law: (1) it applies only to insurance plans for public and private sector employers having over 50 employees, leaving many without coverage, and (2) the law does not require insurers to offer benefits for mental health and substance abuse disorders; it only requires it to match other benefits offered.

2.4.3. Individual's attitudes towards mental health

Another significant barrier to mental health care is attitudinal barriers. The most common attitudinal barrier to mental health care utilization is the low perceived need for professional treatment especially among moderate and mild mental illness cases than severe mental illness cases (Andrade et al., 2014; van Beljouw et al., 2010; Mojtabai et al., 2011). Approximately 64 percent (Andrade et al., 2014) and 73 percent (Mojtabai et al., 2011) of people with a DSM-5 disorder in the past 12 months listed the desire to handle one's own problem as the most common barrier for not seeking mental health treatment. This barrier could be caused by the stigma of mental health and individuals might want to deal with their mental issues themselves.

2.4.4. Clinical definition of mental health illness and health insurance coverage as a limitation

Determining how many people have a mental illness can be difficult and because of this reason the prevalence rates vary from one study to another. Prevalence rates for mental health might be under-reported due to the stigma of mental health. Most population surveys include questions related to mental health status, and a few questions inquire about diagnosed mental illness. In practice, mental health professionals diagnose mental illness based on the criteria in the American Psychiatric Association's DSM (Diagnostic and Statistical Manual) along with a specified indicator of severity. This diagnosis is what is covered by the insurance companies. This discrepancy in the self-reported prevalence of mental illness versus diagnosed prevalence of mental illness creates a gap or barrier in mental health care utilization.

In conclusion, there are many barriers to mental health care utilization, including the level of insurance coverage, criteria for coverable mental illness, and beliefs and biases about mental health treatments. Because of all these limitations to mental health care utilization, people may be turning to other ways of coping with mental health issues such as alcohol and cigarette smoking among others (Lasser et al., 2000; Le Cook, et al., 2014).

2.5. SMOKING AND ALCOHOL CONSUMPTION IN UNITED STATES

Tobacco use, and excessive alcohol consumption remain major health issues in the United States. The health consequences of tobacco use, and excessive alcohol consumption are numerous: heart disease, different types of cancer, and the increase in the severity of chronic health conditions (USDHHS, 2010). These two risk behaviors are prevalent in the population and are modifiable, but for people with mental health issues using these behaviors as a coping mechanism for mental illness, these risky behaviors might not be as easily modifiable (Lasser et al., 2000; Le Cook, et al., 2014).

2.5.1. Alcohol Consumption and Tobacco Use in the United States

Alcohol Consumption

According to the National Center for Health Statistics (MMWR, 2012) 23.4% of adults 18 years old and over had at least one incident of heavy drinking in the past year. Heavy drinking is defined as drinking 8 or more drinks per week for women and 15 or more drinks per week for men. In addition, in 2010 more than half of the US adult population had consumed alcohol in the past 30 days, and 17.1% of men reporting binge drinking, a higher frequency and intensity than women (5.0 episodes per month and 9.0 drinks on one occasion compared to women, 3.2 episodes per month and 5.9 drinks per occasion) (MMWR, 2012). Binge drinking is defined as 4 or more drinks for a woman or 5 or more drinks for a man on an occasion during the past 30 days. The figures below show the prevalence and the intensity of binge drinking in the United States in 2009 - 2011. For all three years, all the states have a prevalence of at least 10 percent with most of the mid-northern states having the highest prevalence between 18 and 25 percent of its population binge drinking.

Excessive alcohol consumption also impacts the economy via workplace productivity (72% of the total cost), health care expenses (11% of the total cost), and additional costs caused by Driving under the Influence (DUIs) or Driving While Intoxicated (DWIs): accident costs, criminal costs, and property damage (Sacks et al., 2015).

Figures 2.4 -2.6 show the prevalence of binge drinking in the United States in 2009, 2010, and 2011 with the higher severity shown in darker shades. Three quarters of the United States show a moderate to high prevalence of binge drinking (17 percent of the adults and higher). On the other hand, figure 2.5 shows the change in average largest number of drinks consumed by binge drinkers on any occasion in the United States.

Again, more than three quarters of the Unites States show a moderate to high average largest number of drinks consumed on any occasion (7 drinks or more).

Tobacco Use

As shown in figure 2.6 below the percentage of US adults who smoke cigarettes has declined from 2005 to 2015 from 20.9% to 15.1%, but the number of smokers is still high at 36.5 million Americans (Jamal et al., 2015). Even though the adjusted smoking rates declined, Le Cook et al., 2014 showed that the decrease was significantly less for the people with mental illness for the period of 2004 to 2011.

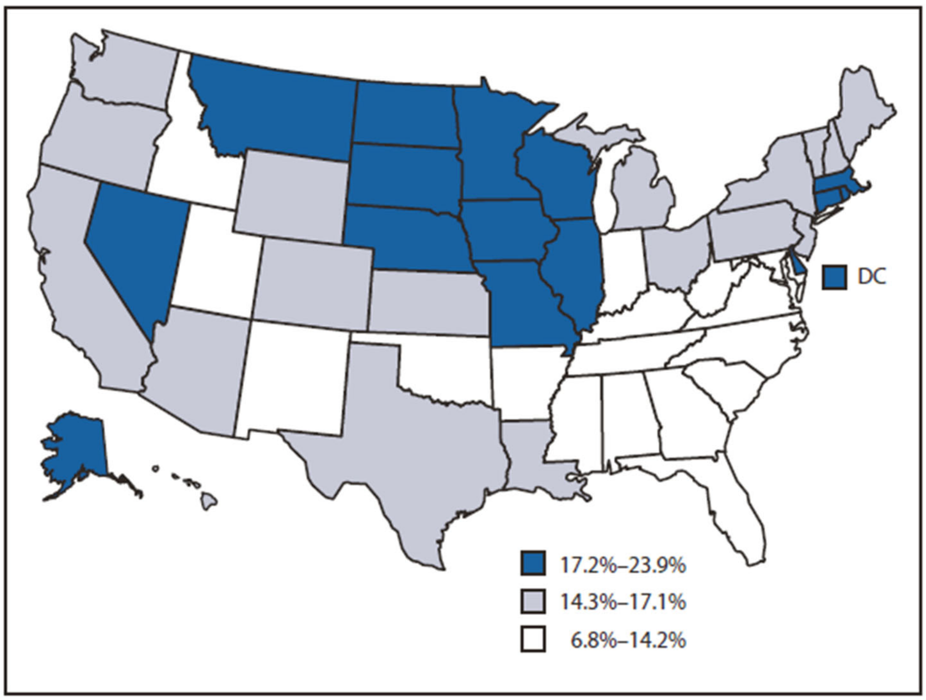

Figure 2.4 Population prevalence of Binge Drinking among Adults, 2009 – Behavioral Risk Factor Surveillance System

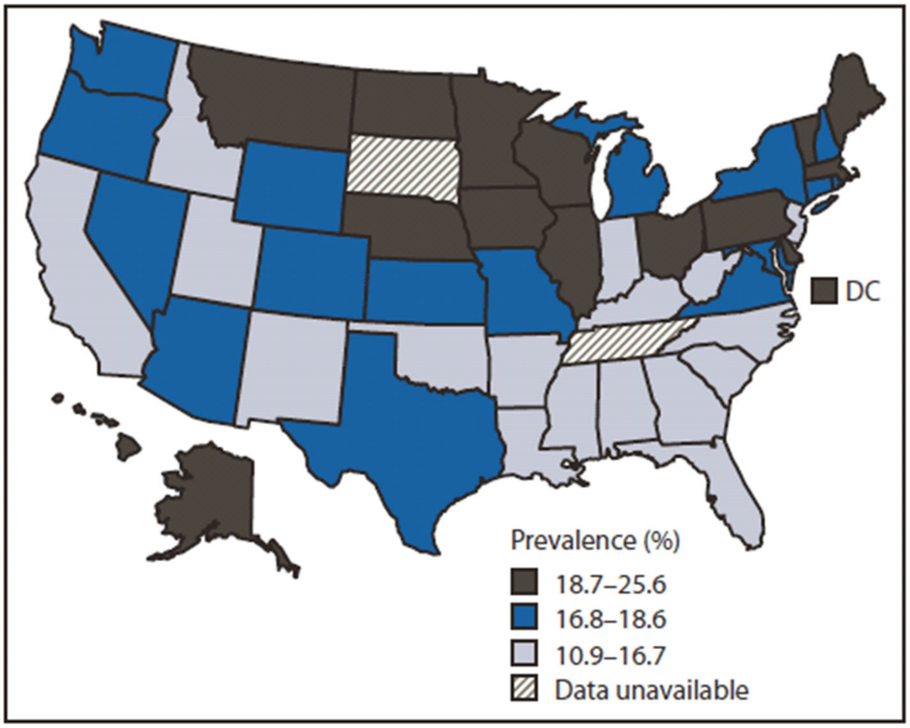

Figure 2.5 Age-adjusted to the 2000 US Census standard population prevalence of Binge Drinking among Adults, 2010 – Behavioral Risk Factor Surveillance System

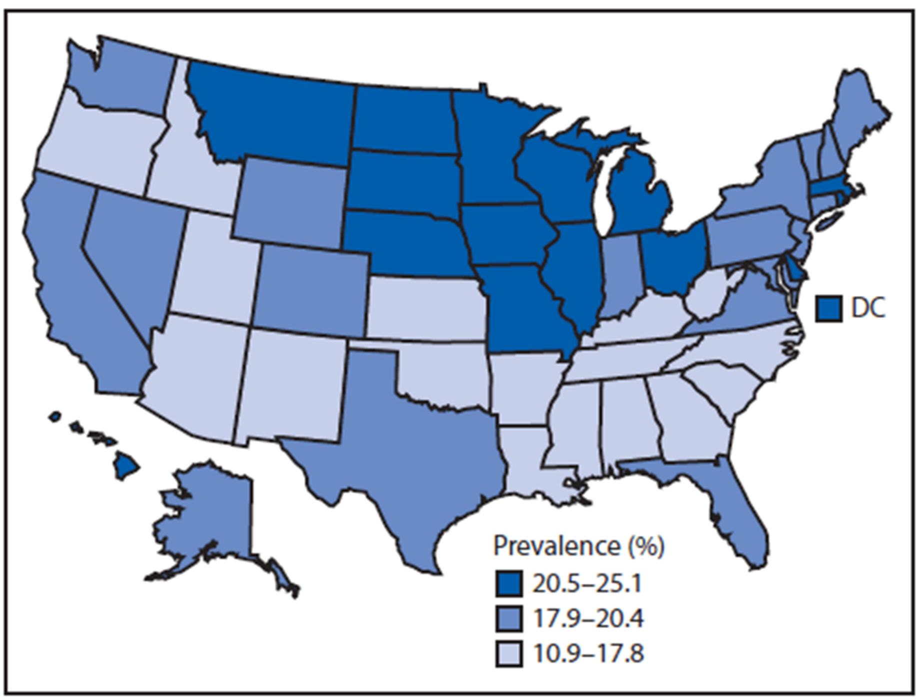

Figure 2.6 Population prevalence of Binge Drinking among Adults, 2011 – Behavioral Risk Factor Surveillance System

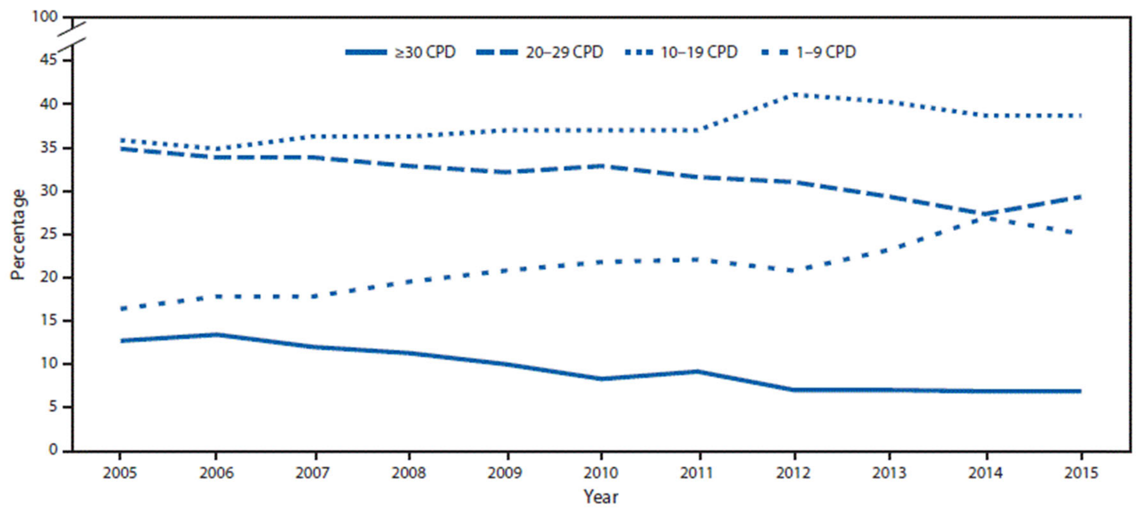

Figure 2.7. Percentage of daily smokers* aged ≥18 years, by number of cigarettes smoked per day (CPD) – National Health Interview Survey, United States 2005-2015
*Persons who reported smoking more than 100 cigarettes during their lifetime and who, at the time of interview, reported smoking every day or some days.

2.5.2. Smoking and Alcohol Consumption among people with Mental Illness

As explained above, smoking and excessive drinking are still prevalent in the United States even though they are modifiable risk factors and, as for their healthy counterparts, research has found that individuals with mental illness are at greater risk for engaging in multiple risk behaviors such as smoking, alcohol abuse, sedentary behavior, obesity, and many other indirectly related diseases/disorders (Lasser et al., 2000; Miles et al., 2003; Sanchez-Villegas et al., 2008). Research has found that people with mental health problems smoke at a higher rate compared to healthy counterparts, and have more difficulty quitting (Lasser et al., 2000; Le Cook et al., 2014).

Lasser et al., 2000 found that in the past month the quit rates (defined as the proportion of lifetime smokers who were not current smokers) were lower in smokers with major depression at 26.0% ($p<0.001$), generalized anxiety disorder at 28.9% ($p<0.005$), and alcohol abuse or dependence at 16.9% ($p<0.0001$) compared with smokers without mental illness at 42.5%. Le Cook et al., 2014 showed even though the smoking rates declined in the non-smoking population, the decrease was significantly less for the people with mental illness (23.5%, 95% CI: 24.2% - 26.3% to 24.9%, 95% CI: 23.8% - 26.0%, p = 0.50) for the period of 2004 to 2011.

Alcohol consumption and cigarette smoking are still a health issue in the United States (Jamal et al., 2015; MMWR, 2012). Even through these risky behaviors are modifiable, research has found that it is much harder for persons with mental health issues to change these behaviors (Lasser et al., 2000; Le Cook et al., 2014; Miles et al., 2003; Sanchez-Villegas et al., 2008). This study is unique in the sense that it examines if risky behaviors such as alcohol consumption and smoking have an effect on psychological distress during times of economic hardship.

2.6. LIMITATIONS OF THE EXISTING LITERATURE

The literature indicates that mental issues are a major problem in Unites States and those with mental issues are at a higher risk of having physical health issues, creating a complex and expensive health care treatment. The literature review also indicates that people with mental health issues are at a higher risk for abusing alcohol and tobacco. Foreclosures and unemployment are associated with significant decrease in self-rated physical and mental health, increases in activity limitations, and increased reports of anxiety and depression (Schaller & Stevens, 2015). Because either the lack of mental health care or lack of insurance coupled with the association between mental issues and abuse of alcohol and tobacco, this investigation examines the impact of the recession and the population's usage of alcohol and tobacco on the level of psychological distress. The chosen approach utilizes BRFSS data collected before, during, and after the economic downturn to determine if there is a significant association between the uninsurance level, income, and risky behaviors (represented by heavy and binge drinking and smoking) and psychological distress. As far as we know, we are the first to examine the effect that severe economic recession has on the level of psychological distress as evidenced by risky behaviors represented by smoking and alcohol consumption, insurance status and income level.

CHAPTER 3

METHODS

3.1. CONCEPTUAL FRAMEWORK

Prior work has shown that people with mental health issues have a higher prevalence of alcohol and tobacco use (Lasser et al, 2000; Le Cook, 2014). The proposed work used a population-based approach to construct a model to determine: 1) if insurance status and income has an effect on psychological distress during economic hardship; 2) if alcohol consumption and/or smoking level is associated with psychological distress during economic hardship; and 3) if the level of psychological distress in the population has an effect on alcohol consumption and smoking when faced with economic hardship during the period of 2007 through 2012.

3.2. STUDY DESIGN

The study is a secondary data analysis of the BRFSS survey, that is a large-scale, nationwide dataset containing questions on general health, specific disease status, mental health, alcohol and tobacco use. The survey is a retrospective annual cross-sectional survey that began collecting data in 1984. This study used information from the surveys before, during, and after the 2007 U.S. Recession: 2007, 2009, 2012, and 2013, respectively. These years were chosen based on the data availability of the Mental Illness and Stigma module. The module was an option module in the annual BRFSS survey, and it was available only for the above four years. One of the limitations of the BRFFS survey is that as a national survey, the questions are general in nature across multiple health issues, making it difficult to isolate specific mental health issues. In addition, the survey does not follow the same population over time, limiting the sample to a cross sectional view. One advantage of the BRFSS survey is that the survey is nationally representative, allowing for generalizable results across the nation. In addition, the survey questions are consistent throughout the years, with some changes in the optional modules and rotating cores. This allows researchers to examine trends in the population.

This study leverages the stated limitations and advantages by studying psychological distress trends over time at the population level using latent growth modeling (LGM) analyses. Latent growth modeling is part of the structural equation modeling (SEM) framework and is commonly used to estimate any changes of mental health over time (McArdle, 1988). LGM is a longitudinal design which estimates change at the individual level over a period of time. The unique feature of this research is the use of latent growth modeling with pooled cross-sectional data to examine the trend of mental health at the population level. The nature of the data allows for such statistical analysis, as the same questions in the mental health and stigma module are used in multiple years.

3.3. HYPOTHESES AND RESEARCH QUESTIONS

This work examines trends and potential linkage between psychological distress, use of alcohol and tobacco, and insurance status during the years 2007 to 2013, with a focus on the 2007-2009 recession. This is accomplished by examining the trend of psychological distress and its levels as described by Kessler et al., 2003, Prochaska et al., 2012, and Forman-Hoffman et al., 2014. The models tested incorporated important covariates such as smoking, drinking, alcohol consumption, mental healthcare utilization, level of insurance, and other variables such as sex, race, and SES.

SPECIFIC AIMS:

The specific aims of this work are as follows:

1) Identify (EFA) and confirm (CFA) the factor structure of K6 in order to confirm that the meaning and structure of the K6 survey holds for the datasets.
2) Examine the trend of psychological distress during and after the 2007 recession.
3) Examine the trend of psychological distress before, during, and after the 2007 recession by income and health insurance status.
4) Investigate the association between mental healthcare utilization and psychological distress and the mediation inference of health insurance on mental healthcare utilization.
5) Investigate the effect of alcohol and/or tobacco consumption on psychological distress and the effect of psychological distress on alcohol and/or tobacco consumption.

3.4. DATA AND DATA SOURCE

Survey Design and Eligible Population

The Mental Health and Stigma module is an optional module as part of the Behavioral Risk Factor Survey System (BRFSS). It includes six questions that make up the Kessler-6 (K6) psychological distress scale. The module uses the K-6 scale not only because it is a short survey (6 questions), but also because it has been validated in the current literature and it has been designed as an index of psychological distress which has been shown to be a reliable predictor for serious mental illness (Kessler et al, 2002; Kessler et al, 2003; Prochaska et al, 2012). The K-6 survey is also part of the U.S. National Health Interview Survey (NIHS) as well as in the National Household Survey on Drug Abuse, and many other large-scale surveys around the world.

3.5. THE KESSLER-6 (K6) SCALE

The K6 scale is used in population surveys such as the BRFSS survey as part of the Mental Illness and Stigma module and the National Survey of Drug Use and Health (NSDUH). The K6 instrument offers a useful tool for states to assess the potentially unmet mental health needs of a large population of adults.

The K6 items assess the level of psychological distress (depression and anxiety symptoms) over the past 30 days. The responses range from "all" = 1, "most"= 2, "some"=3, "a little"= 4, "none"= 5. For the purpose of our research, the scale was reversed with "none" = 1 to "all" = 5. Summing the six items provides a possible range of 6 to 30, with a cutoff point of 19 and higher denoting severe psychological distress (Kessler et al., 2003) and with the range of 11-18 denoting moderate psychological distress (Prochaska et al., 2012). Based on Kessler et al., 2003 and Prochaska et al., 2012 individuals can be classified into one of the three categories: low to no psychological distress (total score < 11), moderate psychological distress (11≤ total score < 19), or serious psychological distress (total score ≥ 19). Since Kessler et al, 2003 has shown that serious PD is strongly associated with serious mental health, the three categories and cutoff points were used in the present study.

Table 3.1 presents the first 6 questions which comprise the K6 (Kessler 6) scale. The K6 survey questions were reverse recoded to help with interpretation from lower to higher frequency (none=1 to all=5).

Table 3.1. Mental Illness and Stigma Module to Capture Self-Reported PD

Survey Questions	**Answer**	**Coding**	**Recode**	**Variable**
About how often during the past 30 days you feel **nervous**?	1 = All 2 = Most 3 = Some 4 = A little 5 = None	MISERVS	NERVS 1 = None 2 = A little 3 = Some 4 = Most 5 = All	V1
During the past 30 days, how often did you feel **hopeless**?	1 = All 2 = Most 3 = Some 4 = A little 5 = None	MISHOPLS	HOPLS 1 = None 2 = A little 3 = Some 4 = Most 5 = All	V2
During the past 30 days, about how often did you feel **restless** or **fidgety**?	1 = All 2 = Most 3 = Some 4 = A little 5 = None	MISRSTLS	RESTLS 1 = None 2 = A little 3 = Some 4 = Most 5 = All	V3
During the past 30 days, about how often did you feel **so** depressed **that nothing could cheer you up**?	1 = All 2 = Most 3 = Some 4 = A little 5 = None	MISDEPRD	DEPRD 1 = None 2 = A little 3 = Some 4 = Most 5 = All	V4
During the past 30 days, about how often did you feel that **everything was an effort?**	1 = All 2 = Most 3 = Some 4 = A little 5 = None	MISEFFRT	EFFRT 1 = None 2 = A little 3 = Some 4 = Most 5 = All	V5
During the past 30 days, about how often did you feel **worthless?**	1 = All 2 = Most 3 = Some 4 = A little 5 = None	MISWTLES	WORTH/WTLES 1 = None 2 = A little 3 = Some 4 = Most 5 = All	V6

During the past 30 days, for about how many days did a mental health condition or emotional problem keep you from doing your work or other usual activities?	1-30 Number	MISNOWRK	DAYSMISS None A week (1-5 days) Two Weeks (6-10 days) More than two weeks (11+ days)	
Are you taking medicine or receiving treatment from a doctor or other health professional for any type of mental health condition or emotional problem?	1= Yes 2= No	MISTMNT	NOTREATMENTMH	
Treatment can help people with mental illness lead normal lives.	1= Agree Strongly 2=Agree slightly 3=Neither agree or disagree 4=Disagree slightly 5=Disagree strongly	MISTRHLP	STIGMATREAT	V9
People are generally caring and sympathetic to people with mental illness.	1= Agree Strongly 2=Agree slightly 3=Neither agree or disagree 4=Disagree slightly 5=Disagree strongly	MISPHLPF	STIGMACARING	V10

3.6. STUDY VARIABLES

This study uses responses to the K6 questions in the Mental Illness and Stigma module. In addition to the K6, the study includes the following demographic variables from the fixed core module: sex, race, age, and income as well as insurance coverage (represented by the existence or absence of health insurance coverage), healthcare utilization (represented by the respondents' positive or negative response to the question: "Are you now taking medicine or receiving treatment from a doctor or other health professional for any type of mental health condition or emotional problem?"). Before the datasets can be linked to mimic a longitudinal dataset, composite observations were created (a unique combination of 5 demographic variables as listed and explained in section 3.9. Unit of Analysis listed below) to allow for one-to-one population level correspondence.

Dependent variables of interest:

Since Aim 1 and Aim 2 evaluate model structures necessary for the analyses in Aims 3-5, the dependent variables are described for Aims 3-5. For Specific Aim 3 and Specific Aim 4, the dependent variable of interest will be Psychological Distress. This is defined as the self-reported psychological distress status using the K6 questionnaire that is part of the Mental Illness and Stigma Module. The survey items are rated on a 5-point Likert type scale (1=All of the time to 5= None of the time) and contains the following items: "How often during the past 30 days you feel: (1) nervous, (2) hopeless, (3) restless or fidgety, (4) so depressed that nothing could cheer you up, (5) that everything was an effort, (6) worthless?". To match the original survey scoring scheme by Kessler et al, the BRFSS Module items were reverse coded to show 1=None of the time to 5=All of the time (NCS, 2005). The level of psychological distress (PD) is calculated by summing the responses to the 6 items. Based on the sum scores, individuals are classified into one of three separate categories: no or low PD (scores between 6 and 10), moderate PD (scores between 11 and 18), and serious PD (scores 19 or higher). For specific aim 5, the dependent variables of interest are Smoking and Alcohol use.

Tobacco Smoking is studied in the first linear regression. Cigarette smoking level was asked as part of the Tobacco Use section in the Core survey. This variable included 4 levels of smoking status: (a) current smoker, now smokes every day, (b) current smoker, now smokes some days, (c) former smoker, (d) never smoked. The variable is a combination of questions "Have you smoked at least 100 cigarettes in your entire life?" and "Do you now smoke cigarettes every day, some days, or not at all?" The answer to the first questions is dichotomous: Yes or No, and the answer to the second question is a choice of "Every day", "Some days", or "Not at all". In order to capture the most accurate answer for each population level observation and allow for a much wider range of answers (continuous instead of categorical), the smoking variable was transformed into mean proportions for each composite observation. The first step was to create two categories for smoking: non/former smoker = 0 (c and d) and smoker =1 (a and b). The next step was to create one smoking score per composite observation for each year. The variable could range from 0 to 1 and measures the percentage of individuals that are smoking. A score closer to 1 indicates more individuals smoke, while a score closer to 0 indicates fewer individuals smoke. This transformation was necessary to allow for easy interpretation of the results.

Alcohol consumption is studied in the second linear regression. Alcohol consumption was elicited using the Alcohol Consumption section in the Core survey. This research uses a computed variable that has 3 levels of alcohol consumption based on the frequency and severity of the consumption: (1) none and moderate drinking (1 drink per day for women, and 2 drinks per day for men), (2) binge drinking (4 or more drinks per occasion for women and 5 or more drinks per occasion for men), and (3) heavy drinking (8 or more drinks per week for women and 15 or more drinks per week for men). In order to get the computed variable, the following questions are used: "During the past 30 days, how many days per week or per month did you have at least one drink of alcoholic beverage such as beer, wine, a malt beverage, or liquor?", "Considering all types of alcoholic beverages, how many times during the past 30 days did you have X (x=5 for men, x=4 for women) or more drinks on an occasion?". The answers to these questions are continuous from 0-7 days per week, 0-30 days in the past 30 days, number of drinks, and number of times. Similar to the smoking variable, the alcohol consumption variable has been transformed for this research. The same steps were performed as described for the smoking variable. In the first step the drinking variable was recoded into two categories: no/moderate drinking =0 and heavy/binge drinking=1. This allowed us to calculate the mean drinking proportion for each composite observation for each year. The values for the drinking variable for each observation ranges from 0 (no to low alcohol consumption) to 1 (heavy/binge alcohol consumption).

Independent variables of interest

There are multiple independent variables examined in this study. Aims 1 and 2 examine latent structures of hypothesized relationships (examines the factor structure underlying the K6 survey via exploratory and confirmatory factor analysis). Independent variables are defined for Aims 3-5. Insurance status is the independent variable for specific Aim 3 and a mediator variable for specific Aim 4 to estimate the association between PD and Healthcare (HC) Utilization. (Does receiving mental health treatment change the effect on PD when it is mediated by insurance status?) HC Utilization is another independent variable measured based on responses to the Mental Illness and Stigma module question: “Are you now taking medicine or receiving treatment from a doctor or other health professional for any type of mental health condition or emotional problem?” The answer for this item is dichotomous: Yes = 1 or No = 2. This variable has been recoded to Yes = 0 and No = 1 for easier interpretation. The final variable used in the analysis represents the proportion of individuals not taking medicine or receiving treatment from a doctor or other professional for any type of mental health condition or emotional problem (notreatmentMH).

Health Insurance (HlthPln) is an independent variable in specific Aim 3, and a mediator in specific Aim 4. The variable was part of the Health Care Access section in the Core survey: “Do you have any kind of health insurance coverage, including health insurance, prepaid plans such as HMOs, government plans such as Medicare, or Indian Health Services?” The answer for this item is dichotomous: Yes = 1 or No =2. The variable has been recoded to Yes = 0 and No = 1 for easier interpretation. The final variable used in the analysis is the proportion of uninsured. For Aim 5, PD is independent variable in both linear regressions.

Examining trends across time:

For the time trends investigation (specific Aims 1 and 2) the variable studied is PD.

Demographic Variables

The following demographic variables are used in the analysis: age (measured in years, 18 an older), gender (male or female), race (white, non-white), income (annual household income from all sources), education level (never attended school or only attended kindergarten, grades 1 through 8 (Elementary), grades 9 through 11 (Some high school), grade 12 or GED (High school graduate), college 1 year to 3 years (Some college or technical school), college 4 years or more (College graduate)), employment status (employed for wages, self-employed, out of work for more than 1 year, out of work for less than 1 year, a homemaker, a student, retired, unable to work).

Other covariates (control variables)

The following covariates are used in specific Aim 5: self-reported general health indicator (5-point scale from poor to excellent), chronic health conditions (myocardial infraction, angina or coronary disease, stroke, asthma, and diabetes), and attitudes towards mental health (2 questions on a 5-point Likert scale).

3.7. EXPLANATORY VARIABLES OF INTEREST

The table below lists all the variables used in the analysis, split by categories, type and for which specific Aim each variable was used. The first 5 variables were used to create composite observations. Since the datasets were cross-sectional, composite observations were created to allow the datasets to be converted into a longitudinal dataset from a set of cross-sectional datasets. The total number of observations of 192 were calculated based on the unique combinations of the categories for each variable. One example is male, white, 18-29 years old with 0-25K income and did not graduate high school.

Table 3.2. Variable Description and Variable Levels to be Used in the Analysis

Variable Name	Categories	Type	Specific Aim in which Variable will be used
Sex	Male Female	Categorical: nominal	Aims 2-5 (part of composite observations)
Race	White Non-White	Categorical: nominal	Aims 2-5 (part of composite observations)
Age	18-29 30-44 45-64 64+	Continuous: interval	Aims 2-5 (part of composite observations)
Income	0-25,000 25,001-50,000 50,001+	Continuous: interval	Aims 2-5 (part of composite observations)
Education Level	Did not Graduate High School Graduated High School Attended College or Technical School Graduated from College or Technical School	Categorical: ordinal	Aims 2-5 (part of composite observations)
Health Insurance Status	Yes No	Categorical: dichotomous	Aim 3: Independent Variable: Aim 4: Moderator Aim 4: Covariate
Cigarette Smoking Level	Never Smoked Former Smoker Current Smoker - some days Current Smoker - everyday	Categorical: ordinal	Aim 5: Independent Variable in latent growth modeling (LGM) Aim 5: Dependent Variable (1^{st} regression) Covariate (2^{nd} regression)
Alcohol Consumption Level	None/ Moderate: 1/day women and 2/day men	Continuous: interval	Aim 5: Independent Variable (LGM) Aim 5: Dependent Variable (2^{nd} regression)

	Binge: 4+/occasion women and 5+/occasion men Heavy: 8+/week women and 15+/week men		Covariate (1st regression)
Mental Healthcare Utilization	Yes No	Categorical: dichotomous	Aim 4: Independent Var. Aim 5: Covariate
Employment Status	Employed Not Employed	Categorical: dichotomous	Aim 5: Covariate
Chronic Health Conditions Status	Yes Chronic Disease No Chronic Disease	Categorical: dichotomous	Aim 5: Covariate
Days missed of work	None A week (1-5 days) Two Weeks (6-10 days) More than two weeks (11+ days)	Continuous: interval	Aim 5: Covariate
Treatment can help people live with mental illness live normal lives	Agree strongly Agree slightly Neither agree nor disagree Disagree slightly Disagree strongly	Categorical: Ordinal	Ain 5: Covariate
People are generally caring and sympathetic to people with mental illness	Agree strongly Agree slightly Neither agree nor disagree Disagree slightly Disagree strongly	Categorical: Ordinal	Ain 5: Covariate
General Health Indicator Status	Excellent Very Good Good Fair Poor	Categorical: ordinal	Aim 5: Covariate

3.8. EMPIRICAL MODELS AND ANALYSIS

The table below shows the summary of each aim, the role of each variable, and the statistical procedure used for each specific Aim.

Table 3.3. Summary – Variable Description and Variable Levels to be Used in the Analysis

Specific Aim	Dependent or Key Variable	Independent Variable of Interest	Mediator	Control Variable/ Covariate	Statistical Procedure Used
Aim 1	Psychological distress				Exploratory Factor Analysis and Confirmatory FA
Aim 2	Psychological Distress	Year (pre-recession and post-recession)			Latent Growth Modeling
Aim 3	Psychological Distress	Income Health Insurance Status			LGM with time dependent covariate
Aim 4	Psychological Distress	Mental Health Utilization	Insurance Status		LGM with time dependent covariate
Aim 5	Psychological Distress	Alcohol Consumptio n Smoking Status			LGM with time dependent covariates
Aim 5	1. Alcohol Consumption 2. Smoking Status	1. Psychol Distress 2. Year		1. Insurance Status 2. MH Utilization 3. Chronic Health Conditions Status 4. General Health Indicator Status 5. Attitude towards MI treatment 6. Caring and sympathetic towards people with MI	Linear Regression

3.9. UNIT OF ANALYSIS

The unit of analysis is the composite observation created from the BRFSS survey respondents to reflect different combinations of age-sex-race-income-education strata, each stratum representing one observation that is longitudinally observed for group level changes. The computed observation process is described below.

Computed Observations:

We use computed observations because longitudinal data with repeated measures on the same subject are not available. In general, trend analysis is performed using longitudinal data (collecting data at different points in time for the same individual). Since the BRFSS surveys are cross-sectional, a one-to-one individual correspondence over time is not possible (data is collected one time at a single point in time). As a result, it is necessary to create composite observations to act as individual observations based on demographic and socio-economic status variables. Each composite observation is unique and is comprised of the following variables: sex, race, age, income level, and education level. For example, one composite variable is: female, white, 18-24 years of age, income less than 15K, did not graduate high school; another is male, black, 30-44 years of age, income of 15-25K that graduated high school. The total number of unique computed observations is 2 (gender) x 2 (race) x 4 (age) x 3 (income levels) x 4 (education levels) = 192. These composite observations are computed for each year of study and used in the analysis. Creating composite population level observations were based on the paired samples design by common factor at the record level (demographics and socioeconomic characteristics); more specifically, matched samples, where individuals are matched on personal characteristics (in this case 5 characteristics) (Stuart, 2010).

3.10. MANAGEMENT OF MISSING DATA

Missing data: An analysis was performed on the composite observations with missing data to assess if the missing data is at random or non-random (meaning that the missingness is correlated with other measured variables: is (non-random) or is not (random)). There are two types of data missing mechanisms for missing at random: missing completely at random (MCAR; Rubin, 1976) and missing at random (MAR) and one missing data mechanism for missing not at random (NMAR). MCAR can be tested (Little, 1988), a p-value of less than 0.05 is interpreted as missing data not being MCAR or not missing at random. MAR means the missingness can be explained by the variable for which we have full information. For MCAR the data missing is completely at random (there is no relationship between a missing data point and any values in the dataset, missing or observed). If the missing data are MCAR, then a LISTWISE deletion was used, such that observations with missing data are removed from the sample. If the missing data are MAR (the data loss depends on other variables) or not missing at random, the missing information can be recovered via imputation where missing scores are replaced by predicted scores via multiple imputation (MI) procedure.

A decision on deleting observations with the missing data via LISWISE method or using the MI procedure is made once the analysis is complete.

3.11. SPECIFIC AIM-WISE DESCRIPTION OF ANALYTIC METHODS

3.11.1. SPECIFIC AIM 1: Identify (EFA) and Confirm (CFA) the Factor Structure of K6 in order to Confirm that the Meaning and Structure of the K6 Survey Holds for the Datasets

Structural Equation Modeling (SEM)

Instead of the traditional bivariate analysis, the selected approach uses structural equation modeling (SEM) to test the hypotheses. SEM is preferable because it incorporates latent (unobserved) variables and is much more flexible at analyzing multiple latent dependent variables as compared to regression. As part of the SEM family the following statistical methods will be used: exploratory factor analysis (EFA), confirmatory factor analysis (CFA), measurement invariance, and calculation of factor scores. EFA analysis determines the number of factors that exist within the K6 survey and CFA analysis confirms the number of factors within the K6 survey. Measurement invariance evaluates latent variable model structures across multiple groups, settings, and time periods. Factor scores are numerical scores and are calculated as part of EFA and/or CFA. It indicates an individual's relative standing/score on a latent factor (factor scores are analogous to the Y hat scores in the regression analysis).

Because the Mental Illness and Stigma module was not created for the specific purpose of this research, this study includes a preliminary analysis to validate the factor structure underlying the set of data via EFA. Based on this result, a CFA is used to validate the factor structure found in the EFA.

SEM equations model relationships between endogenous and exogenous variables, and the relationships among endogenous variables. Endogenous variables (i.e., variables predicted by other latent variables) act as a dependent variable in at least one SEM equation. In this research study the endogenous variables are psychological distress, depression symptoms, and anxiety symptoms. Exogenous variables are always independent variables (i.e., not predicted by other variables) in the SEM equations. In this research the exogenous variables are cigarette smoking level, health care utilization, alcohol consumption, and health insurance status.

This work used SEM instead of the traditional bivariate analysis because all dependent variables are considered latent variables. SEM is much more flexible for analyzing multiple latent dependent variables simultaneously. In addition, part of the statistical analysis is exploratory factor analysis, followed by the confirmatory analysis, and factor scores for each individual. SEM gives us the option to analyze unobservable variable. The most common analysis in SEM includes Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), and Path Analysis.

SEM Notation

EFA and CFA models are represented by path diagrams. A path diagram consists of nodes representing the variables and arrows showing relations among these variables. By convention, in a path diagram, latent variables (e.g., depression symptoms) are represented by a circle or ellipse and observed variables (e.g., a score on a rating scale) are represented by a rectangle or square. Arrows are generally used to represent relationships among the variables. A single straight arrow indicates a causal relation from the base of the arrow to the head of the arrow. Two straight single-headed arrows in opposing directions connecting two variables indicate a reciprocal causal relationship. A curved two-headed arrow indicates there may be some association between the two variables. Error terms for a variable are inserted into the path diagram by drawing an arrow from the value of the error term to the variable with which the term is associated (Gunzler et al., 2013).

<u>Exploratory Factor Analysis (EFA)</u>

By performing EFA on the K6 survey, the number of constructs that are underlying by this questionnaire were determined. Figure 3.1 shows a path model for 6 variables ($V_1 - V_6$) with 2 factors (F_1 and F_2). There are 3 factors that affect V_1: factors F_1 and F_2 and e_1 (or the specific item error).

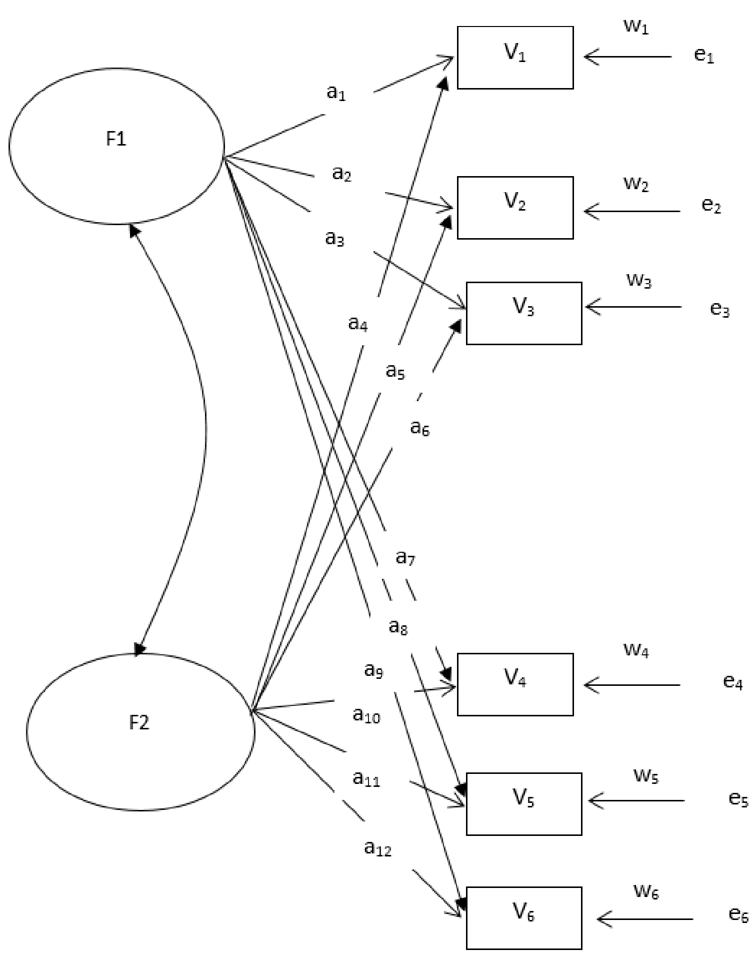

Figure 3.1. Path Model for a 6 Variable, 2 Factor Model, Oblique Factors

By multiplying these factors by specific weights (represented by a_1) it is possible to calculate any subject's score on V_1 (Hatcher, 1994). EFA assumes that the observed variables are a linear combination of the underlying factors as listed below:

$$V_1 = (a_1)*F_1 + (a_4) *F_2 + w_1*e_1$$

The regression weights ($a_1 – a_{12}$) are calculated from the factor pattern matrix. The factor pattern matrix is represented by relationship of each variable and the underlying factor or factor loadings and describes the relationship between the latent variables and their indicator variables (Hatcher, 1994).

<u>Confirmatory Factor Analysis (CFA)</u>

In this study, only the K6 survey questions were used with CFA analysis. The K6 scale also known as the psychological distress scale is our variable of interest. Based on current literature (Ko & Harrington, 2015; Kessler, 2003), along with the SEM notations mentioned in the prior section, the proposed CFA model is shown below, where $\lambda_{1\text{-}6}$ represent the latent loading factors for each item, $e_{1\text{-}6}$ represent the random error associated with each item. Anxiety and depression represent the latent variables, and nervs, restls, worth, hopls, deprd, effrt are the survey items (questions). The results from the EFA were used to guide the CFA analysis. Multiple models were run in order to find the best fitted model. As shown in figure 3.2 the CFA model has two components: one is the structural model and the other is the measurement model. CFA is a measurement model that

<u>Model Evaluation Criteria</u>

CFA model results follow the same comparison, but because the models are comprised of two components there are a set of criteria that needs to be followed. There are two types of testing the fit of a CFA model: global fit and local fit. The global fit statistics measures the average model fit and it is only one step in analyzing the fit of the model. The most common global fit test statistic is the Chi-square fit index. The Chi-square test statistics assesses the overall fit and the difference between the sample covariance matrix and the fitted covariance matrix. The lower the chi-square value the better the model fit (the cut-off for the good fit is p-value>0.05), but because the chi-square fit index is sensitive to the sample size it is causing the index to be statistically significant even if the model is not a good fit. As a result, additional indices will be used.

These indices are local fit indices and are not significance tests. These indices are intended as continuous model data-correspondence (Kline, 2011). The local fit indices include four categories: 1) absolute fit indices (measures how well a priori model explains the data. The priori model is the researcher's model because there is no other point of reference (Kline, 2011)). Selected absolute indices are root-mean-square error of approximation (RMSEA), standardized root mean square residual (SRMR), and goodness of fit index (GFI). These fit indices are relatively independent of sample size. RMSEA represents the closeness of fit between the proposed model and the population model (residual covariance matrix) and should be 0.05 or less while the 90% confidence interval (CI) around RMSEA should contain 0.05 (Browne & Cudeck, 1993). SRMR is the average of standardized residuals between the specified and obtained variance-covariance matrices and should be 0.08 or less (Joreskog & Sorbom, 1996; Hu & Bentler, 1999). GFI is the proportion of variance accounted for by the estimated population covariance and the values should be more than 0.95 (Schumaker & Lomax, 2010); 2) relative fit indices compare a Chi-square for the model tested to a null model/baseline model.

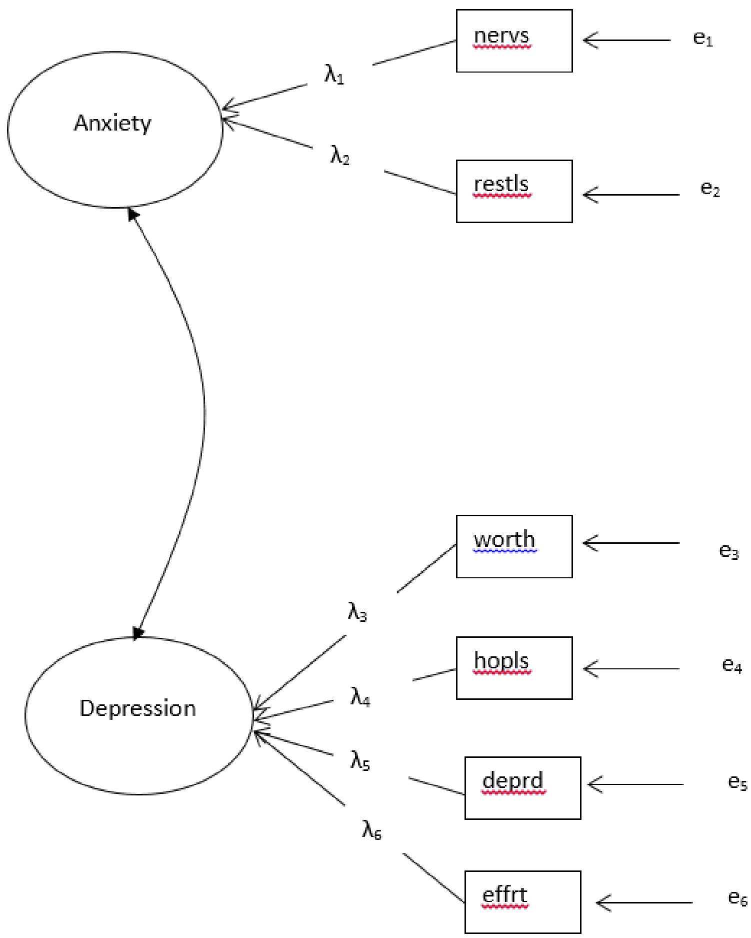

Figure 3.2. Proposed CFA Model

A baseline model is a model in which there are no latent variables (all measured variables are uncorrelated). The baseline model should have a poor fit. Selected relative fit indices are: normed fit index (NFI), nonnormed fit index (NNFI) or Tucker-Lewis Index (TLI), and Comparative Fit Index (CFI; Bentler, 1990). NFI (Bentler & Bonett, 1980) measures the proportion of total observed covariance as explained by the hypothesized model as compared to the baseline model. Values of 0.95 or higher are recommended (Schumaker & Lomax, 2010). NNFI (Bentler & Bonett, 1980) or TLI

(Tucker-Lewis, 1973) measures the relative improvement of fit per degree of freedom of the proposal model over the baseline model. The recommended cut of is 0.95 (Hu & Bentler, 1998; 1999). NNFI or TLI is recommended for smaller samples. CFI measures a relative improvement of a model over the baseline model. The higher the value the better; the recommended threshold is 0.95 (Hu & Bentler, 1998; 1999). Table 3.4 below shows a summary of the indices and their cut off values.

Table 3.4. Model Evaluation Criteria

Index	**Cut Off Value**	**Explanation and Reference**
RMSEA	0.05	Closer to zero better fit (Steiger 1989; Browne & Cudeck, 1993)
NFI (Bentler & Bonett, 1980)	0.95	Closer to one better fit (Schumaker & Lomax, 2010)
NNFI (Bentler & Bonett, 1980)/ TLI (Tucker-Lewis, 1973)	0.95	Closer to one better fit (Hu & Bentler, 1998; 1999)
CFI (Bentler, 1990)	0.90	Closer to one better fit (Hu & Bentler, 1998; 1999)
SRMR	0.08	Closer to zero better fit (Joreskog & Sorbom, 1996; Hu & Bentler, 1999)
GFI	0.95	Closer to one better fit (Schumaker & Lomax, 2010)

3.11.2. <u>SPECIFIC AIM 2: Examine the Trend of Psychological Distress During and After the 2007 Recession</u>

This objective examines the trends of depression symptoms and anxiety symptoms using the levels of psychological distress as described by Kessler et al., 2003, Prochaska et al., 2012, and Forman-Hoffman et al., 2014. The questions were coded from 1 “none of the time” to 5 “all of the time” allowing for a psychological distress score between 6 and 30. Kessler, Prochaska, and Forman-Hoffman did research on the different levels of psychological distress and came up with three distinct groups: individuals with total scores between 6 and 10 are classified as having low psychological distress, individuals with total scores between 11-18 are classified as having moderate psychological distress, and individuals with scores above 19 are classified as having serious psychological distress. The psychological distress is the sum of the item scores for each individual. Psychological distress adds the responses for both depression symptoms and anxiety symptoms.

The goal of this objective is to estimate the growth or trend in psychological distress along with depression and anxiety symptoms over time. The first step in this objective is to combine all the datasets into one to make it act as a longitudinal dataset. Once that is done the latent growth modeling (LGM) analysis can be performed.

The Integrative data analysis (IDA) framework assists by providing a structure for collection of data from multiple sources. IDA was developed by Curran et al. in 2008 and they define it as the analysis of multiple independent data sets that have been pooled into one dataset for investigations. There are two ways to merge the datasets: (1) by common data elements (such as multi-item scales or indices); and (2) by a common factor at the record level (demographics and SES in this case).

Since the data was collected from different sub-samples of the population in different years, a robust analysis on equivalence is needed. This was done by eliminating the between-sample sources of variability. The sources of variability can be methodological (ex: variation in the methods of collecting data, variation in measures of variables, variation in construct) or be related to population differences (geography, history, culture, etc.). Combining the 4 cross-sectional datasets into one is done by using the composite observations described in the variables section. These composite observations act as individual observations.

Eliminating sources of variability:

The same K6 questionnaire used in all the years which makes it easier to show that the analysis is examining the same construct (same latent variables). In general, EFA and CFA can be run on the same dataset split in half, one half used for EFA and one half used for CFA. In this case, since multiple years are available the 2012 dataset was used for EFA analysis and the 2013 dataset was used for CFA analysis. Once the structure is confirmed CFA was run on the remaining two datasets to confirm the latent structure.

a) Measurement equivalence

Measurement equivalence is testing to see if the latent variable is invariant across groups and was done via invariance testing or measurement invariance (MI). There were multiple steps to test MI (Cheung & Rensvold, 2002; Little, 1988):

1) Configural invariance: a model with no constrains or baseline model. It examined if the underlying factor structure was the equivalent across groups (if the number and type of factors were the same across groups). In this analysis gender (the distribution between males and females were equal) was used as the group and if configural invariance hold then there is no difference in the way that males and females answered the K6 questionnaire.
2) Metric invariance: a model where the factor loadings were equal across groups, but the intercepts were allowed to differ between groups (tested whether the meaning of the construct was the same across groups). The metric invariance was testing that all factor loadings were equal across groups. It tested the strength of these relationships

and was achieved by evaluating the fit of the model when the loading parameters were equal across the two groups. (Davis & Finney, 2003).

3) Scalar invariance: a model where only the intercepts and loadings were equal across groups (groups were compared on their scores on the latent variable).
4) Means invariance: a model where the loading, intercepts, and means were constrained to be equal across groups.

<u>Criteria to compare the models:</u>

Initial studies recommended the use of chi-square tests to decide if the increase in fit is substantial (Byrne et al., 1989). Because the traditional check of global model fit, the chi-square test (Cochran, 1952), is dependent on the sample size (it rejects reasonable models if the samples size is large), Cheung & Rensvold, 2002 suggested using a difference in CFI to examine the invariance (a difference in CFI larger than 0.1 indicates that the invariance should not be rejected). We used the difference in CFI method since it is most widely used and best supported method for invariance testing (Chen et al., 2001; Cheung & Rensvold, 2012; Meade et al., 2008).

Since the goal of the analysis is to compare the groups on their scores on the latent variable the analysis stops with the last model or at scalar invariance. In general, in invariance testing groups are considered based on population characteristics (i.e. gender). In this case "groups" will be considered the multiple survey years.

b) <u>Sampling Mechanism</u>

BRFSS survey is a national telephone survey taken yearly by CDC. In 2007 about 93% of the United States as a whole had a landline, but since then an increasing portion of interviews were performed via cell phone. The percentage of cell phone-only households increased to 39% in 2013 (Blumberg and Luke, 2013). The shift from landline to cellular phone in the United States caused the CDC to change the protocol of collection samples beginning in 2012 (addition of cell phone data and revised weighing methodology among others). The first step in accounting for the differences in sample mechanisms is to verify that the sampling mechanism for each year of survey have been maintained throughout the study period. This was verified and therefore determined that there is no need for remedial measures such as the method suggested by Curran and Hussong, 2009 (incorporating sampling information into a model-based framework that either meets Fisher's (1992) nonprobability sampling conditions, or Neyman's (1934) probability sampling conditions).

c) Geographical differences

In addition to the revised sampling mechanism, the mental health and stigma module, since it is an optional module, was not administered uniformly throughout the United States. Twenty-seven states used the module in 2007 (Alaska, Arkansas, California, Connecticut, District of Columbia, Georgia, Hawaii, Illinois, Indiana, Iowa, Kentucky, Louisiana, Massachusetts, Minnesota, Mississippi, Missouri, Montana, Nevada, New Hampshire, New Mexico, Oklahoma, Puerto Rico, Rhode Island, South Carolina, Vermont, Virginia, and Wyoming), 8 states in 2009 (Georgia, Hawaii, Mississippi, Missouri, Nevada, South Carolina, Vermont, and Wyoming), 11 states in 2012 (Illinois, Iowa, Minnesota, Missouri, Montana, Nevada, New Mexico, New York, Oregon, Puerto Rico, and Washington), and 5 states in 2013 (Colorado, Minnesota, Nevada, Tennessee, and Washington). These geographical differences could potentially create between-study heterogeneity.

d) Difference in time (history)

The data points are pre-, during, and post- economic recession history, which can play an important role in the analysis. The proposed analysis examined if societal changes (changes in stigma of mental health), and other characteristics had an influence on the trend of mental health status. As a result, the model included questions 9 and 10 of the of the Mental Illness and Stigma module as a "stigma" variable. In longitudinal studies, maturation is an important confounder. By integrating multiple cross-sectional studies, maturation should not be an issue since the sample in each study is not the same over time.

Latent growth modeling

The outcome of the latent growth model in this study (LGM) is the item responses to the K6 questionnaire represented by the latent variables (psychological distress, and depression and anxiety symptoms). LGM examines the growth trajectory of attributes for an individual at two or more time periods (McArdle, 1988). LGM models the level (intercept) and rate of change (slope). The intercept is a constant for any individual across time and has the same meaning as the intercept for a regression (Duncan & Duncan, 2009). The slope represents the trajectory of an individual over time.

Latent growth modeling is done in two steps. In the first step, a basic model is analyzed as shown in figure 3.2. This is necessary to explain the covariances and the means of the variables and to make sure that the model is acceptable.

In this step, we create a basic model showing the mean scores for each year. The type of shape of the mean scores will dictate what type of model will be created: linear or non-linear.

In figure 3.3 the intercept is analogous to the intercept in a latent variable regression equation and represents the initial status (baseline information) about the latent variable under study. The variance of the intercept represents the variability about individuals' baseline values. The unstandardized loading factors for intercept (λs) are fixed to 1.0.

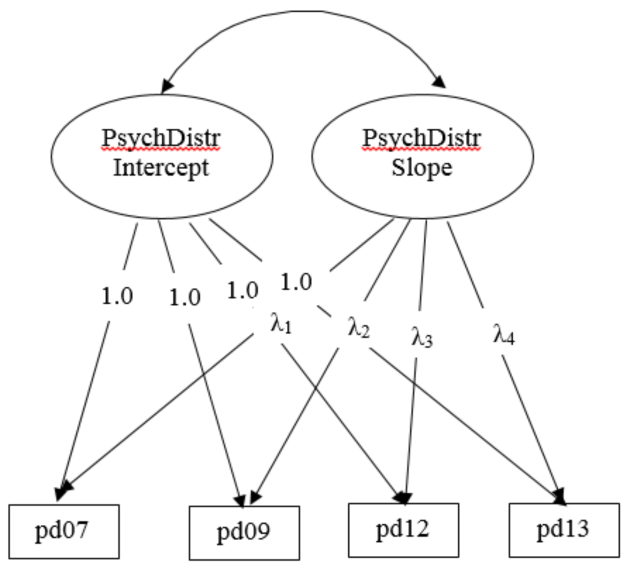

Figure 3.3. Proposed Latent Growth Model, no covariates, where pd07, pd09, pd12, and pd13 are the responses for each year for psychological distress.

Each composite observation's score can be expressed as following:

Score at a point of time t = score at time t-1(intercept) + (change in score per unit time score at time t minus score at t-1) x (time elapsed between t and t-1) + error

In this case the expression above can be written as:

pd07 = PD intercept + (PD slope) x 0 + error1*

Where PD for each composite observation is the average Psychological Distress score of the individuals within each composite observation. PD07 is the PD average for the age-sex-race-income-education stratum (total 192 strata or composite observations each year). The LGM model statements for the rest of the years are shown below:

Model statement for 2009: PD09= PD score at 07 + difference of (PD09-PD07) + error

Expected Result: Predicted PD09 = PD intercept + (PD slope) + error2*

Model statement for 2012: PD12= PD score at 09 + difference of (PD12-PD09) + error

Expected Result: Predicted PD12 = PD intercept + (PD slope) + error3*

Model statement for 2013: PD13 = PD score at 12 + difference of (PD13-PD12) + error

Expected Result: Predicted PD13 = PD intercept + (PD slope) + error4*

(*The error terms are not shown in figures 3.2 and 3.3).

The slope for each year is analogous to the slope in a latent variable regression equation, and the shape of the plotted slopes represents the variable growth over time. Depending on the type of statistical analysis the loading factors can be linear, where the fixed coefficients represent linear growth from the baseline (set at zero) where the values represent measurement in time past the baseline (in this case 0, 2, 5, 6 representing all the available years 2007, 2009, 2012, and 2013 respectively), or non-linear where the only anchored loading factor is the 2007 set at 0. The remaining parameters will be left free meaning that the model will calculate the slope. For this study we are using non-linear loading factors. The mean of the slope growth factor for non-linear models is the change in the outcome variable for a one-unit change in the time score (not a constant rate of change). In this case the time scores are the years: 2007, 2009, 2012, and 2013. As a result, in order to identify the model, at least one-time score must be fixed to a non-zero value, in addition to the time score that is fixed at zero (anchor point).

In linear growth models, a covariance between initial intercept and slope is calculated and interpreted as the extent to which the value of the intercept predicts the rate of the subsequent change. In the case of non-equal linear growth rates such covariance does not exist: it is included in the time scale of the growth model.

The second step is to explore covariates that might predict the change by year, by adding covariates the model, as described under SA 3 below.

3.11.3. SPECIFIC AIM 3: Examine the Trend of Psychological Distress Before, During, and After the 2007 Recession by Income and Health Insurance Status

The second step in the LGM analysis is to predict change, by introducing covariates in the model. The introduction of the covariates in the model show the effect of the covariate on the intercept and slope. Income and health insurance are the two variables that are affected the most during economic recession (Barr, 2012; McLaughlin, 2012; Holahan, 2011). Schaller and Stevens, 2015 used the Medical Expenditure Panel Survey (MEPS) from 1996 through 2012 and found that displaced workers (workers who lost their jobs involuntarily) were more likely to be male, black or Hispanic, and slightly younger than their continually-employed counterparts. They were also less educated, with a fair or poor general health and more likely to experience poor mental health.

Figures 3.4 shows the proposed Latent Growth Models including the two covariates.

Because health plan and income are time variant, a time scale has to be incorporated in the model. The time scale found in SA2 will be used in SA3 (0 for year 07 – anchor year, “a” for 2009, “b” for 2012 and “c” for 2013).

Predicted model:

Predicted Intercept Mean = Predicted Intercept (of Intercept) + Predicted Intercept (HlthPln)*Sample Mean (HlthPln) + Predicted Intercept (Income)*Sample Mean (Income)

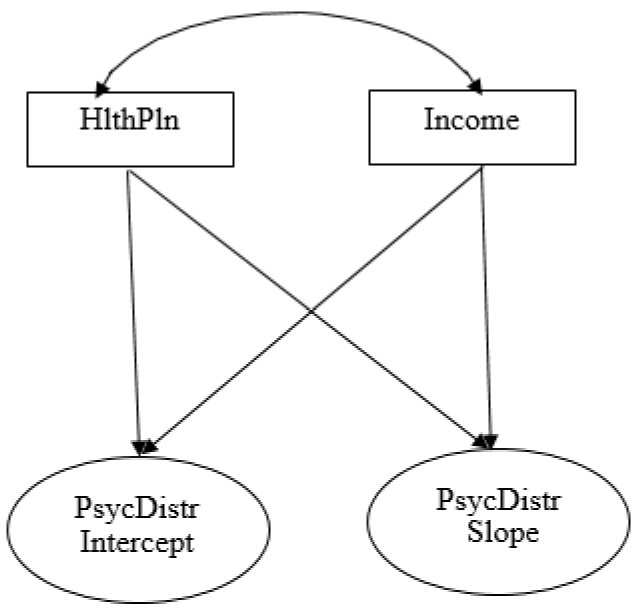

Figure 3.4. Proposed Latent Growth Model including Health Plan and Income as Covariates

Predicted Slope Mean = Estimated Intercept (of Slope) + Predicted Slope (HlthPln)*Sample Mean (HlthPln) + Predicted Slope (Income)*Sample Mean (Income)

General model:

Predicted PD mean at time t = Predicted Intercept Mean + Predicted Slope Mean* (Time Score at Timepoint t)

Predicted PD mean 07 = Predicted Intercept Mean (calculated above) + Predicted Slope Mean (calculated above) * (0) since 07 is the anchor year.

Predicted PD mean 09 = Predicted Intercept Mean (calculated above) + Predicted Slope Mean (calculated above) * (a) calculated in SA2.

Predicted PD mean 12 = Predicted Intercept Mean (calculated above) + Predicted Slope Mean (calculated above) * (b) calculated in SA2.

Predicted PD mean 13 = Predicted Intercept Mean (calculated above) + Predicted Slope Mean (calculated above) * (c) calculated in SA2.

We expect the absence of health insurance to have a negative impact on psychological distress and we expect lower levels of income to have a negative impact on psychological distress.

3.11.4. SPECIFIC AIM 4: Investigate the Association Between Mental Healthcare Utilization and Psychological Distress and the Mediation Inference of Health Insurance on Mental Healthcare Utilization

Aim 4 examines the association between health utilization and psychological distress, via the health insurance status. Individuals with health care insurance have a higher propensity for use of health care services compared to individuals that do not have health insurance (Institute of Medicine, 2001). In this aim the independent variable is healthcare utilization (HC Utilization), the dependent variable is psychological distress (PsycDistr), with the health insurance status as a mediator (HlthPln). It is expected that HlthPln will lower the effect of PD.

A variable is a mediator when there is a causal relationship between the variables. In this case the presence or absence of health insurance causes the level of health care utilization. The mediator is assumed to cause the outcome (MacKinnon, Fairchild, and Fritz, 2007). The presence or absence of health insurance increases or decreases the level of psychological distress. Calculating the effect of a variable as a mediator is done via direct and indirect effects. A direct effect is when the variable directly affects the outcome variable. The direct effect is the effect of HC Utilization on PsycDist. An indirect effect is when the variable indirectly or with the help of another variable affects the outcome variable. The indirect effect is the effect between HC Utilization times the effect of HlthPlan on PsycDistr. The total effect of an independent variable on the outcome variable is the sum of direct plus indirect effects. The total effect is the effect of HC Utilization on PsycDistr + HC Utilization * HlthPlan on PsycDistr.

There are 3 necessary but not sufficient conditions to establish mediation (Baron & Kenny, 1986; Little et al., 2007):

1. Show that exogenous causal influence is related to endogenous causal influence or mediator. In figure 3.5., the exogenous causal influence is HC Utilization and the endogenous causal influence or mediator is HlthPln.
2. Show that the mediator is significantly related to the dependent variable or outcome. In figure 3.5., the mediator is HlthPln, and the outcome is PsychDistr.
3. Since HC Utiliation, and HlthPln are time variant variables, same steps taken in SA3 will be taken in SA4 to estimate the association via LGM method.

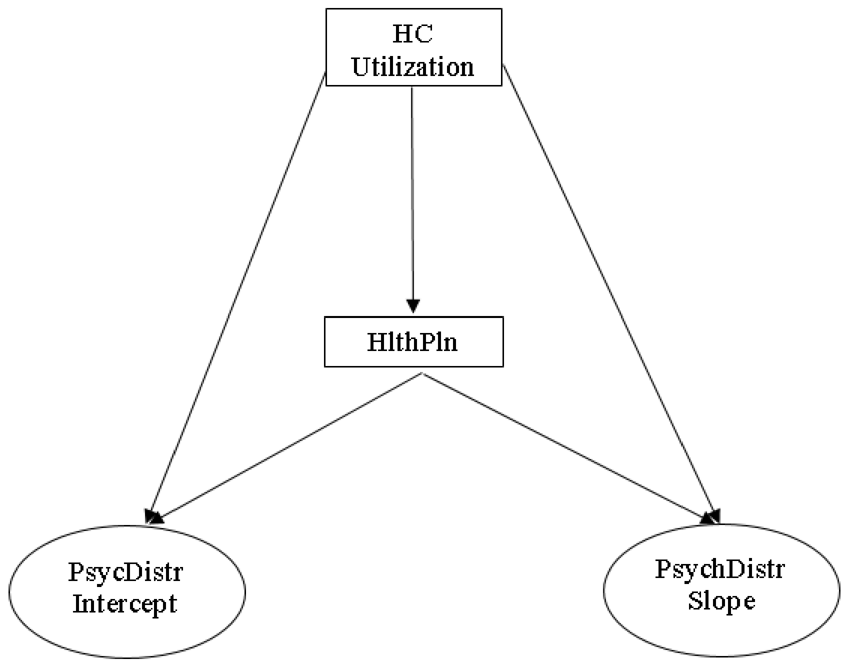

Figure 3.5. Proposed Latent Growth Models including Health Plan as Covariate and Healthcare Utilization as a Mediator

3.11.5. SPECIFIC AIM 5: Investigate the Effect of Alcohol and/or Tobacco Consumption on Psychological Distress and the Effect of Psychological Distress on Alcohol and/or Tobacco Consumption

Specific Aim 5 is analyzed in two steps:

First step using LGM, the first step in the analysis is to add the final covariates: risky behaviors represented by the level of cigarette smoking and the level of alcohol consumption before, during, and after the recession. Since both variables are time variant variables, same steps taken in SA3 and SA4 will be taken in SA5 to estimate the association via LGM method.

Prior research investigating the association between mental health and smoking and/or alcohol consumption found that depression and anxiety are a risk factor for smoking and alcohol consumption (Lasser et al., 2000; Le Cook, et al., 2014). Bell and Britton, 2014 looked at the temporality between mental health and risky behaviors and found that the mental health status influenced changes in alcohol consumption and not vice versa. The proposed analysis expands on these findings, by examining the association between the level of psychological distress (independent variable) and the level of smoking and alcohol consumption (dependent variable) during times of economic hardship.

The first model examines the time-dependent association between PD and the 4 levels of severity of cigarette smoking. The second model examines the time-dependent association between PD and the 4 levels of severity of alcohol consumption. In both models it is expected that smoking/alcohol use severity increased with PD severity.

In the second step, linear regression is used to assess the association between smoking status (dependent variable) and the severity of PD (independent variable, 3 levels: low, moderate, serious) over time while accounting for all the covariates (alcohol consumption, demographics, income, HC utilization, employment status, chronic health conditions, general health indicator).

a) Smoking status = PD severity each year + Alcohol Consumption + Sex + Race + Income +Age + Education Level + Health Insurance + HC Utilization + Employment Status + Chronic Health Conditions + General Health Indicator

Similar linear regression is used to assess the association between alcohol consumption and the severity of PD.

b) Alcohol Consumption = PD severity each year + Smoking Status + Sex + Race + Income +Age + Education Level + Health Insurance + HC Utilization + Employment Status + Chronic Health Conditions + General Health Indicator +Region

CHAPTER 4

RESULTS

4.1. INTRODUCTION

The analyses of the models included structural equation modeling and regression modeling to assess the outcome of interest. Structural equation modeling (SEM) was used in Aims 1-4. The first step in Aim 1, was to analyze the K6 survey answers for each survey years to find the optimal survey structure. In order to find the structure of the K6 survey, EFA was run on the 2013 data, followed by CFA for year 2012 to confirm the structure. The detailed analysis of the EFA and CFA is explained in Aim 1. The EFA and CFA included only the K6 survey answers with no covariates. For Aims 2-5 covariates were added to the models. Since the data for all 4 years is not longitudinal, composite observations were created before the data were merged. The following section describes the steps taken to clean and arrive to the final dataset.

4.2. DATA CLEANING

In order to analyze the final sample, the full dataset was cleaned and transformed to conduct the study. The final sample included the variables listed in the methods section.

Notes: While cleaning and transforming the data, discrepancies were noticed while running frequencies. Frequencies were run on the K6 items to find the discrepancy and it was noted some states listed the CDC website as participants did not use the optional module in their 2007 and/or 2009 BRFSS survey. The following states did not participate in 2007 survey: Alabama, Arizona, Delaware, Florida, Idaho, Maryland, North Carolina, North Dakota, South Dakota, Tennessee, Utah, and West Virginia. In 2013 the following state did not participate in the survey: Colorado. Since these states did not have any responses to any of the K6 questions, they were deleted from the dataset, leaving 2007 with 15 states and 2013 with 4 states. No states were missing from 2009 and 2012 datasets.

Missing Data: The first step was to look at the pattern of the missing data in order to examine if the missing pattern is at random or not. The SAS 9.4 program was used to examine the pattern of the missing data, more specifically the proc MI procedure that outputs a table of the missing data. The table also provides the means for each variable. The decision of the random missingness was made after looking at the means for each variable along with the pattern of the missing data. The pattern of the missingness of the data looked to be random, no imputation was needed. The randomness of missing data coupled with the high response numbers for each year the missing data was listwise deleted. The listwise deletion throws away any rows that have at least one missing value (Allison, 2000; Baraldi & Enders, 2010). A second missing data analysis was performed. Multivariate imputation by fully conditional specification method was run on each dataset and the results of the imputation were close to the results obtained with a listwise deleted dataset. As a result, the listwise dataset was used in subsequent analysis.

The initial (all responses for entire BRFSS survey) and final sample used in the analysis (N) are listed below by year. The initial sample includes all the observations in the full BRFSS survey. The final sample includes only the variables used in the analysis.

Table 4.1. Data Sample by Year

	2007	2009	2012	2013
Initial Sample	430,912	432,607	475,687	491,773
Final Sample	97,610	40,232	63,838	24,662

Descriptive Statistics

Table 4.2 provides descriptive statistics for the Kessler 6 survey questions and psychological distress including the two factors (depression symptoms and anxiety symptoms) for each year. Values of skewness over 3.0 and values of kurtosis below 10 with few just above 10.0 may illustrate problems with normality (Kline, 2011). The level of skewness and/or kurtosis indicates if the data is normal or non-normal. Skewness can be positive or negative and indicates if the data is asymmetrical around its mean, with a longer tail to the left (positive) or to the right (negative) side of the distribution. Kurtosis can also be positive or negative and describes the shape of the curve (how spread the data is around its mean). Most variables have values of skewness and kurtosis below 3 and 10, respectively, with only two variables (How often during the past 30 days you feel worthless?" in 2009, and "How often during the past 30 days you feel worthless?" in 2013). Because only two of the variables are non-normally distributed the ML or maximum likelihood estimate was used in the analysis, but the results did not converge. As a result, Kline (2011) suggested using a different estimation technique to help accommodate for the non-conversion which can be brought on by non-normality: the MLM (maximum likelihood estimation) with robust standard errors and a

Satorra-Bentler scaled statistic (Satorra & Bentler, 1998a). The Satorra-Bentler scaled statistic is a mean-adjusted chi-square that helps adjust the chi-square for non-normality. With this estimation the parameter estimates are the same as those estimated with "standard" ML (maximum likelihood) techniques; only the standard errors and the chi-square are adjusted to produce results that are more in line with what would be obtained if normality was presented.

Table 4.2. Data Descriptive

Variables	Mean	SD	Skewness	Kurtosis	Range
Variables of Interest					
Kessler 6 Survey Questions					
2007					
How often during the past 30 days you feel…					
Nervous	1.90	0.98	1.02	0.64	1-5
Hopeless	1.38	0.79	2.29	5.07	1-5
Restless of Fidgety	1.89	1.04	1.05	0.48	1-5
So depressed that nothing could cheer you up	1.30	0.73	2.76	7.63	1-5
That everything was an effort	1.77	1.29	1.4	1.06	1-5
Worthless	1.29	0.75	2.95	8.74	1-5
2009					
How often during the past 30 days you feel…					
Nervous	1.90	1.02	1.03	0.49	1-5
Hopeless	1.40	0.76	2.01	3.68	1-5
Restless of Fidgety	1.93	1.01	1.07	0.69	1-5
So depressed that nothing could cheer you up	1.26	0.65	2.83	8.15	1-5
that everything was an effort	1.78	1.14	1.47	1.30	1-5
Worthless	1.23	0.67	3.20	10.35	1-5
2012					
How often during the past 30 days you feel…					
Nervous	1.84	0.99	1.2	1.05	1-5
Hopeless	1.31	0.73	2.64	7.1	1-5
Restless of Fidgety	1.81	1.00	1.23	1.08	1-5
So depressed that nothing could cheer you up	1.33	0.76	2.51	5.9	1-5
that everything was an effort	1.64	1.06	1.68	2	1-5
Worthless	1.29	0.74	2.86	8.12	1-5
2013					
How often during the past 30 days you feel…					

Nervous	1.77	0.91	1.11	0.8	1-5
Hopeless	1.33	0.84	2.83	7.62	1-5
Restless of Fidgety	1.80	1.00	1.06	0.32	1-5
So depressed that nothing could cheer you up	1.22	0.65	3.11	9.23	1-5
that everything was an effort	1.66	1.07	1.66	1.95	1-5
Worthless	1.26	0.78	3.26	10.15	1-5
2007					
Psychological Distress	9.52	4.05	1.83	3.85	6-30
Depression Symptoms	5.73	2.75	2.19	5.19	4-20
Anxiety Symptoms	3.79	1.76	1.02	0.78	2-10
2009					
Psychological Distress	9.49	3.88	1.75	3.61	6-30
Depression Symptoms	5.66	2.53	2.12	4.93	4-20
Anxiety Symptoms	3.83	1.76	1.05	0.99	2-10
2012					
Psychological Distress	9.22	4.09	1.91	3.99	6-30
Depression Symptoms	5.57	2.73	2.33	5.86	4-20
Anxiety Symptoms	3.65	1.78	1.27	1.51	2-10
2013					
Psychological Distress	9.06	4.03	2.24	5.41	6-30
Depression Symptoms	5.49	2.76	2.69	7.46	4-20
Anxiety Symptoms	2.57	1.69	1.08	0.76	2-10
Smoking Status	1.78	1.04	1.18	0.10	1-4
Alcohol Consumption	1.70	0.92	1.60	0.30	1-4
Insurance Status	1.28	0.33	2.34	3.00	1-2
Co-variates					
General Health	2.58	1.1	0.34	-0.53	1-5
Employment Status	1.46	0.50	0.17	-1.97	1-2
Chronic Disease Status	1.84	0.36	-1.89	1.57	1-2
How many days a mental health condition keep you from doing work	0.96	0.79	1.78	5.09	1-4
Taking medicine or receiving treatment for mental health condition	1.85	0.35	-1.99	1.99	1-2

4.3. SPECIFIC AIM 1: IDENTIFY (EFA) AND CONFIRM (CFA) THE FACTOR STRUCTURE OF K6 IN ORDER TO CONFIRM THAT THE MEANING AND STRUCTURE OF THE K6 SURVEY HOLDS FOR THE DATASETS

4.3.1. Exploratory Factor Analysis (EFA)

A preliminary analysis was performed on the 2013 K6 survey questions to find the factor structure underlying the dataset. As mentioned in the Methods section, EFA is appropriate when the objective of the analysis is to identify the number and nature of the underlying factors/dimensions that are responsible for the variability of the data. Based on the definitions of depression symptoms and anxiety symptoms it was expected to find two factors (dimensions): one depression symptoms (worth, hopls, deprd, and effrt), and the other one anxiety symptoms (restls, and nervs).

A step-by-step EFA analysis:

Based on the recommendations of EFA researchers (Williams et al., 2010; Yong & Pearce, 2013; Costello & Osborne, 2005), the following steps were conducted:

1) Chose an extraction method: multiple solutions were run to find the best factor structure for the data. Extraction means that observed measures are brought together by underlying common factors and unique factors. There are multiple extraction methods and the most commonly used are: Maximum Likelihood (analyzes the maximum likelihood of sampling the observed correlation matrix and it requires multivariate normality; Tabachnick & Fidell, 2007; Pett et al., 2003), Principal Axis Factor (all variables belong to the first group when the factor is extracted and a residual matrix is calculated and does not require the data to be normally distributed

(Wood, et al., 1996; Fabrigar et al., 1999), and Principal Components (used to extract the maximum variance from the data set with each component reducing the number of variables into smaller numbers of components (Tabachnick & Fidell, 2007). SAS 9.4 statistical software was used for conducting the EFA analysis. The method of extraction used was the principal factor method. The iterated principal factor method is an extension of the principal factor method that yields better estimates of the communalities. The communality is the proportion of total variance that is shared between the factors, and the higher the communality the stronger the influence of the item to the factor. Additionally, promax oblique rotation was employed because it was hypothesized the extracted factors would be correlated. The rotation is necessary in EFA to make the interpretation of the factors that are considered relevant. It works by reweighting the initial solution (the factor axes are shifted) with a target outcome of a solution where each factor explains as much of a variance as possible (Kline, 2011). There are two types of rotation: orthogonal when the axes are orthogonal to each other and oblique when they are not. Promax is a type of oblique rotation and the most widely used rotation in EFA analysis (Gorsuch, 1983).

Table 4.3. K6 Data Descriptive

Item	N	Mean	Std. Deviation	Skewness	Kurtosis
NERVS	24,662	1.773	0.944	1.240	1.238
HOPLS	24,662	1.297	0.715	2.768	7.974
RESTLS	24,662	1.731	0.991	1.358	1.344
DEPRD	24,662	1.211	0.628	3.429	12.464
EFFRT	24,662	1.569	0.997	1.860	2.841
WORTH	24,662	1.210	0.651	3.606	13.746

2) Determined the number of factors to extract:

- Looked at the scree plot to gain a sense of the number of factors needed to summarize the data. In addition, extraction based on eigenvalues of higher than one was used as a starting point for the analysis.
- As the scree plot may be considered subjective, a parallel analysis (Horn, 1965) was performed. Parallel analysis compares plots of eigenvalues from the actual data to extracted values from a correlation matrix of randomly generated uncorrelated variables with the same dimensions as the original dataset (Distefano & Dombrowski, 2006). The parallel analysis suggested that the number of factors was two.

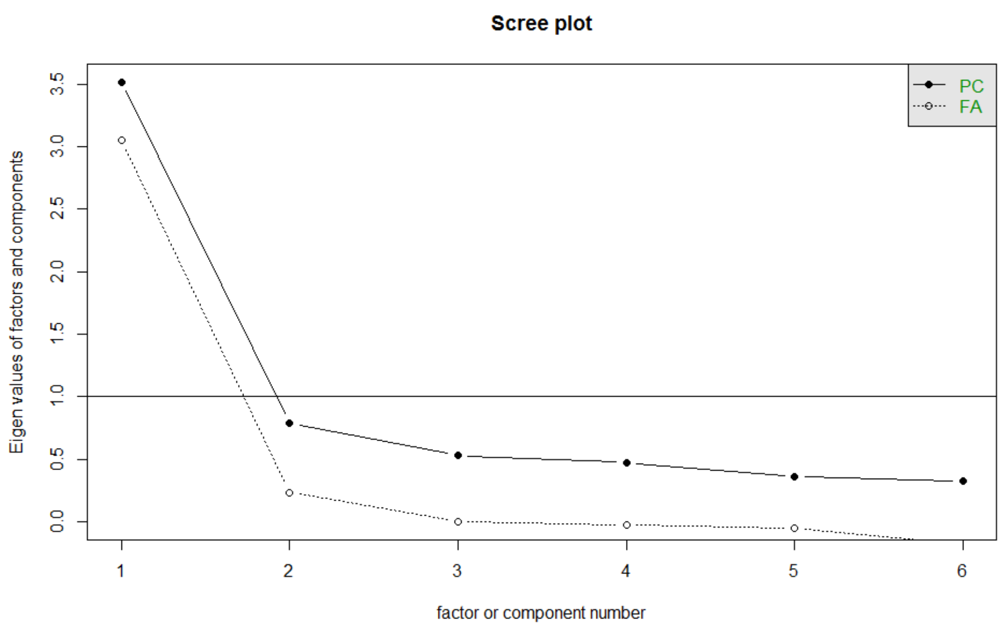

Figure 4.1. Scree Plot

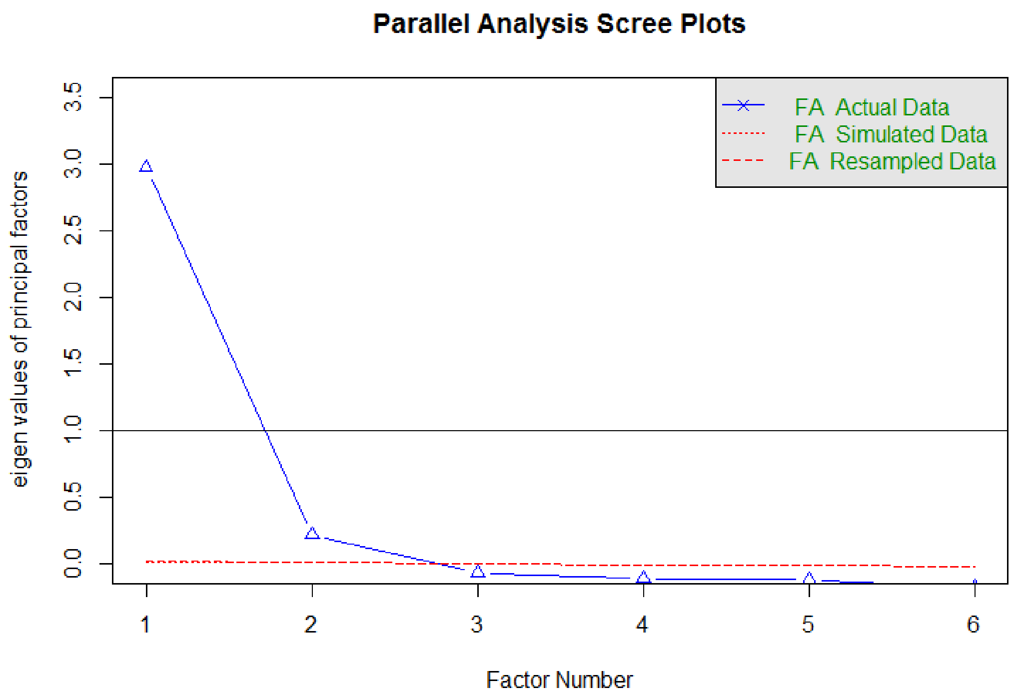

Figure 4.2. Parallel Analysis

Table 4.4. Factor Structure (Correlations)

Factor Structure (Correlations)		
	Factor 1	Factor 2
WORTH	0.79	0.51
DEPRD	0.81	0.55
HOPLS	0.82	0.58
EFFRT	0.62	0.59
RSTLS	0.47	0.75
NERVS	0.54	0.69

For the EFA, the assumption was made that the factors (Depression symptoms and Anxiety symptoms) were correlated; the assumption was correct (Table 4.4). The correlation between the two factors using the promax oblique rotation is 0.64. Examining the loading values in Table 4.4 indicates that the first three variables load higher on

Factor 1, variable 3 loads about equally on both factors, and the last two variables loading higher on Factor 2. We can see similar results with the rotated factor pattern in Table 4.5. below.

Table 4.5. Rotated Factor Pattern (Standardized Regression Coefficients)

Item	Component/Dimension			
	Factor 1 (Depression)	Factor 2 (Anxiety)	Communalities	Residuals
WORTH	**0.79**	0.10	0.63	0.37
HOPLS	**0.77**	0.05	0.68	0.32
DEPRD	**0.78**	0.09	0.65	0.35
EFFRT	**0.41**	0.33	0.45	0.55
RSTLS	-0.10	**0.75**	0.56	0.44
NERVS	0.16	**0.59**	0.49	0.51

- Examined the residual matrix to see if any of the residuals were large compared to the rest of the residuals (checked for outliers). Residual matrix is the difference between the observed matrix and the predicted matrix. Large residuals imply additional factors may need to be extracted. The residual matrix did not have any large residuals confirming the two-factor solution.

Table 4.6. Raw Residual Matrix

Raw Residual Matrix						
	NERVS	HOPLS	RSTLS	DEPRD	EFFRT	WORTH
NERVS	0.00	0.02	0.00	-0.01	0.00	-0.01
HOPLS	0.02	0.00	-0.01	0.00	-0.01	0.00
RSTLS	0.00	-0.01	0.00	0.01	0.00	0.00
DEPRD	-0.01	0.300	0.01	0.00	0.00	0.00
EFFRT	0.00	-0.01	0.00	0.00	0.00	0.01
WORTH	-0.01	0.00	0.00	0.00	0.01	0.00

Conclusion

Through EFA, it was found that the K6 has two a factor structure consisting of Depression symptoms (worth, hopls, deprd, and effrt) and Anxiety symptoms (restls, nervs). Our two-factor solution also confirms the findings by Ko and Harrington, 2015.

4.3.2. Confirmatory Factor Analysis (CFA)

Theoretical Model

CFA was used to confirm the two-factor structure found in the EFA analysis. The 2012 survey data was for the CFA analysis since we had multiple years for the K6 survey. In general, if only 1 dataset is available and sample size is sufficient, it is recommended that half of the dataset is used to run the EFA analysis and an independent half is used to run the CFA analysis. Since the proposal research is using multiple available years, 2012 was selected at random. In addition, performing EFA and CFA on two separate years will confirm the hypothesis that the EFA structure holds from one year to another. As the missing data for each of the K6 variables is around 6,000 which is less than 8 percent and the pattern of the missingness of the data looks to be at random, no imputation was conducted. Missing data were listwise deleted, and total number of observations used in the analysis 63,838. Before the CFA model was created, a theoretical model was created based on the EFA results. This was necessary to confirm that the CFA model is exactly as the EFA model. The R- statistical package was used to run the CFA model. The R-statistical package was used throughout the entire analysis with the exception of data cleaning and EFA analysis. The theoretical model is listed below (a representation of the expectation of the CFA model to look like).

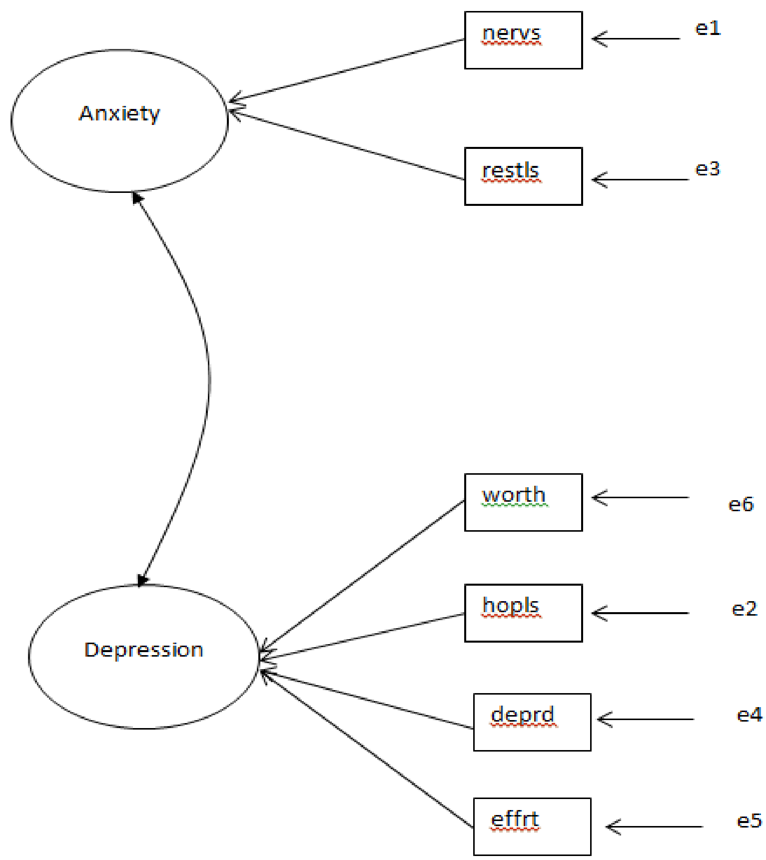

Figure 4.3. Confirmatory Factor Analysis Theoretical Model

<u>Final Model</u>

As shown in the final model below there are two factors: depression symptoms and anxiety symptoms with 4 variables driving the depression symptoms factor (worth, hopls, deprd, and, effrt), and 2 variables driving the anxiety symptoms factor (nervs and restls). This structure matches the theoretical model presented above. All the factor loadings are standardized and are statistically significant (p-value < .0001) which means that each variable is a significant contributor to the measurement of the latent factor.

Factor loading values higher than 0.7 are considered excellent (Comrey & Lee, 1992).

(Cut offs for loadings: http://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/thresholds)

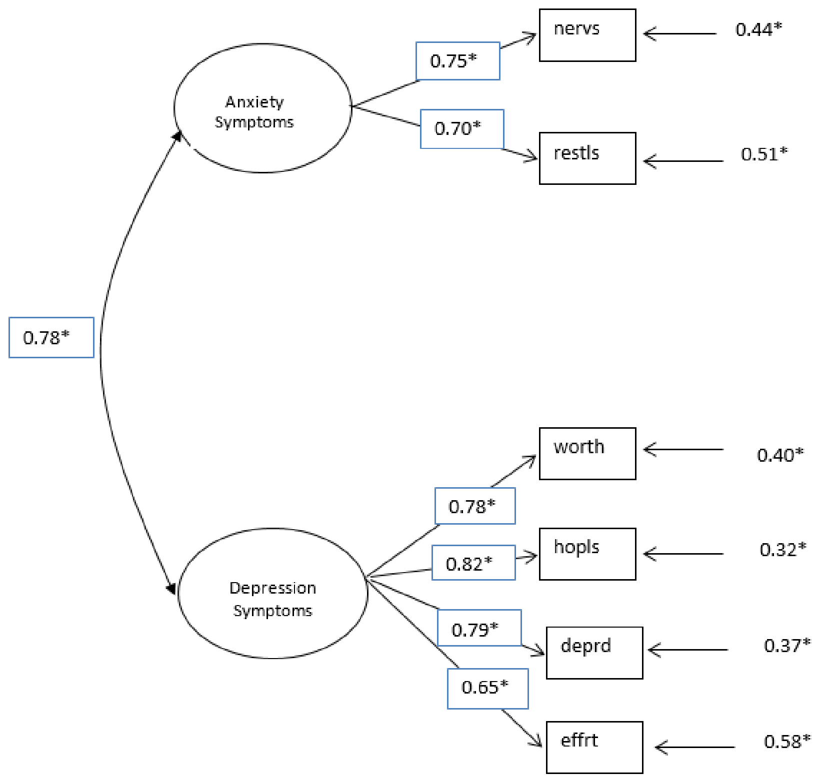

Figure 4.4. Confirmatory Factor Analysis on 2012 Survey Data
DF=8, Chi-Square = 2021.27. RMSEA=0.063, CFI=0.99, SRMR=0.025, GFI=0.99
Note: All path estimates are standardized. Error variances and covariances for stable individual differences variables were calculated but not shown in the figure.

The raw residual matrix shows really low residuals (the expected covariance matrix – the calculated covariance matrix) confirming that the model has a good model fit.

Table 4.7. CFA 2012 Raw Residual Matrix

Raw Residual Matrix						
	NERVS12	HOPLS12	RSTLS12	DEPRD12	EFFRT12	WORTH12
NERVS12	0.0000	0.0106	0.0000	-0.0130	0.0574	-0.0324
HOPLS12	0.0106	0.0000	-0.0111	0.0059	-0.0258	0.0077
RSTLS12	0.0000	-0.0111	0.0000	-0.0037	0.0782	-0.0261

DEPRD12	-0.0130	0.0059	-0.0037	0.0000	-0.0136	0.0059
EFFRT12	0.0574	-0.0258	0.0782	-0.0136	0.0000	0.0050
WORTH12	-0.0324	0.0077	-0.0261	0.0059	0.0050	0.0000

Conclusion

Via CFA we have shown that the two-factor structure found in the EFA: depression symptoms (worth, hopls, deprd, and effrt) and anxiety symptoms (nervs, restls) was confirmed.

4.3.3. Conclusions and Theoretical Implications

The exploratory factor analysis yielded two dimensions/factors: one is depression symptoms and the other is anxiety symptoms. Depression symptoms includes 4 items (worth, hopls, deprd, and effrt) and anxiety includes two items (nervs and restls). Worth, hopls, deprd load high on depression (0.78, 0.82, and 0.79 respectively) and effrt loaded medium (0.65) on depression and restls and nervs loaded medium to high on anxiety (0.70 and 0.75).

The confirmatory analysis results are showing that the model is a good fit to the data. Even though chi-square is statistically significant (meaning the models is not a good fit) it is influenced by the sample size (N=63,838, Bollen, 1990). The RMSEA is moderate (Steiger, 1989; Browne & Cudeck, 1993), with very good NNFI (Bentler & Bonnet, 1980), very good CFI (Bentler, 1990), good SRMR (Joreskog & Sorbom, 1996; Hu & Bentler, 1999), and a very good GFI (Schumaker & Lomax, 2010). Based on these results we can say that the model provides a good fit to the data, supporting the structure of the survey: two dimensions/factors one depression symptoms and the second one anxiety symptoms. This result not only confirmed the theory, but also confirmed the structure found by Ko and Harrington, 2015.

4.4. DATA MERGING AND TRANSOFRMATION STEPS

Merging the Datasets

The first step in in Specific Aims 2 was to merge the three datasets (2007, 2009, and 2012). Since the datasets are not longitudinal an individual-to-individual merge is not possible. As a result, composite observations were created in order to mimic the individual-to-individual merge. The following variables are part of the composite observations: sex (2 levels), race (2 levels), age (4 levels), income (3 levels), and education level (4 levels) for an n=192. The match up was created by grouping individuals based on the same demographic background. For example, the first group included all the white male individuals that are between age 18 to 29, with low income ($0 - $25k) and did not graduate High School. The response for the "group" was calculated averaging the individual responses within each group. The assumption made was that even though each dataset does not follow the responses for the same individual over time by using these composite observations we created an approximation of an average individual for each category. For each variable used an average score based on individuals' responses was created and these scores or composite responses were used in the analysis. Sensitivity analysis was performed and found that each composite observation had a tight distribution around the mean supporting the use of the K6 for an "average" individual of a given type.

Longitudinal Measurement Equivalence. As previously mentioned, examining measurement equivalence can be used to determine whether the survey items assess the same question meanings across time (Horn & McArdle, 1992). If the survey does not measure the same construct over time then the interpretation of changes in mean scores and correlations between the time points might be ambiguous (Horn & McArdle, 1992). The survey asks the same questions for each year and measurement equivalence confirms if the construct analyzed holds the same meaning for all the years. The assumption for analyzing measurement equivalence is that because of the transformation of the datasets from individual response into slices of population it mimics the characteristics of a longitudinal survey. Confirming invariance, also confirms the stability of construct meaning in the population across the different time points.

The first step in computing measurement equivalence was to examine the distribution of the data via the skewness and kurtosis. Kline, 2011 recommended a value of 3.0 and above for "severe" skewness and values above 10.0 suggesting a problem with kurtosis. Looking at the results of our data no severe skewness or serious problem with kurtosis was found with one exception for worth09 (skewness value of 3.48 and kurtosis value of 21.63). As mentioned in the EFA analysis in Specific Aim 1, because only one of the variables is non-normal the same MLM estimation (maximum likelihood estimation) with robust standard errors and a Satorra-Bentler scaled statistic (Satorra, & Bentler, 2010) was used in this analysis.

The second step in the analysis was to specify the model, estimate parameters and impose equality constraints across time periods. The model tested is a CFA model with the K6 items for each year.

The following models were run: configural invariance model (no constraints or baseline model), metric invariance or equal loadings invariance model, scalar invariance or equal loadings and intercepts model, and means model or equal loadings, intercepts, and means model (Cheung & Rensvold, 2002; Little, 1988).

Criteria to compare the models:

Initial studies recommended the use of chi-square tests to decide if the increase in fit is substantial (Byrne et al., 1989). Because the traditional check of global model fit is the chi-square test (Cochran, 1952) that is dependent on the sample size (it rejects reasonable models if the samples size is large), Cheung & Rensvold, 2002 suggested using a difference in CFI to examine the invariance (a difference in CFI larger than 0.1 indicates that the invariance should not be rejected). We used the difference in CFI method since it is most widely used and best supported method for invariance testing (Chen et al., 2001; Cheung & Rensvold, 2002; Meade et al., 2008).

Testing for longitudinal measurement equivalence steps:

Configural invariance examines if the underlying factor structure is the equivalent across groups (if the number and type of factors are the same across groups). In this analysis gender was used as the group and if configural invariance holds then there is no difference in the way that males and females answered the K6 questionnaire. The metric invariance constrained the configural model by examining the equality between items and factors across gender. For the weak invariance or loadings test invariance the model comparisons indicate that the factor loadings are assumed to be equal between males and females. The results show that this model has a moderate fit to the data (chi-square of 1106, df= 15, CFI = 0.860, RMSEA = 0.092, SRMR = .063).

Metric invariance is testing that all factor loadings are equal across groups. The difference between CFI, TLI, and RMSEA for the baseline and metric models is not significant meaning the loadings are invariant across groups. The results show that this model has a moderate fit of the data, since the chi-square test is not significant and the difference in CFI is 0.001, which is smaller than the cutoff point of 0.01 (Cheung & Rensvold, 2002). The metric invariance was thought to hold.

Scalar invariance is testing that all factor loading parameters and intercepts are equal. When constraining the intercepts to be equal for both males and females the difference in CFI is 0.019 which is higher than the cutoff point of 0.01 (Cheung & Rensvold, 2002) indicating that the scalar invariance cannot be met. The next step in the analysis was to find out why the scalar invariance did not hold even though the initial model showed both the loading parameters and the intercepts being equal. As a result, after running the modification indices test, by constraining the correlation between restles07 and restles09 the model was improved and the difference between the CFI between the metric and scalar models was brought down to 0.005 which is below the Cheung & Rensvold, 2002 cutoff point.

The results of the configural, metric and scalar models are listed in table 4.8 below.

Table 4.8. Measurement Invariance

	DF	Chi-square	Diff Chi-square	Diff DF	p-value	CFI	RMSEA	Diff CFI	Diff RMSEA
Configural	264	1106				0.860	0.092		
Metric	279	1133	14.12	15	0.5167	0.859	0.090	0.001	0.002
Scalar	294	1172	39.10	15		0.840	0.093	0.019	0.003
Significant codes: *** p=<.0001									

Based on these results we can conclude there is no difference between males and females regarding the measure of the K6 survey for all the years. These results allow the analysis to move forward, as invariance was confirmed. Thus, the data may be considered as a longitudinal representation of mental health perspectives in the population.

4.5. SPECIFIC AIM 2: EXAMINE THE TREND OF PSYCHOLOGICAL DISTRESS BEFORE, DURING, AND AFTER THE 2007 RECESSION

In Specific Aim 2, the goal was to model the item responses to the K6 questionnaire via the latent growth modeling (LGM). However, LGM yielded a negative variance for the slope. Negative variances are specific to SEM and are known as "Haywood Cases" and can be attributed to specification error (not setting one of the loading factors to 1, low factor loadings, paths omitted), outliers, non-convergence, or model misspecification (Kline, 2011). First, all the errors possible was examined: one of the loading factors was set to 1, the factor loadings were not low, all the paths were accounted for, no outliers, and the model was not misspecified. The model was not converging. Second, to try to correct for the negative variance, we increased the number of iterations, but still the model did not converge. The third approach was to set the slope variance to a small fixed numeric variable (Chen et al., 2001) and the model was re-estimated. However, once the variance of the slope was constrained, the variance for the psychological distress for year 2007 became negative. These results suggested that LGM might not be appropriate for the data. As a result, a linear panel model (with MLM as the estimator) was used to model the data. Panel models measure changes that occur between two or more points in time and it provides paths between the latent variables (Hannan, M.T., and Young, A.A., 2018). The paths between the variables represent stability over time. Compared to LGM, panel models measure relationships between two successive time points (i.e. relationship between PD07 to PD09 and the relationship between PD09 to PD12, where LGM measures overall change of constructs over time as a growth/model).

The MLM (robust maximum likelihood estimation) with robust standard errors and a Satorra-Bentler scaled statistic (Satorra, & Bentler, 2010) technique was used for the model in this specific Aim 2 as well as the other specific aims (3 through 5). Using the robust estimation method to assess the fit of the model yielded a Satorra-Bentler scaled Chi-square value of 247.134, with 133 degrees of freedom, and $p<0.001$. The fit indices for the model are good with: CFI = 0.924, TLI = 0.912, RMSEA = 0.067 (95% CI RMSEA [0.059, 0.075], SRMR = 0.059. These values indicate a good fit between the model and the observed data.

The stability coefficient is defined as the correlation of measurement results from Time 1 with measurement results from Time 2, where the subjects being measured and the measuring instrument remain precisely the same (Lewis-Beck et al., 2004). The coefficient of stability ranges from 0 to 1, with higher values signifying stronger stability (0.7 and above), and lower values signifying instability across the two time periods (0.6 to 0.7 being considered questionable stability and below 0.6 being considered poor stability). The stability coefficients between PD07 to PD09 equals to 0.977 and PD09 to PD12 equals to 0.741 and are statistically significant (p-value ≤ 0.05), with standardized estimates at 0.874 and 0.795, respectively. The higher the stability coefficient (closer to 1 is better) the more stable the relationship between PD07 and PD09, meaning that PD in 2009 stayed at the same level with PD in 2007. On the other hand, the coefficient of stability between years 2009 and 2012 is much lower, at 0.741, which means that the level of PD is not the same across years. The change in stability from 2009 to 2012 is expected since the recession ended mid-2009. Figure 4.5 below shows the panel model, as 3 CFA models with the stability coefficients from PD07 to PD09 and PD09 to PD12.

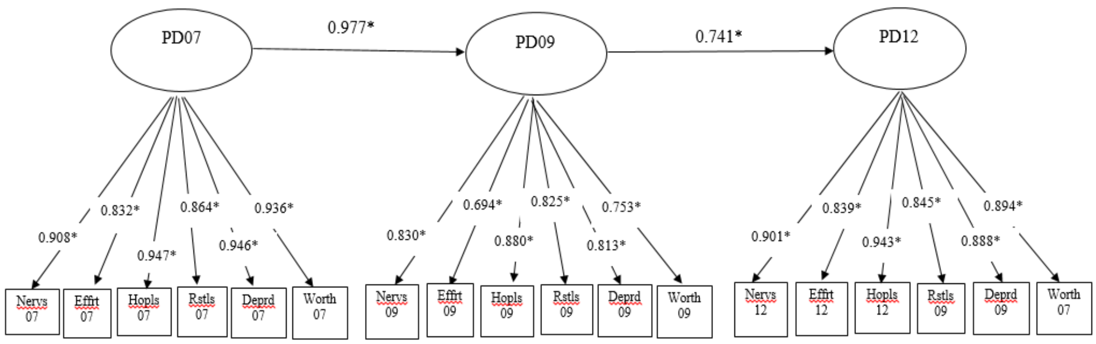

Figure 4.5. Panel model, no covariates
DF=133, Chi-Square = 247.134. RMSEA=0.067, CFI=0.924, SRMR=0.059, TLI=0.912

The path model's coefficients of stability showed that the level of PD between 2007 and 2009 stayed the same and the level of PD between 2009 and 2012 has changed, but it does not show the direction or magnitude of the change. As a result, sum scores by PD severity level were created for each year in order to see where the change in PD might have happened. Table 4.9 below shows the change in percentage of the average individual by severity level by year. The three severity levels (low, moderate, and severe) were calculated by summing the individual K6 responses (on a scale from 1 to 5) and split into categories based on the previous research completed by Kessler et al., 2002 and Prochaska et al., 2012. Low PD represents a total score of 6-10, moderate PD represents a total score of 11 to 18, and severe PD represents a total score of 19 to 30.

Table 4.9. Percentage of Composite Observations Within Each Severity Level by Year

	Year		
	2007	2009	2012
Low	69%	67%	73%
Moderate	26%	28%	22%
Severe	5%	5%	5%

From the table, the percentage of the population with low PD decreased, and moderate PD increased from 2007 to 2009 by 2%. This result was on par with the hypothesis where individuals will experience a higher level of PD from 2007 to 2009 and a lower level of PD from 2009 to 2012.

4.6. SPECIFIC AIM 3: EXAMINE THE TREND OF PSYCHOLOGICAL DISTRESS BEFORE, DURING, AND AFTER THE 2007 RECESSION BY INCOME AND HEALTH INSURANCE STATUS

In specific Aim 3, we added income and health insurance status to the previous model. Health insurance status variable is represented in the model by the percentage of uninsured and it is continuous, and the income variable is categorical and ordered (1 = 0-25K, 2=25-50K, 3=50K+). The model was constructed as follows: panel model from specific Aim 2 was used, with the addition of the covariates uninsured and income. The income and uninsured variables were not split by year, but were considered as one variable since we examine the impact of the two covariates on PD. By adding covariates to the model, we want to assess the impact of income and/or level of uninsured on psychological distress over the three time periods. We expected income to have a negative effect on the level of PD, meaning that with increasing level of income we expect the level of PD to decrease; and we expected a positive effect of uninsured on PD, meaning that with 1 percent increase in uninsured level we expect PD to increase by 1 percent. Since we expect income to have a negative effect on PD and level of uninsured to have a positive effect on PD, multiple models were run to assess the effect of each variable on PD separately, as well model which included the effects of both variables on PD.

<u>The first model</u> tested the effect of income on the stability of PD07. The fit of the robust model is good with the following fit indices: CFI = 0.926, TLI = 0.915, RMSEA = 0.065, 90% CI RMSEA = [0.057, 0.073], and SRMR = 0.057.

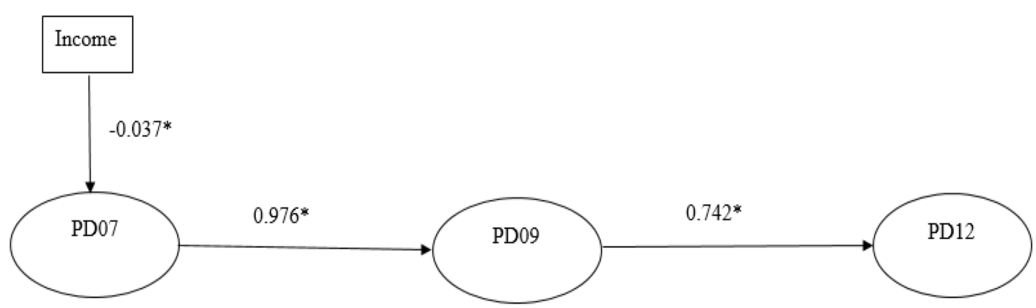

Figure 4.6. Panel Model, Income as Covariate
DF=150, Chi-Square = 272.591, RMSEA=0.065, CFI=0.926, SRMR=0.057, TLI=0.915

The effect of income on PD07 is statistically significant. An increase in income by one unit (i.e. from low income to middle income or middle income to high income) decreases PD07 by 3.7%. In addition, the coefficient of stability between 2007 and 2009 shows that the relationship between the two time periods is stable with an unstable relationship between 2009 and 2012. The introduction of income did not change the relationships between the PD years. Separate panel models showing the effect of income on PD09 and income on PD12 were run with similar results. Here, we look at the mean for income by year to see if the distribution of income changed for 2009 and/or 2012 and it did not, the mean for all three years for income was between $25K – $50K.

The second model looked at the effect of level of uninsured on PD and the stability of PD over time. The fit of the robust model is good with the following fit indices: CFI = 0.901, TLI = 0.888, RMSEA = 0.076, 90% CI RMSEA = [0.069, 0.084], and SRMR = 0.065. Note that uninsured is an average percentage of uninsured for all three years. Here, we look at the mean of percentage of uninsured for all three years to see if the distribution changed over time and it changed by one percentage point from 2007 to 2009 when the mean percentage of uninsured was 19%, with the mean percentage for 2009 and 2012 at 18%.

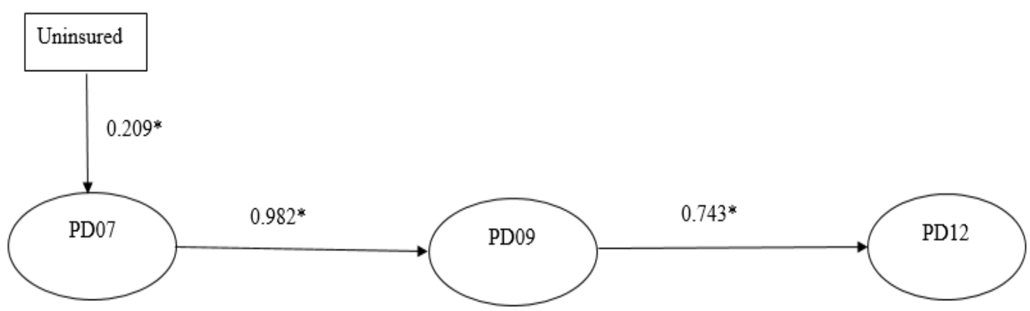

Figure 4.7. Panel Model, Uninsured as Covariate
DF=150, Chi-Square = 316.995, RMSEA=0.076, CFI=0.901, SRMR=0.065, TLI=0.888

The effect of the percentage of uninsured on PD is statistically significant and it has a positive increase in PD. One percent in increase in the uninsured increased PD07 by almost 21%. The coefficient of stability from PD07 to PD09 is close to 1 showing a stable relationship (PD level stayed the same from 2007 to 2009), where the relationship between PD09 and PD12 is not stable. Alternative models examined the effect of the uninsured on PD09 and PD12. The effect of the uninsured on PD09 changed the relationship between PD07 and PD09 from a high/strong coefficient of stability of 0.982 to a lower/weaker coefficient of stability of 0.780. In addition, the effect of uninsured on PD12 on the coefficient of stability between PD09 and PD12 was lower at 0.635 as compared to 0.743 (effect on PD07) and 0.753 (effect on PD09). Even though the level of uninsured on PD07 (20.9%) has a much larger impact on psychological distress as compared to PD09 (8.0%) and PD12 (4.7%), it did not change the stability between the years, where in the other years the effect of uninsured did change it. The change in the stability between the years could be caused by the change in the effect of the level of uninsured on PD over time from 20.9% in 2007 to 8% in 2009 to 4.7% in 2012.

The third and final model looked at the effect of both variables, income and uninsured on the stability of PD. The fit of the robust model is good with the following fit indices: CFI = 0.903, TLI = 0.890, RMSEA = 0.075, 90% CI RMSEA = [0.067, 0.082], and SRMR = 0.063. Income is statistically significant and has a negative effect on PD07 by 2.3%. Uninsured is statistically significant and has a positive effect on PD07 of 14.8%. The level of uninsured has a much larger impact on psychological distress as compared to income. Research found that the recession causes lower per capital income with individuals moving from private insurance to public insurance (Herring & Trish, 2015; Holohan & McMorrow, 2013). The results in Aim 3 show that the mean income did not change over time as previous research suggested.

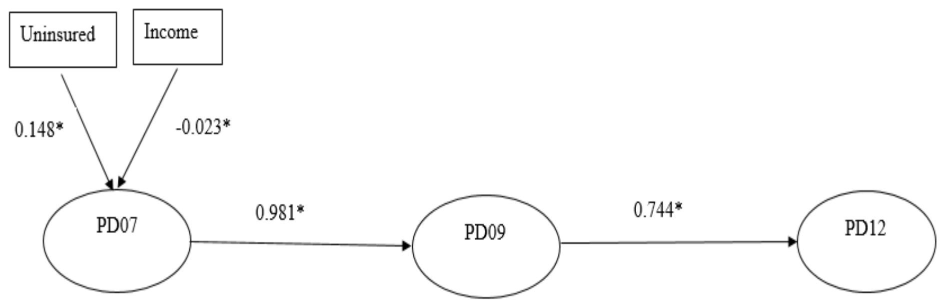

Figure 4.8. Panel Model, Income and Uninsured as Covariates
DF=167, Chi-Square = 345.008, RMSEA=0.075, CFI=0.903, SRMR=0.063, TLI=0.890

Adding the two covariates in the model did not have a significant impact on the coefficients of stability and did not change the relationships among PD07 over time. Alternative models examined the effect of the covariates on PD09 and PD12. The effect of the covariates on PD09 changed the relationship between PD07 and PD09 from a high/strong coefficient of stability of 0.981 to a lower/weaker coefficient of stability of 0.756. In addition, the coefficient of stability between PD09 and PD12 was lower for the effect of the covariates on PD12 at 0.608 as compared to 0.744 (effect on PD07) and 0.753 (effect on PD09). These results are consistent with the results found with the uninsured as the sole covariate to the panel model, showing that uninsured status has an effect on PD over time.

4.7. SPECIFIC AIM 4: INVESTIGATE THE ASSOCIATION BETWEEN MENTAL HEALTHCARE UTILIZATION AND PSYCHOLOGICAL DISTRESS AND THE MEDIATION INFERENCE OF HEALTH INSURANCE ON MENTAL HEALTHCARE UTILIZATION

The goal in specific Aim 4 was to investigate the association between mental health care utilization represented by the percentage of individuals not receiving mental health treatment (notreatmentMH) on psychological distress with the percentage of uninsured as a mediator. A variable is a mediator when there is a causal relationship between the variables. In this case the presence or absence of health insurance causes the level of health care utilization. The mediator is assumed to cause the outcome (MacKinnon, Fairchild, and Fritz, 2007). We assumed that the presence or absence of health insurance is associated with mental healthcare utilization which in turns it impacts the level of psychological distress.

The fit of the mediation model is good with the following fit indices: CFI = 0.877, TLI = 0.860, RMSEA = 0.083, 90% CI RMSEA = [0.076, 0.090], and SRMR = 0.070.

The direct effect of uninsured on psychological distress is 17.4% and is statistically significant. One percent increase in uninsured population increases the psychological distress by 17.4%. The indirect effect of uninsured on PD is 16.0% (the effect of uninsured on notreatmentMH times the effect of notreatmentMH on PD07). The percentage of individuals not receiving treatment has a small impact on the relationship between uninsured and PD07. The difference is 1.4%.

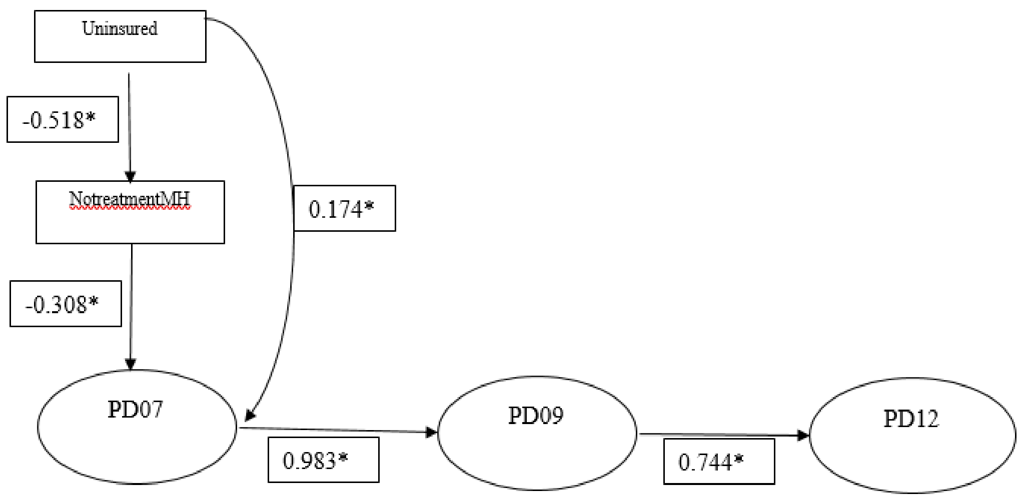

Figure 4.9. Mediation Panel Model
DF=167, Chi-Square = 389.449, RMSEA=0.083, CFI=0.877, SRMR=0.070, TLI=0.860

The total effect of uninsured on PD, direct and indirect, at 33.3% is statistically significant. The introduction of the covariates does not change the direction of the relationship of PD over time compared with the panel model with no covariates.

Alternate models were run to examine the effect of the mediation on the relationship between PD over time. The first alternative model looked at mediation effect on PD09. Results show that the mediation has changed the relationship between PD07 to PD09 with a coefficient of stability of 0.610. The indirect effect of the uninsured on PD09 is 7.4% with a total effect (direct plus indirect) of 17.4%. The average percentage of uninsured for 2007 was 1 percent higher than 2009 and 2012 at 19 percent which could be the cause for the change in the relationship between PD07 and PD09. The second model run was to see the effect of the mediation on PD12. The indirect effect of the percentage of uninsured on PD12 was 11.2%, with a total effect (direct plus indirect) of 19.2%. Mediation changed the relationship between PD09 and PD12 with a coefficient of stability of 0.389.

The changes in the relationships could be caused by the increase of one percent in the average percentage of individuals that seek mental health treatment from 2009 to 2012 (from 12 percent to 13 percent). The uninsured rate decreased by one percent from 2007 to 2009 (from 19 percent to 18 percent) and stayed constant from 2009 to 2012.

4.8. SPECIFIC AIM 5: INVESTIGATE THE EFFECT OF ALCOHOL AND/OR TOBACCO CONSUMPTION ON PSYCHOLOGICAL DISTRESS AND THE EFFECT OF PSYCHOLOGICAL DISTRESS ON ALCOHOL AND/OR TOBACCO CONSUMPTION

The goal of Specific Aim 5 was to investigate the association between psychological distress and tobacco use and alcohol consumption over time. Since we do not know the directionality of the association, the panel model was used to look at the effect of smoking and alcohol consumption on psychological distress, and linear regressions were used to look at the effect of psychological distress on smoking and alcohol consumption.

<u>Definitions</u>

Smoking: percentage of smokers

Alcohol Consumption: percentage of heavy and binge alcohol consumption

<u>The first panel model</u> looked at the effect of alcohol consumption (drinking) on psychological distress. The fit of the model is good with the following fit indices: CFI =

0.907, TLI = 0.893, RMSEA = 0.073, 90% CI RMSEA = [0.066, 0.081], and SRMR = 0.065.

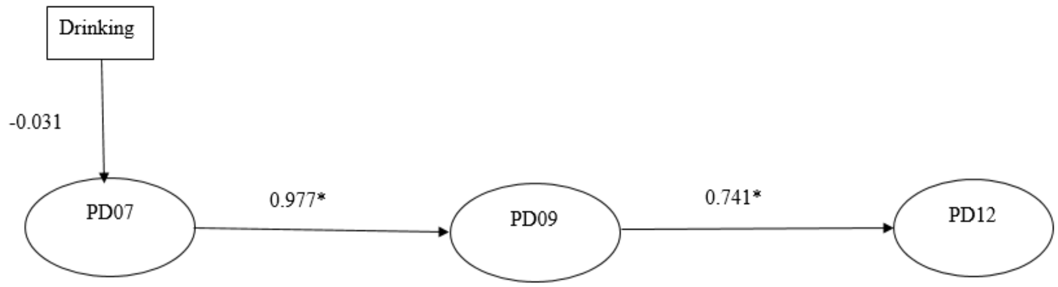

Figure 4.10. Panel Model, Alcohol Consumption as a Covariate
DF=150, Chi-Square = 304.560, RMSEA=0.073, CFI=0.907, SRMR=0.065, TLI=0.893

The heavy and binge alcohol consumption has no effect on PD07, the coefficient is not statistically significant, p-value=0.158.

Alternative panel models were run to see the effect of drinking on PD09. Heavy and binge alcohol consumption has no effect on PD09, and the coefficients of stability are similar to the first panel model (0.973 and 0.740, respectively). Similar results can be seen when running the panel model with the effect on alcohol consumption and PD12.

<u>The second panel model</u> looked at the effect of smoking on psychological distress. The fit of the model is good with the following fit indices: CFI = 0.908, TLI = 0.895, RMSEA = 0.073, 90% CI RMSEA = [0.066, 0.081], and SRMR = 0.068.

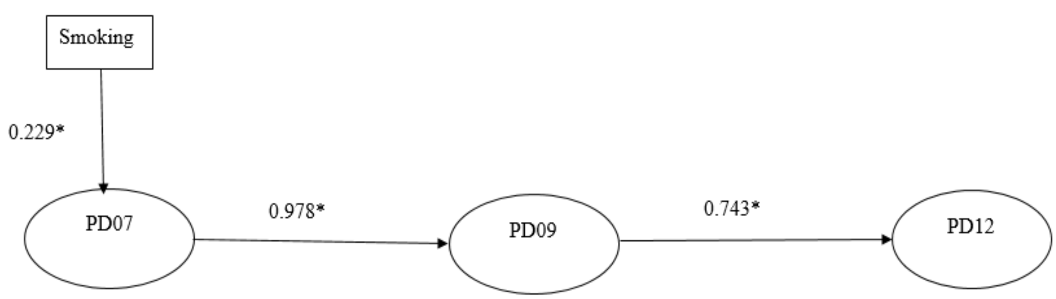

Figure 4.11. Panel Model, Smoking as a Covariate

DF=150, Chi-Square = 304.424, RMSEA=0.073, CFI=0.908, SRMR=0.068, TLI=0.895

The effect of smoking on PD07 is positive and statistically significant. With 1 percent increase in the smoking status, there is an increase in PD 2007 status by 22.8%. Even though the effect of smoking on PD is significant it does not change the relationship between the PD over time. Alternative panel models were run to see the effect of smoking on PD09, and while the effect of smoking on PD09 was lower at 5.6%, it did not change the relationship between PD over time. The effect of smoking on PD12 is significant at 11.4% and it did change the relationship between PD09 and PD12. The coefficient of stability is lower at 0.543. The average smoking rate for each year is 0.24 for 2007, 0.23 for 2009, and 0.21 for 2012. The decrease in the percent of individuals smoking had an influence on the relationship between PD09 and PD12.

The third and final panel model looked at the effect of both variables, drinking and smoking on PD. The fit of the model is good with the following fit indices: CFI = 0.896, TLI = 0.881, RMSEA = 0.076, 90% CI RMSEA = [0.068, 0.083], and SRMR = 0.072.

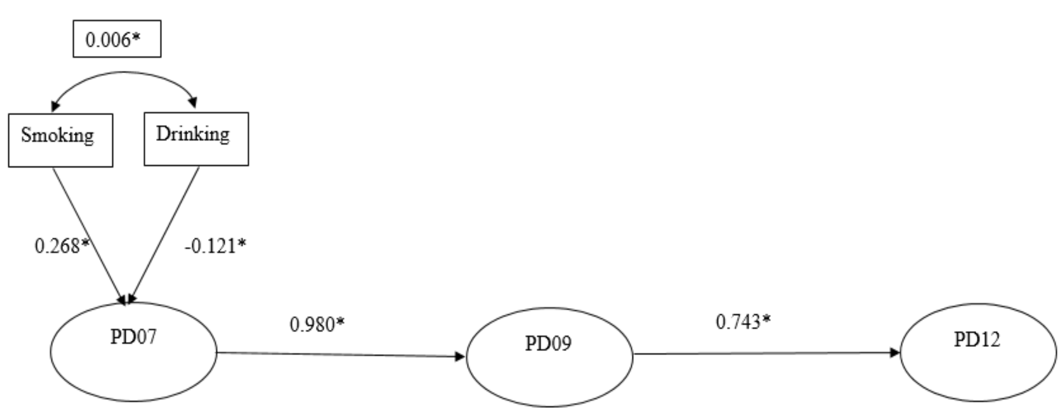

Figure 4.12. Panel Model, Alcohol Consumption and Smoking as Covariates
DF=167, Chi-Square = 350.218, RMSEA=0.076, CFI=0.896, SRMR=0.072, TLI=0.881

When smoking was added to the panel model alcohol consumption's effect on PD07 is statistically significant, with a positive effect of 26.8% for smoking and a negative effect of 12.1% for alcohol consumption (drinking) on PD07. Even though smoking and drinking had a significant effect on PD the coefficients of stability did not change over time as compared to the no covariates panel model. Drinking and smoking's effect on PD07 did not change the magnitude or direction of the relationship between PD over time. Alternate panel models showed the effect of the two covariates on PD09: smoking had a significant positive effect on PD09 at 9.1% and alcohol consumption had a negative significant effect on PD09 at 5.2%. The two covariates together changed the relationship between PD07 and PD09 at 0.789. The effect of smoking on PD12 was significant and positive at 15.7%, with the effect of alcohol consumption on PD12 being significant and decreased PD12 by 6.5%. The effect of the two covariates has changed the relationship between PD09 and PD12 as compared with the initial panel model, but as seen in the panel model with smoking only, the change in the relationship was caused by the decrease in the percentage of smoking and not alcohol consumption. Note that the correlation between smoking and alcohol consumption is statistically significant, but very small.

The <u>second part of Specific Aim 5</u> looked at the effect of psychological distress on smoking and the third part of Specific Aim 5 looked at the effect of psychological distress on alcohol consumption. Both sets of results were achieved by running linear regressions.

Multiple regressions were run to assess the <u>effect of PD on alcohol consumption</u>.

a) Base model with alcohol consumption as a dependent variable and PD by year as independent variables.

Model: Drinking = intercept + PD07 + PD09 + PD12 + error

Results: Drinking = 0.17* + 0.015*PD07 + 0.002*PD09 – 0.013*PD12

None of the PDs are statistically significant, and the model R-square is very low at 0.008.

Additional analyses were performed to examine the effect of low, moderate, or high PD on drinking.

Drinking = 0.25* - 0.10*lowPD07 - 0.11*lowPD09 + 0.14*lowPD12

Drinking = 0.13* + 0.20*modPD07 + 0.19*modPD09* – 0.12*modPD12

Drinking = 0.23* + 0.24*highPD07 - 0.30*highPD09 – 0.32*highPD12

The only PD that is statistically significant is the moderate PD09 at $p=0.1$ with all the other PDs being not statistically significant. Even though we cannot draw any significant conclusion based on the results, we can see that low PD has a negative impact on alcohol consumption in years 2007 and 2009 with a positive impact in 2012. On the other hand, moderate PD has a positive impact in 2007 and 2009 with a negative impact in 2012 and high PD has a positive impact on alcohol consumption in 2007 and a negative impact in 2009 and 2012.

b) Add smoking to the base model

Separate models were run one with smoking as a general variable and one with smoking by year.

Model: Drinking = intercept + PD07 + PD09 + PD12 + smoking + error

Results: Drinking = 0.49* + 0.004*PD07 – 0.011*PD09 – 0.04*PD12* + 0.64*smoking*

* Statistically significant $p<0.05$.

With the inclusion of smoking in the model, PD12 became statistically significant with a negative association. With one point increase in psychological distress it decreased alcohol consumption by 4 percent in 2012. Smoking as positively associated with alcohol consumption (statistically significant at $p<0.001$). One percent increase in smoking yielded about 64 percent increase in alcohol consumption. This result is consistent with current literature (De Leon et al., 2007; Beard et al., 2017). R-squared for the model is 0.21 which is higher than the based model, but still low.

Additional models were run to see the effect of the severity of PD on alcohol consumption. The results are listed below:

Drinking = -0.17 + 0.02*lowPD07 +0.02*lowPD09 + 0.32*lowPD12* + 0.55*smoking*

Drinking = 0.14* + 0.09*modPD07 + 0.08*modPD09 – 0.23*modPD12 + 0.34*smoking*

Drinking = 0.13* - 0.79*highPD07*- 0.19*highPD09 – 0.97*highPD12* + 0.79*smoking*

Examining the association of the severity of PD on alcohol consumption we can say that high or serious PD has a significant negative effect on alcohol consumption in 2007 and 2012 with a not significant negative effect in year 2009. These results are not what we hypothesized. We expected with increased PD, alcohol consumption would increase. The negative association between severe PD and alcohol consumption indicates that individuals with high PD do not resort to alcohol to deal with the increased PD.

c) Adding covariates to the model (demographics and the rest of the covariates)

By adding covariates to the model, it did increase the goodness of fit to an acceptable fit of model r-square 0.8698 and 0.8943 for the demographics only model and for the full model, respectively. For both models, psychological distress is not statistically significant and hence was not associated with alcohol consumption pre, during, or after recession.

Multiple regressions were run assess the effect of PD on smoking.

a) Base model with smoking as a dependent variable and PD by year as independent variables.

Model: Smoking = intercept + PD07 + PD09 + PD12

Results: Smoking = -0.49* + 0.017*PD07 + 0.02*PD09* + 0.04*PD12*

PD07 is not statistically significant, but all other coefficients are statistically significant, $p < 0.001$. The residual standard error is small at 0.09, and R-squared is 0.53. Psychological distress for 2009 and 2012 has a positive effect on smoking, with a higher effect on 2007 compared to 2012. Individuals smoked at a higher rate at the beginning of the recession (2007) compared to no recession (2012). R-squared is low, the model does not fit the data very well.

Additional analysis was performed to assess the effect of the PD severity on smoking. The results are listed below:

Smoking = 0.76* - 0.21*lowPD07* - 0.23*lowPD09* - 0.32*lowPD12*

Smoking = -0.02 + 0.33*modPD07* + 0.32*modPD09* + 0.33*modPD12*

Smoking = 0.12* + 1.30*highPD07* - 0.13*highPD09 + 0.81*highPD12*

All PDs are statistically significant at p=<0.05 with the exception of highPD09 which is not statistically significant. Based on the results, low PD decreases the level of smoking, while moderate and high PD increases the level of smoking. These results confirm our hypothesis that increased PD increases the level of smoking.

b) Inclusion of alcohol consumption.

In the panel model we showed that both smoking and drinking influence psychological distress. The model below investigates if drinking and PD and the severity of PD influence smoking.

Model: Smoking = intercept + PD07 + PD09 + PD12 + drinking

Results: Smoking = -0.54* + 0.01*PD07 – 0.02*PD09* + 0.04*PD12 + 0.31*drinking

Smoking = 0.69* - 0.18*lowPD07* – 0.20*lowPD09* - 0.36*lowPD12 + 0.29*drinking

Smoking = -0.05* + 0.29*modPD07* + 0.27*modPD09* + 0.35*modPD12* + 0.22*drinking*

Smoking = 0.02* + 1.20*highPD07* + 0.00*highPD09 + 0.95*highPD12* + 0.43*drinking*

*PD07 and highPD09 are not statistically significant, but all other coefficients are statistically significant, p = <0.05 with the exception of lowPD07 and high intercept which are statistically significant at 0.1.

The residual standard error is small at 0.080, and R-squared is 0.66. When alcohol was included in the model the coefficients for the PDs saw a small decrease except for all the severity levels of PD12 which saw a small increase.

c) Inclusion of demographic variables

Model: Smoking = intercept + PD07 + PD09 + PD12 + drinking + sex + race + income + age + education + error

Results: Smoking = -0.39* + 0.02*PD07* + 0.01*PD09* + 0.04*PD12*

+0.26*drinking* -0.03*sex* + 0.02*race* + 0.02*income* + 0.01*age – 0.03*education*

Smoking = 0.90* - 0.30*lowPD07* - 0.17*lowPD09* - 0.35*lowPD12* +0.26*drinking*

-0.03*sex* - 0.07*race* + 0.02*income* + 0.01*age – 0.03*education*

Smoking = 0.32* + 0.20*modPD07 + 0.19*modPD09* + 0.31*modPD12*

+0.18*drinking - 0.03*sex - 0.08*race* - 0.01*income - 0.00*age – 0.05*education*

Smoking = 0.31* + 1.19*highPD07* - 0.18*highPD09 + 0.69*highPD12*

+0.27*drinking* -0.02*sex + 0.05*race* + 0.00*income* - 0.01*age – 0.04*education*

Inclusion of demographic and socioeconomic variables had no effect on the association between psychological distress and smoking or the association between smoking and alcohol consumption before, during, or after recession. Income and education are statistically significant across all the models. Sex is statistically significant for all the models except moderate PD, and race is statistically significant for all the models except for high PD, and age is not statistically significant across all the models. The impact of each demographic on smoking is minimal across the board. *All variables statistically significant at $p<0.05$. Model R-square = 0.7687, residual standard error=0.6687.

d) Full model (model in part c plus the rest of the covariates)

Model: Smoking = intercept + PD07 + PD09 + PD12 + drinking07 + drinking09 + drinking12 + sex + race + income + age + education + uninsured+ unemployed07 + unemployed09 + unemployed12+ chronic07+ chronic09 + chronic12 + generalhealth07 + generalhealth09 + generalhealth12 + daysmiss07+ daysmiss09 + daysmiss09 + notreatmentMH + generalhealth07+ generalhealth09+ generalhealth12 + stigmatreat07+

stigmatreat09 + stigmatreat12 + stigmacaring07 + stigmacaring09 + stigmacaring12 + error

When including the rest of the covariates in the model, PD is not statistically significant for all the years, except lowPD in 2009 with a positive effect on smoking of 30% and a negative effect on smoking of 36% for moderatePD the same year.

The results in Specific Aim 5 found that heavy and binge alcohol consumption did not influence PD and it does not change the relationship between PD over time and conversely PD does not have an effect on heavy and binge drinking over time. For most of the models run, PD is not statistically significant meaning that there no relationship between the severity of PD and alcohol consumption.

On the other hand, smoking has a positive effect on PD and did not change the relationship between the PDs (recession did not change the relationship between the variables) over time and PD has a positive effect on smoking. PD is statistically significant and shows a negative effect on smoking for no or low PD. The effect of moderate and high PD on smoking is positive with a lower positive effect for moderate PD as compared to high PD. These results were as we expected, with higher severity of PD associated with increased level of smoking.

As part of sensitivity analysis, additional models were run in order to examine if all the variables of interest (variables from Aims 3-5) would impact the coefficients of stability. Results show that adding all the variables into the model did not change the conclusions drawn in Aims 3 through 5.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1. CONCLUSIONS

The overarching research question was to investigate the role of tobacco use (smoking), alcohol consumption, and insurance status on psychological distress in non-institutional adult population before, during, and after the 2007-2009 U.S. economic recession. In order to answer this research question, we looked at the effects of income, health insurance coverage (represented by percentage of uninsured), and severity of smoking and alcohol consumption on psychological distress pre-, during and post-recession. The relationship between PD07 and PD09 remained stable, while the relationship between PD09 and PD12 has changed. That change could be cause by the 6 percent increase in low PD and 6 percent decrease in moderate PD. Income had a small effect on psychological distress pre-, during, and post- recession with no disruption in the association between PD over time. On the other hand, increased level of uninsured had a significant positive effect on psychological distress on all the panel models including the mediation model. Ward and Martinez (2015) found that adults with public/other health coverage had the highest levels of psychological distress, followed by the uninsured and then private coverage. Alcohol consumption had a small negative effect on PD with no disruption in the association between the PD over time. Smoking had a larger positive effect on PD with a decrease in the percentage of individuals smoking from 2009 to 2012 which could have played a part in the disruption of the relationship between the years. In addition, the level of smoking increased with the level of severity of PD, especially for the moderate and severe PD showing a two-way association between smoking and PD, especially during hard economic times.

In addition, we have shown that the K6 survey instrument can show trends in PD over time at the population level and not individual level. This was achieved via IDA analysis. The first step in the analysis was to merge the three datasets (2007, 2009, and 2012). Since the datasets were cross-sectional and individual-to-individual matching was not possible, composite observations were created to mimic the individual-to-individual longitudinal matching. The composite observations are not at individual level, but a grouping of individuals based on the same demographic background. Each individual from each dataset was distributed into one of the unique combinations comprised of: sex (2 levels), race (2 levels), age (4 levels), income (3 levels), and education level (4 levels). The total number of composite observations used in the analysis was 192. The response for each composite observation (group) was calculating by averaging the individual responses within each group. Based on this data transformation the results of the analysis was at the population level and not at individual level. This data transformation is unique to this research and it was created in order to be able to analyze multiple cross-sectional datasets as a longitudinal dataset. As far as we know, we are the first ones to attempt such transformation and if confirmed by additional research it can revolutionize the world of Public Health research, giving the community a new tool to analyze trends in large cross-sectional datasets at the population level.

Exploratory Factor Analysis and the Confirmatory Factor Analysis results suggested that the psychological distress questionnaire has two factors/dimensions: anxiety symptoms and depression symptoms and the two-factor structure held for all the years: 2007, 2009, and 2012 meaning that the questions were answered in the same manner over time by different individuals (i.e., held the same meaning over time). These results are consistent and confirm the results found by Ko and Harrington, 2015. This first step in the data analysis paved the way for conducting a trend analysis using cross-sectional data. The second step in the analysis was to create composite observations for each dataset. These two steps along with analysis of variability allowed us to perform trend analysis on PD on cross-sectional datasets and is adding to the current knowledge on IDA methodology. IDA was used on merging different longitudinal datasets and, as far as we know, this research is the first to attempt merging data from national cross-sectional datasets.

The results in Specific Aim 2 indicate that the economic recession did not have an influence on the level of PD from 2007 to 2009 but did have an influence on the level of PD from 2009 to 2012 in the non-institutionalized adult population. This result is different from what we expected. We expected PD to increase from 2007 to 2009 and decrease from 2009 to 2012, but the coefficient of stability between 2007 and 2009 showed no change in PD across the two years. This result could be caused by multiple factors such as: (a) timeframe of the recession: it began in the middle of 2007 and ended in the middle of 2009 which could have diminished psychological distress for those years (US Labor Statistics, 2016); (b) the Mental Illness and Stigma module questions asks the individual about anxiety and depression symptoms in the past 30 days which can skew the results since the telephone interviews were conducted during each calendar month (CDC, 2014); (c) economic factors such as employment status, income status, housing status, economic outlook, etc.; (d) status of individual or family member's physical or mental health; (e) major changes in life such as graduating from college, marriage, divorce, etc. The change in coefficient of stability from 2009 to 2012 was expected, with economy improving and unemployment rate steadily decreasing. The unemployment rate peaked in 2010 at 9.6 percent and then steadily decreased to 8.1 percent in 2012 (U.S. Bureau of Labor Statistics, 2016). One of the reasons why PD stayed the same from 2007 to 2009 might be that the recession started mid-2007 and ended mid-2009. Another reason could be the way that K6 questions were asked: "in the past 30 days…" which is a short time frame, and individuals might have felt the same in 2007 and 2009 when the survey was taken. The changes in responses from 2009 to 2012 were as expected: a shift from moderate PD into low PD, as the recession ended, and the economy was improving.

Examining the PD level by severity for each year, the low PD decreased, and moderate PD increased from 2007 to 2009 by 2 percentage points. The low PD level increased, and moderate PD level decreased by 6 percentage points from 2009 to 2012. It is important to note that the changes in PD happened at the low or moderate level and not the severe level (severe PD is used as a screening tool for serious mental illness (Kessler et. al., 2002). Based on these results we confirmed that the recession had an effect on low and moderate psychological distress and that additional research on the effects of these two levels of PD on the population is necessary.

The results in Aim 3 show that higher level of income decreased the level of PD and higher percentage of uninsured increases the level of PD. These results are consistent with the current literature. It is interesting to see that income has a minimal effect on the stability of PD over time and could be caused by the framing of the K6 questionnaire, for most people income is constant for more than the 30-day time frame in the K6 survey. On the other hand, percentage of uninsured had a larger effect on the stability of PD during recession (2007 to 2009) and a minimal effect on the stability of PD after recession (2009 to 2012). The average level of income for the 3 years stayed constant for the three years at 25-50K. Even though this result contradicts prior literature (income levels decline during economic recession especially for low to moderate income individuals (Edmiston, 2013)), the average income level may not reflect the proportion of individuals who might have experienced a decline in income. The sharpest decline was in the average income for low income families has followed between 9.6 percent and 12.3 percent from 2001 (lowest 40 percent of the income distribution) (Edmiston, 2013). Since we had 3 major income levels, we did not see a major shift income from the high income to middle income or the middle income to lower income. The large effect of uninsured on psychological distress could be cause by the fact that individuals that have no insurance might have a lower level of income or no income. That along with the high unemployment rate caused by the recession limiting the success of earning an income and ultimately having insurance could be driving the magnitude of the uninsured rate on psychological distress. Herring and Trish, 2015, Holohan and McMorrow, 2013, and Dranove et al., 2014 showed that the recession not only caused the income per capita to fall, but also showed that the health care expenditures also fell which could be a cause for increased psychological distress (low health care utilization, which will be discussed in the next aim).

The results in Aim 4 show that PD increases by more than 17 percent with an increase by one percent of uninsured (direct effect and indirect effect) which is a slightly lower percentage than found in Aim 3. The difference is caused by the introduction of the treatment of mental health variable as a mediator in the model. The direct effect and the indirect effect of uninsured is almost the same with only 0.3 percent difference (the direct effect is higher) meaning that the introduction of mental health treatment as a mediator does not have much of an impact on the PD level. The total effect (direct and indirect) does have larger impact on the PD coefficients over time as compared with the panel model or the panel + uninsured model (Aim 3). The mediation on PD07 did not change the magnitude of the coefficients of stability, but the mediation's effect on PD07 did change the relationship between PD07 and PD09 (coefficient of stability decreased from 0.983 to 0.610) and on PD12 changed the relationship between PD09 and PD12 (coefficient of stability decreased from 0.744 to 0.389). The changes in the relationship could be caused by (1) lower average percentage of uninsured and/or (2) higher percentage of individuals taking medicine or receiving treatment for a mental health condition.

In the final aim, Specific Aim 5, we found that heavy and binge drinking had no effect on PD and PD had no effect on heavy and binge drinking. The coefficients of stability have remained constant when assessing the effect of alcohol consumption on PD07, PD09, and PD12. In addition, results show that smoking increases PD before, during, and after recession for all the years. It changed the relationship between PD09 to PD12 only when assessing its effect on PD12 (the coefficient of stability decreased from 0.741 to 0.543). On the other hand, low PD had a significant negative effect on smoking, and moderate and high PD had a positive effect on smoking with high PD showing higher effect as compared with moderate PD. In addition, we found that low psychological distress decreased smoking while moderate and severe psychological distress increased smoking. This is an important finding for smoking cessation researchers, policy makers, and medical personnel to review their policies and programs on smoking cessation.

When both variables were added into the model, the results mimicked the results of the models when variables were added individually to the model. It is interesting to note that the correlation between smoking and alcohol consumption in the panel model is significant, but very small, indicating that there is no relationship between alcohol consumption and smoking. The results in the regression models tells us a different story, where alcohol consumption has a significant positive effect on smoking, and vice versa.

These results are a stepping stone to an important mental health condition that it is often overlooked by individuals, clinicians, employees, and policy makers.

5.2. IMPLICATIONS AND RECOMMENDATIONS

This research is exploratory in nature to assess, during an economic recession, the effect of insurance, income, smoking, and alcohol consumption on psychological distress over time and it can serve as a basis upon which future research can build. The work highlighted the importance of psychological distress on risky health behaviors and use of mental health services. PD is not a diagnosable mental health illness and not covered by insurance even though individuals experience it at one time or another. Our research has shown that during high periods of stress such as economic recession individuals shift from low PD to moderate PD. One policy recommendation would be for the introduction of government mental health programs designed to target uninsured individuals that experience moderate to high psychological in order to address the higher use of risky behaviors.

Our research has found that smoking has a positive effect on PD and that PD has a positive effect on smoking for moderate and high PD. In addition, we found that low psychological distress decreased smoking while moderate and severe psychological distress increased smoking. This is a significant finding because it could be the stepping stone for additional smoking cessation research. Previous research has found that individuals with serious mental illness have lower rates of quitting as compared to smokers with no mental illness (Lasser at al., 2000; Le Cook et al., 2014). In light of our finding, additional research should be conducted to assess the level of quitting rates for individuals with PD. These results could have significant impact on tobacco secession strategies for individuals with moderate and severe PD.

The main goal of this research was to bring psychological distress, a form or mild to moderate mental illness, to the forefront of policymakers, lawmakers, and clinicians. PD is a very important topic that has been much talked about, but not much has been done with regards to policy. As shown in our research, as well as previous research, psychological distress exists within the adult noninstitutionalized population and if not treated it can lead to severe mental illness. The first step in the process is to have psychological distress recognized as a form of mental illness and added as a coverable illness under insurance policies. The K10/K6 scale developed by Dr. Kessler (Kessler et al., 2012) can be used as a screening tool for the severity of PD of the population and intervention programs could be put in place for individuals with moderate or severe PD.

5.3. STUDY LIMITATIONS

The biggest limitation of this research is the use of cross-sectional datasets. The statistical analysis was not performed at individual level, but at population level. This is a limitation if the research needs to be done at individual level.

One advantage of the merging of the datasets besides creating one dataset to mimic longitudinal dataset is that the matching was done for an "average" person created by using covariates which strengthen the link between the time periods (this type of matching is not novel, it was done before such as propensity scoring). One disadvantage of creating composite observations is that the "average" person is not the same person over time (even though it exhibits the same covariate characteristics), as a result we do not know if the results reflect any one individual's actual behavior.

Prior literature cautions the researchers against using a solution with one or two items in any one factor, three or more are preferred (Kenny, 1979). By having one or two items in a factor it makes it difficult to estimate error correlations which lead to specification error (Kline, 2011). The results from EFA and CFA have shown that anxiety symptoms factor has only 2 items which was consistent with the results found by Ko and Harrington, 2015.

Another limitation is that we did not have surveys in consecutive years to fully assess the effect of the covariates on PD. The results do not show the effect on 2008 (a full year in recession), or the results right after the recession in 2010 or even 2011.

The research was limited in scope and examined the association between psychological distress and risky behaviors (smoking and alcohol use) during economic recession and did not account for the severity of the economic loss on the population (i.e.

unemployment level, salary suppression, housing foreclosure level, etc.) or other confounders.

This study used a novel methodological design, integrative data analysis, to identify general population trends on cross-sectional datasets by creating composite observations in order to link the datasets and make it mimic a longitudinal dataset. As far as we know this is the first time this was done in such manner. Additional research should be done to confirm the findings.

Even though creating composite observations gave us the opportunity to analyze the PD trend in multiple cross-sectional datasets it also decreased the variability of the data by not allowing each individual to contribute to the analysis. This limitation can be eliminated by using longitudinal data.

5.4. FUTURE RESEARCH

The next step in this research is for the authors to look at the use of NIHS K6 datasets for additional data points or confirmation of methodology and results.
Since this research limited in scope, additional analyses could be performed to assess the effect of insurance type on PD, of PD by U.S. region, etc.

5.5. SUMMARY

A goal of the research was to bring psychological distress on the front of decision-makers (policy makers, employers, insurance companies, etc.). PD is an important part of the mental health of an individual and it is often overlooked. Smoking has positive effect on psychological distress and psychological distress has a positive effect on smoking. Smoking a modifiable behavior, but as research has shown (Lasser et al., 2000;

Le Cook, et al., 2014) in conjunction with mental illness it is harder to quit. With the new information gathered from this research the smoking cessation programs should take into account the associations found in this research. Insurance status is a fixed factor and it has a negative effect on psychological distress. Policymakers along with mental health providers and insurance companies should investigate the possibility of recognizing PD as a mental illness and include moderate to severe psychological distress as a coverage option. Even though PD did not change from 2007 to 2009 it does not mean that the economic distress did not have an effect on PD over time. The 2007-2009 recession begun in the middle of 2007 and ended in the middle of 2009 and the 2007 survey and the 2009 survey might not have caught the full effect. A change in PD was seen from 2009 to 2012 potentially showing that the recession did have an effect on PD.

In addition, by testing a new methodology of linking cross-sectional datasets, our hope is that this research will open the door to future research that previously has not been done because of the limitations of the longitudinal design.

REFERENCES

Adams S.H., Park M.J., Irwin C.E., Jr. (2015). Adolescent and Young Adult Preventive Care: Comparing National Survey Rates. *Am J Prev Med* 2015;49(2):238-47.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control.* 19(6): 716-723.

Alley, D.E., Lloyd J., Pagan J.A., Pollack C.E., Shardell M., Cannusion C. (2011). Mortgage delinquency and changes in access to health resources and depressive symptoms in a nationally representative cohort of Americans older than 50 years. *Am J Public Health*, 101(12):22293-2298.

Allison, P.D. (2000). Multiple imputation for missing data: a cautionary tale. *Sociological Methods & Research,* 28(3):301-309.

American Psychiatric Association. (1987). Diagnostic and Statistical Manual of Mental, Disorders, Third Edition, Revised. Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2000). Quick Reference to the Diagnostic Criteria from DSM-IV. Washington, DC. American Psychiatric Press, Inc.

American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders. 5th Edition, Text Revision. Arlington, VA: American Psychiatric Association.

Andersen, E.M., Catlin, T.K., Wyrwich K.W., Jackson-Thompson, J. (2003). Retest reliability of surveillance questions on health related quality of life. J. *Epidemio Commun Health*, 57:339-373.

Andrade, L.H., Alonso, J., Mneimneh, Z., Wells, J.E., Al-Hamzawi, A., Borges, G., Bromet, E., Bruffaerts, R., de Girolamo, G., de Graaf, R., Florescu, S., Gureje, O., Hinkov, H.R., Hu, C., Huang, Y., Hwang, I., Jin, R., Karam, E.G., Kovess-Masfety, V., Levinson, D., Matschinger, H., O'Neill, S., Posada-Villa, J., Sagar, R., Sampson, N.A., Sasu, C., Stein, D., Takeshima, T., Viana, M.C., Xavier M., Kessler, R.C. (2014). Barriers to mental health treatment: results from the WHO World Mental Health (WMH) Survey. *Psychol Med*., 44(6):1303-1317

Angst, J., Sellaro, R., Merikangas, K.R. (2000). Depressive spectrum diagnosis. *Compr Psychiatry,* 41(Suppl 1):39-47.

Baraldi, A.N., Enders, C.K. (2010). An introduction to modern missing data analysis. *Journal of School Psychology,* 48:5-37.

Barr, B., Taylor-Robinson, D., Scott-Samuel, A., McKee, M., Stuckler D. (2012). Suicides associated with the 2008-10 economic recession in England: time trend analysis. *BMJ*, 345:e5142 [PubMed: 22893569].

Baron, R.M., and Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology,* 51:1173-1182.

Beard E., West, R., Michie S., Brown, J. (2017). Association between smoking and alcohol-related behaviours: a time-series analysis of population trends in England. *Society for the Study of Addiction.* Vol. 112(10) p. 1832-1841

Bell, S., Britton, A. (2014). An exploration of the dynamic longitudinal relationship between mental health and alcohol consumption: a prospective cohort study. *BMC Medicine,* 12:91. DOI: 10.1186/1741-7015-12-91.

Bentler, P.M, & Bonett, D.G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin,* 88:588-606.

Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin,* 107:238-246.

Berkout, O.V., Gross, A.M., Young, J. (2014). Why so many arrows? Introduction to structural equation modeling for the novitiate user. *Clin Child Fam Phsychol Rev.,* 17: 217-229. doi: 10.1007/s10567-014-0165-3.

Blumberg S.J., Luke J.V. (2013). Wireless substitution: Early release of estimates from the National Health Interview Survey, January–June 2013. National Center for Health Statistics. December 2013. Available from http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201312.pdf

Bollen, K.A. (1989). *Structural Equations with Latent Variables*. New York, NY: Wiley.

Bollen, K.A. (1990). Overall fit in covariance structure models: Two types of sample size effects. *Psychological Bulletin,* 107:256-259.

Browne, M.W. and Cudeck, R. (1993). *Alternative ways of assessing model fit.* Testing Structural Equation Models (Bollen, K.A. and Long, J.S. (Eds.)):136-162. Sage, Newbury Park, CA.

Burger A.E., Reither E.N. (2014). Monitoring receipt of seasonal influenza vaccines with BRFSS and NHIS data: challenges and solutions. *Vaccine*, 32(31):3950-4.

Byrne, B.M., Shavelson, R.J., Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychological Bulletin,* 3:456-466.

Cannuscio, C.C., Alley D.E., Pagan, J.A., Soldo, B., Krasny, S., Shardell, M., Asch, D.A., Limpan, T.H. (2012). Housing strain, mortgage foreclosure, and health. *Nursing Outlook,* 60(3):134-142.

Center for Behavioral Health Statistics and Quality. (2015). *Behavioral health trends in the United States: Results from the 2014 National Survey on Drug Use and Health* (HHS Publication No. SMA 15-4927, NSDUH Series H-50). Retrieved from http://www.samhsa.gov/data/

Centers for the Disease Control and Prevention. (2013, October 4). *Mental Health Basics.* Retrieved from: http://www.cdc.gov/mentalhealth/basics.htm.

Centers for Disease Control and Prevention. *Behavioral Risk Factor Surveillance System Overview*. August 15, 2014. Retrieved from http://www.cdc.gov/brfss/annual_data/annual_2013.html on 3/10/2015.

Centers for Disease Control and Prevention. Program Performance and Evaluation Office. March 30, 2016. Retrieved from http://www.cdc.gov/mentalhealth/basics/mental-illness/depression.htm.

Centers for Disease Control and Prevention. (2012). Mental illness surveillance among adults in the United States. *Morbidity and Mortality Weekly Report; 60(Suppl.)*.

Centers for Disease Control and Prevention. (2012). Vital signs: binge drinking prevalence, frequency, and intensity among adults- United States, 2010. *Morbidity and Mortality Weekly Report,* 61(01):14-19. Retrieved from http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6101a4.htm

Chen, F., Bollen, K.A., Paxton, P., Curran, P., & Kirby, J. (2001). Improper solutions in structural equation modes: Causes, consequences, and strategies. *Sociological Methods and Research,* 29:468-508.

Cheung, G.W., & Rensvold, R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling,* 9:233-255.

Clum, G.A., Rice, J.C., Broussard, M., Johnson, C.C., Webber, L.S. (2014). Associations between depressive symptoms, self-efficacy, eating styles, exercise and body mass index in women. *Journal of Behavioral Medicine*, 37:577-586. DOI 10.1007/s10865-013-9526-5

Cochran, W.G. (1952). The Chi-square test of goodness of fit. *The Annals of Mathematical Statistics,* 23(3):315-345.

Collins, L.M., Schafer, J.L, Kim, C.M. (2002). A comparison of restrictive strategies in modern missing data procedures. *Psychological Methods.* 6:330-351.

Colton, C.W., Manderscheid, R.W. (2006). Congruencies in increased mortality rates, years of potential life lost, and causes of death among public mental health clients in eight states. *Preventing Chronic Disease*, 3(2), A42.

Compton, W.M., Gfroerer, J., Conway, K.P., Finger, M.S. (2014). Unemployment and substance outcomes in the United States 2012-2010. *Drug Alcohol Depend*., September 1:0:350-353. [PubMed: 25042761]

Comrey, A.L., Lee, H.B. (1992). Interpretation and application of factor analytic results. In: *A First Course on Factor Analysis,* 2nd edition. Hillsdale, NJ: Lawrence Erlbaum.

Costello, A.B., Osborne, J.W. (2005). Best practices in Exploratory Factor Analysis: four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation,* 10(7):1-9.

Couch, K.A., Placzek, D.W. (2010). Earnings losses of displaced workers revisited. *The American Economic Review*, 100(1):572-589.

Cujipers, P., Aartjan, T.F., Beekman, T.F., Reynolds C.F. (2012). Preventing depression: a global priority. *JAMA*, 307(10):1033-1034.

Curran, P.J., & Hussong, A.M., Cai, L., Huang, W., Chassin, L., Sher, K.J., Zucker, R.A. (2008). Pooling data from multiple longitudinal studies: the role of item response theory in integrative data analysis. *Dev Psychology,* 44(2):365-380.

Curran, P.J., & Hussong, A.M. (2009). Integrative Data Analysis: the simultaneous analysis of multiple data sets. *Psychol Methods,* 14(2):81-100.

Curran, P.J., West, S.G., Finch, J.F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods,* 1, 16-29.

Davalos, M.E., Fang, H., French, M.T. (2012). Easing the pain of an economic downturn: macroeconomic conditions and excessive alcohol consumption. *Health Econ.,* 11:1318-1335 [PubMed: 21913282]

Davis, S.L. & Finney, S.J. (April, 2003). Examining the psychometric properties of the cross-cultural adaptability inventory. *American Educational Research Association,* Chicago, IL.

DeCoster, J. (1998). Overview of factor analysis. Retrieved October 12, 2018 from www.stat-help.com/notes.html

Dee, T.S. (2001). Alcohol abuse and economic conditions: evidence from repeated cross-sections of individual data. *Health Econ*., 10:257-270 [PubMed: 11288191]

De Leon J., Rendon D.M., Baca-Garcia E., Aizpuru F., Gonzalez-Pinto A., Anitua C., Diaz, F.J. (2017). Association between smoking and alcohol use in the general population: stable and unstable odds ratios across two years in two different countries, *Alcohol and Alcoholism*, Volume 42(3), p. 252–257

DeNavas-Walt, C., Proctor, B.D., Smith, J.C., U.S. Census Bureau. (2010). Current Population Reports, Income, Poverty, and Health Insurance Coverage in the United States: 2009, *U.S. Government Printing Office,* Washington, DC. Retrieved from: https://www.census.gov/prod/2010pubs/p60-238.pdf on 7/5/2018.

DiStefano, C., Dombrowski, S.C. (2006). Investigating the Theoretical Structure of the Stanford-Binet-Fifth Edition. *Journal of Psychoeducational Assessment.* 24(2):123-136.

Dranove, D., Garthwaite, C., Ody, C. (2014). Health spending slowdown is mostly due to economic factors, not structural in change in the health care sector. *Health Affairs,* 33(8): 1399-1406.

Druss, B.G., Walker, E.R. (2011). *Mental disorders and medical comorbidity.* Princeton (NJ): Robert Wood Johnson Foundation.

Duncan, T.E., Duncan, S.C. (2009). The ABC's of LGM: An introductory guide to latent variable growth curve modeling. *Social and Personality Psychology Compass,* 3(6):979-991.

Dwyer-Lindgren L, Freedman G, Engell RE, Fleming TD, Lim SS, Murray CJ, et al. (2013). Prevalence of physical activity and obesity in US counties, 2001-2011: a road map for action. *Popul Health Metr*;11:7.

Edge, L.E., Zheng, D., Simpson, K. (2002). Comorbid depression is associated with increased health care use and expenditures in individuals with diabetes. D*iabetes Care*. 25:464-70.

Edmiston, K. (2013). The low- and moderate-income population in recession and recovery: results from a new survey. Federal Reserve Bank of Kansas City*, Economic Review,* First Quarter.

Elliott, M.R., West, B.T. (2015). "Clustering by Interviewer": A source of variance that is unaccounted for in single-stage health surveys: Facebook likes. *J Med Internet Res,* 17(4):e98.

Enders, C.K. (2013). Dealing with missing data in developmental research. *Child Development Perspectives.* Vol. 7(1): 27-31.

Endicott, J., Spitzer, R.L., Fleiss, J.L., Cohen, J. (1976). The Global Assessment Scale: a procedure for measuring overall severity of psychiatric disorders. *Archives of General Psychiatry.* 33:766-771.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., Strahan, E.J. (1999). Evaluating the use of Exploratory Factor Analysis in Psychological Research. *Psychological Methods,* 4(3):272-299.

Fahimi, M., Link, M., Schwartz, D.A., Levy, P., Mokdad, A. (2008). Tracking chronic disease and risk behavior prevalence as survey participation declines: statistics from the Behavioral Risk Factor Surveillance System and other national surveys. *Prev Chronic Dis*.; 5(3).

Finney, S.J., Distefano, C. (2013). Non-normal and categorical data in structural equation modeling. *Structural Equation Modeling: A second course, 2nd edition.*, (pp. 439-492). Charlotte, NC: Information Age Publishing.

Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, Series A,* 222:309-368.

Forero, C.G., Maydeu-Olivares, A.M., Gallardo-Pujol, D.G. (2009). Factor Analysis with Ordinal Indicators: A Monte Carlo Study Comparing DWLS and ULS Estimation. *Structural Equation Modeling.* 16:625-641.

Forman-Hoffman, V.L., Muhuri, P.K., Novak S.P., Pemberton M.R., Ault K.L., Mannix D. (2014) Psychological Distress and Mortality among Adults in the U.S. Household Population. Substance Abuse and Mental Health Services Administration (SAMHSA). Retrieved from www.samhsa.gov/data.

Forsell, Y. (2007). A three-year follow-up of major depression, dysthymia, minor depression and subsyndromal depression: results from a population-based study. *Depression and Anxiety.* 24:62-65. DOI 10.1002/da.20231

Furukawa, T.A., Kessler, R.C., Slade, T., Andrews, G. (2003). The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. *Psychological Medicine,* 33:357-362.

Gameroff, M.J., Olfson, M. (2006). Major depressive disorder, somatic pain, and health care costs in an urban primary care practice. *J Clin Psychiatr*, 67:132-9.

Gittelman S., Lange V., Gotway Crawford C.A., Okoro C.A., Lieb E., Dhingra S.S., Trimarchi, E. (2015). A new source of data for public health surveillance: Facebook likes. *J Med Internet Res*;17(4):e98.

Gorsuch, R.L. (1983). *Factor Analysis (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum.

Greenberg, P.E., Fournier, A.A., Sisitsky, T., Pike, C.T., Kessler, R.C. (2015). The economic burden of adults with major depressive disorder in United States (2005 and 2010). *J Clin Psychiatry*, 76(2):155-62.

Gunzler, D., Chen, T., Wu, P., Zhang, H. (2013). Introduction to mediation analysis with structural equation modeling. *Shanghai Archives of Psychiatry,* 25(6): 390-394, doi: 10.3969/j.issn.1002-0829.2013.06.009.

Hanley, J., & McNeil, B. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology,* 143:29-36.

Hannan, M.T. & Young, A.A. (2018). Estimation in panel models: results on pooling cross-sections and time series. *Sociological Methodology,* Vol. 8 (1977), 52-83.

Hasson-Ohayon, I., Levy, I., Kravetz, S., Vollanski-Narkis, A., Roe, D. (2011). Insight into mental illness, self-stigma, and the family burden of parents of persons with a severe mental illness. *Comprehensive Psychiatry*, 52(1):75-80.

Hatcher, L. (1994). A Step-by-Step approach to using the SAS System for Factor Analysis and Structural Equation Modeling. *SAS Institute, Inc.*, Cary, NC, USA.

Herring B., & Trish, E. (2015). Explaining the growth in US health care spending using state-level variation in income, insurance, and provider market dynamics. *The Journal of Health Care Organization, Provision, and Financing.* I-II. DOI: 10.1177/0046958015618971

Hert, M.D., Correll, C.U., Bobes, J., Cetkovich-Bakmas, M., Cohen, D., Asai, I., Detraux, J., Gautam, S., Moller, H.J., Ndetei, D.M., Newcomer, J.W., Uwakwe, R., Leucht, S. (2011). Physical illness in patients with severe mental disorders. I. Prevalence, impact of medications and disparities in health care. *World Psychiatry*, 10(1):52-77.

Holahan, J. (2011). The 2007-09 Recession and health insurance coverage. *Health Affairs.* 330(1):145-152.

Holohan, J., and McMorrow, S. (May 2013). What drove the recent slowdown in health spending growth and can it continue? *Urban Institute Report,* Washington DC. Retrieved from:

https://www.urban.org/sites/default/files/publication/23596/412814-What-Drove-the-Recent-Slowdown-in-Health-Spending-Growth-and-Can-It-Continue-.PDF on 7/5/2018.

Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrica,* 30:179-185.

Horn, J.L. and McArdle, J.J. (1992). A practical and theoretical to measurement invariance in aging research. *Experimental Aging Research.* 18:117-144.

Horwath, E., Johnson, J., Klerman G.L., Weissman M.M. (1992). Depressive symptoms as relative and attributable risk factors for first-onset major depression. *Arch Gen Psychiatry,* 26:117-126.

Houle, J.N., & Light, M.T. (2014). The home foreclosure crisis and rising suicide rates, 2005 to 2010. *American Journal of Public Health,* 104(6):1073-1079.

Hox, J.J., & Bechger, T.M. (1998). Introduction to structural equation modeling. *Family Science Review*, 11:354-373.

Hu, L. & Bentler, P.M., Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin,* 112:351-362.

Hu, L, & Bentler, P.M. (1998). Fit indices in covariance structure modeling: sensitivity to underparameterized model specification. *Psychological Methods,* 3(4):424-433.

Hu, L. & Bentler, P.M. (1999). Cutoff criterion for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6:1-55.

Hu, S.S., Pierannunzi, C., Balluz, L. (2011). Integrating a multimode design into a national random-digit–dialed telephone survey. *Prev Chronic Dis*., 8(6):A145.

Insel, T.R. (2008). Assessing the economic costs of serious mental illness. *Am J Psychiatry*, 165(6):663-665.

Institute of Medicine. (2001). *Coverage matters: Insurance and health care*. Washington, DC: National Academies Press.

Jacobson, L.S., LaLonde, R.J., Sullivan, D.G. (1993). Earnings losses of displaced workers. *The American Economic Review,* 83(4):685-709.

Jamal, A., Homa, D.M., O'Connor, E., Babb, S.D., Caraballo, R.S., Singh, T., Hu S.S., King, B.A. (2015). Current Cigarette Smoking Among Adults – United States, 2005-2014. *MMWR Morb Mortal Wkly Rep*; 64(44):1233-1240.

Jones, D.R., Macias, C., Barreira, P.J., Fisher, W.H., Hargreaves, W.A., Harding, C.M. (2004). Prevalence, severity, and co-occurrence of chronic physical health problems of persons with serious mental illness. *Psychiatr Serv*., 55(11):1250-1257.

Joreskog, K.G. and Sorbom, D. (1996). *LISREL 8: User's reference guide.* Scientific Software International, Chicago, IL.

Kapp, J.M., Jackson, T., Petroski, G.F., Schootman, M. (2009). Reliability of Health related quality of life indicators in cancer survivors from a population based sample, 2005, BRFSS. *Public Health*, 123:321-325.

Karg, R.S., Bose, J., Batts, K.R., Fortman-Hoffman, V.L., Liao, D., Hirsch, E., Pemberton, M.R., Colpe, L.J., Hedden, D.L. (2014). Past year mental disorders among adults in the United States: Results from the 2008-2012 Mental Health Surveillance Study. *NCBI Bookshelf*. CBHSQ Data Review. Rockville (MD). PIMD: 27748100.

Kenny, D.A. (1979). *Correlation and Causality*. New York: Wiley.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L. T., Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalence and trends in non-specific psychological distress. *Psychological Medicine*, 32:959 –976. http://dx.doi.org/ 10.1017/S0033291702006074.

Kessler, R.C., Barker, P.R., Colpe, L.J., Epstein, J.F., Gfroerer, J.C., Hiripi, E., Howes, M.J, Normand, S-L.T., Manderscheid, R.W., Walters, E.E., Zaslavsky, A.M. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry*. 60(2):184-189.

Kessler, R.C., Herringa, S., Lakoma M.D., Petukhova M., Rupp, A.E., Schoenbaum, M., Wang, P.S., Zasalvsky, A.M. (2008). Individual and societal effects of mental disorders on earnings in the United States: results from the National Co-morbidity Survey Replication. *Am J Psychiatry*, 165:703-711.

Kessler, R.C., Green, J.G., Gruber, M.J., Sampson, N.A., Bromet, E., Cuitan, M., Furukawa, T.A., Gureje, O., Hinkov, H., Hu, C-I., Lara, C., Lee, S., Mneimneh, Z., Myer, L., Oakley-Browne, M., Posada-Villa, J., Sagar, R., Viana, M.C., Zaslavsky, A.M. (2010). Screening for serious mental illness in the general population with the K6 screening scale: results from the WHO World Mental Health (WMH) survey initiative. *International Journal of Methods in Psychiatric Research,* 19(Supplement 1):4-22.

Kim, H., Lee J. (2006). The impact of comorbidity on wealth changes in later life. *J Gerontol Series B: Psychol Sci Soc Sci,* 61:S307-14.

Kline, R.B. (2011). *Principles and practice of structural equation modeling.* (3rd ed.). New York, NY: Guilford Press.

Ko, J. and Harrington, D. (2015). Factor structure and validity of the K6 scale for adults with suicidal ideation. *Journal of the Society for Social Work and Research*, 7(1):43-63.

Kowalski, J., Tu, X.M. (2007). *Modern Applied U Statistics.* New York, NY: Wiley.

Lasser, K., Boyd, J.W., Woolhandler, S., Himmelstein, D.U., McCormick, D., Bor, D.H. (2000). Smoking and mental illness: A population-based prevalence study. *JAMA*, 284(20):2606-2610.

Le Cook, B., Wayne, G.F., Kafali, N.E., Liu, Z., Shu, C., Flores, M. (2014). Trends in smoking among adults with mental illness and association between mental health treatment and smoking cessation. *JAMA.* 311(2):172-182.

Lewis-Beck, M. S., Bryman, A., & Futing Liao, T. (2004). *The SAGE encyclopedia of social science research methods* Thousand Oaks, CA: Sage Publications, Inc. doi: 10.4135/9781412950589

Li, C., Balluz, L.S., Ford, E.S., Okoro, C.A., Zhao, G., Pierannunzi, C. (2012). A Comparison of Prevalence Estimates for Selected Health Indicators and Chronic Diseases or Conditions from the Behavioral Risk Factor Surveillance System, the National Health Interview Survey, and the National Health and Nutrition Examination Survey, 2007-2008. *Preventive Medicine*, 54(6):381-7.

Libman, K., Fields D., Saegert S. (2012). Housing and health: a social ecological perspective on the US foreclosure crisis*. Housing Theory Soc.,* 29(1):1-24.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83:1198–1202.

Little, T. D., Card, N. A., Bovaird, J. A., Preacher, K. J., & Crandall, C. S. (2007). Structural equation modeling of mediation and moderation with contextual factors. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 207-230). Mahwah, NJ: Lawrence Erlbaum Associates.

Lutterman, T., Ganju V., Schacht, L., Shaw, R., Monihan, K, Huddle, M. (2003). *Sixteen state study on mental health performance measures.* HHS Publication No. SMA 03-3835. DHHS pub no Rockville, Md, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, 2003.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation Analysis. *Annual Review of Psychology*, *58*, 593. http://doi.org/10.1146/annurev.psych.58.110405.085542

McArdle, J.J. (1988). Dynamic but Structural Equation Modeling of Repeated Measures Data. In: Nesselroade J.R., Cattrell R.B. (eds). *Handbook of Multivariate Experimental Psychology*. Perspectives on Individual Differences. Springer, Boston, MA.

McHugh, M.L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica,* 22(3):276-282.

McInnerney, M., Mellor, J.M., Hersch Nicholas, L. (2013). Recession depression: Mental health effects of the 2008 market crash. *Journal of Health Economics,* 32:1090-1104.

McLaughlin, K.A., Nandi A., Keyes K.M., Uddin M., Aiello A.E., Galea S., and Koenen K.C. (2012). Home foreclosure and risk of psychiatric morbidity during the recent financial crisis. *Psycol Med*., 42(7):1441-1448 [22099861]

Meade, A.W., Johnson, E.C., & Braddy, P.W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology,* 93(3): 568-592.

Miles, H., Johnson, S., Amponsah-Afuwape, S., Finch, E., Leese, M., Thornicroft, G. (2003). Characteristics of subgroups of individuals with psychotic illness and comorbid substance use disorder. *Psychiatric Services,* 54(4):544-561.

Miller, B.J., Paschall, C.B., 3rd, Svendsen, D.P. (2006). Mortality and medical comorbidity among patients with serious mental illness. *Psychiatric Services*, 57(10):1482-1487.

Miyazaki Y., Raudenbush, S.W. (2000). A test for linkage of multiple cohorts from an accelerated longitudinal design. *Psychological Methods,* 5:44-63. Pubmed: 10937322.

Mojtabai, R., Olfson, M., Sampson, N.A., Jin, R., Druss, B., Wang, P.S., Wells, K.B., Pincus, H.A., Kessler, R.C. (2011). Barriers to mental health treatment: results from the National Comorbidity Survey Replication (NCS-R). *Psychological Methods*, 41(8):1751-1761.

Mortensen, K., & Chen, Jie. (2013). The great recession and racial and ethnic disparities in health services use. *JAMA Intern Medicine,* 173(4):315-317

National Institute of Mental Health. (2016, January 28). Mental Health Information. Retrieved from: http://www.nimh.nih.gov/health/topics/index.shtml

National Comorbidity Survey (2005). Harvard Medical School. Retrieved from: https://www.hcp.med.harvard.edu/ncs/

Nelson, D.E., Powell-Griner, E., Town, M., Kovar, M.G. (2003). A comparison of national estimates from the National Health Interview Survey and the Behavioral Risk Factor Surveillance System. *American Journal of Public Health*, 93:1335–1341.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection*. Journal of the Royal Statistical Society*, 109:558-606.

Office of the Management and Budget, Executive Office of the President (1974). *Circular No. A-015.* Retrieved from http://www.gao.gov/assets/120/119653.pdf

Pett, M., Lackey, N. & Sullivan, J. (2003). *Making sense of factor analysis.* Thousand Oaks: Sage Publications, Inc.

Pierannunzi, C., Hu, S.S., Balluz L. (2013). A systematic review of publications assessing reliability and validity of the Behavioral Risk Factor Surveillance System (BRFSS), 2004-2011. *BMC Medical Research Methodology,* 13:49.

Pollack, C.E., Kurd, S.K., Livishits, A., Weiner, M., Lynch, J. (2011). A case-control study of home foreclosures, health conditions, and health care utilization. *J Urban Health,* 88(3):469-478.

Pratt, L.A., Brody, D.J. (2010). Depression and smoking in the U.S. household population aged 20 and over, 2005-2008*. NCHS Data Briefs*, (34):1-8.

Pratt, L.A., Brody, D.J. (2014). Depression in the U.S. household population, 2009-2012. *NCHS Data Brief*, No. 172. Hyattsville, MD: National Center for Health Statistics 2014.

Prochaska, J.J., Sung H.Y., Max W., Shi Y., Ong M. (2012). Validity Study of the K6 Scale as a Measure of Moderate Mental Distress based on Mental Health Treatment Need and Utilization*. Int J Methods Psychiatr Res*, 21(2): 88-97 doi: 10.1002/mpr.1349.

Rehm, J., Ustun, T.B., Saxena, S., Nelson, C.B., Chatterji, S., Ivis, D., Adlaf, E. (1999). On the development and psychometric testing of the WHO screening instrucment to assess disablement in the general population. *International Journal, of Methods in Psychiatric Research,* 8(2):110-122.

Riumallo-Herl C., Basu, S., Stuckler, D., Courtin, E., Avendano, M. (2014). Job loss, wealth and depression during the Great Recession in the USA and Europe*. Int J Epidemiol,* 43(5):1508-1517. PMID: 24942142.

Rowan, K., McAlpine D.D., Blewett L.A. (2013). Access and cost barriers to mental health care, by insurance status, 1999-2000. *Health Affairs,* 32:1723-1730.

Rummel, R.J. (1970). *Applied factor analysis*. Evanston, IL: Northwestern University Press.

Sacks, J.J., Gonzales, K.R., Bouchery, E.E., Tomedi, L.E. Brewer, R.D. (2014). 2010 National and state costs of excessive alcohol consumption. *Am J Preve Med*., 49(5):e73-e79.

Sanchez-Villegas, A., Serrano-Martinez, M., Alonso, A., de Irala, J., Tortosa, A., Martinez-Gonzales, M.A. (2008). Role of tobacco use on the incidence of depression in the SUN cohort study. *Medicina Clinica,* 130(11);405-409.

Satorra, A., & Bentler, P.M. (1988a). Scaling corrections for chi-square statistics in covariance structure analysis. *ASA 1988 Proceedings of the Business and Economic Statistics,* Section (308-313). Alexandria, VA: American Statistical Association.

Satorra, A., & Bentler, P.M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C.C. Clogg (Eds.), *Latent Variable Analysis: Applications to Developmental Research.* (pp. 399-419). Thousand Oaks, CA, US: Sage Publications, Inc.

Satorra, A., & Bentler, P.M. (2010). Ensuring positiveness of the scaled difference Chi-square test statistic. *Psychometrika,* 75:243-248.

Schaller, J., Stevens A.H. (2015). Short-run effects of job loss on health conditions, health insurance, and health care utilization. *Journal of Health Economics*, 43:190-203.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics.* 6(2): 461-464.

Scott, K.M., Bruffaerts, R., Simon, G.E., Alonso, J., Angermeyer, M., de Girolamo, G., Demyttenaere K., Gasquet I., Haro J.M., Karam E., Kessler R.C., Levinson D., Medina Mora M.E., Oakley Browne M.A., Ormel J., Villa J.P., Uda H., Von Korff M. (2008). Obesity and mental disorders in the general population: results from the world mental health surveys. *Int J Obes (Lond.),* 32(1):192-200. PMID: 2736857.

Schumaker, R.E., & Lomax, R.G. (2010). *A beginners guide to structural equation.* New York: Statsline.

Simon, G.E., Arterburn D., Rhode, P., Ludman, E.J., Linde, J.A., Operskalski B.H., Jeffrey, R.W. (2011). Obesity, depression, and health services costs among middle-aged women. *J Gen Intern Med,* 26(11):1284-90. PMCID: 3208460.

Sullivan, D., Von Wachter, T. (2009). Job displacement and mortality; an analysis using administrative data. *The Quarterly Journal of Economics,* 124(3):1265-1306.

Steiger, J.H. (1989). EZPATH: *A supplementary module for SYSTAT and SYGRAPH.* Evanston, IL: SYSTAT.

Stuart, E.A. (2010). Matching methods for causal inference: A review and look forward. *Stat. Sci.* 25(1):1-21.

Szultecka-Debek, M., Miernik, K., Stelmachowski J., Jakovljevic M., Jukic, V., Aadamsoo, K., Hanno, S., Bitter, I., Tolna, J., Jarema, M., Janovic, S., Pecenak, J., Vavrusova, L., Tavcar, R., Walczak, J., Talbot, D., Augustynska, J. (2016). Schizophrenia causes significant burden to patients' and caregivers' lives. *Psychiatr Danub*, 28(2):104-10.

Tabachnick, B.G., & Fidell, L.S. (2007). *Using multivariate statistics*. (5th. Ed.). Boston, MA: Allyn & Bacon.

Travers, J.L., Cohen, C.C., Dick, A.W., Stone, P.W. (2017). The Great American Recession and forgone healthcare: Do widened disparities between African-Americans and Whites remain?, *PLsoS ONE,* 12(12):e0189676. https://doi.org/10.1371/journal.pone.0189676

Tucker, L.R., & MacCallum, R.C. (1997). *Exploratory factor analysis*. Unpublished Manuscript, Ohio State University, Columbus.

U.S. Bureau of Labor Statistics. (2016). Labor Force Statistics from the Current Population Survey. Data extracted on July 18, 2018 from https://data.bls.gov/timeseries/LNS14000000

U.S. Census Bureau. (2015). *Census Regions and Divisions of the United States.* Retrieved from http://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf

U.S. Department of Health and Human Services. (1999). Mental Health: A Report of the Surgeon General. Rockville, MD: U.S. Department of Health and Human Services; Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, National Institute of Mental Health.

US Department of Health and Human Services. (2010) How tobacco smoke causes disease: the biology and behavioral basis for smoking-attributable disease: a report of the Surgeon General. Atlanta, GA: US Department of Health and Human Services, CDC. Available at http://www.cdc.gov/tobacco/data_statistics/sgr/2010/index.htm.

US Department of Health and Human Services. (2013) *ASPE Research Brief*. Atlanta, GA: US Department of Health and Human Services, CDC. Available at https://aspe.hhs.gov/sites/default/files/pdf/76591/rb_mental.pdf.

U.S. Bureau of Labor Statistics. (2012). BLS Spotlight on statistics THE RECESSION OF 2007-2009. Retrieved from http://www.bls.gov/spotlight/2012/recession/pdf/recession_bls_spotlight.pdf on May 24, 2016.

van Beljouw, I., Verhaak, P., Prings, M., Cuikpers P., Pennix, B., Bensing, J. (2010). Reasons and determinants for not receiving treatment for common mental disorders. *Psychiatric Services*, 61(3):250-257.

Walker E.R., Cummings, J.R., Hockenberry J.M., Druss B.G. (2015). Insurance status, use of mental health services, and unmet need for mental health car in the United States. *Psychiatric Services*, 66(6):578-584.

Williams, B., Onsman, A., Brown, T. Exploratory factor analysis: A five-step guide for novices. *Journal of Emergency Primary Health Care,* 8(3):1-13.

Wood, A.J., Harrington, R.C., & Moore, A. (1996). Controlled trial of a brief cognitive-behavioral intervention in adolescent patients with depressive disorder. *Journal of Child Psychology Psychiatry,* 37:737-746.

World Health Organization. (2001). *Strengthening mental health promotion.* Geneva, World Health Organization (Fact sheet no. 220).

Yong, A.G., Pearce, S. (2013). A beginner's guide to factor analysis: focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology.* 9(2):79-94.

APPENDIX A

METHODS BACKGROUND

The BRFSS survey

The Behavioral Risk Factor Surveillance System (BRFSS) is a publicly available, health-related, telephone administered survey that collects data across all states to profile the non-institutionalized U.S. adult population (18 years of age and older) living in households regarding their health-related risk behaviors (CDC, 2014). The survey began in 1984 in 15 states and now collects data in all 50 states and U.S. territories.

The questionnaire is comprised of three parts:

Core Components:

The core components, consisting of the fixed core, rotating core, and emerging core.

<u>The fixed core</u> has a standard set of questions about demographic characteristics, questions on current health behaviors (health status, health care access, alcohol consumption, tobacco use, fruits and vegetable consumption, demographics). The questions in this portion of the survey are asked each year and mandatory for each state.

<u>The rotating core</u> is made up of two distinct sets of questions each asked in alternating years for each state. These questions address different topics such as inadequate sleep, oral health, falls, hypertension awareness, cholesterol awareness, cancer screening (breast, prostate, colorectal). The questions in this portion of the survey are asked by all states every other year.

The emerging core is a set of up to five questions that are added to the fixed and rotating cores. The questions typically focus on "late breaking" issues and are part of the core for only one year (e.g. the H1N1 flu questions added in 2009). After one year, these questions are either discontinued or incorporated into the fixed core, rotating core, or optional modules.

Optional Modules:

Optional modules are made up of standardized questions on various topics that each state may select and include in the questionnaire. For example, the 2015 BRFSS survey included 25 optional modules (pre-diabetes, diabetes, cognitive decline, shingles, adult human papillomavirus, sexual orientation and gender identity, etc.).

State-added questions:

The survey encourages states to add certain questions based their specific health priorities through the use of extra questions they choose to add to their questionnaire. For example, in 2018, the state of South Carolina added the following modules: marijuana use, recovery from substance use, opioid use, children's health assessment survey script. Another example, the state of Virginia added the following modules: e-cigarettes, tetanus diphtheria (TDAP) (Adults), shingles (Zostavax or ZOS), and sexual orientation and gender identity.

Numerous studies have examined data quality and national estimates, showing high reliability and validity (Pierannunzi et al., 2013; Li et al., 2012; Fahimi et al., 2008; Nelson et al., 2003). Pierannunzi et al, 2013 reviewed and summarized existing literature published on the reliability and validity of the survey data from 2004-2011. The overall findings indicated that BRFSS prevalence rates were comparable with other national self-reported surveys. Fahimi et al, 2008 and Li et al., 2012 compared selected health and risk factors estimates with two large self-reported surveys, the National Health Interview Survey (NHIS) and the National Health and Nutrition Examination Survey (NHANES). The results of the study found that the estimates from the BRFSS survey mirror the estimates from the two surveys. Nelson et al., 2003 compared NIHS and BRFSS data on smoking, height, weight, BMI (Body Mass Index), hypertension, immunization, lack of insurance coverage, cost as a barrier to medical care, and health status and found both to yield similar national estimates. The BRFSS data is collected each year, but not from the same individuals, collecting on average about 400 adult interviews each year.

Sample Selection for the Study

Respondents for the BRFSS are identified through telephone-based methods. Starting in 2011, the survey also included cellular telephone-based survey along with the landline phone survey. The disproportionate stratified sampling (DSS) has been used in landline sampling, and random generation from a sampling frame of confirmed cellular area codes and prefix combination was used in cellular phone sampling. The cellular phone respondents have equal chance of being contacted. In conducting the landline telephone survey, the data is collected from a randomly selected adult present in the household, while with the cellular telephone survey the data is collected from the individual using the cellular phone.

The design was used by the participating states. With the DSS design, telephone numbers are divided into two groups that are sampled separately. The high and medium-density groups contain telephone numbers that are expected to belong to households. The choice of high or medium density group is determined by the number of household numbers in its hundred block (a set of one hundred telephone numbers with the same area code, prefix, and first two digits of the suffix and all possible combinations of the last two digits). The two groups (high and medium-density) are sampled to obtain a probability sample of all households with telephones. The landline sampling ratio of high to medium density is 1:1.5. In 2007 and 2009 Guam, Puerto Rico, and the U.S. Virgin Islands did not use DSS design but used a simple random sample design. In 2012 Guam and Puerto Rico used the simple random design.

The validity and reliability of the BRFSS survey is a popular topic in the literature. There are several studies that concluded that the validity of its data is equivalent to other national surveys (National Health and Nutrition Examination Survey NHANES and National Health Interview Survey NIHS). The results of these studies are summarized in a systematic review published by Pierannunzi, Hu, and Baluz in 2013. With regards to mental health status they found two studies that yielded similar results of moderate to excellent reliability across quality of life measures with Kappa statistics of 0.67 for poor mental health days, 0.58 for frequent mental distress (Andersen et al., 2003), 0.57 social and emotional support, 0.61 life satisfaction (Kapp et al., 2009). Cohen's Kappa coefficient is a statistic that is frequently used to measure interrater reliability (the extent to which the raters assign the same score to the same variable) (McHugh, 2012).

The Mental Health Survey Instrument

The Mental Illness and Stigma module was introduced into the BRFSS in 2007 as an optional module, and repeated as an optional module in 2009, 2012, and 2013. Because the module was optional, not all the states opted to use it in their questionnaire.

As a result, 27 states used the module in 2007 (Alaska, Arkansas, California, Connecticut, District of Columbia, Georgia, Hawaii, Illinois, Iowa, Kentucky, Louisiana, Massachusetts, Minnesota, Mississippi, Missouri, Montana, Nevada, New Hampshire, New Mexico, Oklahoma, Puerto Rico, Rhode Island, South Carolina, Vermont, Virginia, and Wyoming), 8 states in 2009 (Georgia, Hawaii, Mississippi, Missouri, Nevada, South Carolina, Vermont, and Wyoming), 11 states in 2012 (Illinois, Iowa, Minnesota, Missouri, Montana, Nevada, New Mexico, New York, Oregon, Puerto Rico, and Washington), and 5 states in 2013 (Colorado, Minnesota, Nevada, Tennessee, and Washington). The state variable was not included in the analysis for two reasons. First, the analysis was performed at the population-level and not individual-level. Second, the distribution of the states by region (North East, North Central, South, and West) looked to be uniform.

The module is comprised of 10 questions, of which the first 6 make up the Kessler Psychological Distress Scale (K6; Kessler, 2003). The K6 scale measures psychological distress; it does not cover questions related to major mental health issues, such as schizophrenia, bipolar disorder, etc. However documented studies have shown that the K6 scale is a strong predictor of serious mental illness, more specifically mood and anxiety disorders (Kessler et al., 2003; Furukawa et al., 2003). Kessler et al., 2013 and Furukawa et al., 2013 used a truncated World Health Organization (WHO) Composite International Diagnostic Interview Short-Form (CIDI-SF) scales designed to assess 30-day DSM-IV disorders and K10/K6 scales designed to assess psychological distress in order to evaluate the efficiency of the K10/K6 scales in screening for mood and anxiety disorders.

Forman-Hoffman et al, 2014 has found that adults with serious psychological distress (K6 score of 13 or greater) have more than double the mortality rate of adults with no serious psychological distress (SPD). The authors also found that even after controlling for sociodemographic characteristics, and chronic health conditions, SPD remained a significant risk factor for mortality (Forman-Hoffman et al, 2014).

In the original study, Kessler et al. 2012, the one factor structure was supported for the K6 scales, but Ko and Harrington, 2016 looked at the possibility of two factor model (anxiety and depression) and a second-order two factor model (psychological distress by depression and anxiety) besides the single-factor model (psychological distress).

The authors used K6 scale on the 2011 NSDUH survey to validate the scale on adult individuals with suicidal ideation. The sample included 2,098 adults with suicidal ideation (seriously thought about killing themselves). The authors found that while the single factor model possessed good fit (RMSEA = 0.099, CFI = 0.974, TLI=0.957) the two-factor model, and the second order two-factor model had excellent fit (RMSEA= 0.048, CFI = 0.994, TLI = 0.990 for both models). This research examined the two factor and the second order two factor structure found by Ko and Harrington, 2016 in a different population.

Structural Equation Modeling (SEM)

Structural Equation Modeling allows for examination of the relationships stated in a hypothesized model of relationships. In SEM, a latent variable is defined as a variable that cannot be measured directly but is believed to influence responses to one or more survey questions. In the context of this dissertation, psychological distress, depression and anxiety symptoms are considered latent variables.

SEM is a very general, very powerful multivariate technique. It uses a conceptual model and path diagram to illustrate complex and dynamic relationships between observed and unobserved variables. This research examines the relationship between the latent/unobserved variables (psychological distress, depression symptoms, and anxiety symptoms) derived via the first 6 questions in the mental illness and stigma module (K6).

SEM is fundamentally different from regression. In a regression model, there exists a clear distinction between dependent and independent variables. In SEM, a dependent variable in one model equation can become an independent variable in other components of the SEM system. SEM models include both *endogenous* and *exogenous* variables and assume that each of the observed variables being analyzed is measured on an interval or ratio scale (most commonly used are the Likert type scales).

Assumptions and Limitations of SEM

Since SEM is a multivariate statistical method, it requires multivariate normality. The data should be examined for univariate and multivariate outliers. Another assumption has to do with the sample size. Because of the complex relationships between the variables, the sample size has to me more than 200 observations or at least 50 more than 8 times the numbers of variables in the model. Usually a larger sample is desired. In our analysis the sample size is 192 which is close to this assumption. In addition, the residual covariances (the difference between proposed model covariance and estimated model covariance) needs to be small and centered about zero.

In order to construct an SEM model a researcher has to know the number of parameters to be estimated including path coefficients, covariances, and variances. All the relationships in the model have to be specified before the model can be analyzed.

Another limitation of the SEM is that it looks only at linear relationships between variables (the linear relationship can be explored by examining bivariate plots of the variables of interest).

What does EFA do? The primary objective of EFA is to determine the number and nature of the common factors (unobservable latent variables that share common variance) pattern of their influences on the surface attributes (Tucker & MacCallum, 1997; Fabrigar et al., 1999). EFA is appropriate in cases where there are obtained measures on a number of variables and where the objective is to identify the number and nature of the underlying factors/dimensions that can reproduce the variability in the data. By performing EFA on a set of questions it is possible to determine the number of constructs that are underlying by this questionnaire.

A step-by-step view of an EFA analysis:

1) Choose an extraction method: multiple solutions will be run to find the best factor structure for the data. There are many extraction methods available, but the most common used ones are the Maximum Likelihood (ML), Principal Axis Factor (PAF), and Principal Components (PC). ML analyzes the maximum likelihood of sampling the observed correlation matrix (Tabachinick & Fidell, 2007). This method is used when the data is assumed to exhibit the multivariate normality. With the PAF method all variables belong to the first group and when the factor is extracted, a residual matrix is calculated. The process is repeated (the second factor next most variance, etc.) until a predefined maximum number of iterations are performed and there is a large enough of variance accounted for in the correlation matrix (Tucker & McCallum, 2007). This method is used when the data violates the assumption of

multivariate normality. PC analysis is used to extract the maximum variance from the dataset with each component reducing the number of variables into smaller number of components (Tabachnick & Fidell, 2007).

2) Determine the number of factors to extract: examine the scree plot of eigenvalues (Kaiser) by number of factors extracted (amount of variance associated with each component); examine parallel axis analysis (compares a scree plot of factors of the observed data with another scree plot of factors from a correlation matrix of randomly generated, uncorrelated variables with the same dimensions as the original dataset (DiStefano & Dombrowski, 2006; Horn, 1965)); analyze the very simple structure (each item should associate strongly with only one factor and low loadings on the other factors; Gorsuch, 1983).

3) Rotation: the rotation method redistributes the extracted variance to make the solution interpretation much easier. The goal of the rotation is to attain a simple structure in which each variable loads on as few factors as possible and maximizes the number of high loading on each variable (Rummel, 1970). There are 2 families of rotation methods: orthogonal (uncorrelated factors) or oblique (correlated factors). In the orthogonal rotation the factors are rotated 90 degrees from each other, and it assumed that the factors are uncorrelated (DeCoster, 1998; Rummel, 1970). There are two orthogonal rotation techniques: Quartimax minimizes the number of factors needed to explain each variable (Gorsuch, 1983), and Varimax minimizes the number of variables that have high loadings on each factor and makes the small loadings even smaller (Yong & Pearce, 2013). In the oblique rotation it is assumed that the factors are correlated. The oblique rotation is more complex, since it involves one or two

coordinate systems: primary axes and reference axes (Rummel, 1970). There are two oblique rotation techniques: Direct Oblimin simplifies the structure of the output and Promax results in greater correlations among the factors and achieves a simple structure (Gorsuch, 1983); it is computationally faster and used with larger datasets.

4) Deciding the number of factors to retain:

When variables are factored the total number of factors equals the total number of variables. Many of these factors may not contribute substantially to the overall solution, as a result some of them might not be useful to retain in the analysis. Since the goal of the EFA is to reduce the number of factors that explain the most variance, it is important to extract the correct number of factors. Below there are couple of ways in which to decide how many factors to retain.

- Eigenvalue criterion (Kaiser criterion): retain any factor with values of higher than 1;
- scree plot analysis looks for a break between the factors with relatively large eigenvalues: the factors appearing before the break are retained for rotation;
- Proportion of variance accounted for: retain any factors that account for at least 5 or 10 percent of the common variance.

$$\text{Proportion} = \frac{\text{Eigenvalue for the factor of interest}}{\text{Total eigenvalues of the correlation matrix}}$$

- In the very simple structure analysis the cut off point for the loading factors is 0.40. In very simple structure analysis all loadings that are less than the maximum factor loading (of an item to a factor) are set to zero making the results easier to interpret.

- Parallel axis analysis is another method used to confirm the number of factors underlying the data.
- Residual matrix = Implied/hypothesized correlation matrix – Observed correlation matrix.

5) Explain the solution based on the results from steps 1-4.

<u>Confirmatory Factor Analysis (CFA)</u>

CFA is appropriate when researcher is trying to confirm the underlying structure of the data. For a CFA, the researcher has to have a wide knowledge of the current research done on the number of factors on the survey used to collect the data. When using CFA, the researcher has a strong understanding of the underlying theory and uses CFA to test the theory. The researcher knows or suspects the structure of the data, the number of factors, and which items are loaded on which factors.

<u>A step-by-step view of CFA analysis</u>:

- Model identification: cannot estimate more parameters compared to the existing pieces of information. A CFA model is identified if and only every single parameter has a unique solution. The key rule is that there need to be at least two indicators/parameters per latent variable and that their errors are uncorrelated (Kline, 2011). The total amount of information is the total number of elements in the variance – covariance matrix. It can be calculated as:

  Available Information = [# items * (#items +1)] / 2

  - If the number of parameters equals number of information available, then the model is just identified.

- If the number of parameters are less than number of information available, then the model is over-identified.
- If the number of parameters are greater than number of information available, then the model is under-identified.

If the model is under-identified, then the CFA cannot be run. The model as to be exactly identified or over-identified.

- Test a series of three alternative models: (a) one factor model where all the items will load on one factor; (b) two factor model based on the theory; and (c) three factor model to see if there are any additional factors underlying the data. The next step is to examine the fit for all the alternative models to assess which one is the best model.
- Once the best model that fits the data has been chosen, the next step is to interpret the estimates for that model:
  - Factor loadings are interpreted as regression coefficients. Increase of 1 unit in the unstandardized factor loading yields an increase of 1 unit in the factor.
  - A value of 0.40 for standardized factor loading is the minimum value for the item to be considered a significant contribution to the factor. A value higher than 0.70 is considered excellent factor loading (Comrey & Lee, 1992).
  - Standardized factor loadings squared are proportions of explained variance. For example, if the standardized factor loading for the item (λ) is 0.70 then $(0.70)^2 = 0.49$, then the factor explains 49% of the observed

variance for that item (Kline, 2011). The ideal explained variance for the CFA model is above 50%.

APPENDIX B

SAS AND R CODE

Data Cleaning SAS Code (only code for 2007 shown, same code run for each year)

```
RUN 2007 DATA*;
data CFA2007;
	set CDBRFS07;
	keep SEX _RACEGR2 _AGEG5YR INCOME2 EMPLOY _STATE
_SMOKER3 _RFBING4 _RFDRHV3 _DRNKDY3 _EDUCAG GENHLTH
HLTHPLAN DIABETE2 CVDINFR4 CVDCRHD4 CVDSTRK3 ASTHMA2
MISNERVS MISHOPLS MISRSTLS MISDEPRD MISEFFRT MISWTLES
MISNOWRK MISTMNT MISTRHLP MISPHLPF;
	run;
Proc contents data=CFA2007;
run;
Proc freq data=CFA2007;
run;
data CFA2007;
	set CFA2007;
	if _STATE=9 then STATE=9;
	if _STATE=23 then STATE=23;
	if _STATE=25 then STATE=25;
	if _STATE=33 then STATE=33;
	if _STATE=44 then STATE=44;
	if _STATE=50 then STATE=50;
	if _STATE=17 then STATE=17;
	if _STATE=18 then STATE=18;
	if _STATE=26 then STATE=26;
	if _STATE=39 then STATE=39;
	if _STATE=55 then STATE=55;
	if _STATE=19 then STATE=19;
	if _STATE=20 then STATE=20;
	if _STATE=27 then STATE=27;
	if _STATE=29 then STATE=29;
	if _STATE=31 then STATE=31;
	if _STATE=13 then STATE=13;
	if _STATE=45 then STATE=45;
	if _STATE=51 then STATE=51;
	if _STATE=11 then STATE=11;
```

```
    if _STATE=21 then STATE=21;
    if _STATE=28 then STATE=28;
    if _STATE=5 then STATE=5;
    if _STATE=22 then STATE=22;
    if _STATE=40 then STATE=40;
    if _STATE=48 then STATE=48;
    if _STATE=8 then STATE=8;
    if _STATE=30 then STATE=30;
    if _STATE=32 then STATE=32;
    if _STATE=35 then STATE=35;
    if _STATE=56 then STATE=56;
    if _STATE=2 then STATE=2;
    if _STATE=6 then STATE=6;
    if _STATE=15 then STATE=15;
    if _STATE=41 then STATE=41;
    if _STATE=53 then STATE=53;
    if _STATE=72 then STATE=72;
run;
data CFA2007;
    set CFA2007;
    if nmiss(STATE) = 0;
    run;
*create categorical variable for MISNOWORK;
data CFA2007;
    set CFA2007;
    if MISNOWRK = 88 then DAYMISS = 1; *none;
    if MISNOWRK = 1 then DAYMISS = 2; *1-5 or a week;
    if MISNOWRK = 2 then DAYMISS = 2; *1-5 or a week;
    if MISNOWRK = 3 then DAYMISS =2; *1-5 or a week;
    if MISNOWRK = 4 then DAYMISS = 2; *1-5 or a week;
    if MISNOWRK = 5 then DAYMISS = 2; *1-5 or a week;
    if MISNOWRK =6 then DAYMISS =3; *6-10 or 2 weeks;
    if MISNOWRK =7 then DAYMISS =3; *6-10 or 2 weeks;
    if MISNOWRK =8 then DAYMISS =3; *6-10 or 2 weeks;
    if MISNOWRK =9 then DAYMISS =3; *6-10 or 2 weeks;
    if MISNOWRK =10 then DAYMISS =3; *6-10 or 2 weeks;
    if MISNOWRK =11 then DAYMISS = 4; *11 or more;
    if MISNOWRK =12 then DAYMISS = 4; *11 or more;
    if MISNOWRK =13 then DAYMISS = 4;*11 or more;
    if MISNOWRK =14 then DAYMISS = 4;*11 or more;
    if MISNOWRK =15 then DAYMISS = 4;*11 or more;
    if MISNOWRK =16 then DAYMISS = 4;*11 or more;
    if MISNOWRK =17 then DAYMISS = 4;*11 or more;
    if MISNOWRK =18 then DAYMISS = 4;*11 or more;
    if MISNOWRK =19 then DAYMISS = 4;*11 or more;
    if MISNOWRK =20 then DAYMISS = 4;*11 or more;
```

```
		if MISNOWRK =21 then DAYMISS = 4;*11 or more;
		if MISNOWRK =22 then DAYMISS = 4;*11 or more;
		if MISNOWRK =23 then DAYMISS = 4;*11 or more;
		if MISNOWRK =24 then DAYMISS = 4;*11 or more;
		if MISNOWRK =25 then DAYMISS = 4;*11 or more;
		if MISNOWRK =26 then DAYMISS = 4;*11 or more;
		if MISNOWRK =27 then DAYMISS = 4;*11 or more;
		if MISNOWRK =28 then DAYMISS = 4;*11 or more;
		if MISNOWRK =29 then DAYMISS = 4;*11 or more;
		if MISNOWRK =30 then DAYMISS = 4;*11 or more;
		else if MISNOWRK = 77 then DAYMISS = .;
		else if MISNOWRK = 99 then DAYMISS = .;
run;
*create new variable for race*;
data CFA2007;
		set CFA2007;
		if _RACEGR2 = 1 then RACE = 1; *white;
		if _RACEGR2 = 2 then RACE = 2; *non-white;
		if _RACEGR2 = 3 then RACE = 2;
		If _RACEGR2 = 4 then RACE = 2;
		if _RACEGR2 = 5 then RACE = 2;
		else if _RACEGR2 = 9 then RACE = .;
Run;
*Create new variabe for age*;
data CFA2007;
		set CFA2007;
		if _AGEG5YR = 1 then AGE = 1; *18-29;
		if _AGEG5YR = 2 then AGE = 1;
		if _AGEG5YR = 3 then AGE = 2; *30-44;
		if _AGEG5YR = 4 then AGE = 2;
		if _AGEG5YR = 5 then AGE = 2;
		if _AGEG5YR = 6 then AGE = 3; *45-64;
		if _AGEG5YR = 7 then AGE = 3;
		if _AGEG5YR = 8 then AGE = 3;
		if _AGEG5YR = 9 then AGE = 3;
		if _AGEG5YR = 10 then AGE = 4; *65+;
		if _AGEG5YR = 11 then AGE = 4;
		if _AGEG5YR = 12 then AGE = 4;
		if _AGEG5YR = 13 then AGE = 4;
		if _AGEG5YR = 14 then AGE = .;
Run;
*Create new variable for income*;
data CFA2007;
		set CFA2007;
		if INCOME2= 1 then INCOME = 1;  *0-25k;
		if INCOME2= 2 then INCOME = 1;
```

```
    if INCOME2= 3 then INCOME = 1;
    if INCOME2= 4 then INCOME = 1;
    if INCOME2= 5 then INCOME = 2; *25-50;
    if INCOME2= 6 then INCOME = 2;
    if INCOME2= 7 then INCOME = 3; *50+;
    if INCOME2= 8 then INCOME = 3;
    else if INCOME2= 77 then INCOME = .;
    else if INCOME2= 99 then INCOME = .;
Run;
*create variable for region;
    data CFA2007;
        set CFA2007;
    if _STATE=9 then REGION=1; *North East;
    if _STATE=23 then REGION=1;
    if _STATE=25 then REGION=1;
    if _STATE=33 then REGION=1;
    if _STATE=44 then REGION=1;
    if _STATE=50 then REGION=1;
    if _STATE=34 then REGION=1;
    if _STATE=36 then REGION=1;
    if _STATE=42 then REGION=1;
    if _STATE=17 then REGION=2; *north central;
    if _STATE=18 then REGION=2;
    if _STATE=26 then REGION=2;
    if _STATE=39 then REGION=2;
    if _STATE=55 then REGION=2;
    if _STATE=19 then REGION=2;
    if _STATE=20 then REGION=2;
    if _STATE=27 then REGION=2;
    if _STATE=29 then REGION=2;
    if _STATE=31 then REGION=2;
    if _STATE=38 then REGION=2;
    if _STATE=46 then REGION=2;
    if _STATE=10 then REGION=3;*south;
    if _STATE=12 then REGION=3;
    if _STATE=13 then REGION=3;
    if _STATE=24 then REGION=3;
    if _STATE=37 then REGION=3;
    if _STATE=45 then REGION=3;
    if _STATE=51 then REGION=3;
    if _STATE=11 then REGION=3;
    if _STATE=54 then REGION=3;
    if _STATE=1 then REGION=3;
    if _STATE=21 then REGION=3;
    if _STATE=28 then REGION=3;
    if _STATE=47 then REGION=3;
```

```
	if _STATE=5 then REGION=3;
	if _STATE=22 then REGION=3;
	if _STATE=40 then REGION=3;
	if _STATE=48 then REGION=3;
	if _STATE=4 then REGION=4;*west;
	if _STATE=8 then REGION=4;
	if _STATE=16 then REGION=4;
	if _STATE=30 then REGION=4;
	if _STATE=32 then REGION=4;
	if _STATE=35 then REGION=4;
	if _STATE=49 then REGION=4;
	if _STATE=56 then REGION=4;
	if _STATE=2 then REGION=4;
	if _STATE=6 then REGION=4;
	if _STATE=15 then REGION=4;
	if _STATE=41 then REGION=4;
	if _STATE=53 then REGION=4;
	if _STATE=72 then REGION=1;
run;
*alcohol consumption;
data CFA2007;
	set CFA2007;
	if _DRNKDY3 = 0 then DRINK=1; *none;
	if _DRNKDY3 = 1 and SEX=2 then DRINK=2; *moderate;
	if _DRNKDY3 = 1 and SEX=1 then DRINK=2;
	if _DRNKDY3 = 2 and SEX=1 then DRINK=2;
	if _RFBING4 = 2 then DRINK=3;*binge;
	if _RFDRHV3 = 2 then DRINK=4;*heavy;
	else if _DRNKDY3= 990 then DRINK=.;
	run;
*employment status;
data CFA2007;
	set CFA2007;
	if EMPLOY=1 then EMPLOYED=1;*yes;
	if EMPLOY=2 then EMPLOYED=1;
	if EMPLOY=3 then EMPLOYED=2;
	if EMPLOY=4 then EMPLOYED=2;*no;
	if EMPLOY=5 then EMPLOYED=2;
	if EMPLOY=6 then EMPLOYED=2;
	if EMPLOY=7 then EMPLOYED=2;
	if EMPLOY=8 then EMPLOYED=2;
	else if EMPLOY=9 then EMPLOYED=.;
run;
*create smoking variables;
data cfa2007;
	set CFA2007;
```

```
	if _SMOKER3=1 then SMOKE=4;*current smoker - every day;
	if _SMOKER3=2 then SMOKE=3;*current smoker-some days;
	if _SMOKER3=3 then SMOKE=2;*former smoker;
	if _SMOKER3=4 then SMOKE=1;*never smoked;
	if _SMOKER3=9 then SMOKE=.;
run;

*chronic diseases;
data CFA2007;
	set CFA2007;
	if DIABETE2=1 then Chronic=1;*yes chronic disease;
	if DIABETE2=2 then Chronic=2;*no chronic disease;
	if DIABETE2=3 then Chronic=2;
	if DIABETE2=4 then Chronic=2;
	if CVDINFR4=1 then Chronic=1;
	if CVDINFR4=2 then Chronic=2;
	if CVDCRHD4=1 then Chronic=1;
	if CVDCRHD4=2 then Chronic=2;
	if CVDSTRK3=1 then Chronic=1;
	if CVDSTRK3=2 then Chronic=2;
	if DIABETE2=7 and DIABETE2=9 then Chronic=.;
	if CVDINFR4=7 and CVDINFR4=9 then Chronic=.;
	if CVDCRHD4=7 and CVDCRHD4=9 then Chronic=.;
	if CVDSTRK3=7 and CVDSTRK3=9 then Chronic=.;
run;
*recode v1-v6*;
data CFA2007;
	set CFA2007;
	if MISNERVS = 1 then NERVS07 = 5;
	if MISNERVS = 2 then NERVS07 = 4;
	if MISNERVS = 3 then NERVS07 = 3;
	if MISNERVS = 4 then NERVS07 = 2;
	if MISNERVS = 5 then NERVS07 = 1;
	if MISNERVS = 7 then NERVS07 = .;
	if MISNERVS = 9 then NERVS07 = .;

	if MISHOPLS = 1 then HOPLS07 = 5;
	if MISHOPLS = 2 then HOPLS07 = 4;
	if MISHOPLS = 3 then HOPLS07 = 3;
	if MISHOPLS = 4 then HOPLS07 = 2;
	if MISHOPLS = 5 then HOPLS07 = 1;
	if MISHOPLS = 7 then HOPLS07 = .;
	if MISHOPLS = 9 then HOPLS07 = .;

	if MISRSTLS = 1 then RSTLS07 = 5;
	if MISRSTLS = 2 then RSTLS07 = 4;
```

```
    if MISRSTLS = 3 then RSTLS07 = 3;
    if MISRSTLS = 4 then RSTLS07 = 2;
    if MISRSTLS = 5 then RSTLS07 = 1;
    if MISRSTLS = 7 then RSTLS07 = .;
    if MISRSTLS = 9 then RSTLS07 = .;

    if MISDEPRD = 1 then DEPRD07 = 5;
    if MISDEPRD = 2 then DEPRD07 = 4;
    if MISDEPRD = 3 then DEPRD07 = 3;
    if MISDEPRD = 4 then DEPRD07 = 2;
    if MISDEPRD = 5 then DEPRD07 = 1;
    if MISDEPRD = 7 then DEPRD07 = .;
    if MISDEPRD = 9 then DEPRD07 = .;

    if MISEFFRT = 1 then EFFRT07 = 5;
    if MISEFFRT = 2 then EFFRT07 = 4;
    if MISEFFRT = 3 then EFFRT07 = 3;
    if MISEFFRT = 4 then EFFRT07 = 2;
    if MISEFFRT = 5 then EFFRT07 = 1;
    if MISEFFRT = 7 then EFFRT07 = .;
    if MISEFFRT = 9 then EFFRT07 = .;

    if MISWTLES = 1 then WORTH07 = 5;
    if MISWTLES = 2 then WORTH07 = 4;
    if MISWTLES = 3 then WORTH07 = 3;
    if MISWTLES = 4 then WORTH07 = 2;
    if MISWTLES = 5 then WORTH07 = 1;
    if MISWTLES = 7 then WORTH07 = .;
    if MISWTLES = 9 then WORTH07 = .;
run;

*create composite observations;
data CFA2007;
    set CFA2007;
    if sex=1 and race=1 and age=1 and income=1 and _EDUCAG=1 then obs=1;
    if sex=1 and race=1 and age=1 and income=1 and _EDUCAG=2 then obs=2;
    if sex=1 and race=1 and age=1 and income=1 and _EDUCAG=3 then obs=3;
    if sex=1 and race=1 and age=1 and income=1 and _EDUCAG=4 then obs=4;
    if sex=1 and race=1 and age=1 and income=2 and _EDUCAG=1 then obs=5;
    if sex=1 and race=1 and age=1 and income=2 and _EDUCAG=2 then obs=6;
    if sex=1 and race=1 and age=1 and income=2 and _EDUCAG=3 then obs=7;
    if sex=1 and race=1 and age=1 and income=2 and _EDUCAG=4 then obs=8;
    if sex=1 and race=1 and age=1 and income=3 and _EDUCAG=1 then obs=9;
    if sex=1 and race=1 and age=1 and income=3 and _EDUCAG=2 then obs=10;
    if sex=1 and race=1 and age=1 and income=3 and _EDUCAG=3 then obs=11;
    if sex=1 and race=1 and age=1 and income=3 and _EDUCAG=4 then obs=12;
```

```
if sex=1 and race=1 and age=2 and income=1 and _EDUCAG=1 then obs=13;
if sex=1 and race=1 and age=2 and income=1 and _EDUCAG=2 then obs=14;
if sex=1 and race=1 and age=2 and income=1 and _EDUCAG=3 then obs=15;
if sex=1 and race=1 and age=2 and income=1 and _EDUCAG=4 then obs=16;
if sex=1 and race=1 and age=2 and income=2 and _EDUCAG=1 then obs=17;
if sex=1 and race=1 and age=2 and income=2 and _EDUCAG=2 then obs=18;
if sex=1 and race=1 and age=2 and income=2 and _EDUCAG=3 then obs=19;
if sex=1 and race=1 and age=2 and income=2 and _EDUCAG=4 then obs=20;
if sex=1 and race=1 and age=2 and income=3 and _EDUCAG=1 then obs=21;
if sex=1 and race=1 and age=2 and income=3 and _EDUCAG=2 then obs=22;
if sex=1 and race=1 and age=2 and income=3 and _EDUCAG=3 then obs=23;
if sex=1 and race=1 and age=2 and income=3 and _EDUCAG=4 then obs=24;
if sex=1 and race=1 and age=3 and income=1 and _EDUCAG=1 then obs=25;
if sex=1 and race=1 and age=3 and income=1 and _EDUCAG=2 then obs=26;
if sex=1 and race=1 and age=3 and income=1 and _EDUCAG=3 then obs=27;
if sex=1 and race=1 and age=3 and income=1 and _EDUCAG=4 then obs=28;
if sex=1 and race=1 and age=3 and income=2 and _EDUCAG=1 then obs=29;
if sex=1 and race=1 and age=3 and income=2 and _EDUCAG=2 then obs=30;
if sex=1 and race=1 and age=3 and income=2 and _EDUCAG=3 then obs=31;
if sex=1 and race=1 and age=3 and income=2 and _EDUCAG=4 then obs=32;
if sex=1 and race=1 and age=3 and income=3 and _EDUCAG=1 then obs=33;
if sex=1 and race=1 and age=3 and income=3 and _EDUCAG=2 then obs=34;
if sex=1 and race=1 and age=3 and income=3 and _EDUCAG=3 then obs=35;
if sex=1 and race=1 and age=3 and income=3 and _EDUCAG=4 then obs=36;
if sex=1 and race=1 and age=4 and income=1 and _EDUCAG=1 then obs=37;
if sex=1 and race=1 and age=4 and income=1 and _EDUCAG=2 then obs=38;
if sex=1 and race=1 and age=4 and income=1 and _EDUCAG=3 then obs=39;
if sex=1 and race=1 and age=4 and income=1 and _EDUCAG=4 then obs=40;
if sex=1 and race=1 and age=4 and income=2 and _EDUCAG=1 then obs=41;
if sex=1 and race=1 and age=4 and income=2 and _EDUCAG=2 then obs=42;
if sex=1 and race=1 and age=4 and income=2 and _EDUCAG=3 then obs=43;
if sex=1 and race=1 and age=4 and income=2 and _EDUCAG=4 then obs=44;
if sex=1 and race=1 and age=4 and income=3 and _EDUCAG=1 then obs=45;
if sex=1 and race=1 and age=4 and income=3 and _EDUCAG=2 then obs=46;
if sex=1 and race=1 and age=4 and income=3 and _EDUCAG=3 then obs=47;
if sex=1 and race=1 and age=4 and income=3 and _EDUCAG=4 then obs=48;
if sex=1 and race=2 and age=1 and income=1 and _EDUCAG=1 then obs=49;
if sex=1 and race=2 and age=1 and income=1 and _EDUCAG=2 then obs=50;
if sex=1 and race=2 and age=1 and income=1 and _EDUCAG=3 then obs=51;
if sex=1 and race=2 and age=1 and income=1 and _EDUCAG=4 then obs=52;
if sex=1 and race=2 and age=1 and income=2 and _EDUCAG=1 then obs=53;
if sex=1 and race=2 and age=1 and income=2 and _EDUCAG=2 then obs=54;
if sex=1 and race=2 and age=1 and income=2 and _EDUCAG=3 then obs=55;
if sex=1 and race=2 and age=1 and income=2 and _EDUCAG=4 then obs=56;
if sex=1 and race=2 and age=1 and income=3 and _EDUCAG=1 then obs=57;
if sex=1 and race=2 and age=1 and income=3 and _EDUCAG=2 then obs=58;
```

if sex=**1** and race=**2** and age=**1** and income=**3** and _EDUCAG=**3** then obs=**59**;
if sex=**1** and race=**2** and age=**1** and income=**3** and _EDUCAG=**4** then obs=**60**;
if sex=**1** and race=**2** and age=**2** and income=**1** and _EDUCAG=**1** then obs=**61**;
if sex=**1** and race=**2** and age=**2** and income=**1** and _EDUCAG=**2** then obs=**62**;
if sex=**1** and race=**2** and age=**2** and income=**1** and _EDUCAG=**3** then obs=**63**;
if sex=**1** and race=**2** and age=**2** and income=**1** and _EDUCAG=**4** then obs=**64**;
if sex=**1** and race=**2** and age=**2** and income=**2** and _EDUCAG=**1** then obs=**65**;
if sex=**1** and race=**2** and age=**2** and income=**2** and _EDUCAG=**2** then obs=**66**;
if sex=**1** and race=**2** and age=**2** and income=**2** and _EDUCAG=**3** then obs=**67**;
if sex=**1** and race=**2** and age=**2** and income=**2** and _EDUCAG=**4** then obs=**68**;
if sex=**1** and race=**2** and age=**2** and income=**3** and _EDUCAG=**1** then obs=**69**;
if sex=**1** and race=**2** and age=**2** and income=**3** and _EDUCAG=**2** then obs=**70**;
if sex=**1** and race=**2** and age=**2** and income=**3** and _EDUCAG=**3** then obs=**71**;
if sex=**1** and race=**2** and age=**2** and income=**3** and _EDUCAG=**4** then obs=**72**;
if sex=**1** and race=**2** and age=**3** and income=**1** and _EDUCAG=**1** then obs=**73**;
if sex=**1** and race=**2** and age=**3** and income=**1** and _EDUCAG=**2** then obs=**74**;
if sex=**1** and race=**2** and age=**3** and income=**1** and _EDUCAG=**3** then obs=**75**;
if sex=**1** and race=**2** and age=**3** and income=**1** and _EDUCAG=**4** then obs=**76**;
if sex=**1** and race=**2** and age=**3** and income=**2** and _EDUCAG=**1** then obs=**77**;
if sex=**1** and race=**2** and age=**3** and income=**2** and _EDUCAG=**2** then obs=**78**;
if sex=**1** and race=**2** and age=**3** and income=**2** and _EDUCAG=**3** then obs=**79**;
if sex=**1** and race=**2** and age=**3** and income=**2** and _EDUCAG=**4** then obs=**80**;
if sex=**1** and race=**2** and age=**3** and income=**3** and _EDUCAG=**1** then obs=**81**;
if sex=**1** and race=**2** and age=**3** and income=**3** and _EDUCAG=**2** then obs=**82**;
if sex=**1** and race=**2** and age=**3** and income=**3** and _EDUCAG=**3** then obs=**83**;
if sex=**1** and race=**2** and age=**3** and income=**3** and _EDUCAG=**4** then obs=**84**;
if sex=**1** and race=**2** and age=**4** and income=**1** and _EDUCAG=**1** then obs=**85**;
if sex=**1** and race=**2** and age=**4** and income=**1** and _EDUCAG=**2** then obs=**86**;
if sex=**1** and race=**2** and age=**4** and income=**1** and _EDUCAG=**3** then obs=**87**;
if sex=**1** and race=**2** and age=**4** and income=**1** and _EDUCAG=**4** then obs=**88**;
if sex=**1** and race=**2** and age=**4** and income=**2** and _EDUCAG=**1** then obs=**89**;
if sex=**1** and race=**2** and age=**4** and income=**2** and _EDUCAG=**2** then obs=**90**;
if sex=**1** and race=**2** and age=**4** and income=**2** and _EDUCAG=**3** then obs=**91**;
if sex=**1** and race=**2** and age=**4** and income=**2** and _EDUCAG=**4** then obs=**92**;
if sex=**1** and race=**2** and age=**4** and income=**3** and _EDUCAG=**1** then obs=**93**;
if sex=**1** and race=**2** and age=**4** and income=**3** and _EDUCAG=**2** then obs=**94**;
if sex=**1** and race=**2** and age=**4** and income=**3** and _EDUCAG=**3** then obs=**95**;
if sex=**1** and race=**2** and age=**4** and income=**3** and _EDUCAG=**4** then obs=**96**;

if sex=**2** and race=**1** and age=**1** and income=**1** and _EDUCAG=**1** then obs=**97**;
if sex=**2** and race=**1** and age=**1** and income=**1** and _EDUCAG=**2** then obs=**98**;
if sex=**2** and race=**1** and age=**1** and income=**1** and _EDUCAG=**3** then obs=**99**;
if sex=**2** and race=**1** and age=**1** and income=**1** and _EDUCAG=**4** then obs=**100**;
if sex=**2** and race=**1** and age=**1** and income=**2** and _EDUCAG=**1** then obs=**101**;
if sex=**2** and race=**1** and age=**1** and income=**2** and _EDUCAG=**2** then obs=**102**;
if sex=**2** and race=**1** and age=**1** and income=**2** and _EDUCAG=**3** then obs=**103**;

```
if sex=2 and race=1 and age=1 and income=2 and _EDUCAG=4 then obs=104;
if sex=2 and race=1 and age=1 and income=3 and _EDUCAG=1 then obs=105;
if sex=2 and race=1 and age=1 and income=3 and _EDUCAG=2 then obs=106;
if sex=2 and race=1 and age=1 and income=3 and _EDUCAG=3 then obs=107;
if sex=2 and race=1 and age=1 and income=3 and _EDUCAG=4 then obs=108;
if sex=2 and race=1 and age=2 and income=1 and _EDUCAG=1 then obs=109;
if sex=2 and race=1 and age=2 and income=1 and _EDUCAG=2 then obs=110;
if sex=2 and race=1 and age=2 and income=1 and _EDUCAG=3 then obs=111;
if sex=2 and race=1 and age=2 and income=1 and _EDUCAG=4 then obs=112;
if sex=2 and race=1 and age=2 and income=2 and _EDUCAG=1 then obs=113;
if sex=2 and race=1 and age=2 and income=2 and _EDUCAG=2 then obs=114;
if sex=2 and race=1 and age=2 and income=2 and _EDUCAG=3 then obs=115;
if sex=2 and race=1 and age=2 and income=2 and _EDUCAG=4 then obs=116;
if sex=2 and race=1 and age=2 and income=3 and _EDUCAG=1 then obs=117;
if sex=2 and race=1 and age=2 and income=3 and _EDUCAG=2 then obs=118;
if sex=2 and race=1 and age=2 and income=3 and _EDUCAG=3 then obs=119;
if sex=2 and race=1 and age=2 and income=3 and _EDUCAG=4 then obs=120;
if sex=2 and race=1 and age=3 and income=1 and _EDUCAG=1 then obs=121;
if sex=2 and race=1 and age=3 and income=1 and _EDUCAG=2 then obs=122;
if sex=2 and race=1 and age=3 and income=1 and _EDUCAG=3 then obs=123;
if sex=2 and race=1 and age=3 and income=1 and _EDUCAG=4 then obs=124;
if sex=2 and race=1 and age=3 and income=2 and _EDUCAG=1 then obs=125;
if sex=2 and race=1 and age=3 and income=2 and _EDUCAG=2 then obs=126;
if sex=2 and race=1 and age=3 and income=2 and _EDUCAG=3 then obs=127;
if sex=2 and race=1 and age=3 and income=2 and _EDUCAG=4 then obs=128;
if sex=2 and race=1 and age=3 and income=3 and _EDUCAG=1 then obs=129;
if sex=2 and race=1 and age=3 and income=3 and _EDUCAG=2 then obs=130;
if sex=2 and race=1 and age=3 and income=3 and _EDUCAG=3 then obs=131;
if sex=2 and race=1 and age=3 and income=3 and _EDUCAG=4 then obs=132;
if sex=2 and race=1 and age=4 and income=1 and _EDUCAG=1 then obs=133;
if sex=2 and race=1 and age=4 and income=1 and _EDUCAG=2 then obs=134;
if sex=2 and race=1 and age=4 and income=1 and _EDUCAG=3 then obs=135;
if sex=2 and race=1 and age=4 and income=1 and _EDUCAG=4 then obs=136;
if sex=2 and race=1 and age=4 and income=2 and _EDUCAG=1 then obs=137;
if sex=2 and race=1 and age=4 and income=2 and _EDUCAG=2 then obs=138;
if sex=2 and race=1 and age=4 and income=2 and _EDUCAG=3 then obs=139;
if sex=2 and race=1 and age=4 and income=2 and _EDUCAG=4 then obs=140;
if sex=2 and race=1 and age=4 and income=3 and _EDUCAG=1 then obs=141;
if sex=2 and race=1 and age=4 and income=3 and _EDUCAG=2 then obs=142;
if sex=2 and race=1 and age=4 and income=3 and _EDUCAG=3 then obs=143;
if sex=2 and race=1 and age=4 and income=3 and _EDUCAG=4 then obs=144;
if sex=2 and race=2 and age=1 and income=1 and _EDUCAG=1 then obs=145;
if sex=2 and race=2 and age=1 and income=1 and _EDUCAG=2 then obs=146;
if sex=2 and race=2 and age=1 and income=1 and _EDUCAG=3 then obs=147;
if sex=2 and race=2 and age=1 and income=1 and _EDUCAG=4 then obs=148;
if sex=2 and race=2 and age=1 and income=2 and _EDUCAG=1 then obs=149;
```

```
if sex=2 and race=2 and age=1 and income=2 and _EDUCAG=2 then obs=150;
if sex=2 and race=2 and age=1 and income=2 and _EDUCAG=3 then obs=151;
if sex=2 and race=2 and age=1 and income=2 and _EDUCAG=4 then obs=152;
if sex=2 and race=2 and age=1 and income=3 and _EDUCAG=1 then obs=153;
if sex=2 and race=2 and age=1 and income=3 and _EDUCAG=2 then obs=154;
if sex=2 and race=2 and age=1 and income=3 and _EDUCAG=3 then obs=155;
if sex=2 and race=2 and age=1 and income=3 and _EDUCAG=4 then obs=156;
if sex=2 and race=2 and age=2 and income=1 and _EDUCAG=1 then obs=157;
if sex=2 and race=2 and age=2 and income=1 and _EDUCAG=2 then obs=158;
if sex=2 and race=2 and age=2 and income=1 and _EDUCAG=3 then obs=159;
if sex=2 and race=2 and age=2 and income=1 and _EDUCAG=4 then obs=160;
if sex=2 and race=2 and age=2 and income=2 and _EDUCAG=1 then obs=161;
if sex=2 and race=2 and age=2 and income=2 and _EDUCAG=2 then obs=162;
if sex=2 and race=2 and age=2 and income=2 and _EDUCAG=3 then obs=163;
if sex=2 and race=2 and age=2 and income=2 and _EDUCAG=4 then obs=164;
if sex=2 and race=2 and age=2 and income=3 and _EDUCAG=1 then obs=165;
if sex=2 and race=2 and age=2 and income=3 and _EDUCAG=2 then obs=166;
if sex=2 and race=2 and age=2 and income=3 and _EDUCAG=3 then obs=167;
if sex=2 and race=2 and age=2 and income=3 and _EDUCAG=4 then obs=168;
if sex=2 and race=2 and age=3 and income=1 and _EDUCAG=1 then obs=169;
if sex=2 and race=2 and age=3 and income=1 and _EDUCAG=2 then obs=170;
if sex=2 and race=2 and age=3 and income=1 and _EDUCAG=3 then obs=171;
if sex=2 and race=2 and age=3 and income=1 and _EDUCAG=4 then obs=172;
if sex=2 and race=2 and age=3 and income=2 and _EDUCAG=1 then obs=173;
if sex=2 and race=2 and age=3 and income=2 and _EDUCAG=2 then obs=174;
if sex=2 and race=2 and age=3 and income=2 and _EDUCAG=3 then obs=175;
if sex=2 and race=2 and age=3 and income=2 and _EDUCAG=4 then obs=176;
if sex=2 and race=2 and age=3 and income=3 and _EDUCAG=1 then obs=177;
if sex=2 and race=2 and age=3 and income=3 and _EDUCAG=2 then obs=178;
if sex=2 and race=2 and age=3 and income=3 and _EDUCAG=3 then obs=179;
if sex=2 and race=2 and age=3 and income=3 and _EDUCAG=4 then obs=180;
if sex=2 and race=2 and age=4 and income=1 and _EDUCAG=1 then obs=181;
if sex=2 and race=2 and age=4 and income=1 and _EDUCAG=2 then obs=182;
if sex=2 and race=2 and age=4 and income=1 and _EDUCAG=3 then obs=183;
if sex=2 and race=2 and age=4 and income=1 and _EDUCAG=4 then obs=184;
if sex=2 and race=2 and age=4 and income=2 and _EDUCAG=1 then obs=185;
if sex=2 and race=2 and age=4 and income=2 and _EDUCAG=2 then obs=186;
if sex=2 and race=2 and age=4 and income=2 and _EDUCAG=3 then obs=187;
if sex=2 and race=2 and age=4 and income=2 and _EDUCAG=4 then obs=188;
if sex=2 and race=2 and age=4 and income=3 and _EDUCAG=1 then obs=189;
if sex=2 and race=2 and age=4 and income=3 and _EDUCAG=2 then obs=190;
if sex=2 and race=2 and age=4 and income=3 and _EDUCAG=3 then obs=191;
if sex=2 and race=2 and age=4 and income=3 and _EDUCAG=4 then obs=192;
        Run;
*recode the non-responses and unknown data 7 and 9s;
data CFA2007;
```

```
	set CFA2007;
if HLTHPLAN = 1 then HLTHPLN1=1;
if HLTHPLAN = 2 then HLTHPLN1=2;
if HLTHPLAN = 7 then HLTHPLN1= .;
if HLTHPLAN = 9 then HLTHPLN1= .;
if _EDUCAG = 9 then _EDUCAG = .;
if MISTMNT = 7 then MISTMNT= .;
if MISTMNT = 9 then MISTMNT= .;
if GENHLTH = 7 then GENHLTH= .;
if GENHLTH = 9 then GENHLTH= .;
if MISTRHLP = 7 then MISTRHLP= .;
if MISTRHLP = 9 then MISTRHLP= .;
if MISPHLPF = 7 then MISPHLPF= .;
if MISPHLPF = 9 then MISPHLPF= .;
run;
data working2007;
	set CFA2007;
	keep obs sex race age income _EDUCAG HLTHPLN1 DAYMISS NERVS07
HOPLS07 RSTLS07 DEPRD07 EFFRT07 WORTH07 SMOKE GENHLTH DRINK
MISTMNT EMPLOYED CHRONIC REGION MISTRHLP MISPHLPF;
	run;
data working2007;
	set working2007;
	if nmiss (obs,sex, race, age, income, HLTHPLN1, DAYMISS, NERVS07,
HOPLS07, RSTLS07, DEPRD07, EFFRT07, WORTH07, SMOKE, GENHLTH,
DRINK, MISTMNT, EMPLOYED, CHRONIC, REGION, MISTRHLP, MISPHLPF) =
0;
	run;
```

AIM 1 CODE

EFA SAS CODE

```
*run descriptives;
proc univariate data=working2013_R1 NORMAL PLOT;
var NERVS13 HOPLS13 RSTLS13 DEPRD13 EFFRT13 WORTH13;
run;
* FINAL EFA CODE;
*run EFA*;
*printit - principal axis factor analysis;
proc factor data=Working2013_R
    simple
    method=prinit
    priors=smc
    nfact=2
    corr
```

```
scree
rotate=promax
reorder
round
flag=.40
out=d2;
var NERVS13 HOPLS13 RSTLS13 DEPRD13 EFFRT13 WORTH13;
run;
```

CFA CODE

```
* RUN CFA
*run covariance matrix;
proc corr data=CFA2012 COV OUTP=corr_cov nomiss noprob;
    var NERVS12 HOPLS12 RSTLS12 DEPRD12 EFFRT12 WORTH12 ;
    run;
*run CFA*;
data Path2012 (Type=COV);
        input _Type_$ _NAME_ $ V1-V6;
        label
            V1 = 'NERVS12'
            V2 = 'HOPLS12'
            V3 = 'RSTLS12'
            V4=  'DEPRD12'
            V5 = 'EFFRT12'
            V6 = 'WORTH12';
        cards;
 n      .       63837   63837   63837   63837   63837   63837
 COV    V1      0.8766  0.3409  0.4801  0.2827  0.4148  0.2733
 COV    V2      0.3409  0.5447  0.3169  0.3256  0.3804  0.3304
 COV    V3      0.4801  0.3169  0.9499  0.2809  0.4283  0.2691
 COV    V4      0.2827  0.3256  0.2809  0.4454  0.3393  0.2874
 COV    V5      0.4148  0.3804  0.4283  0.3393  1.0137  0.3561
 COV    V6      0.2733  0.3304  0.2691  0.2874  0.3561  0.4782

    ;
run;
```

AIM 2 R-CODE

Measurement Invariance

```
invariance<- '
```

```
pd07 =~ nervs07nmean + effrt07nmean + hopls07nmean + rstls07nmean +
deprd07nmean + worth07nmean
pd09 =~ nervs09nmean + effrt09nmean + hopls09nmean + rstls09nmean +
deprd09nmean + worth09nmean
pd12 =~ nervs12nmean + effrt12nmean + hopls12nmean + rstls12nmean +
deprd12nmean + worth12nmean
'
measurementInvariance(model=invariance, data=merged1, group="sexmean",
estimator="MLM", std.lv = FALSE, strict = FALSE, quiet = FALSE, fit.measures =
"default", method = "satorra.bentler.2010")
describeBy(merged, group="sexmean", mat="FALSE", type=3)
#Model 1 configural model
model1<-cfa(invariance, data=merged, estimator = "MLM", group="sexmean")
summary(model1, standardized=TRUE, rsquare=TRUE, fit.measures= TRUE)
#Model 2 Metric (weak) invariance model
model2<-cfa(invariance, data=merged, group="sexmean",estimator = "MLM",
group.equal=c("loadings"))
summary(model2, fit.measures=TRUE)
#Model 3 scalar invariance (loading and intercepts)
invariance1<- '
pd07 =~ nervs07nmean + effrt07nmean + hopls07nmean + rstls07nmean +
deprd07nmean + worth07nmean
```

```
pd09 =~ nervs09nmean + effrt09nmean + hopls09nmean + rstls09nmean +
deprd09nmean + worth09nmean
pd12 =~ nervs12nmean + effrt12nmean + hopls12nmean + rstls12nmean +
deprd12nmean + worth12nmean
rstls07nmean ~~ rstls09nmean
'
model3<-cfa(invariance1, data=merged, group="sexmean", estimator =
"MLM",group.equal=c("loadings", "intercepts"))
summary(model3, fit.measures=TRUE)
MI<-modindices(model3)
MI
MI[MI$op =="~1",]
#model 4 Strict invariance (loadings, intercepts, and residuals)
model4<-cfa(invariance, data=merged, group="sexmean", estimator =
"MLM",group.equal=c("loadings", "intercepts", "residuals"))
summary(model4, fit.measures=TRUE)
```

PANEL MODEL

```
#a specify model
longpanel<- '
score07 =~ nervs07nmean + effrt07nmean + hopls07nmean + rstls07nmean +
deprd07nmean + worth07nmean
score09 =~ nervs09nmean + effrt09nmean + hopls09nmean + rstls09nmean +
deprd09nmean + worth09nmean
```

```
score12 =~ nervs12nmean + effrt12nmean + hopls12nmean + rstls12nmean +
deprd12nmean + worth12nmean
score09 ~ score07
score12 ~ score09
'
#b estimate the model & summarize
longpanel.fit<-sem(longpanel, data=merged, estimator="MLM" )
summary(longpanel.fit, standardized=TRUE, fit.measures=TRUE)
fitMeasures(longpanel.fit)
MI<-modindices(longpanel.fit)
MI
semPaths(longpanel.fit,intercepts=TRUE)
```

AIM 3 R-CODE

Panel Model with Income Only

```
#run model with income only
hyp3panelincome<- '
score07 =~ nervs07nmean + effrt07nmean + hopls07nmean + rstls07nmean +
deprd07nmean + worth07nmean
score09 =~ nervs09nmean + effrt09nmean + hopls09nmean + rstls09nmean +
deprd09nmean + worth09nmean
score12 =~ nervs12nmean + effrt12nmean + hopls12nmean + rstls12nmean +
deprd12nmean + worth12nmean
score09~score07
```

```
score12~score09
score07~income
'
#b estimate the model & summarize
hyp3panelincome.fit<-sem(hyp3panelincome, data=merged, estimator="MLM")
summary(hyp3panelincome.fit, standardized=TRUE, fit.measures=TRUE)
semPaths(hyp3panelincome.fit,intercepts=TRUE)
```

<u>Panel Model with Uninsured Only</u>

```
#run model with uninsured only
hyp3paneluninsured<- '
score07 =~ nervs07nmean + effrt07nmean + hopls07nmean + rstls07nmean +
deprd07nmean + worth07nmean
score09 =~ nervs09nmean + effrt09nmean + hopls09nmean + rstls09nmean +
deprd09nmean + worth09nmean
score12 =~ nervs12nmean + effrt12nmean + hopls12nmean + rstls12nmean +
deprd12nmean + worth12nmean
score09~score07
score12~score09
score07~uninsured
'
#b estimate the model & summarize
hyp3paneluninsured.fit<-sem(hyp3paneluninsured, data=merged, estimator="MLM")
summary(hyp3paneluninsured.fit, standardized=TRUE, fit.measures=TRUE)
```

```
semPaths(hyp3paneluninsured.fit,intercepts=TRUE)
```

Panel Model with income and uninsured

```
hyp3panel<- '
score07 =~ nervs07nmean + effrt07nmean + hopls07nmean + rstls07nmean +
deprd07nmean + worth07nmean
score09 =~ nervs09nmean + effrt09nmean + hopls09nmean + rstls09nmean +
deprd09nmean + worth09nmean
score12 =~ nervs12nmean + effrt12nmean + hopls12nmean + rstls12nmean +
deprd12nmean + worth12nmean
score09~score07
score12~score09
score07~uninsured
score07~income
'
#b estimate the model & summarize
hyp3panel.fit<-sem(hyp3panel, data=merged, estimator="MLM")
summary(hyp3panel.fit, standardized=TRUE, fit.measures=TRUE)
semPaths(hyp3panel.fit,intercepts=TRUE)
```

AIM 4 R-CODE

```
hyp4panelnew<- '
score07 =~ nervs07nmean + effrt07nmean + hopls07nmean + rstls07nmean +
deprd07nmean + worth07nmean
```

```
score09 =~ nervs09nmean + effrt09nmean + hopls09nmean + rstls09nmean +
deprd09nmean + worth09nmean
score12 =~ nervs12nmean + effrt12nmean + hopls12nmean + rstls12nmean +
deprd12nmean + worth12nmean
score09~score07
score12~score09
score07~b*notreatmentMH
uninsured~a*notreatmentMH
score07~c*uninsured
ab:=a*b
totalab :=c +(a*b)
'
#b estimate the model & summarize
hyp4panelnew.fit<-sem(hyp4panelnew, data=merged, estimator="MLM")
summary(hyp4panelnew.fit, standardized=TRUE, fit.measures=TRUE)
semPaths(hyp4panelnew.fit,intercepts=TRUE)
```

<u>AIM 5 R-CODE</u>

<u>Panel Model with Smoking Only</u>

```
#model with smoking only
hyp5panelsmoke<- '
score07 =~ nervs07nmean + effrt07nmean + hopls07nmean + rstls07nmean +
deprd07nmean + worth07nmean
```

```
score09 =~ nervs09nmean + effrt09nmean + hopls09nmean + rstls09nmean +
deprd09nmean + worth09nmean
score12 =~ nervs12nmean + effrt12nmean + hopls12nmean + rstls12nmean +
deprd12nmean + worth12nmean
score09~score07
score12~score09
score07~smoking
'
#b estimate the model & summarize
hyp5panelsmoke.fit<-sem(hyp5panelsmoke, data=merged, estimator="MLM")
summary(hyp5panelsmoke.fit, standardized=TRUE, fit.measures=TRUE)
semPaths(hyp5panelsmoke.fit,intercepts=TRUE)
```

Panel Model with Alcohol Consumption Only

```
#model with drinking only
hyp5paneldrink<- '
score07 =~ nervs07nmean + effrt07nmean + hopls07nmean + rstls07nmean +
deprd07nmean + worth07nmean
score09 =~ nervs09nmean + effrt09nmean + hopls09nmean + rstls09nmean +
deprd09nmean + worth09nmean
score12 =~ nervs12nmean + effrt12nmean + hopls12nmean + rstls12nmean +
deprd12nmean + worth12nmean
score09~score07
score12~score09
```

```
score07~drinking
'
#b estimate the model & summarize
hyp5paneldrink.fit<-sem(hyp5paneldrink, data=merged, estimator="MLM")
summary(hyp5paneldrink.fit, standardized=TRUE, fit.measures=TRUE)
semPaths(hyp5paneldrink.fit,intercepts=TRUE)
```

<u>Panel Model with Smoking and Alcohol Consumption</u>

```
#model with smoking and drinking
hyp5panel1<- '
score07 =~ nervs07nmean + effrt07nmean + hopls07nmean + rstls07nmean +
deprd07nmean + worth07nmean
score09 =~ nervs09nmean + effrt09nmean + hopls09nmean + rstls09nmean +
deprd09nmean + worth09nmean
score12 =~ nervs12nmean + effrt12nmean + hopls12nmean + rstls12nmean +
deprd12nmean + worth12nmean
score09~score07
score12~score09
score07~smoking
score07~drinking
smoking~~drinking
'
#b estimate the model & summarize
hyp5panel1.fit<-sem(hyp5panel1, data=merged, estimator="MLM")
```

```
summary(hyp5panel1.fit, standardized=TRUE, fit.measures=TRUE)

semPaths(hyp5panel1.fit,intercepts=TRUE)
```

Regression Models

```
#part 2 of the hypothesis

model_fit= lm(smokingmean ~ pd07mean+pd09mean+pd12mean, data=merged)

summary(model_fit)

model_fit= lm(smokingmean ~ pd07lowmean+pd09lowmean+pd12lowmean,
data=merged)

summary(model_fit)

model_fit= lm(smokingmean ~ pd07modmean+pd09modmean+pd12modmean,
data=merged)

summary(model_fit)

model_fit= lm(smokingmean ~ pd07highmean+pd09highmean+pd12highmean,
data=merged)

summary(model_fit)

#add drinking

model_fit= lm(smokingmean ~ pd07mean+pd09mean+pd12mean +drinkingmean,
data=merged)

summary(model_fit)

model_fit= lm(smokingmean ~ pd07lowmean+pd09lowmean+pd12lowmean
+drinkingmean, data=merged)

summary(model_fit)
```

```
model_fit= lm(smokingmean ~ pd07modmean+pd09modmean+pd12modmean
+drinkingmean, data=merged)
summary(model_fit)
model_fit= lm(smokingmean ~ pd07highmean+pd09highmean+pd12highmean
+drinkingmean, data=merged)
summary(model_fit)
#add demographics
model_fit= lm(smokingmean ~ pd07mean+pd09mean+pd12mean
+drinkingmean+sexmean+racemean+incomemean+agemean+educagmean, data=merged)
summary(model_fit)
model_fit= lm(smokingmean ~ pd07lowmean+pd09lowmean+pd12lowmean
+drinkingmean+sexmean+racemean+incomemean+agemean+educagmean, data=merged)
summary(model_fit)
model_fit= lm(smokingmean ~ pd07modmean+pd09modmean+pd12modmean
+drinkingmean+sexmean+racemean+incomemean+agemean+educagmean, data=merged)
summary(model_fit)
model_fit= lm(smokingmean ~ pd07highmean+pd09highmean+pd12highmean
+drinkingmean+sexmean+racemean+incomemean+agemean+educagmean, data=merged)
summary(model_fit)
#add rest of the variables
model_fit= lm(smokingmean ~ pd07mean+pd09mean+pd12mean
+drinkingmean+sexmean+racemean+incomemean+agemean+educagmean+
uninsuredmean+daymiss07mean+daymiss09mean+daymiss12mean+notreatmentmhmean
```

```
+unemployed07mean+unemployed09mean+unemployed12mean+yeschronic07mean+yes
chronic09mean+yeschronic12mean+genhlth07mean+genhlth09mean+genhlth12mean+sti
gmatreat07mean+stigmatreat09mean+stigmatreat12mean+stigmacaring07mean+stigmac
aring09mean+stigmacaring12mean+regionmean, data=merged)
summary(model_fit)
model_fit= lm(smokingmean ~ pd07lowmean+pd09lowmean+pd12lowmean
+drinkingmean+sexmean+racemean+incomemean+agemean+educagmean+
uninsuredmean+daymiss07mean+daymiss09mean+daymiss12mean+notreatmentmhmean
+unemployed07mean+unemployed09mean+unemployed12mean+yeschronic07mean+yes
chronic09mean+yeschronic12mean+genhlth07mean+genhlth09mean+genhlth12mean+sti
gmatreat07mean+stigmatreat09mean+stigmatreat12mean+stigmacaring07mean+stigmac
aring09mean+stigmacaring12mean+regionmean, data=merged)
summary(model_fit)
model_fit= lm(smokingmean ~ pd07modmean+pd09modmean+pd12modmean
+drinkingmean+sexmean+racemean+incomemean+agemean+educagmean+uninsuredme
an+daymiss07mean+daymiss09mean+daymiss12mean+notreatmentmhmean+unemploye
d07mean+unemployed09mean+unemployed12mean+yeschronic07mean+yeschronic09m
ean+yeschronic12mean+genhlth07mean+genhlth09mean+genhlth12mean+stigmatreat07
mean+stigmatreat09mean+stigmatreat12mean+stigmacaring07mean+stigmacaring09mea
n+stigmacaring12mean+regionmean, data=merged)
summary(model_fit)
model_fit= lm(smokingmean ~ pd07highmean+pd09highmean+pd12highmean
+drinkingmean+sexmean+racemean+incomemean+agemean+educagmean+uninsuredme
```

```
an+daymiss07mean+daymiss09mean+daymiss12mean+notreatmentmhmean+unemploye
d07mean+unemployed09mean+unemployed12mean+yeschronic07mean+yeschronic09m
ean+yeschronic12mean+genhlth07mean+genhlth09mean+genhlth12mean+stigmatreat07
mean+stigmatreat09mean+stigmatreat12mean+stigmacaring07mean+stigmacaring09mea
n+stigmacaring12mean+regionmean, data=merged)
summary(model_fit)
#part 3 of the hypothesis
model_fit= lm(drinkingmean ~ pd07mean+pd09mean+pd12mean, data=merged)
summary(model_fit)
model_fit= lm(drinkingmean ~ pd07lowmean+pd09lowmean+pd12lowmean,
data=merged)
summary(model_fit)
model_fit= lm(drinkingmean ~ pd07modmean+pd09modmean+pd12modmean,
data=merged)
summary(model_fit)
model_fit= lm(drinkingmean ~ pd07highmean+pd09highmean+pd12highmean,
data=merged)
summary(model_fit)
#add smoking
model_fit= lm(drinkingmean ~ pd07mean+pd09mean+pd12mean +smokingmean,
data=merged)
summary(model_fit)
```

```
model_fit= lm(drinkingmean ~
pd07lowmean+pd09lowmean+pd12lowmean+smokingmean, data=merged)
summary(model_fit)
model_fit= lm(drinkingmean ~
pd07modmean+pd09modmean+pd12modmean+smokingmean, data=merged)
summary(model_fit)
model_fit= lm(drinkingmean ~
pd07highmean+pd09highmean+pd12highmean+smokingmean, data=merged)
summary(model_fit)
#add demographics
model_fit= lm(drinkingmean ~ pd07mean+pd09mean+pd12mean
+smokingmean+sexmean+racemean+incomemean+agemean+educagmean,
data=merged)
summary(model_fit)
model_fit= lm(drinkingmean ~ pd07lowmean+pd09lowmean+pd12lowmean
+smokingmean+sexmean+racemean+incomemean+agemean+educagmean,
data=merged)
summary(model_fit)
model_fit= lm(drinkingmean ~ pd07modmean+pd09modmean+pd12modmean
+smokingmean+sexmean+racemean+incomemean+agemean+educagmean,
data=merged)
summary(model_fit)
```

```
model_fit= lm(drinkingmean ~ pd07highmean+pd09highmean+pd12highmean
+smokingmean+sexmean+racemean+incomemean+agemean+educagmean,
data=merged)
summary(model_fit)
#add rest of the variables
model_fit= lm(drinkingmean ~ pd07mean+pd09mean+pd12mean
+smokingmean+sexmean+racemean+incomemean+agemean+educagmean+uninsuredme
an+daymiss07mean+daymiss09mean+daymiss12mean+notreatmentmhmean+unemploye
d07mean+unemployed09mean+unemployed12mean+yeschronic07mean+yeschronic09m
ean+yeschronic12mean+genhlth07mean+genhlth09mean+genhlth12mean+stigmatreat07
mean+stigmatreat09mean+stigmatreat12mean+stigmacaring07mean+stigmacaring09mea
n+stigmacaring12mean+regionmean, data=merged)
summary(model_fit)
model_fit= lm(drinkingmean ~ pd07lowmean+pd09lowmean+pd12lowmean
+smokingmean+sexmean+racemean+incomemean+agemean+educagmean+uninsuredme
an+daymiss07mean+daymiss09mean+daymiss12mean+notreatmentmhmean+unemploye
d07mean+unemployed09mean+unemployed12mean+yeschronic07mean+yeschronic09m
ean+yeschronic12mean+genhlth07mean+genhlth09mean+genhlth12mean+stigmatreat07
mean+stigmatreat09mean+stigmatreat12mean+stigmacaring07mean+stigmacaring09mea
n+stigmacaring12mean+regionmean, data=merged)
summary(model_fit)
model_fit= lm(drinkingmean ~ pd07modmean+pd09modmean+pd12modmean
+smokingmean+sexmean+racemean+incomemean+agemean+educagmean+uninsuredme
```

```
an+daymiss07mean+daymiss09mean+daymiss12mean+notreatmentmhmean+unemploye
d07mean+unemployed09mean+unemployed12mean+yeschronic07mean+yeschronic09m
ean+yeschronic12mean+genhlth07mean+genhlth09mean+genhlth12mean+stigmatreat07
mean+stigmatreat09mean+stigmatreat12mean+stigmacaring07mean+stigmacaring09mea
n+stigmacaring12mean+regionmean, data=merged)
summary(model_fit)
model_fit= lm(drinkingmean ~ pd07highmean+pd09highmean+pd12highmean
+smokingmean+sexmean+racemean+incomemean+agemean+educagmean+
uninsuredmean+daymiss07mean+daymiss09mean+daymiss12mean+notreatmentmhmean
+unemployed07mean+unemployed09mean+unemployed12mean+yeschronic07mean+yes
chronic09mean+yeschronic12mean+genhlth07mean+genhlth09mean+genhlth12mean+sti
gmatreat07mean+stigmatreat09mean+stigmatreat12mean+stigmacaring07mean+
stigmacaring09mean+stigmacaring12mean+regionmean, data=merged)
summary(model_fit)
```