2021

Medical Expenditures and Health Care Utilization: A comparison among U.S. Adult Tobacco Consumers (i.e. Never Tobacco Users, Cigarette Smokers and Smokeless Tobacco Users)

Raheema s. Muhammad
Virginia Commonwealth University

MEDICAL EXPENDITURES AND HEALTH CARE UTILIZATION: A COMPARISON AMONG U.S. ADULT TOBACCO CONSUMERS (I.E. NEVER TOBACCO USERS, CIGARETTE SMOKERS AND SMOKELESS TOBACCO USERS)

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by

Raheema Saleema Muhammad-Kah, MSPH
Master of Science in Public Health, Emory University, Atlanta, GA, 2004

Advisor:
Norman V. Carroll, PhD, RPh
Professor
Department of Pharmacotherapy and Outcomes Science

Virginia Commonwealth University
Richmond, VA
April, 2021

DEDICATION

To my mother and father who have always encouraged me to go above and beyond and that anything is possible with hard work, honesty, patience and pray.

ACKNOWLEDGEMENT

On reaching this milestone in my academic career, I find it important to show gratitude to all who have helped me along the way. Above all I say thank you God for giving me life, abundant blessings, guidance and perseverance to complete this endeavor. I thank my parents for always believing in me and encouraging me to excel. My belated father would be extremely proud of my accomplishment.

The unwavering support of my husband Ousman was crucial in getting through the tough times . He afforded me the time and quiet space I needed to get my work done. Much thanks to my siblings and children who brought moments of joy and humor to brighten the long arduous days.

Heartfelt thanks to my advisor Dr. Norman V Carroll for accepting me as his student and imparting the knowledge and guidance I needed to reach this milestone. Special thanks to Dr. Mohamadi Sarkar for his mentoring, experience and support, I really appreciate the advice and encouragement over the years. I am most grateful to my committee members, Dr. Mary Peace McRae, Dr. Edward Boone, Dr. Mohamadi Sarkar and Dr. Julie Patterson for their time and support. I would also like to thank the remarkable present and former faculty members in our department for their guidance, knowledge and support over the years special shout outs to Dr. David Holdford , Dr. Moczygemba and Dr. Patricia Slattum.

Last but not least I would like to thank my managers at Altria; Yezdi Pithawalla, Ryan Black, and Edward Largo for their understanding and willingness to allow me the time to meet my commitments at work and school. I thank the School of Pharmacy, for accepting me into the Doctoral program.

It was a pleasure knowing and working with the McGuire 218 crew who welcomed me with open arms.

In closing it is significant to acknowledge the year 2020, and the positive role the lockdown played in allowing me to sit still and complete this dissertation.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AHRQ	Agency for Healthcare Research and Quality
AIC	Akaike Information Criterion
BMI	Body Mass Index
BIC	Bayesian Information Criterion
CBO	Congressional Budget Office
CINAHL	Cumulative Index of Nursing and Allied Health Literature
CI	Confidence Interval
COPD	Chronic obstructive pulmonary disease
DUPERSID	MEPS sample person ID
DSM-V	Diagnostic and Statistical Manual of Mental Disorders, 5th Edition
ED	Emergency department
ER	Emergency room
FDA	Food and Drug Administration
FDA CTP	FDA Center for Tobacco Products
FMX	Family Sequence Number
FPX	Person Sequence Number
GLM	Generalized linear model
HBV	Hepatitis B virus
HC	Household Component
HCV	Hepatitis C virus
HHX	Household Serial Number
HIV	Human immunodeficiency virus
IC	Insurance Component
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
IOM	Institute of Medicine
MCDA	Multi-Criteria Decision Analysis
MCS	Mental Component Summary
MEPS	Medical Expenditure Panel Survey
MEPS-HC	Medical Expenditure Panel Survey- Household Component
MI	Myocardial Infarction

NASEN	North American Syringe Exchange Network
NDI	National Death Index
NCHS	National Center for Health Statistics
NHIS	National Health Interview Survey
NLMS	National Longitudinal Mortality Survey
NSDUH	National Survey on Drug Use and Health
OR	Odds Ratio
PATH	Population Assessment of Tobacco and Health
PCE	Personal Consumption Expenditure
PCS	Physical Component Summary
PERWT	Person-level weights
PUF	Public Use File
RR	Relative Risk
SE	Standard Error
SF-12v2	Short-Form 12 Version 2
SMDM	Society for Medical Decision Making
SRVY_YR	NHIS survey year
TUS-CPS	Tobacco Use Supplement to the Current Population Survey
U.S.	United States
USHHS	U.S. Department of Health and Human Services
VARPSU	Primary sampling unit
VARSTR	Variance estimation strata
ZIP	Zero Inflated Poisson regression

ABSTRACT

MEDICAL EXPENDITURES AND HEALTH CARE UTILIZATION: A COMPARISON AMONG U.S. ADULT TOBACCO CONSUMERS (I.E. NEVER TOBACCO USERS, CIGARETTE SMOKERS AND SMOKELESS TOBACCO USERS)

By Raheema Muhammad-Kah, MSPH

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

Virginia Commonwealth University, 2021

Advisor: Norman V. Carroll, PhD
Professor, Department of Pharmacotherapy and Outcomes Science

Introduction

Cigarette smoking poses a great economic cost on the U.S. health care system by incurring high health care expenditure and health care utilization in the adult population. Although there is a sufficient amount of information on the health care cost associated with cigarette smoking there is a gap in how other forms of tobacco use may impact healthcare expenditures and utilization in comparison to cigarette smoking. The main objectives of this study were: 1) to estimate and compare health care expenditure and health care utilization across different tobacco use groups (i.e. current exclusive cigarette smokers, current exclusive smokeless tobacco users and never tobacco users) using nationally representative data 2) to estimate and compare health care utilization across different tobacco use groups.

Methods

Data used in this study was obtained from linking individual-level data from two nationally representative data sources, 2009–2016 Medical Expenditure Panel Survey (MEPS) with 2011–2017 National Health Interview Survey (NHIS). The outcomes of interest were health care expenditure and health care utilization. Sociodemographic, health status, selected comorbidity and tobacco use status variables were collected from the NHIS and MEPS data. Tobacco use groups were categorized into three groups based on self-reported tobacco use history as exclusive current cigarette smoker, exclusive current smokeless tobacco use and never tobacco users. Econometric models were used to estimate annual mean total health care expenditures and health care utilization by tobacco use status adjusting for several covariates.

Results

Current exclusive cigarette smokers tended to be more likely to have one or more ER visits than current exclusive smokeless tobacco users (p-value =0.0161) and never tobacco users (p-value =0.0009). Modeling results adjusting for sociodemographic and health status variables indicated that current exclusive smokeless tobacco users and current exclusive cigarette smokers were not statistically significantly different than never tobacco users in their utilization of the health care services measured (i.e. with at least one ER visit, office-based visit, hospital outpatient visit, hospital inpatient visit, home health care visit and dental visit in the past 12 months).

Although no statistical association was found between total health care expenditure and tobacco use status at a 5% significant level. The highest annual mean total health care expenditure in US 2017 dollars was observed in the current exclusive cigarette smoking group $5627.64 (95% CI = $4068.50, $7186.78) followed by current exclusive smokeless users $4478.33 (95% CI = $3035.05, $5921.62) and never tobacco users had the lowest annual mean cost $4426.89 (95% CI = $3514.19, $5339.59).

Conclusions

Cigarette smoking has a heavy economic burden on the U.S. population. This study is the first to compare medical expenditure and health care utilization associated with current exclusive use of different tobacco products of in U.S. adults' over time. More data may be needed to obtain more conclusive results. This study shows the promise in the potential of reducing healthcare expenditure and utilization by aiding cigarette smokers down the continuum of risk of nicotine containing products to non-combustible tobacco products with less associated risk.

Chapter 1: Introduction

Introduction

Although adult cigarette smoking prevalence has declined over the past decades, tobacco use remains the leading preventable cause of death and disease in the United States.[1] The most recent Surgeon General's report on the health consequences of smoking estimated that for the years 2009–2012, annual smoking-attributable economic costs in the United States were between $289–332.5 billion, and $132.5–175.9 billion was spent for direct medical care of adult cigarette smokers.[2] Indirect cost due to lost productivity was estimated at approximately $157 billion from 2005-2009.[2]

Several nationally representative studies found higher rates of health care utilization (i.e. hospitalizations and outpatient visits) in adult cigarette smokers compared to never smokers.[3-5]

During the last two decades there has been a vast amount of literature estimating the healthcare utilization and/or medical expenditures attributable to cigarette smoking.[2,6-12] The current literature tends to focus on health care expenditure of cigarette smokers and/or former cigarette smokers.[12,13]

While there is a sufficient amount of information that indicates a heavy economic impact of adult cigarette consumption on the U.S. healthcare system, there is still a gap in how other forms of tobacco use may impact healthcare expenditures and utilization in comparison to cigarette smoking. Very little is known about the healthcare expenditures related to the use of other forms

of tobacco, like smokeless tobacco products. These are non-combustible tobacco products that have high use prevalence and have been available in the U.S. for decades.

This study fills a gap in the literature by providing a better understanding of health care expenditures and health care utilization in the U.S. population across the spectrum of current adult tobacco consumers using nationally representative data. This study would be the first to investigate and compare health care expenditures and use between current exclusive use of cigarettes and smokeless tobacco products, which have both been on the market for decades. This study is based on the payer's perspective and may aid in both health and insurance policy making. These analyses also have potential societal implications by determining whether the economic cost across the spectrum of tobacco products is reflective of the Tobacco Harm Reduction Model.

Background information, theoretical framework, study rationale and specific aims are provided in the remainder of Chapter 1. A systematic review of the literature on health care expense and utilization and tobacco usage is provided in Chapter 2. The methods and results for this study are presented in Chapters 3 and 4, respectively. The study conclusions, discussion of the study results, study limitations, and suggestions for future research, are presented in Chapter 5.

Background

Economic Cost of Cigarette smoking

In the U.S., cigarette smoking remains the leading cause of preventable diseases, disabilities and death and is associated with increased risk of lung cancer, cardiovascular disease, and chronic respiratory conditions.[2] Even with declines in the prevalence of current cigarette smoking, the annual burden of smoking-attributable mortality in the United States has remained above 400,000 for more than a decade and millions more live with smoking-related diseases.[2]

According to the latest Surgeon General report on the health consequences of smoking, cigarette smoking causes about one in every five deaths annually and the life expectancy of cigarette smokers is about 10 years shorter than non-cigarette smokers in the U.S.[2] The Centers for Disease Control and Prevention reported that, "smoking is the primary causal factor for at least 30% of all cancer deaths, for nearly 80% of deaths from chronic obstructive pulmonary disease(COPD), and for early cardiovascular disease and deaths."[14]

The dominant cause of COPD in men and women in the United States has been found to be associated with cigarette smoking. COPD causes emphysema and damage to the airways.[2,15] Increased risk for pulmonary infections is also associated with cigarette smoking through the weakening of the smoker's immune system. A causal relationship has also been established between active cigarette smoking and exacerbation of asthma in adults in the U.S.[2,15]

A vast body of literature reported in the most recent U.S. Surgeon General's report on the health consequences of smoking has shown that cigarette smokers are at a greater risk than non-cigarette smokers for cardiovascular diseases (i.e. diseases that affect the heart and blood vessels), that lead to coronary heart disease, myocardial infarction (MI) and stroke.[2]

The above morbidities associated with cigarette smoking have a substantial impact on health care costs resulting in a high economic burden. Annual smoking-attributable economic costs in the United States estimated for the years 2009–2012 were between $289–332.5 billion, including $132.5–175.9 billion for direct medical care of adults. Indirect cost due to lost productivity (due to premature death and exposure to secondhand smoke) was estimated at approximately $157 billion from 2005-2009.[2]

A causal relationship between cigarette smoking and diminished overall health has been established. This reduction in overall health among cigarette smokers is showed through self-reported poor health, increased absenteeism from work, and increased health care utilization and cost.[2] Although the prevalence of smoking continues to decline in the U.S., smoking-related health care expenditures were found to still account for an estimated 5–14% of the total health care expenditures as reported by the Congressional Budget Office (CBO) in 2012 and Levy and Newhouse.[2,16,17]

The CBO estimated annual per capita health care spending among adults ≥ 18 years of age. They found that spending tended to be highest among former smokers, and that current smokers had greater expenditures than never smokers. In their report they showed an example of adults 45–64 years of age, where annual health care spending was $7,650 for recent quitters, $5,540 for

current smokers, and $5,040 for never smokers. They also report that never smokers had the lowest spending in each age group, except for the oldest age groups (i.e. adults ≥ 75 years of age) where spending was $1,060 less for current smokers than for never smokers. As noted in the CBO report the above findings suggest that current smokers who survive to older ages may be in good health regardless of their cigarette smoking or may have a lower propensity to use health care. The CBO report and other studies have indicated that cigarette smoking increases the use and cost of health care.[2,6,12,13,16]

Smoking Cessation

Smoking cessation has been found to reduce the risk of smoking-related disease, reduce mortality, increase lifespan and improve well-being, including higher quality of life and improved health status.[18] Smoking cessation interventions have also been found to be cost-effective. Quitting smoking at any age is beneficial. However, it has been found that smokers who quit by the time they are 35–44 years of age avoid most of the risk of dying from a smoking-related disease.[19,20]

The Centers for Disease Control and Prevention (CDC) reported that 80 percent of all people who smoke see a physician each year and 70 percent of those smokers report that they want to quit. However, only about 32 percent attempted to do so using evidence-based counseling and/or medication.[21] According to the recent Surgeon General's report on smoking cessation, the use of tobacco cessation resources among persons who use tobacco remains low. Of adults 18 years of age and older, only 29.0% used cessation medication, 6.8% used any counseling, and only 4.1% used a telephone-based quit line. All states provide the latter resource free of charge.[18] Babb

and colleagues also reported that the use of counseling and/or medication was lower among young adults (16.6%) than among all adults (31.2%).[22]

Studies have also found that rather than quitting cigarettes all at once a gradual reduction in the number of cigarettes smoked per day leading up to a quit attempt may be a preferred approach to quitting by smokers who were unwilling/ unable to quit smoking abruptly.[23] An analysis of nationally representative data from the 2010–2011 Tobacco Use Supplement to the Current Population Survey (TUS-CPS) indicated that over 40% of adult cigarette smokers in the U.S. who had tried to quit smoking in the past year reported gradually cutting down on their cigarette use as a cessation strategy.[24]

The recent Surgeon General's report on smoking cessation states that, although evidence-based cessation counseling and/or medications have increased among adult cigarette smokers since 2000, still over two-thirds of adult cigarette smokers who tried to quit during the past year did not use an evidence-based cessation treatment. This report also found a large proportion of adult cigarette smokers reported using non-evidence-based approaches when trying to quit smoking like switching to other tobacco products such as e-cigarettes and smokeless tobacco.[2]

Harm Reduction Strategies

Given the low usage rate of evidence based cessation treatments and the large proportion of cigarette smokers who use other tobacco products as a means to quit smoking, a harm reduction model may be another viable option to help reduce the economic cost of adult cigarette consumption on the U.S. healthcare system. A harm reduction model refers to policies,

regulations and actions focused on reducing health risks, usually by providing safer forms of hazardous products or encouraging less risky behaviors, rather than simply banning products or behaviors.[25,26]

The harm reduction model is a public health strategy that was first developed in the 1980s for adults with substance abuse disorders for whom abstinence was not feasible. Over time, harm reduction strategies have been effective in reducing morbidity and mortality associated with risky health behaviors.[27]

For example, syringe exchange programs in the U.S. that began in the late 1980s at the state and local levels have been shown to be extremely effective in reducing human immunodeficiency virus (HIV) transmission among persons who inject drugs. Funding for these programs primarily comes from state and local governments and the support of the North American Syringe Exchange Network (NASEN). There are currently approximately 200 programs for syringe exchange in the U.S. Additional services are also offered at these sites for drug users, which include condom distribution, referrals to substance abuse treatment, HIV, hepatitis C virus (HCV) and hepatitis B virus (HBV) counseling and testing, overdose education, and naloxone distribution to reverse overdose.

Harm reduction programs also have an impact on overall sociomedical health. For example, access to and use of methadone maintenance programs have been found to be strongly related to decreased mortality from both natural causes and overdoses.[29]

Theoretical Framework

The harm reduction model in the tobacco space is referred to as the Tobacco Harm Reduction Model, which is a public health strategy to lower the health risks to individuals and the broader society by using forms of tobacco products other than combustible tobacco products like smoking cigarettes.[30] A strong public health consensus has formed that not all tobacco products present the same risk. Public health authorities agree that there is a broad continuum of risk among tobacco products, with cigarettes at the highest end of that spectrum recognizing that most of the harm caused by tobacco results from the burning of tobacco.[31-33] Completely quitting tobacco use is the best option. Figure 1 below shows the continuum of risk for nicotine containing products. The Surgeon General Report on the health consequences of smoking has acknowledged that the greatest burden of disease and disability arises from combustible tobacco products, especially cigarettes, and that moving adult tobacco consumers away from combustible tobacco products is a needed outcome.[2] Other tobacco products such as smokeless tobacco (particularly, low-nitrosamine Swedish snus) are not safe but are up to 90 percent less harmful than cigarettes. [34] The case for lower risk for individual users is well established for smokeless tobacco use given that it is non- combustible. Smokeless tobacco has not been found to be associated with lung cancer or other respiratory diseases, which account for most cigarette-caused deaths.[35]

My study will evaluate whether the economic cost across the spectrum of tobacco products is reflective of the Tobacco Harm Reduction Model. This is a reduction in health care expenditure with the usage of tobacco products on the lower continuum of risk compared to those on the higher end of the spectrum, smokeless tobacco product use and cigarette smoking respectively.

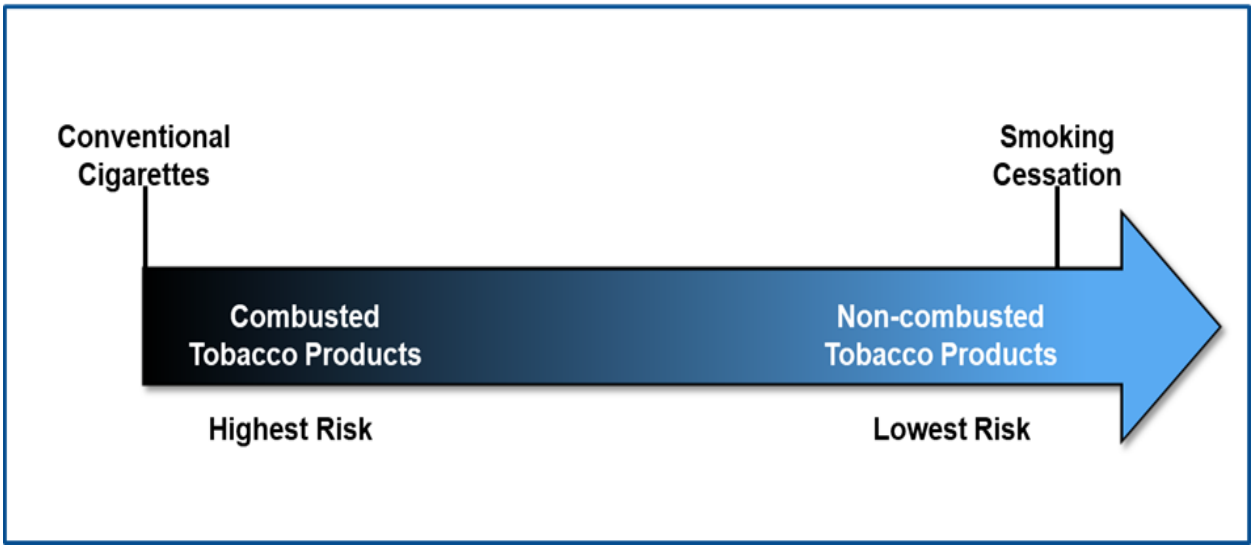


Figure 1: Tobacco Harm Reduction Model: Continuum of Risk for Nicotine containing Products

This proposal hypothesizes that current exclusive cigarette smokers will have the highest mean medical expenditure and health care utilization, followed by current exclusive smokeless tobacco users, and never tobacco users (reference group) will have the lowest mean values after adjusting for potential covariates.

Rationale

Sparse peer-reviewed literature on the economic cost of other forms of tobacco products has been published. To date there are only two peer-reviewed articles that estimate the healthcare utilization and expenditures attributable to other forms of tobacco use: smokeless tobacco (2018)

and cigars (2018) respectively.[36,37] These two articles focus specifically on smokeless tobacco and cigar user populations respectively with the primary objective of estimating health care utilization and expenditure attributable to these specific tobacco user populations. In these studies, a single point in time was used to calculate the unit costs for health care expenditure for major health care utilization measures. A more accurate methodology would be to link personal-level data on tobacco usage and health care expenditure and utilization, which allows the matching of an individual's services used to their specific costs rather than an estimated average cost.

The focus of my study is to investigate the difference in health care expenditure and use across current exclusive tobacco users which allows for a more accurate estimation of health care expenditures and utilization for current exclusive users of a specific tobacco product. Former users of tobacco (i.e. former smokeless users and former smokers) are not included in this study. The inclusion of the former tobacco user groups may confound health care estimates because when an individual quits tobacco use - particularly cigarettes – this may impact health care costs given the residual disease risk that former smokers carry after quitting.

Given the heavy economic cost of direct medical expenditure for adult cigarette smokers and the public health consensus on a Tobacco Harm Reduction model, comparing health care expenditure and utilization in the U.S. across the spectrum of current users of tobacco products is of great relevance. This comparison will provide a better understanding of the association between tobacco use status and health care expenditure and utilization. Cigarettes and smokeless tobacco are the two products that were used for my research, given they are the most commonly used in the U.S. and have been on the market for decades. By comparison, the more novel

tobacco products, such as vaping, have only been on the market a short while and would not provide a very large dataset.

The main purpose of my study was to estimate and compare health care expenditures and health care utilization associated with current exclusive cigarette smoking, current exclusive smokeless tobacco and never tobacco use. The results of this study can be important in health policy decision making. This analysis has potential societal implications by determining whether the economic cost across the spectrum of tobacco products is reflective of the Tobacco Harm Reduction Model. (i.e. a reduction in health care expenditure with the usage of tobacco products on extreme ends of the continuum of risk) and is an important advancement to the literature.

Specific Aims

This study aims to compare health care expenditures and health care utilization across different tobacco use groups (i.e. current exclusive cigarette smokers, current exclusive smokeless tobacco users and never tobacco users) using nationally representative data. Below are the specific aims for this study:

Specific Aim 1:

- Estimate the prevalence of tobacco use by tobacco use status by year of Medical Expenditure Panel Survey collection
- Estimate the prevalence of tobacco use status by age category
- Assess the distribution of the study sample of adults by tobacco use status and sociodemographic characteristics

- Assess the distribution of the study sample of adults by tobacco use status and selected comorbidities
- Estimate mean annual medical expenditures by tobacco use status.

Specifics of Aim 2:

- Assess the distribution of health care utilization by type of health care service and tobacco use status
- Estimate mean annual health care utilization by tobacco use status

Chapter 2: Literature Review

Literature Review of Healthcare Expenditure and Utilization and Tobacco use

A comprehensive review of the existing body of literature was conducted using PubMed/MEDLINE, Cumulative Index of Nursing and Allied Health Literature (CINAHL) and Google Scholar. The search strategy combined multiple search terms and MeSH terms to retrieve relevant articles including : “Healthcare Expenditure” or “Medical Expenditures”; “Healthcare Utilization” or “ Health Care Services”; “Tobacco Use” or “Tobacco” or “Nicotine” or “Smoking” or “Tobacco Smoking” or “Smokeless” or “Cigarette” or “Cigar”. The following inclusion and exclusion criteria were applied after screening through the titles and abstracts:

Inclusion Criteria :

1. Evaluates health care expenditure and /or utilization associated with tobacco product use
2. Articles between 1980 to date (i.e. 2020)
3. Published in a peer-review journal and in English language

Exclusion Criteria :

1. Studies conducted on populations outside of the United States
2. Novel tobacco products like electronic cigarettes

The refined search yielded 39 articles. A total of 4 studies were identified for full-text review after screening titles and abstracts and applying the inclusion and exclusion criteria, see Figure 2 below. To evaluate if any other studies that met the inclusion criteria were missed in the initial search the references of the 4 studies were reviewed. Table 1 summarizes the 4 studies.

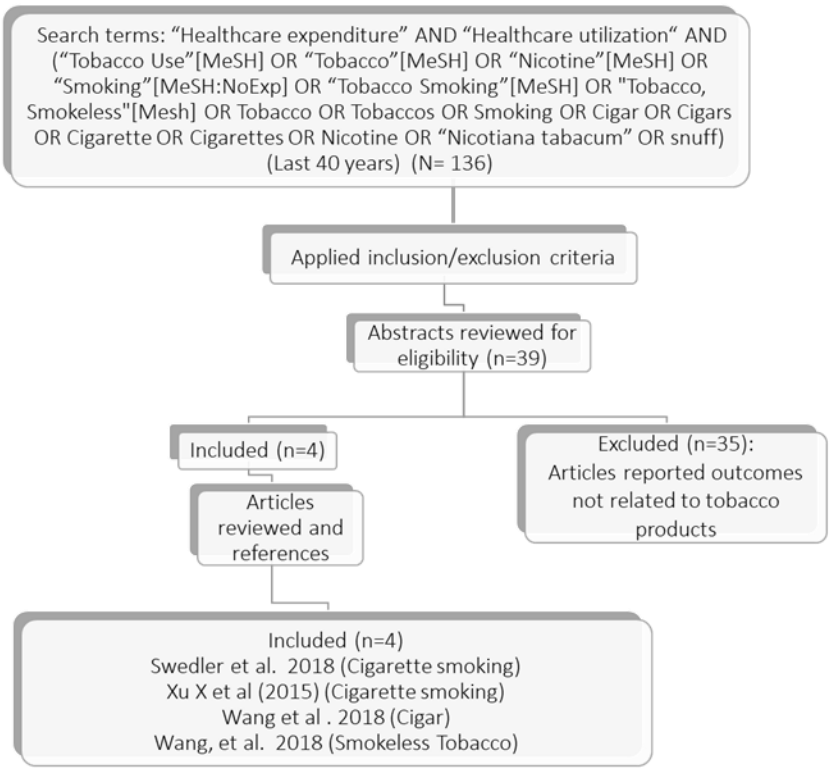


Figure 2 : Literature Search Strategy

Table 1: Summary of Articles Identified from the Literature review

Author	Outcome	Tobacco product	Sample Size	Data Source & Time Frame	Findings
Xu et al. 2015[12]	Annual smoking-attributable Healthcare Spending	**Cigarette Smokers** 1.Current Smoker 2. Former Smoker who quit within the last 5 years 3. Former Smoker who quit greater than 5 years 4. Never Smoker	~ 41,000 adults (unweighted sample) (NHIS & MEPS)	NHIS (2004 -2009) MEPS (2006-2010)	A total of 8.7% (95%CI(6.8%, 11.2%) of annual healthcare spending was attributed to smoking between 2006 and 2010. Approximately 60% of the attributable spending was paid by public programs.
Swedler et al. 2019[13]	Medical Expenditure	**Cigarette Smokers** 1. Current Smoker 2. Former Smoker 3. Never Smoker	~ 250 million adults (weighted sample) (NHIS & MEPS)	NHIS (2009-2014) MEPS (2011-2015)	Never smokers had statistically significantly lower expenditures than current and former smokers. Former smokers had the highest expenditure but was not significantly different than current smokers.
Wang et al. 2018[36]	Health Care Utilization Health Care Expenditure	**Smokeless Tobacco users** 1. Current Smokeless tobacco user 2. Former Smokeless tobacco user 3. Non- Smokeless tobacco users 4. Never tobacco users	134,451 adults (NHIS)	NHIS (2012-2015) MEPS 2014 (to calculate unit cost of health care services)	Current smokeless tobacco users significantly differed from never tobacco users in ER visits in the past 12 months but did not differ in the number of hospital nights, doctor visits, and home care visits. Smokeless tobacco use annual excess expenditures were estimated total 3.4 billion across all measured health care services.
Wang et al. 2018[37]	Health Care Utilization Health Care Expenditure	**Cigars users** 1. Current sole cigar smokers 2. Current poly cigar smokers 3. Former sole cigar smokers 4. Former poly cigar smokers 5. Other tobacco users 6.Never tobacco users	84,178 adults (NHIS)	NHIS (2000,2005,2010 and 2015) MEPS 2014 (to calculate unit cost of health care services)	Current and former sole cigar smokers were not significantly different from never tobacco users in their utilization of the health care services measured. Sole cigar smoking attributable annual health care expenditures were estimated to be ($625 per sole cigar smoker) $284 million.

Summary of Literature

Xu et al. (2015)[12]

Xu et al. conducted an analysis to estimate nationally representative cigarette smoking–attributable and associated healthcare spending for U.S. adults 18 years of age and older. Data used in this study was obtained from linking individual-level data from two nationally representative data sources, 2006–2010 Medical Expenditure Panel Survey (MEPS) with 2004–2009 National Health Interview Survey (NHIS). MEPS, a subsample of NHIS, is a survey of civilian non-institutionalized families and individuals, their medical providers, and employers, that collects information on individual healthcare utilization and medical expenditures and tracks respondents for 2 years after their NHIS interview. NHIS is a nationally representative cross-sectional survey of approximately 35,000 households in the US civilian non-institutionalized population which collects information on their health and behaviors; including individuals' sociodemographic, tobacco use information, health conditions, health care utilization, and health insurance coverage.

The study data was grouped into four categories based on self-reported cigarette smoking history: (1) current cigarette smokers, (2) former cigarette smokers who quit smoking within the last 5 years, (3) former cigarette smokers who quit smoking > 5 years ago and (4) never cigarette smokers (reference). Current cigarette smokers were those who had smoked 100 cigarettes in their lifetime and smoked some days or every day at the time of the interview.

The main outcome of interest for this study was annual healthcare spending. Estimates from two-part models were combined to predict the share of annual healthcare spending that could be attributable to cigarette smoking, adjusting for selected covariates.

Current cigarette smokers made up 21.5% of the final sample of adult respondents , 22.6% were former cigarette smokers where 6.0% quit within the last 5 years and 16.6% had quit for longer than 5 years, and 56.0% were never cigarette smokers. Current smokers were more likely to be younger, male and non-Hispanic white compared to never smokers. The authors measured other markers of risk for their potential association with increased health expenditures. They found that current cigarette smokers were more likely to believe in overcoming illness without medicine, less likely to have health insurance, more likely to report being an excessive drinker and more inclined to take risks compared to never smokers.

The modeling results indicated that a total of 8.7% (95% CI (6.8%, 11.2%) of annual healthcare spending (up to $ 170 billion per year) was attributed to cigarette smoking (i.e. current and former smokers) between 2006 and 2010 and approximately 60% of attributable spending was paid by public programs (i.e. Medicare, other federal paid programs and Medicaid).

Swedler et al. (2019)[13]

Swedler et al.[13]conducted a retrospective study to assess medical expenditures by smoking status among US adults age 18 and older. An objective of this study was to provide the most updated information (i.e. 2015) on medical expenditure by cigarette smoking status. Data used in this study was obtained from 2011–2015 MEPS linked with 2009–2014 NHIS. Like Xu et al[12] individual level information on medical expenditures was obtained from MEPS and self-reported smoking related history was obtained from NHIS.

The study data was grouped into three categories based on self-reported cigarette smoking history: (1) never smokers (reference), (2) current smokers and (3) former smokers . Current

smokers were those who had smoked 100 cigarettes in their lifetime and did not report that they quit smoking. Former smokers were those who had smoked 100 cigarettes in their lifetime and reported that they had quit smoking. The authors further categorized former smokers by years since quitting (i.e. 1, 2 and 5 years-since-quitting). Never smokers were those who had not smoked at least 100 cigarettes in their lifetime.

The main outcome of interest for this study was medical expenditure. Estimated average expenditures per individual and marginal costs for individuals by smoking status were obtained using a two-part model adjusting for selected covariates such as sociodemographics, health status characteristics and comorbidities.

Of the approximately 250 million weighted sample of adults in the US from 2011-2015 , 19.7 million identified themselves as current smokers and 23.9 million as former smokers. Of the former smokers, 24.4% quit within the prior 5 years, 8.7% quit within the prior 2 years and only 4.3% quit within 1 year prior to taking the survey.

Model results estimated that the average medical expenditures for an adult in the US was $4830 in 2015 US dollars. Never smokers, $4360 (95% CI 4154.3 to 4566.3), had lower medical expenditures than current smokers, $5244 (95% CI 4707.9 to 5580.3) and former smokers, $5590 (95% CI 5267.4 to 5913.5) . Former smokers had the highest medical expenditure but were not significantly different than current smokers. Years-since-quitting in the former smoker group did not impact medical expenditures.

The definition of the never smoking group in this study allows for the inclusion of users of other tobacco products and cigarette smokers that did not meet the lifetime criteria of having smoked at least 100 cigarettes. Given that never smokers were found to have the lowest medical

expenditure, the model results may indicate that cigarette smoking (current and former smoker who meet the lifetime criteria of being a smoker) has a bigger impact on medical expenditure than use of other tobacco products.

Wang et al. (2018)[35]

Wang et al[35] conducted a retrospective study to estimate the health care utilization and expenditures attributable to the use of smokeless tobacco among US adults age 18 and older. The primary data used in this study was from NHIS. The final data set comprised pooled 2012–2015 NHIS data containing 139,451 adults age18 years or older. A single point in time from the MEPS database (i.e. 2014 MEPS) was used to calculate the unit costs for health care utilization measured from NHIS.

This study focused on smokeless tobacco use. Tobacco use status included four mutually exclusive groups based on self-reported tobacco use and were defined as follows : (1) current ST users, (2) former ST users, (3) non-ST tobacco users, and (4) never tobacco users (as the reference). Current smokeless tobacco users were those who now use smokeless tobacco every day or some days. Former smokeless tobacco users were those who have used smokeless tobacco products at least once and now do not use smokeless tobacco at all. Non-smokeless tobacco users comprised respondents who have smoked 100 cigarettes (including current and former cigarette smokers) or have ever smoked cigars (regular cigars, little filtered cigars, or cigarillos) or pipes (regular pipes, water pipes, or hookah) at least once in their lifetime but have never used smokeless tobacco.

The main outcome variables in the study were four types of health care utilization obtained from NHIS: Hospital nights: measured by the number of nights spent in a hospital receiving inpatient care in the last 12 months, Emergency department (ED) visits: number of visits to the ED for the respondents' own health in the past 12 months. Doctor visits were determined by the answers to the following two NHIS Family Core questions: "During the last 2 weeks, did [person] see a doctor or other health care professional at a doctor's office, a clinic, an emergency room, or some other place?" and "How many times did [person] visit a doctor or other health care professional during the last 2 weeks? Home care visits were determined as the number of home care visits by a health care professional that the respondent had received in the past 2 weeks. The health care utilization modeling results were used to estimate what the authors refer to as smokeless tobacco-attributable health care utilization and smokeless tobacco-attributable health care expenditures adjusting for selected covariates. Smokeless tobacco-attributable health care utilization was derived by using an "excess utilization" approach which assessed the difference between factual and counterfactual predictions to obtain health care utilization attributable to smokeless tobacco use using a zero inflated Poisson regression model (ZIP model).

Of the 136,035 sampled adults, 51.3% were females between ages 34 and 64, 66.4% were non-Hispanic white, 53.0% were married, 13.4% had less than a High school education, 36.9% lived in the South and 29.6% reported having low income or being poor. The majority of the sample (60%) were overweight or obese, 23.0% were identified as binge drinkers and 14.1% had no health insurance during the past 12 months. The prevalence of adults' current ST use, former ST use, non-ST tobacco use, and never tobacco use through 2012–2015 was 2.1%, 7.9%, 39.8%, and 50.2% respectively.

Based on the modeling results, current smokeless tobacco users had statistically significantly more ER visits in the past 12 months than never tobacco users but did not differ in the number of hospital nights, doctor visits, and home care visits. The authors reported that based on the estimated excess annual utilization, smokeless tobacco use annual excess expenditures were 3.4 billion in 2014 dollars across all measured health care services. They conclude that smokeless tobacco use is associated with excess health care utilization and expenditures.

Obtaining linking person-level data from NHIS and MEPS over a four-year period would be a more accurate methodology in terms of estimating health care utilization and expenditures. This enables a participant's services used to be matched to their specific costs rather than an average cost as used in this study (i.e. a single time point in 2014). Also, other types of health care utilization data like dental care visits were not included because they are not collected in NHIS.

Wang et al. (2018)[36]

This second article by Wang and colleagues conducted an analysis to estimate the health care utilization and expenditures attributable to cigar smoking among US adults age 35 and older.

Primary data used in this study was from NHIS. The final data set included pooled NHIS data from 2000, 2005, 2010 and 2015 and contained 84,178 adults. A single point in time from the MEPS database (i.e. 2014 MEPS) was used to calculate unit costs for health care utilization measured from NHIS.

This study focused on cigar use with six defined tobacco use statuses based on self-reported tobacco use: (1) current sole cigar smokers (i.e. exclusive cigar use), (2) current poly cigar smokers (i.e. smoke cigars and smoke cigarettes or use smokeless tobacco), (3) former sole cigar

smokers (i.e. former exclusive cigar use), (4) former poly cigar smokers (i.e. smoked cigars and smoked cigarettes or used smokeless tobacco), (5) other tobacco users (ever smoked cigarettes or used smokeless tobacco but not cigars), and (6) never tobacco users (never smoked cigars, smoked cigarettes, or used smokeless tobacco: reference). Current sole cigar smokers were those who had smoked at least 50 cigars and currently smoked cigars but did not met the lifetime criteria for being a cigarette smoker or smokeless tobacco user (i.e. smoked ≥ 100 cigarettes or used smokeless tobacco ≥ 20 times) . Current poly cigar smokers were those who had smoked at least 50 cigars and currently smoked cigars and met the lifetime criteria for being a cigarette smoker or smokeless tobacco user. Former sole cigar smokers were those who had smoked at least 50 cigars and currently did not use cigars at all and did not met the lifetime criteria for being a cigarette smoker or smokeless tobacco user. Former poly cigar smokers were those who had smoked at least 50 cigars and currently did not use cigars at all and met the lifetime criteria for being a cigarette smoker or smokeless tobacco user. Other tobacco users were respondents who met the lifetime criteria for being a cigarette smoker or smokeless tobacco user (including current and former users) but did not met the classification for the previously defined groups. Never tobacco users were defined as those who had never smoked 50 cigars and did not met the lifetime criteria for a cigarette smoker or smokeless tobacco user.

The main outcomes of interest in the study were four types of health care utilization obtained from NHIS: hospital nights, emergency department (ED) visits, doctor visits, and home care visits. These were defined in the same way as in the Wang study described earlier.

The health care utilization models used to estimate cigar smoking-attributable health care utilization and cigar smoking-attributable health care expenditures were similar to those detailed

in the authors' paper that investigated smokeless tobacco attributable health care utilization and expenditure.[35]

Of the 84,178 sampled adults, the majority, 75.2%, were between ages 34 and 64, 47.3% were male, 73.4% were non-Hispanic white, 64.4% were married, 15.2% had less than a High school education and 36.1% lived in the South. Most respondents (64.1%) were overweight or obese, 13.4% were identified as binge drinkers and 32.3% had no health insurance during the past 12 months. The prevalence of adult current sole cigar smokers, current poly cigar smokers, former sole cigar smokers, former poly cigar smokers, other tobacco users and never tobacco users were 0.6%, 1.7%, 0.7%, 4.8%, 40.5% and 51.7% respectively. Over 90% of adult respondents from the study sample did not use cigars.

The modeling results indicated that current and former sole cigar smokers were not significantly different from never tobacco users in their utilization of the health care services measured in this study. The authors reported that based on the estimated excess annual utilization, sole cigar smoking attributable annual health care expenditures were estimated to be $284 million in 2014 dollars (i.e. $625 per sole cigar smoker) and the inclusion of poly cigar smoking increased the attributable annual health care expenditures to $1.75 billion in 2014 dollars.

A main limitation of this analysis is the low prevalence of current and former cigar usage (i.e.< 10%) observed from the pooled study sample. Using the recommended methodology of linking person-level data from NHIS and MEPS would further reduce the analysis dataset. Also, other types of health care services like dental care visits were not included because they are not collected in NHIS.

Gap in the literature

A number of studies have examined health care expenditure and smoking status. Few studies have examined the impact on healthcare expenditures related to the use of forms of tobacco use other than cigarettes. To the best of our knowledge, to date only two studies have assessed healthcare utilization and expenditures in relation to other forms of tobacco (i.e. smokeless tobacco and cigars). These two studies are identified as the first peer- reviewed articles to explore the economic cost of other forms of tobacco usage besides cigarette smoking. None of the studies identified compared current exclusive use of more than one type of tobacco product. Also, no studies have compared the health care utilization or expenditure across different forms of current or exclusive tobacco product usage (i.e. combustible product compared to non-combustible).

There is a need to better understand health care expenditure across the different types of tobacco products, given the hefty economic burden of cigarette smoking on the US health care system. The public health strategy of taking cigarette smokers down the continuum of risk of nicotine containing products could also be examined in the context of health care expenditure and use.

Chapter 3: Methods

This chapter describes the methods used to address the specific aims of this study and includes details on study design, data sources, study sample, variables and statistical analyses.

Data Sources

The data used for the study was extracted from nationally representative data bases that are publicly available: The National Health Interview Survey (NHIS) and Medical Expenditure Panel Survey (MEPS). NHIS is an annual, cross-sectional survey designed to monitor health and behaviors of non-institutionalized U.S. civilians (e.g., individuals not in nursing homes, prisons, or the military) living in the 50 states or the District of Columbia.[37] NHIS is sponsored by the Centers for Disease Control and Prevention (CDC) and conducted by the National Center for Health Statistics (NCHS). The NHIS uses a face-to-face interviewing format and geographically clustered sampling techniques to select the sample of dwelling units. NHIS has a continuous data collection process, the sample is designed such that each month's sample is nationally representative and collected throughout the year. The NHIS Supplemental Adult Questionnaire was used in the study to identify tobacco use status data for adults 18 years and older from 2009-2016. The NHIS Supplemental Adult Questionnaire captures information on tobacco product use and behavioral characteristics of adult tobacco consumers.

MEPS is a complex national probability survey of the U.S. civilian non-institutionalized population (i.e. both households and individuals) and is conducted by the Agency for Healthcare Research and Quality (AHRQ), U.S. Department of Health and Human Services (HHS).[38] There are two major components of MEPS: the Household Component (HC) and the Insurance

Component (IC). This study utilized data from MEPS-HC which provides data on demographics, geographic region, access to health care, charges and sources of payments , priority conditions, employment, health status, income, health insurance coverage, health care utilization and expenditures from individual households and their members. The data is also supplemented by a survey of medical providers (physicians, hospitals, home health agencies, and pharmacies) who provide medical care to respondents of the MEPS-HC. medical providers are contacted by telephone if information cannot be accurately provided by the respondents. The MEPS-HC data are collected by computer-assisted personal interviews. The MEPS sampling frame is drawn from respondents in NHIS. Beginning the year after participants' NHIS interviews a nationally representative subsample of these participants is tracked for 24 months through MEPS. MEPS collects five rounds of data per respondent on healthcare visits and expenditures regardless of the payment source. All forms of payment for care are included.

MEPS provides national estimates of health care use and expenditures and was used to obtain data on total health care expenditure and utilization from the individual perspective from 2011-2017 in adults 18 years and older. The two data sets were linked via the Agency for Healthcare Quality Data Center (AHRQ), providing sample data for individuals with a complete set of information from both databases needed to address my specific study aims. This methodology is widely used in the literature when analyzing health care expenditure and utilization and tobacco use.[12,13,39] The final study sample comprised seven years of pooled data. Pooling the data increases the sample size, reduces the standard error of the estimates and enhances the ability to analyze small subgroups. The information on individuals in the final data sample was obtained from the MEPS Full-Year Consolidated files and the NHIS Adult Questionnaire files.

Linking of NHIS and MEPS Public Use files

As previously mentioned, the two Public Use Files (PUF's) were linked via the AHRQ Data Center. The MEPS full-year consolidated PUF's can be linked to the NHIS Core person-level PUF by using a MEPS/NHIS link file which is available from 1996-2017. Each MEPS/NHIS link file contains a crosswalk that enables merging of MEPS full-year PUF's with NHIS person-level PUF's that contain data collected for MEPS respondents in the year prior to their initial year of MEPS participation. The MEPS/NHIS link file is a restricted file that can only be accessed at the AHRQ Data Center.

The MEPS full-year PUF's collect data through an overlapping panel design. This design collects information from each household through in-person interviews over two calendar years, conducted over five rounds, with Round 3 spanning both calendar years.

For each panel, Rounds 1, 2, and part of Round 3 typically contain data from calendar year 1; the remaining part of Round 3, and Rounds 4 and 5 cover calendar year 2. Therefore, MEPS full calendar year PUF's contain data from the first year of a new panel combined with that of the second year of the previous panel.

Example of 2015 MEPS PUF's linkage with 2013/2014 PUF's

Figure 3 illustrates the MEPS/NHIS full calendar year 2015 estimates (i.e. for 2015 MEPS and 2013/2014 NHIS PUF's). Rounds 1, 2, and 3 of Panel 20 (i.e. 2014- 2015 MEPS) uses the 2014 NHIS as its sampling frame and are combined with Rounds 3, 4, and 5 of Panel 19 (i.e. 2015-2016 MEPS) which uses the 2013 NHIS as its sampling frame.

Table *1*Table 2 summarizes the linkages between the two databases. Linkage with the 2013 NHIS data was established for 14,726 of the 16,578 persons in Panel 19 of 2015 MEPS, while

for the 2014 NHIS data 17,249 of the 18,849 persons in Panel 20 of 2015 MEPS were linked. From the two panels a total of 3,452 persons did not link to either 2013 or 2014 NHIS data. cases that were not linked include newborns, newly in-scope persons and a small number of cases where the NHIS identified a household as responding but when fielded in MEPS it was determined to be a nonresponding household.

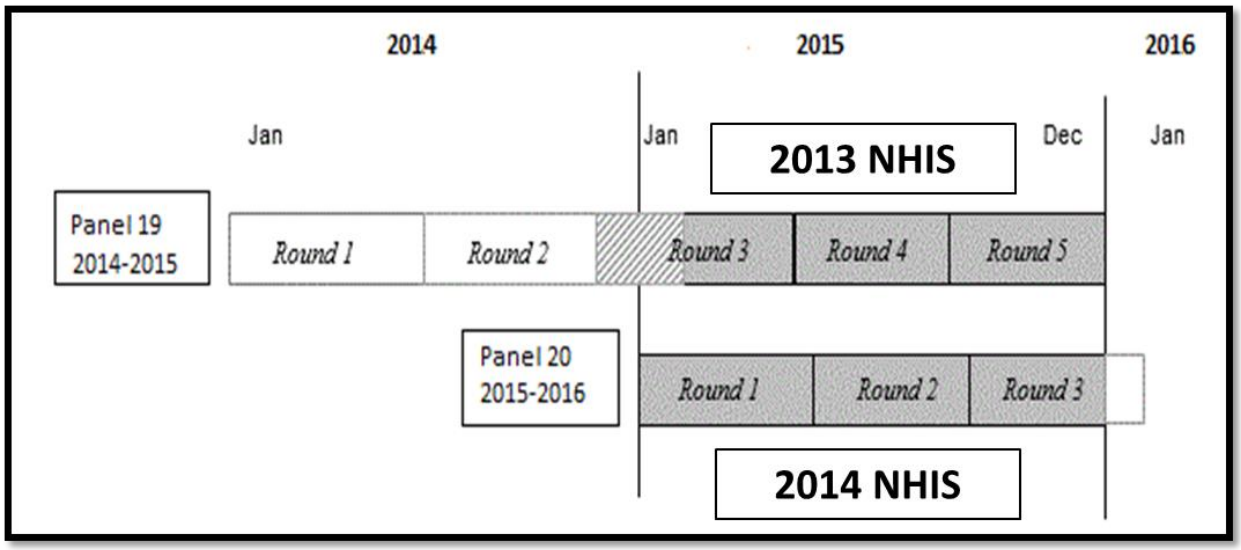


Figure 3: Illustration of linking 2015 MEPS PUF's with 2013/2014 NHIS PUF's (adapted) [40]

Table 2: Linkage file record counts from MEPS and NHIS PUF's (adapted) [40]

2015 MEPS Full-Year Data	Linked to 2013 NHIS PUF (n=104,520)	Linked to 2014 NHIS PUF (n=104,520)	Linked to 2013 or 2014 NHIS PUF (n=216,573)	Not Linked to NHIS	Total
Panel 19 persons	14,726	0	14,726	1,852	16,578
Panel 20 persons	0	17,249	17,249	1,600	18,849
Total	14,726	17,249	31,975	3,452	35,427

The restricted MEPS and NHIS linkage file for this example, NHMEP15X.DAT, permits the data user to merge any of the person-level MEPS 2015 Full Year PUF's with the 2013 and 2014 NHIS person-level PUF's, specifically the Sample Adult PUF for this study. The linkage file (NHMEP15X.DAT) contains 35,427 person-level records and seven variables listed in Table 3.

A record exists in the linkage file for each of the 2015 MEPS full-year persons. Each record contains the MEPS sample person ID (DUPERSID) and the corresponding NHIS unique sample person ID (Household Serial Number (HHX), Family Sequence Number (FMX), and Person Sequence Number (FPX)). A person-level 2015 MEPS Full Year PUF can be linked with the linkage file using the variable DUPERSID. Similarly, the NHIS 2013 or 2014 person-level data files can be linked with the linkage file by HHX, FMX, FPX, and SRVY_YR.

HHX is set to 999999, FMX is set to 99, PX is set to 99, FPX is set to 99, SRVY_YR is set to 9999, and LINKFLAG is set to 0, when a link cannot be established between MEPS sample person and the corresponding NHIS person.

Table 3: Record layout for the person-level MEPS/NHIS Linkage file (NHMEP15X.DAT)[40]

Variable	Column Position	Type	Label and value range
DUPERSID	1 - 8	Character	MEPS encrypted person ID (range=60001101-80571103)
HHX	9 - 14	Character	NHIS household serial number (range=000002 – 065122)
FMX	15 - 16	Character	NHIS family number (range=01-06)
FPX	17 - 18	Character	NHIS person number (range=01-18)
LINKFLAG	19 - 19	Numeric	Linkage status between MEPS and NHIS (1 or 0)
PANEL	20 - 21	Numeric	MEPS panel number (either 19 or 20)
SRVY_YR	22 - 25	Numeric	NHIS survey year (2013 or 2014)

Linkage instructions with sample SAS and STATA programs for adding NHIS variables to a MEPS dataset can be found on the AHRQ website.[42]

For this study the linkage process was repeated 7 times using 2011-2017 MEPS-HC Full Year PUF's and the corresponding 2009-2016 NHIS PUF's with the associated restricted MEPS/NHIS link files (i.e. NHMEP11X.DAT, NHMEP12X.DAT, NHMEP13X.DAT,NHMEP14X.DAT ,NHMEP15X.DAT,NHMEP16X.DAT and NHMEP17X.DAT). NHIS PUF's data for a given calendar year was attained from NCHS(NHIS - 1997-2018 (cdc.gov)).[43] Yearly MEPS PUF's were obtained from Medical Expenditure Panel Survey Download Data Files (ahrq.gov).[42] SAS analytical software (SAS Institute, Cary, North Carolina, USA), Version 9.4 was used for data collection and merging the NHIS files with their corresponding MEPS datasets.

Study Population

Data for this study was obtained from 2011 to 2017. Individuals included in the study had to fall into one of three defined tobacco use status categories: current exclusive cigarette smokers, current exclusive smokeless tobacco user or never tobacco user as identified from the NHIS Supplemental Adult Questionnaire using well established tobacco use definitions.

Study Design

A retrospective, cross-sectional study design was employed using MEPS/NHIS linked data files from 2011-2017. For Specific Aim 1 : I- II, the prevalence was estimated for tobacco use by tobacco use status by year of Medical Expenditure Panel Survey collection and by age category. For Specific Aim 1 : III-IV, the distribution of the study sample of adults by tobacco use status and sociodemographic characteristics, health status and selected comorbidities was evaluated. Bivariate associations between tobacco use status and sociodemographic characteristics, health status and selected comorbidities were also assessed. For Specific Aim 2 : I, the distribution of health care utilization by type of health care service and tobacco use status was evaluated.

Prior to addressing Specific Aim 1: V and Specific Aim 2: II, propensity score matching was used to control for observable differences between the tobacco use status groups. The current exclusive smokeless tobacco users were considered the “Cases” and current exclusive cigarette smokers and never tobacco users were considered the “Controls” respectively (i.e. two separate control groups). An econometric approach, described later in this chapter, was then used to estimate the mean annual health care expenditure and utilization by tobacco use status for Specific Aim 1: V and Specific Aim 2: II respectively.

Perspective

The health care expenditure estimation and health care utilization were based on a payer's perspective. The total health care expenditures captured in MEPS are direct payments for care provided during a given year based on the sum of 12 sources of payment variables. The study data allows the assessment of direct cost component and does not capture indirect cost.

Variables

Outcome variables

The primary outcomes of interest were total health care expenditure and health care utilization obtained from MEPS. Total expenditure was based on the responses of participants in MEPS-HC and was defined as the sum of payments from all sources to hospitals, physicians, other health care providers (including dental care), and pharmacies for services. Total expenditure is based on expenses on all annual health services including the following:

- Hospital inpatient care
- Hospital outpatient care
- Office-based medical provider services
- Emergency room services
- Home health care
- Prescription medicines

Health care utilization (health care services) measures the number of reported visits for a given service within a given year (i.e. the number of visits within the last 12 months). The study investigates 6 types of health care services listed below:

- Hospital outpatient visits
- Hospital inpatient visits
- Emergency room (ER) visits
- Office-Based visits
- Home health care visits
- Dental visits

All positive costs were inflated to 2017 U.S. dollars using the Personal Consumption Expenditure Health (PCE-Health) price index as recommended when pooling two or more years of MEPS total expenditure data.[42,43] The PCE-Health price indexes for 2011-2017 obtained from the Bureau of Economic analysis are shown in Table 4.

Table 4: Personal Consumption Expenditure (PCE) Price Index 2011-2017

Year	2011	2012	2013	2014	2015	2016	2017
Health	98.058	100.00	101.228	102.635	103.748	105.425	107.225

Table 5 displays the ratios used to adjust the raw total expenditure data in the final study sample to 2017 U.S. dollars. These ratios were obtained from the series of formulas below:

Year 2016 adjustment ratio = (2017 PCE index)/(2016 PCE index)

Year 2015 adjustment ratio = (2017 PCE index)/(2015 PCE index)

Year 2014 adjustment ratio = (2017 PCE index)/(2014 PCE index)

.
.
.

Year 2011 adjustment ratio = (2017 PCE index)/(2011 PCE index)

Table 5: Inflation Adjustment Ratios that are multiplied by the Total Expenditure for a given Year

Year	2011	2012	2013	2014	2015	2016
Ratio	1.094	1.072	1.059	1.045	1.033	1.017

Tobacco Use Status variable

As previously mentioned, the two tobacco products compared in this study are cigarettes and smokeless tobacco because they are the most commonly used in the U.S. and have been on the market for decades as compared to the more novel tobacco products such as vaping, and therefore they provide the most comprehensive data. The study also assesses current exclusive tobacco use status at the time of interviewing, allowing for a more accurate estimation of health care expenditures and use for current exclusive users of a specific tobacco product.

Tobacco use status was categorized into three groups: current exclusive smokeless tobacco users, current exclusive cigarette smokers and never tobacco users. Table 6 provides the NHIS questions used to derive the tobacco use status groups.

- **Never tobacco users** were defined as those who have never used cigarettes or smokeless tobacco in their lifetime (i.e. never smoked 100 cigarettes, and never smoked or used smokeless tobacco) at the time of the interview.
- **Current exclusive cigarette smokers** were defined as current cigarette smokers who smoked 100 cigarettes in their lifetime and smoked some days or every day at the time of the interview.
- **Current exclusive current smokeless tobacco users** were defined as current smokeless tobacco users who had used smokeless tobacco at least 20 times and were currently using every day or some days at the time of interview. Smokeless tobacco use included chewing tobacco, snuff, dip, snus, or dissolvable tobacco.

Table 6: Tobacco Use Status Definitions based on NHIS Questionnaires

Tobacco Use Status	Definitions based on NHIS Sample Adult Questionnaires
Current exclusive cigarette smokers	Responded YES to – "Ever smoked 100 cigarettes" AND Smoking Status -Responded as being "Current every day smoker" or " Current some day smoker"
Current exclusive current smokeless tobacco users	**Response for NHIS 2010-** Smoking Status -Responded as being "Never smoker" AND Responded YES to "Used snuff at least 20 times" AND Responded using Snuff "every day" or " some day" Smoking Status -Responded as being "Never smoker" AND Responded YES to "Used chewing tobacco at least 20 times" AND Responded using Chewing tobacco "every day" or " some day" **Response for NHIS 2012-** Smoking Status -Responded as being "Never smoker" AND Responded YES to – "Ever used smokeless tobacco products" AND Responded using smokeless tobacco products "every day" or " some day" **Response for NHIS 2016** Smoking Status -Responded as being "Never smoker" AND Responded YES to – "Ever used smokeless tobacco products, even once" AND Responded using smokeless tobacco products "every day" or " some day"
Never tobacco users	**Response for NHIS 2010-** Smoking Status -Responded as being a "Never smoker" AND Responded NO to – "Ever smoked 100 cigarettes" Responded NO to "Ever used snuff" Responded NO to "Ever used chewing tobacco" **Response for NHIS 2012-** Smoking Status -Responded as being a "Never smoker" AND Responded NO to – "Ever smoked 100 cigarettes" Responded NO to – "Ever used smokeless tobacco products"

Sociodemographic variables

Age was used both as a continuous variable and recoded into a categorical variable. Age was categorized as follows: 18-30 years, 31-40 years , 41-50 years, 51-60 years, 61-70 years and 70+ years. The age variable in MEPS is determined by date of birth and age given at the time of interview. A participant age is also verified during subsequent MEPS interviews and is top coded at 85 years. Only adults 18 years or old were included in the study. Sex was coded as male or female in MEPS. Race/ethnicity was coded into 5 categories in MEPS; Hispanic, non-Hispanic White only, non-Hispanic Black only, non-Hispanic Asian only and non-Hispanic Other or multiple race only. Region of residency was coded as Northeast, Midwest, South and West. A Body Mass Index (BMI) variable is calculated for adults 18 years of age or older. Adult BMI categories are coded as: underweight = BMI is less than 18.5, normal weight = BMI is between 18.5 – 24.9 inclusive, overweight = BMI is between 25.0 – 29.9 inclusive, and obesity = BMI greater than or equal to 30.0. Poverty status variable in MEPS is constructed using information on income, family and poverty categories. Family income is measured as a percent of the poverty line. This variable is classified it into one of five poverty categories: negative or poor (less than 100%), near poor (100% to less than 125%), low income (125% to less than 200%), middle income (200% to less than 400%), and high income (greater than or equal to 400%). Also, MEPS codes insurance type as uninsured, public and private.

Marital status variable was recoded into 4 categories from the original 9 categories in NHIS. These categories are married, widow/divorce/separated, living with partner and never married. Highest educational level attained was recoded into 4 categories; less than high school, high school, some college and college graduate or higher. Self -reported binge drinking status was recoded as Yes or No based on a question from NHIS. The question was " In the past year, on

how many days did you have 5 or more drinks of any alcoholic beverage?". Binge drinkers were considered those who responded that they had 1 or more days of 5 or more drinks.[35]

Health Status variables

A Self-Administered Questionnaire is fielded during MEPS interviews and designed to collect health status and health care quality measures of adults age 18 and older. The Self-Administered Questionnaire contains three measures of health status: Short-Form 12 Version 2,[44] the Kessler Index of non-specific psychological distress,[45] and the Patient Health Questionnaire.[46]

Short-Form 12 Version 2

The Short-Form 12 Version 2 (SF-12v2) questions are listed in Table 7.The SF-12v2 questionnaire comprises two components : Physical Component Summary (PCS) and Mental Component Summary (MCS) and is used as a quality of life measure. The standard approach to assessing data from the SF-12v2 is to form two summary scores based on responses to the questions in Table 7. Summary scores for both the PCS and MCS are obtained through a scoring algorithm which incorporate information from all 12 questions.[47] A score ranging from 0 (the worse health status) to 100 (the best health status) is assigned for each component.

Table 7: Short-Form 12 Version 2: Questions

12 Questions
General health today
During a typical day, limitations in moderate activities
During a typical day, limitations in climbing several flights of stairs
During past 4 weeks, as result of physical health, accomplished less than would like
During past 4 weeks, as result of physical health, limited in kind of work or other activities
During past 4 weeks, pain interfered with normal work outside the home and housework
During the past 4 weeks, felt calm and peaceful
During the past 4 weeks, had a lot of energy
During the past 4 weeks, felt downhearted and depressed
During past 4 weeks, as result of mental problems, accomplished less than you would like
During past 4 weeks, as result of mental problems, did work or other activities less carefully than usual
During the past 4 weeks, physical health or emotional problems interfered with social activities

Non-Specific Psychological Distress

A second measure of health status in the Self-Administered Questionnaire is the Kessler Index of non-specific psychological distress. This measure includes six mental health-related questions, which assesses the person's non-specific psychological distress during the past 30 days. The questions are listed in Table 8.

Table 8: Kessler Index: Questions

6 Questions
During the past 30 days, felt nervous
During the past 30 days, felt hopeless
During the past 30 days, felt restless or fidgety
During the past 30 days, felt so sad that nothing could cheer the person up
During the past 30 days, felt that everything was an effort
During the past 30 days, felt worthless

The summation of the values of the six questions in Table 8 provides an index to measure non-specific psychological distress using the following response values:

0 -None of the Time
1 -A Little of the Time
2- Some of the Time
3 -Most of the Time
4 - All of the Time

The Kessler index value ranges from 0 - 24, where the higher the value the greater the person's tendency towards mental disability.

Patient Health Questionnaire

The final health status measure in the Self-Administered Questionnaire is the Patient Health Questionnaire which assesses the frequency of a person's depressed mood and decreased interest in usual activities. This measure includes two mental health questions listed in the Table 9. This index is measured by summing the values of the two questions in Table 9, the score ranges from 0 - 6. The higher the score the greater a person's tendency towards depression. A score of 3 is suggested to be the optimal cut point for depression screening purposes.[46] This index is not equivalent to a Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-V) diagnosis of depression and is only intended as a screening measure for depression.

Table 9:Patient Health Questionnaire

2 Questions
During the past two weeks, bothered by having little interest or pleasure in doing things
During the past two weeks, bothered by feeling down, depressed, or hopeless

Selected Comorbidities variables

MEPS captures information on self-reported diagnosis history of various diseases. Table 10, list the variables that were included in this study and the associated question asked during the MEPS interviews. The response to the questions in Table 10 were "Yes" or "No".

Table 10:Self-reported Disease Diagnosis

Disease	Questions
Cancer	Have you ever been diagnosed as having cancer or a malignancy of any kind ?
Lung cancer	Have you ever been diagnosed with lung cancer?
Angina	Have you ever been diagnosed as having angina, or angina pectoris?
Coronary heart disease	Have you ever been diagnosed as having coronary heart disease?
Myocardial Infarction	Have you ever been diagnosed as having a heart attack, or myocardial infarction?
Stroke	Have you ever been diagnosed as having had a stroke or transient ischemic attack ?
Emphysema	Have you ever been diagnosed with emphysema?
Asthma	Have you ever been diagnosed with asthma?
Arthritis	Have you ever been diagnosed with arthritis?
Diabetes	Have you ever been diagnosed with diabetes (excluding gestational diabetes)?

Design variables

In order to generate national estimates, the complex sampling design of the MEPS dataset was taken into account for all study analyses by using person-level weights (PERWT), primary sampling unit (VARPSU) and variance estimation strata (VARSTR).

Independent variables

Independent variables were grouped into tobacco use status characteristics, sociodemographic characteristics, health status characteristics and comorbidity characteristics. Tobacco use status characteristics included: current exclusive smokeless tobacco users, current exclusive cigarette smokers and never tobacco users. The set of variables under sociodemographic characteristics were age, sex, race/ethnicity, marital status, BMI, region of residency, education, poverty status, self-reported binge drinking status and insurance type. The variables in the health status characteristics group included; Short-form 12 version 2 summary component scores, Kessler 6 index and Personal Health index. Comorbidity characteristics contained a self-reported diagnosis history for the following diseases; any cancer, lung cancer, angina ,coronary heart disease ,myocardial infarction, stroke, emphysema, asthma , arthritis and diabetes. Table 11 summarizes the predictor variables and their associated coding.

Sociodemographic characteristics such as age and sex are known to be associated with health care expenditure. Healthcare expenditure increases through maturity for both males and females. Females on average have higher expenditures during childbearing age.[48,49] White , non-Hispanics have higher healthcare expenditures than all other race/ethnicity groups.[48] Those with insurance (i.e. private and Medicare) also have been reported to have higher healthcare expenditure.[48] The health status and selected comorbidity characteristics are variables that have been suggested to control for when analyzing administrative medical data.[13,50,51,52,53]

Smokeless tobacco use prevalence is about 2% of the U.S. population and is predominately used by males.[54]

Table 11: Independent Variables

Independent Variables	Coding
Tobacco Use Status Characteristics	
Tobacco Use status	Current exclusive smokeless tobacco users, Current exclusive cigarette smokers and Never tobacco users
Sociodemographic Characteristics	
Age (years)	Continuous:18 through 85
	Categorical :18-30,31-40,41-50,51-60,61-70,70+
Sex	Male, Female
Race/ethnicity	Hispanic White, non-Hispanic Black, non-Hispanic Asian, non-Hispanic Other or multiple race, non-Hispanic
Marital status	Married, Living with Partner, Widow/divorce/separated ,Never Married
Body mass index (BMI)	Normal or Under Weight, Overweight ,Obesity
Education level	Less than high school, High school, Some college, College grad or higher
Region of residency	Northeast, Midwest, South, West
Poverty Status	Poor/Negative, Near Poor, Low income, Middle income, High income
Self-reported binge drinking status	No, Yes
Insurance type	Uninsured, Public, Private
Health Status Characteristics	
Quality of life score (Short-Form 12)	
Physical Component Summary	Score between 0-100
Mental Component Summary	Score between 0-100
Mental Illness score	
Kessler 6 Index	Score between 0-24
Depression score	
Personal Health Index	0,1,2,3,4,5 or 6
Comorbidity Characteristics	
Cancer	No, Yes
Lung cancer	No, Yes
Angina	No, Yes

Coronary heart disease	No, Yes
Myocardial Infarction	No, Yes
Stroke	No, Yes
Emphysema	No, Yes
Asthma	No, Yes
Arthritis	No, Yes
Diabetes	No, Yes

Statistical Analysis
Descriptive and Bivariate analysis

This section describes the statistical approach for Specific Aim 1: I- IV and Specific Aim 2 : I. Descriptive statistics were used to describe the sociodemographic, health status, and comorbidities characteristics of the overall study population and by tobacco use status. Descriptive statistics (frequency, percentage, mean, standard error (SE)) were also calculated to describe health care utilization by type of health care service and tobacco use status. The weighted frequency is also calculated in the descriptive statistics which represents the population size and is reported as the sum of weights (i.e. weighted total of the sample size). Since the data is pooled across several years the sum of weights represents the sum of each year's population for the years pooled. Dividing the weighted frequency by the number of years pooled, allows the reporting of the average population size over the years.[55] For this study the weighted frequency would need to be divided by 7 to obtain the average yearly weighted frequency (i.e. average yearly population size).

Continuous variables were expressed as the weighted mean, SE and the 95% confidence intervals (CI) of the mean. Categorical variables were expressed as the weighted frequency and the row percent and 95% CI of the row percent. The SURVEYMEANS and SURVEYFREQ procedures were used for continuous and categorical variables respectively.[56] These procedures incorporate complex survey sample designs where stratification, clustering, and weights can be applied.

Bivariate associations between tobacco use status groups and the continuous independent variables were tested using domain analysis which computes the means and the difference among the domain means and assesses the statistical significance based on the t-test. The Wald chi-square test was used to test the association between the tobacco use status groups and the

categorical independent variables. Statistical significance level was set at a p-value of 0.05. SAS version 9.4 was used for these analyses.

Propensity Score Matching

Selection bias, which is a result of the lack of randomization, can be a challenge when analyzing observational data. The groups of interest may not be comparable and key characteristics like age, gender, etc. may differ when groups are not randomized. To control for observable differences, the tobacco use status groups were matched using propensity score matching. This technique is used to compare groups while adjusting for group differences.[57] This matched control methodology was used to address potential selection bias and confounding between the tobacco use status groups. The study has three tobacco use status groups, therefore propensity score matching was performed twice, and the results were merged to obtain the final matched dataset.

Propensity score matching was performed using Greedy nearest neighbor matching which selects the control nearest to each case.[56] Greedy nearest neighbor matching is done sequentially for case units and without replacement. This method allows each case to be matched with the most suitable control available for matching at that point in the matching process and then the case and control are removed from the matching process. A 1:1 match where each case was matched with one control was performed. In this analysis the case and controls are the three-tobacco use status groups. The current exclusive smokeless tobacco group was considered as “group 1” , the current exclusive cigarette smoker group was “group 2” and never tobacco user was “group 3”. Specifying the set of confounding variables is a key issue in evaluating propensity scores. All the variables where an observable difference was seen between the current smokeless tobacco

group and one of the other groups were used to fit the logistic regression to estimate the propensity scores.

To maximize the amount of pairs obtained in the matching procedure the covariates in the logistic regression included age (i.e. as a continuous variable), sex (male, female), BMI (normal or underweight, overweight and obese), region of residency (northeast, midwest, south, west) , poverty status (poor/negative, near poor, low income, middle income, high income) and type of insurance (uninsured, public, private). Other variables were explored but due to missing responses the match sample sizes were not maximized.

The current exclusive smokeless tobacco group had a small sample size (1.2% of the final study sample) compared to the current exclusive cigarette smokers (19.5%) and never tobacco user groups (79.3%), which is reflective of the 2% prevalence of smokeless tobacco use in the U.S. population. The descriptive analysis indicated that the current exclusive smokeless tobacco users were predominately white non-Hispanic males (i.e. in line with the U.S. population), younger, in the middle to high income group, had insurance and resided in the south or midwest.

The PROC PSMATCH statements in SAS was used to invoke the PSMATCH procedure.[55] Greedy nearest neighbor matching was used to match observations for participants in the current exclusive smokeless tobacco group with observations for participants in the current exclusive cigarette smoker group or never smoker group The PSMODEL statement specifies the logistic regression model that creates the propensity score for each observation, which is the probability that the participant is a current exclusive smokeless tobacco user. The tobacco use variable was a binary treatment indicator variable where current exclusive smokeless tobacco user (group 1) is considered the case since the goal is to obtain matching pairs with current exclusive smokeless tobacco with the other two group . Group 2 and 3 (current exclusive cigarette smoker and never

tobacco user) were consider the control groups. The logits of propensity scores such that only observations that have propensity scores in the specified support region are used in matching. The logits of the propensity scores were used in computing differences between pairs of observations. Exact matches were used for age and sex.

Propensity score matching diagnostics were assessed using various plots for assessing balance. They include the following plots:

- cloud plots, which are scatter plots in which the points are jittered by adding random noise to prevent overplotting
- box plots for continuous variables
- bar charts for classification variables
- a standardized differences plot that summarizes differences between the case and control groups.

The recommended upper limit for standardized mean difference was set to 0.25,[58,59] although other have used an upper limit of 0.10.[60,61,62] The variance ratios between the case and control were assessed within the recommended range of 0.5 to 2.

The absolute standardized difference for continuous variables was computed using the formula below:

$$d = \frac{|\bar{x}_t - \bar{x}_c|}{\sqrt{\frac{s_t^2 + s_c^2}{2}}},$$

where $\bar{x}_t$ and $\bar{x}_c$ are the means and s_t^2 and s_c^2 are the variances of the variables in the case and control groups.

The absolute standardized difference for categorical variables was computed using the formula,

$$d = \frac{|p_t - p_c|}{\sqrt{\frac{p_t(1-p_t)+p_c(1-p_c)}{2}}},$$

where p_t and p_c denotes the proportions in the groups.

The study has three tobacco use status groups, therefore propensity score matching was performed twice, and the results were merged to obtain the final matched dataset.

Multivariate Analysis

Manning and Basu's[63] two-part modeling approach was used to estimate health care services and expenditures associated with tobacco usage. These types of models are commonly used in modeling health care expenditures, due to their highly skewed distribution (i.e., a large number of individuals with zero expenditure and a small number of individuals with substantial expenditures).[64] The number of participants with zero cost for healthcare expenditure in my sample was 21%, higher than the recommended 10%.[65] In two-part models, health care expenditures are estimated as a product of probability of any medical expenses and the predicted amount of these expenses conditional on the presence of any medical expenses. The two-part model consists of:

(1) A first part which uses a logit or probit regression model to estimate the parameters that determine the threshold between zero and nonzero values of the outcome. The first

regression, in the context of my analysis, models whether an individual had any medical expenditures or utilization in a given year.

(2) For the second part, a generalized linear regression with an appropriate link function and response distribution was used. The second model estimates the costs for individuals who the first model predicted had any medical expenditures or utilization.

An upside of generalized linear regressions is that they explicitly model heteroskedasticity. Also, with the choice of an appropriate distribution, a generalized linear regression allows the variance of the outcome to be a function of its predicted value.

A generalized linear model (GLM) with a Gamma distribution and a log link function was the best fit for the health expenditure data and was used to estimate the adjusted total healthcare expenditure by tobacco use status.

For count data such as health care utilization (i.e. 6 type of health care services) the Hurdle model was employed which is the two-part model used for count data. A logit model was assessed for the first model and the best fit for the second model was a truncated Poisson regression.

Both sets of regressions were controlled for appropriate co-variates/ independent variables and accounted for the MEPS complex survey sampling design.

It is important to understand the strengths and limitations of the models being implemented in any analysis. Statistical tests and model checks were employed using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)) model fit criteria. Different modeling approaches were conducted to obtain the best model fit. The two-part Hurdle models for count

data results were compared with the Zero-Inflated Poisson Regression Models (ZIP) where applicable.

The two-part model was implemented using the "twopm" command or corresponding command for count data (i.e. best fit model used "tpoisson") in Stata which enables the incorporation of the survey weights in the model.[64,66,67] Access to these survey weights are made available by Agency for Healthcare Research and Quality (AHRQ). To generate national estimates, the complex sampling design of the MEPS dataset was taken into account by using pooled person-level weights (PERWT), primary sampling unit (VARPSU) and variance estimation strata (VARSTR). Propensity score matching can significantly reduce the study sample size producing a stratum that only has one sample unit. If this occurs then by default Stata's survey commands will report missing standard errors. To resolve this issue variance estimation will be estimated using the single unit (center) option in Stata.[68] This specifies that strata with one sampling unit are centered at the population mean instead of the stratum mean to estimate the variation (i.e. standard errors). All modeling analyses was conducted using STATA software version 15 (Stata Corp, College Station, Texas, USA)[69] and statistically significant levels were set at 5%.

Sensitivity Analysis

As per best modeling practices, sensitivity analysis was used to determine the extent to which some of the assumptions of the study influenced the estimates. The following sensitivity analyses were conducted :

1. It is well known that healthcare expenditure increases with age, therefore the older population (participants older than 65 years) were excluded from the modeling analysis to assess the change in the estimates by tobacco use status.

2. Excluded female participants from model analysis, since smokeless tobacco products are predominately used by males in the U.S. population and participants greater than 65 years old.
3. Models were also conducted with the healthcare expenditure and utilization data before propensity score matching for health care utilization.

Human subject protection and data privacy

Based on Virginia Commonwealth University's guideline for identifying if research is required to go through their Institutional Review Board, the proposed research is eligible for exemption under 45 CFR 46.101(b)(4) from 45 CFR part 46 requirements because the data is publicly available and cannot be identified.

Chapter 4: Results

Descriptive and Bivariate Results of Final Study Sample

This section describes the results for Specific Aim 1: I- IV

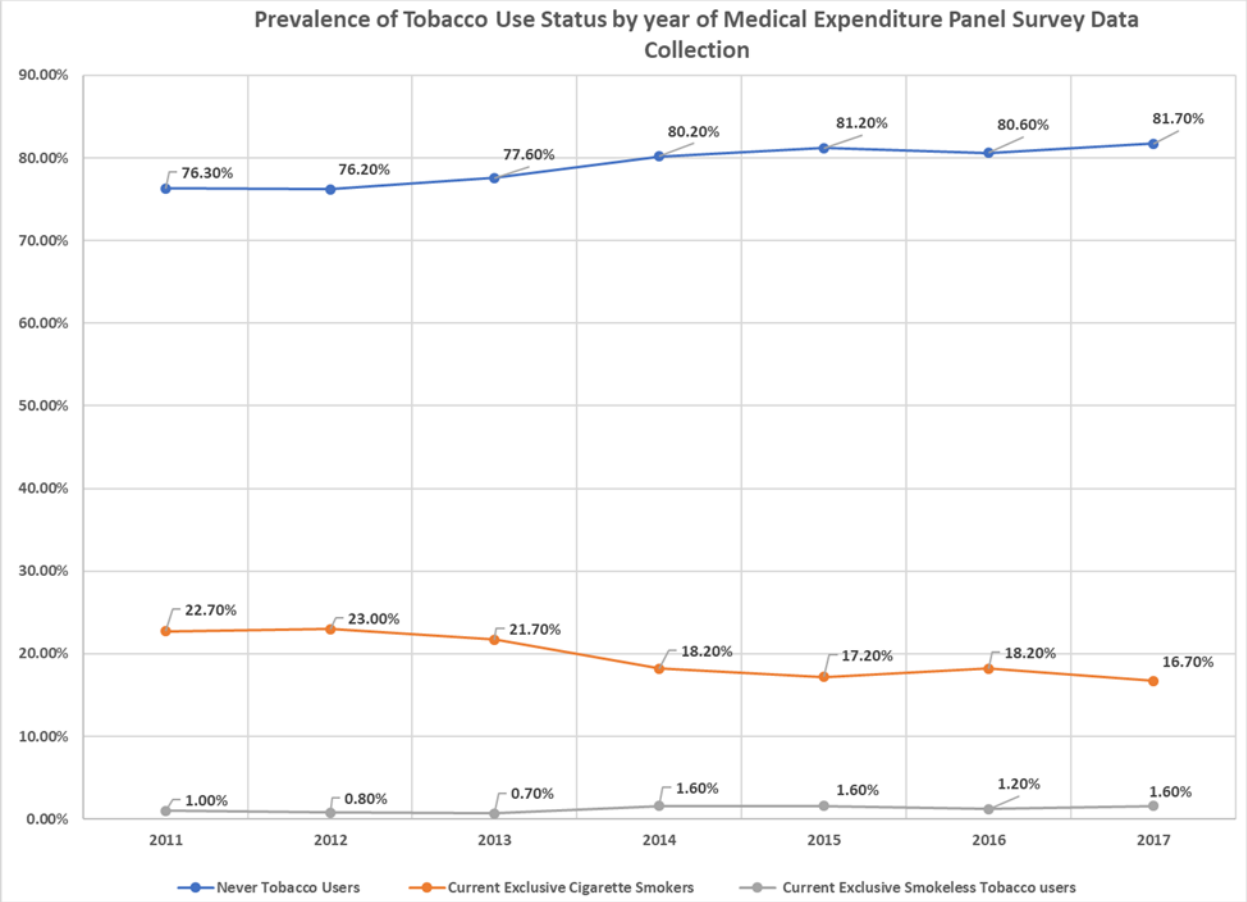
Sample Distribution

The MEPS/NHIS linked 2011-2017 dataset resulted in a total of 207,267 participants. The final pooled study sample, after applying the tobacco use status criteria, included a total of 68,866 participants who were eligible for the study. The annual weighted number of participants ranged between 89,474,000 and 104,640,000 for the sampled years. Six hundred and thirty-three (633) participants (1.2% of the study population) were current exclusive smokeless tobacco users. The annual weighted number of current exclusive smokeless tobacco users ranged between 650,000 and 1,686,000 for the sampled years. Thirteen thousand four hundred and twenty-three (13,423) participants (19.5%) were current exclusive cigarette smokers and the annual weighted number of current exclusive cigarette smokers ranged between 16,878,000 and 20,604,000 for the sampled years. Fifty-four thousand eight hundred and ten (54,810) participants (79.3%) were never tobacco users. The annual weighted number of never tobacco users ranged between 68,160,000 and 85,531,000 for the sampled years.

Table 12 shows the distribution of the final study sample size by full MEPS calendar year.

Table 12:Final Study Sample Size by MEPS Calendar Year

Year	Adults (18 years and Older)
2011	9,239
2012	10,034
2013	9,489
2014	10,144
2015	10,605
2016	9,419
2017	9,936
Pooled Sample Size	68,866



Note: Percentages were developed using survey weights to account for selection probabilities from the complex sampling design and to adjust for survey nonresponse

Figure 4:Prevalence of Tobacco Use Status by year of MEPS Data collection

Figure 4 depicts tobacco use status by study year. A slight increase was observed in current exclusive smokeless tobacco use as a decrease was seen in current exclusive cigarette smoking over the 2011 to 2017 timeline.

Final Study Sample before Matching Tobacco Use Groups

A total of 68,866 participants were eligible for this study. The mean age (SE) of the total population was 47.7 (0.18) years and 55.2% were female. Approximately 60% of the study population identified as white, non-Hispanic, 42.7% were married and 77.7% were in the Middle- or High-income bracket in terms of their poverty status. Most of the total population had some type of health insurance, 22.2% public insurance (i.e. Medicare) and 67.3% had some type of private insurance.

On average participants in the never tobacco user group were older 50.8 years [95% CI: 50.3 to 51.2] than those in the current exclusive smokeless user group 46.2 years [95% CI: 44.6 to 47.8] and the current exclusive cigarette smoker group 46.5 years [95% CI: 46.1 to 46.9] by about 4 years. This difference in mean age was found to be statistically significant between the never tobacco user group vs current exclusive smokeless user group (p-value <0.0001) and the never tobacco user group vs current exclusive cigarette smoker group (p-value <0.0001).

As expected, participants in the current exclusive smokeless user group were predominately male, 94.6% and white, non-Hispanic (79.5%). The current exclusive cigarette smoker group had a higher proportion of male participants (53.1%) compared to females and 62.3% participants identified as being white non-Hispanic. The participants in the never tobacco user groups were reflective of the total sample population, for example this group had a larger percentage of females (58.0%) compared to males.

Approximately 70 % of participants in the current exclusive smokeless user group reside in the South and Midwest and 75% and 73% of the participants in this group was in the middle- or high-income category and had some type of private health insurance. These were higher than the

percentages in the other two tobacco user status groups. Also 73% of the current exclusive smokeless tobacco user group was overweight or obese compare to 62.1% and 63.9% of current exclusive cigarette smokers and never tobacco users respectively.

A statistically significant association between the tobacco use status groups and the categorical sociodemographic characteristic was observed at the 5% statistical significance level, see Table 13 and Table 14. A summary of the study participants sociodemographic characteristics by tobacco use status is presented in Table 13 and Table 14

Table 13:Descriptive Statistics of Study Sample, Among US Adults (≥ 18 years) by Sociodemographic and Tobacco Use Status before propensity score matching, 2011–2017

Characteristics	Total Sample		Never tobacco user		Current exclusive smokeless user		Current exclusive cigarette smoker		P-value
	Unweighted 68,866 (100%)[a]		Unweighted 54,810 (79.3%)[a]		Unweighted 633 (1.2%)[a]		Unweighted 13,423 (19.5%)[a]		
Age(years), mean (SE)	47.68 (0.18)		50.75 (0.24)*		46.18 (0.98)		46.50 (0.26)		<0.0001
Gender									
Male	28,909	(44.8%)	21,537	(42.0%)	570	(94.6%)	6,802	(53.1%)	<0.0001
Female	39,957	(55.2%)	33,273	(58.0%)	63	(5.4%)	6,621	(46.9%)	
Race/ethnicity									
White, non-Hispanic	29,436	(58.9%)	22,419	(57.8%)	464	(79.5%)	6,553	(62.3%)	<0.0001
Black, non-Hispanic	12,719	(11.3%)	9,983	(11.5%)	62	(4.5%)	2,674	(11.2%)	
Asian, non-Hispanic	8,797	(13.8%)	7,111	(13.9%)	56	(9.4%)	1,630	(13.9%)	
Other or multiple race, non-Hispanic	1,589	(2.3%)	1,139	(2.1%)	25	(2.9%)	425	(3.0%)	
Hispanic	16,325	(13.6%)	14,158	(14.8%)	26	(3.7%)	2,141	(9.5%)	
Marital status									
Married	28,105	(42.7%)	23,867	(45.4%)	273	(45.1%)	3,965	(31.5%)	<0.0001
Living with Partner	4,532	(6.0%)	3,126	(5.2%)	41	(6.2%)	1,365	(9.3%)	
Widow/divorce/separated	18,101	(25.7%)	13,859	(24.6%)	150	(21.7%)	4,092	(30.3%)	
Never Married	18,027	(25.5%)	13,868	(24.6%)	169	(27.0%)	3,990	(28.9%)	
Body mass index (BMI)									
Normal or Under Weight	23,504	(36.4%)	18,442	(36.2 %)	160	(26.5 %)	4,902	(37.9 %)	0.0037
Overweight	22,742	(32.9%)	18,155	(32.9 %)	227	(37.4%)	4,360	(32.8 %)	
Obesity	22,620	(30.7%)	18,213	(31.0 %)	246	(36.0%)	4,161	(29.3 %)	

Table 14:Descriptive Statistics of Study Sample, Among US Adults (≥ 18 years) by Sociodemographic and Tobacco Use Status before propensity score matching, 2011–2017 continued

Characteristics	Total Sample		Never tobacco user		Current exclusive smokeless user		Current exclusive cigarette smoker		P-value
	Unweighted 68,866 (100%)[a]		Unweighted 54,810 (79.3%)[a]		Unweighted 633 (1.2%)[a]		Unweighted 13,423 (19.5%)[a]		
Education level									
Less than high school	17,882	(24.6%)	14,009	(22.1%)	142	(18.9%)	3,731	(25.6%)	<0.0001
High school	13,430	(20.3%)	10,435	(18.1%)	156	(22.5%)	2,839	(21.1%)	
Some college	11,007	(17.7%)	8,609	(16.1%)	98	(16.8%)	2,300	(17.8%)	
College grad or higher	21,306	(37.4%)	17,782	(36.7%)	200	(35.9%)	3,324	(26.5%)	
Region of residency									
Northeast	11,263	(17.7%)	9,139	(18.3%)	67	(10.7%)	2,057	(15.8%)	<0.0001
Midwest	14,154	(22.6%)	10,557	(21.4%)	159	(22.9%)	3,438	(27.2%)	
South	26,189	(37.0%)	20,468	(36.3%)	295	(45.8%)	5,426	(39.3%)	
West	17,260	(22.7%)	14,646	(23.9%)	112	(20.7%)	2,502	(17.7%)	
Poverty Status									
Poor/Negative	14,005	(13.6%)	10,012	(11.5%)	90	(8.8%)	3,903	(22.6%)	<0.0001
Near Poor	4,357	(4.8%)	3,252	(4.4%)	37	(5.0%)	1,068	(6.8%)	
Low Income	11,207	(13.8%)	8,730	(13.3%)	80	(12.6%)	2,397	(16.3%)	
Middle Income	19,245	(28.6%)	15,391	(28.3%)	208	(32.6%)	3,646	(29.5%)	
High Income	20,052	(39.1%)	17,425	(42.6%)	218	(41.1%)	2,409	(24.9%)	

Characteristics	Total Sample		Never tobacco user		Current exclusive smokeless user		Current exclusive cigarette smoker		P-value
	Unweighted		Unweighted		Unweighted		Unweighted		
	68,866 (100%)[a]		54,810 (79.3%)[a]		633 (1.2%)[a]		13,423 (19.5%)[a]		
Self-reported binge drinking status[b]									
No	26,187	(40.8%)	21,292	(42.6%)	148	(25.8%)	4,747	(34.5%)	<0.0001
Yes	14,436	(24.2%)	9,223	(20.0%)	297	(49.5%)	4,916	(39.7%)	
Insurance type									
Uninsured	10,158	(10.5%)	7,526	(9.1%)	70	(10.9%)	2,562	(16.5%)	
Public	19,045	(22.2%)	14,334	(20.6%)	143	(15.8%)	4,568	(28.9%)	<0.0001
Private	39,663	(67.3%)	32,950	(70.3%)	420	(73.3%)	6,293	(54.6%)	

[a]Percentages were developed using survey weights to account for selection probabilities from the complex sampling design and to adjust for survey nonresponse

[b]Self-reported binge drinking status does not total 100% due to missing response.

*Statistical significance difference between never tobacco users and current exclusive smokeless user and never tobacco users and current exclusive cigarette smoker

SE= Standard error

The health status measures in the Self-Administered Questionnaire field in MEPS were the Short- Form 12 version 2 used to measure the quality of life, the Kessler Index of non-specific psychological distress and the Patient Health Index which assesses a person's tendency towards depression.

The quality of life scores in both the PCS and MCS tended to be slightly higher in the current exclusive smokeless tobacco group indicating slightly better health status compared to the other two tobacco status groups and the overall population. Current exclusive cigarette smokers had the lowest mean scores. The mean differences in the quality of life scores (i.e. both MCS and PCS) were found to be statistically significantly different between current exclusive cigarette smokers (MCS: 48.3 [95% CI: 48.0 to 48.7] , PCS: 46.8 [95% CI: 46.5 to 47.3]) vs current exclusive smokeless tobacco users (MCS: 52.6 [95% CI: 51.5 to 53.8] , PCS: 50.1 [95% CI: 48.7 to 51.5]) (p-value <0.0001) and current exclusive cigarette smokers vs never tobacco users (MCS: 51.9 [95% CI: 51.7 to 52.0] , PCS: 49.5 [95% CI: 49.3 to 49.7]) (p-value <0.0001). these differences may or may not be meaningful in terms of an individual's health status. A difference of 3 points has been determined to be a clinically meaningful difference for the PCS and MCS on the Short-Form 12.[69,70,71,72] A clinically meaningful difference of 4.3 points and 3.3 points in the MCS and PCS for current exclusive smokeless tobacco users vs current exclusive cigarette smokers was observed. A clinically meaningful difference was only found in the MCS for never tobacco users vs current exclusive cigarette smokers at 3.6 points. No statistically significant or clinically meaningful difference was observed between participants' mean scores in the never tobacco user group and current exclusive smokeless tobacco user group for both quality of life scores (MCS: p-value=0.1902 , PCS: p-value=0.4292).

Current exclusive cigarette smokers were found to have a greater tendency towards mental disability as indicated by their higher Kessler index compared to the never tobacco user group and current exclusive smokeless tobacco user group. These differences were statistically significantly different between exclusive cigarette smokers (4.7 [95% CI: 4.5 to 4.8]) vs current exclusive smokeless tobacco users 2.8 [95% CI: 2.4 to 3.3]) (p-value <0.0001) and current exclusive cigarette smokers vs never tobacco user (2.88 [95% CI: 2.82 to 2.93]) (p-value <0.0001). No statistically significant difference was observed between the Kessler index for never tobacco users and current exclusive smokeless tobacco users (p-value=0.8469).

Findings from the Patient Health index showed that the majority of participants in all three tobacco use groups had little tendency towards depression (i.e. 0 score in 66.1% of never tobacco user group, 65.8% of current exclusive smokeless tobacco user group and 52.5% of current exclusive cigarette). Approximately, 15% of current exclusive cigarette smokers had a Patient Health score of 3 or greater indicative of a positive screening for depression (i.e. greater tendency toward depression in cigarette smokers) compared to 6.2% and 6.1% for never tobacco users and current exclusive tobacco users respectively. Table 15 summarizes the study participants' health status characteristics by tobacco use status.

Of the 10 selected comorbidities identified in MEPS that were explored in this study, only 4 have positive self-reported disease diagnosis greater than 10% (i.e. arthritis, asthma , diabetes and any cancers). Never tobacco users (12.9%) and current exclusive cigarette smokers (10.7%) responded to having been diagnosed with any cancer more than current exclusive smokeless users (6.5%). Participants in the study had the highest positive diagnosis rates for arthritis, 29.8% participants reported having arthritis in never tobacco user group, 30.6 % in the current exclusive cigarette smoker group and 26.5% in the current exclusive smokeless user group.

Current exclusive cigarette smokers reported the highest positive diagnosis rates for emphysema (6.7%) compared to 2.5% in the current exclusive smokeless user group and 1.5% in the never tobacco user group. Similarly, the current exclusive cigarette smokers reported the highest positive diagnosis rates for asthma (11.8%) compared to 9.8% for current exclusive smokeless users and 10.4% in never tobacco users. Although the lung cancer variable had 88% missing data for the overall study population, current exclusive cigarette smoker reported the highest positive diagnosis rates of 4.0%.

There was a statistically significant association between tobacco use status and being diagnosed with any of the following 7 comorbidities (i.e. any cancer, coronary heart disease, myocardial infarction, stroke, emphysema, diabetes and asthma). No association was found between tobacco use status and arthritis and tobacco use status and angina. A statistical significance level was assessed at 5%.

Table 16 summarizes the study participants self-reported selected comorbidity characteristics (i.e. self-reported diagnosis history of selected diseases) by tobacco use status.

Table 15:Descriptive Statistics of Study Sample, Among US Adults (>=18 years) by Health Status and Tobacco Use Status before propensity score matching, 2011–2017

Characteristics	Total Sample		Never tobacco user		Current exclusive smokeless user		Current exclusive cigarette smoker	
	Unweighted		Unweighted		Unweighted		Unweighted	
	68,866 (100%)[a]		54,810 (79.3%)[a]		633 (1.2%)[a]		13,423 (19.5%)[a]	
Quality of life score (Short-Form 12)	**Mean (SE)**		**Mean (SE)**		**Mean (SE)**		**Mean (SE)**	
Physical Component Summary (0-100)	49.31 (0.08)		49.53 (0.11)		50.09 (0.70)		46.89 (0.21)*	
Mental Component Summary (0-100)	51.48 (0.06)		51.85 (0.07)		52.65 (0.59)		48.34 (0.17)*	
Mental illness score	**Mean (SE)**		**Mean (SE)**		**Mean (SE)**		**Mean (SE)**	
Kessler 6 Index	3.10 (0.03)		2.88 (0.03)		2.83 (0.24)		4.66 (0.07)*	
Depression score					**n (%)**			
Patient health Index								
0	42,728	(70.0%)	35,358	(66.1%)	406	(65.8%)	6,964	(52.5%)
1	6,222	(10.1%)	4,841	(9.0%)	49	(8.0%)	1,332	(10.0%)
2	7,348	(11.3%)	5,368	(9.3%)	54	(7.3%)	1,926	(14.6%)
3	1,982	(2.8%)	1,395	(2.2%)	-×	(1.9%)	570	(4.0%)
4	2,093	(3.0%)	1,329	(2.1%)	23	(2.6%)	741	(5.3%)
5	740	(1.0%)	481	(0.7%)	-×	(0.3%)	254	(1.8%)
6	1,270	(1.8%)	773	(1.2%)	-×	(1.3%)	489	(3.4%)

[a]Percentages were developed using survey weights to account for selection probabilities from the complex sampling design and to adjust for survey nonresponse

*Statistical significance difference between current exclusive cigarette smoker status and current exclusive smokeless user status and exclusive current cigarette smoker status and never tobacco user

×Sample size for this cell is < 20 and cannot be disclose based on Agency for Healthcare Research and Quality (AHRQ) guidelines

SE= Standard error

Table 16:Descriptive Statistics of Study Sample, Among US Adults (>=18 years) by Comorbidities and Tobacco Use Status before propensity score matching, 2011–2017

Characteristics	Total Sample		Never tobacco user		Current exclusive smokeless user		Current exclusive cigarette smoker		P-Value
	Unweighted		Unweighted		Unweighted		Unweighted		
	68,866 (100%)[a]		54,810 (79.3%)[a]		633 (1.2%)[a]		13,423 (19.5%)[a]		
Any Cancer Diagnosis									
Yes	7,239	(12.4%)	5,902	(12.9%)	46	(6.5%)	1,291	(10.7%)	<0.0001
No	61,516	(87.6%)	48,824	(86.9%)	586	(93.4%)	12,106	(89.1%)	
Cardiovascular Disease Diagnosis									
Angina									
Yes	1,811	(2.7%)	1,426	(2.7%)	-[x]	(1.6%)	372	(2.8%)	0.1650
No	66,951	(97.3%)	53,306	(97.1%)	619	(98.3%)	13,026	(97.0%)	
Coronary heart disease									
Yes	4,030	(6.0%)	3,223	(6.1%)	27	(3.3%)	780	(5.8%)	0.0076
No	64,726	(94.0%)	51,506	(93.8%)	605	(96.6%)	12,615	(94.1%)	
Myocardial Infarction									
Yes	2,831	(4.2%)	2,118	(4.0%)	25	(3.0%)	688	(5.2%)	0.0010
No	65,933	(95.8%)	52,613	(95.9%)	607	(96.9%)	12,713	(94.6%)	
Stroke									
Yes	3,299	(4.7%)	2,525	(4.6%)	21	(2.4%)	753	(5.2%)	0.0124
No	65,468	(95.3%)	52,209	(95.2%)	611	(97.4%)	12,648	(94.7%)	
Emphysema Diagnosis									
Yes	1,588	(2.5%)	772	(1.5%)	-[x]	(2.5%)	800	(6.7%)	<0.0001
No	67,178	(97.5%)	53,964	(98.4%)	616	(97.4%)	12,598	(93.2%)	

Characteristics	Total Sample		Never tobacco user		Current exclusive smokeless user		Current exclusive cigarette smoker		P-Value
	Unweighted		Unweighted		Unweighted		Unweighted		
	68,866 (100%)[a]		54,810 (79.3%)[a]		633 (1.2%)[a]		13,423 (19.5%)[a]		
Asthma Diagnosis									
Yes	7,248	(10.6%)	5,505	(10.4%)	64	(9.8%)	1,679	(11.8%)	0.0210
No	61,526	(89.4%)	49,237	(89.5%)	568	(90.0%)	11,721	(88.1%)	
Arthritis Diagnosis									
Yes	19,934	(30.0%)	15,623	(29.8%)	109	(26.5%)	4,121	(30.6%)	0.2246
No	48,821	(70.0%)	39,109	(70.1%)	442	(73.4%)	9,270	(69.2%)	
Diabetes Diagnosis									
Yes	8,216	(10.7%)	6,708	(10.9%)	81	(12.0%)	1,427	(9.4%)	0.0018
No	60,555	(89.3%)	44,030	(88.9%)	551	7 (87.8%)	11,974	(90.5%)	
Lung cancer	**n=83,948,815**		**n=69,329,221**		**n= 537,005**		**n=14,082,588**		
Yes	204	(2.2%)	142	(1.9%)	-[×]	(0.0%)	62	(4.0%)	
No	7,035	(97.8%)	5,760	(98.1%)	46	(100.0%)	1,229	(96.0%)	

[a]Percentages were developed using survey weights to account for selection probabilities from the complex sampling design and to adjust for survey nonresponse

[×]Sample size for this cell is < 20 and cannot be disclose based on Agency for Healthcare Research and Quality (AHRQ) guidelines

Propensity Score Matching of Tobacco Use Groups

The study population before propensity score matching consisted of current exclusive smokeless tobacco users (N=633 unweighted) , current exclusive cigarette smokers (N=13,423 unweighted) and never tobacco users (N=54,810 unweighted). For this analysis current exclusive smokeless tobacco users were considered as the "group 1-case", current exclusive cigarette smokers "group 2-control", and never tobacco users "group 3-control ". Two sets of propensity score matching analyses were conducted and the matched datasets were pooled to obtain the final dataset used for the main analysis of the study.

The relatively small sample observed in the group 1 differed in their sociodemographic characteristic - age, sex, race/ethnicity, region of residency, poverty status, insurance type and BMI - compared to the groups 2 and 3. Individuals in group 1 were younger, predominately male white non-Hispanic, largely resided in the South or Midwest, were in a middle to high income bracket, possessed some type of health insurance and were in the overweight or obese BMI category. Propensity scores for the case (i.e. group 1) and controls (i.e. group 2 and group 3) were computed using logistic regression adjusting for age, gender, region of residency, poverty status, BMI category and insurance type. Other variables like binge drinking status, health status and comorbidities were not included in the matching analysis given their high nonresponse rate or not being diagnosed with a particular comorbidity quite a few of the potential match pairs would be lost . Since matching was with grp1 (current exclusive smokeless tobacco users, N=633) a significant portion of the already small matching sample would be lost. All sociodemographic and health status variables were adjusted for in the final econometric models.

Similar matching analysis was conducted including only participants who were 65 years of age or less. This set of matched data was used in a sensitivity analysis to assess health care expenditure without the inclusion of potentially high values based on the increase of healthcare cost with aging.[49]

Propensity score matching diagnostics

Group 1 matching on Group 2

Of the 13,423 observations in Group 2, 12,846 fell within the support region in which matching was assessed. The mean difference between the propensity scores was 0.0001 for the matched observations (N=633). The standardized mean differences were significantly reduced in the matched observations, and the largest of these differences was 0.05118 , which was less than the recommended upper limits of 0.25 or 0.1.[61,62] The group 1-to- group 2 variance ratios were 1.2434 and 1.0012 in the matched observations, which are within the recommended range of 0.5 to 2.

Group 1 matching on Group 3

Of the 54,810 observations in Group 3, 53,845 fell within the support region in which matching was assessed. The mean difference between the propensity scores was 0.00001 for the matched observations (N=633). The standardized mean differences were significantly reduced in the matched observations, and the largest of these differences was 0.02150 in absolute value , which was less than the recommended upper limit of 0.25 or 0.1.[61,62] The group 1-to-group 3 variance ratios were 0.9648 and 1.0000 in the matched observations, which are within the recommended range of 0.5 to 2.

Figure 5 and Figure 6 illustrate the distribution of the propensity scores after matching between the case and two controls. Both show overlapping distributions indicating that the common support assumption holds true.

Figure 7 is a plot of the standardized mean differences in gender, age and the logit of the propensity score for all observations and matched observations. Figure 8 displays box plots that compare the distributions of the logit propensity score for units in the case and control groups, based on all observations, on observations in the support region, and on matched observations. the 4 figures below show the distributions are well balanced for the matched observations, indicating that the matching procedure balanced the covariates across the groups.

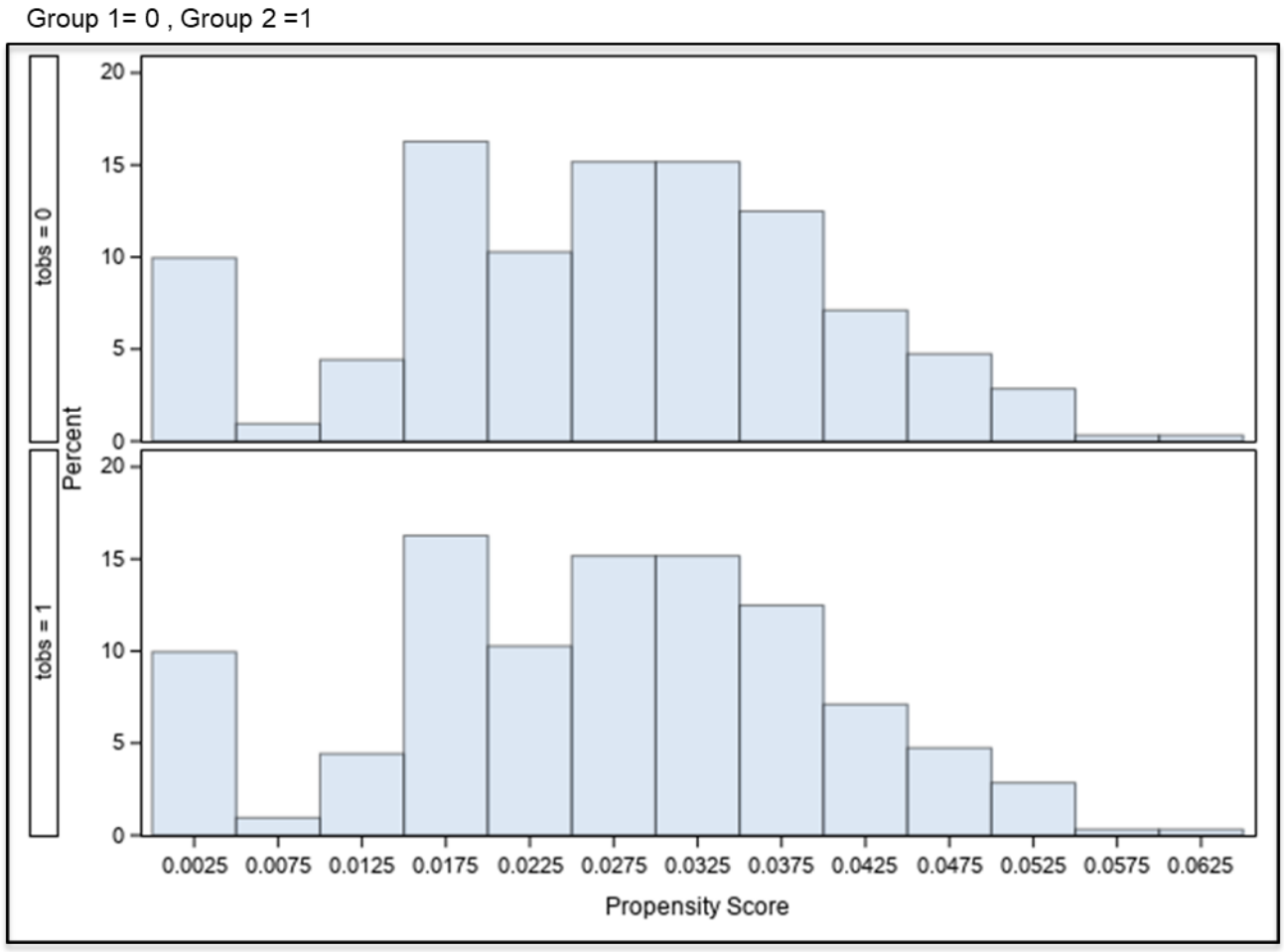


Figure 5: Distribution of propensity scores after matching Group 1 vs Group 2

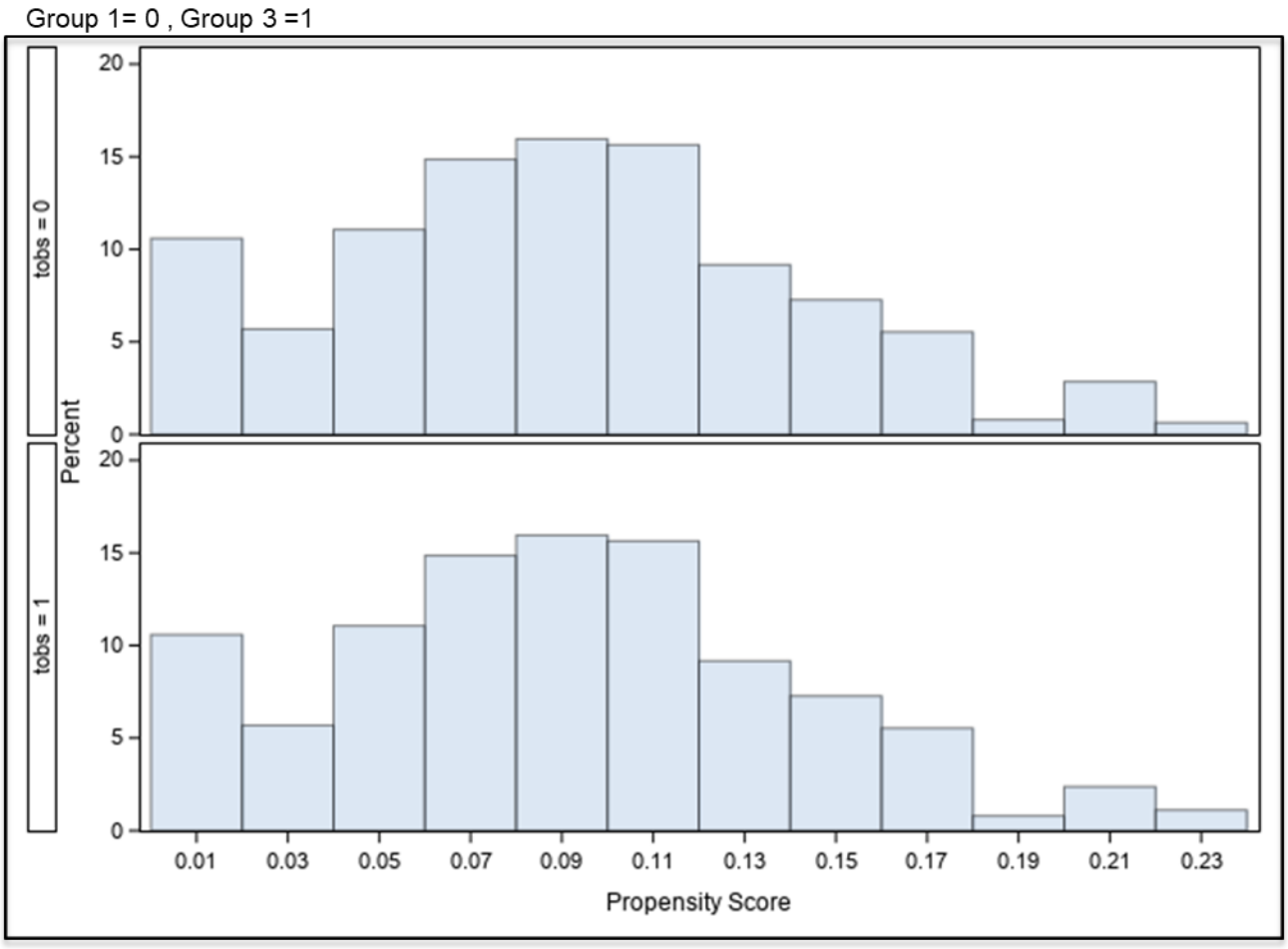


Figure 6: Distribution of propensity scores after matching Group 1 vs Group 3

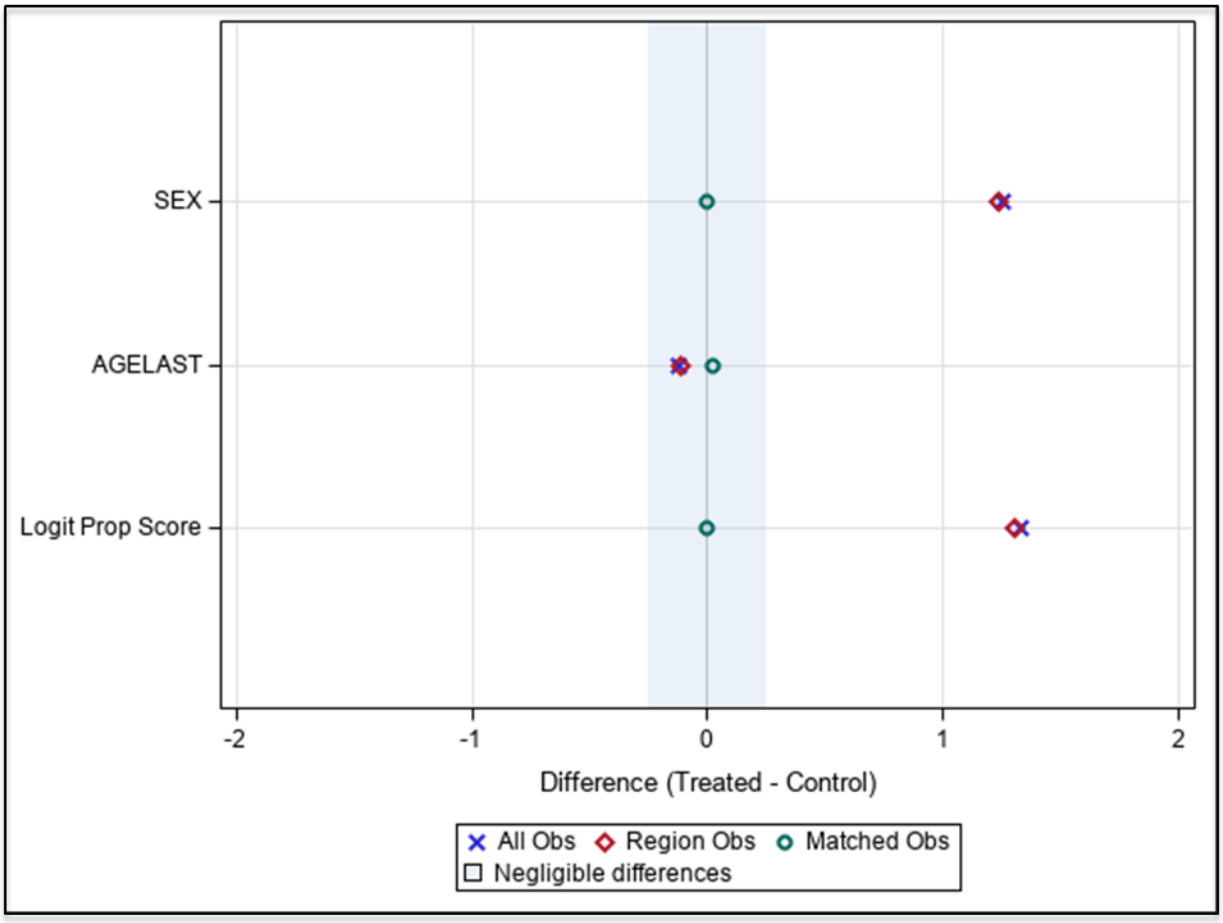


Figure 7: Standardized Mean Differences (Group 1- Group 3) Plot

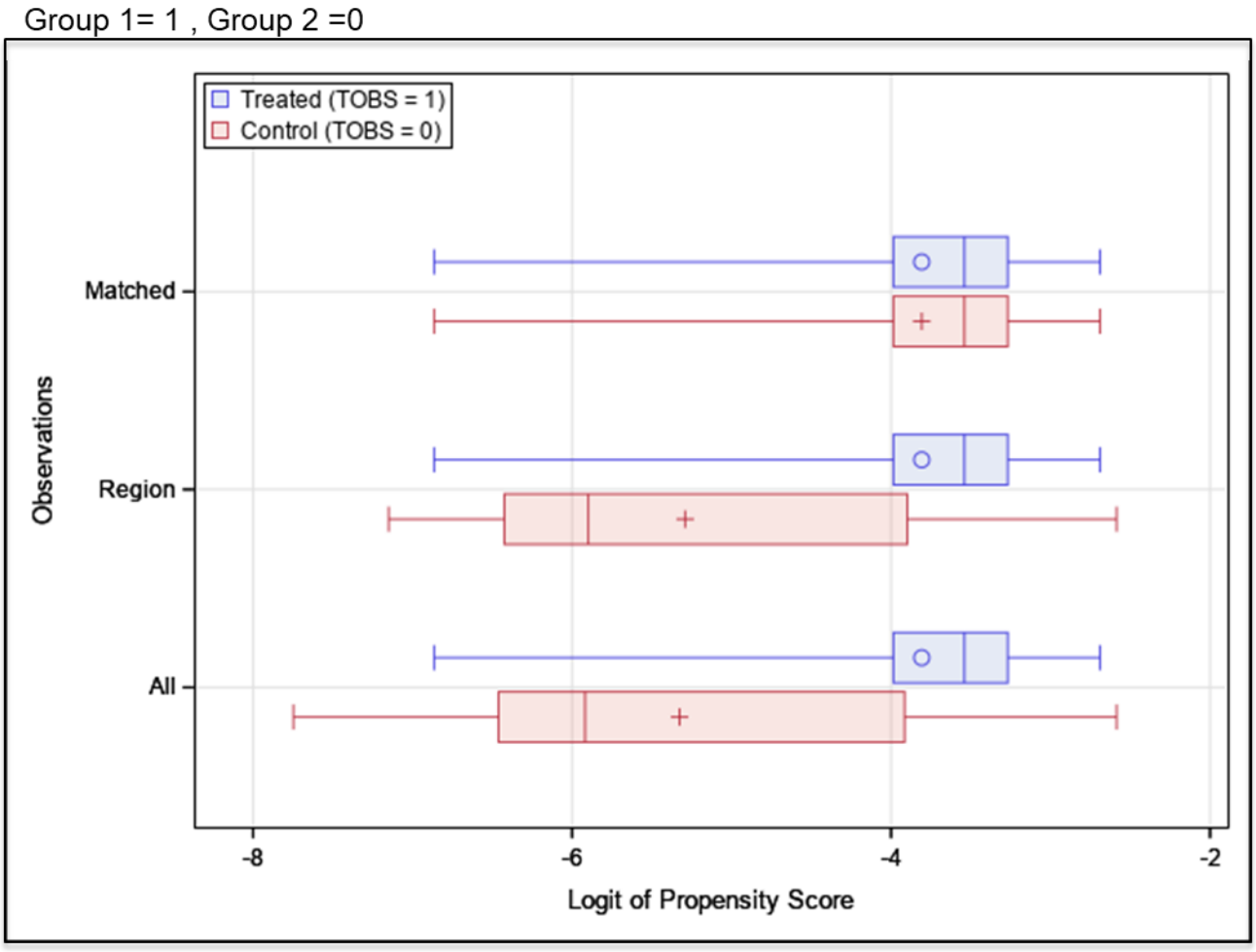


Figure 8: Distribution of Logit of Propensity Score

Study Sample After Matching Tobacco Use Groups

After matching, the sample size reduced significantly as expected given that current exclusive smokeless tobacco users, the smallest tobacco use status group, was used as the case group. The total sample size after matching was N=1899 participants . After matching the three groups had equal number of participants (N=633).

The descriptive and bivariate results showed no statistically significant difference or association between the three tobacco use status groups and most of the sociodemographic characteristics including all the covariates used in the matching analysis. An association between the three tobacco groups with education level, self-reported binge drinking status and race was still observed (see Table 18 and Table 17).

The health status characteristic after matching produced similar trends as observed before matching. The quality of life scores in both the PCS and MCS after matching tended to be slightly higher in never tobacco users indicating slightly better health status compared to the other two tobacco status group and the overall population. Before matching current exclusive smokeless user had slightly higher quality of life scores.

The mean quality of life scores increased for both current exclusive cigarette smokers and never tobacco users, current exclusive cigarette smokers still had the lowest mean scores. The mean differences in the quality of life scores were still found to be statistically significantly different but not clinically meaningful (PCS= 2.1 points and MCS= 1.7 points difference) between current exclusive cigarette smokers and current exclusive smokeless tobacco users ((PCS: 48.0 [95% CI: 46.9 to 49.1] vs PCS: 50.1 [95% CI: 48.7 to 51.5]) (p-value =0.0183)), ((MCS: 50.9 [95% CI: 49.8 to 51.9] vs MCS: 52.6 [95% CI: 51.5 to 53.8]) p-value (0.0282)). They were also

statistically significantly different between the mean scores of current exclusive cigarette smokers and never tobacco users ((PCS: 48.0 [95% CI: 46.9 to 49.1] vs PCS: 50.50 [95% CI: 49.2 to 51.8]) (p-value=0.0044)), ((MCS: 50.9 [95% CI: 49.8 to 51.9] vs MCS: 52.8 [95% CI: 51.8 to 53.7]) (p-value =0.0050)) but no clinically meaningful difference (PCS= 2.5 points and MCS= 1.9 points). No statistically significant or clinically meaningful difference was observed between participants' mean scores in the never tobacco user group and current exclusive smokeless tobacco user group for either quality of life score (MCS: p-value=0.8709 , PCS: p-value=0.6866).

After matching, current exclusive cigarette smokers still had a greater tendency towards mental disability as indicted by their higher Kessler index compared to never tobacco users and current exclusive smokeless tobacco users. These differences were statistically significantly different between current exclusive cigarette smokers (3.8 [95% CI: 3.4 to 4.2]) vs current exclusive smokeless tobacco users 2.8 [95% CI: 2.4 to 3.3]) (p-value =0.0021) and current exclusive cigarette smokers vs never tobacco user (2.7 [95% CI: 2.82 to 2.93]) (p-value <0.0001). No statistically significant difference was observed between the Kessler index for participants in the never tobacco user group and current exclusive smokeless tobacco user group (p=0.5233).

After matching, the Patient Health index showed that the majority of participants in all three tobacco use groups had little tendency towards depression (i.e. 0 score in 71.3% of never tobacco user group, 65.8% of current exclusive smokeless tobacco user group and 58% of current exclusive cigarette smokers). Current exclusive cigarette smokers (8%) still had greater tendency toward depression (i.e. those who had a patient health score of 3 or greater) compared to 4.2% and 4.8% for never tobacco users and current exclusive tobacco users respectively. Table 19

summarizes the study participants health status characteristics by tobacco use status after matching.

The sample size for the selected comorbidities was also reduced after matching as expected. Only 2 of the 10 have positive self-reported disease diagnosis greater than 10% (i.e. arthritis and diabetes) compared to the 4 observed before matching. Participants in the study still had the highest positive diagnosis rates for arthritis, 20.7% participants reported having arthritis in never tobacco user group, 28.8 % in the current exclusive cigarette smoker group and 26.5% in the current exclusive smokeless user group.

Current exclusive cigarette smokers also still reported the highest positive diagnosis rates for emphysema (6.9%) compared to 2.5% in the current exclusive smokeless user group and 0.9% in the never tobacco user group. Similarly, the current exclusive cigarette smokers reported the highest positive diagnosis rates for coronary heart disease (7.4%) compared to 3.3 % for current exclusive smokeless users and 4.3% in never tobacco users.

After matching, a statistically significant association was found between tobacco use status and having reported being diagnosed with emphysema, coronary heart disease or arthritis at a 5% statistical significance level. No association was found between tobacco use status and with having reported being diagnosed with the other 7 selected comorbidities examined in this study.

Table 20 summarizes the study participants self-reported selected comorbidity characteristics (i.e. self-reported diagnosis history of selected diseases) by tobacco use status after matching.

Table 17:Descriptive Statistics of Study Sample, Among US Adults (≥ 18 years) by Sociodemographic and Tobacco Use Status after propensity score matching, 2011–2017

Characteristics	Total Sample		Never tobacco user		Current exclusive smokeless user		Current exclusive cigarette smoker		P-value
	Unweighted : 1,899 (100%)[a]		Unweighted 633 (%)[a]		Unweighted 633 (%)[a]		Unweighted 633 (%)[a]		
Age(years), mean (SE)	46.69 (0.47)		46.65 (0.80)		46.18 (0.98)		47.33 (0.69)		0.5164
Gender									
Male	1710	(93.7%)	570	(94.0%)	570	(94.6%)	570	(92.3%)	0.4063
Female	189	(6.3%)	63	(6.0%)	63	(5.4%)	63	(7.7%)	
Race/ethnicity									
White, non-Hispanic	1048	(67.0%)	262	(59.0%)	464	(79.5%)	322	(60.4%)	<0.0001
Black, non-Hispanic	311	(8.8%)	143	(13.1%)	62	(4.5%)	106	(9.8%)	
Asian, non-Hispanic	234	(13.8%)	79	(12.7%)	56	(9.4%)	99	(17.4%)	
Other or multiple race, non-Hispanic	60	(3.1%)	-[×]	(3.0%)	25	(2.9%)	22	(3.5%)	
Hispanic	264	(8.1%)	136	(12.2%)	26	(3.7%)	84	(9.3%)	
Marital status									
Married	789	(43.1%)	296	(47.1%)	273	(45.1%)	229	(37.0%)	0.1016
Living with Partner	138	(6.7%)	37	(5.6%)	41	(6.2%)	60	(8.3%)	
Widow/divorce/separated	444	(22.5%)	133	(21.0%)	150	(21.7%)	161	(25.1%)	
Never Married	519	(27.6%)	167	(26.3%)	169	(27.0%)	183	(29.6%)	
Body mass index (BMI)									
Normal or Under Weight	491	(25.3%)	155	(24.3%)	160	(26.5 %)	176	(31.5%)	0.8805
Overweight	685	(37.0%)	231	(35.9%)	227	(37.4%)	227	(32.5%)	
Obesity	723	(37.6%)	247	(39.7%)	246	(36.0%)	230	(31.9%)	

[a]Percentages were developed using survey weights to account for selection probabilities from the complex sampling design and to adjust for survey nonresponse , [×]Sample size for this cell is < 20 and cannot be disclose based on Agency for Healthcare Research and Quality (AHRQ) guidelines

SE= Standard error

Table 18: Descriptive Statistics of Study Sample, Among US Adults (≥ 18 years) by Sociodemographic and Tobacco Use Status after matching, 2011–2017, continued

Characteristics	Total Sample		Never tobacco user		Current exclusive smokeless user		Current exclusive cigarette smoker		P-value
	Unweighted 1,899 (100%)[a]		Unweighted 633 (%)[a]		Unweighted 633 (%)[a]		Unweighted 633 (%)[a]		
Education level									
Less than high school	467	(23.1%)	142	(19.5%)	142	(18.9%)	183	(26.4%)	
High school	431	(23.4%)	133	(19.4%)	156	(22.5%)	142	(23.1%)	0.0078
Some college	282	(17.2%)	104	(18.6%)	98	(16.8%)	80	(12.4%)	
College grad or higher	573	(36.4%)	203	(35.6%)	200	(35.9%)	170	(29.7%)	
Region of residency									
Northeast	191	(10.5%)	70	(10.2%)	67	(10.7%)	54	(10.4%)	
Midwest	472	(25.5%)	158	(28.4%)	159	(22.9%)	155	(25.8%)	0.5230
South	902	(46.5%)	295	(45.1%)	295	(45.8%)	312	(48.6%)	
West	334	(17.6%)	110	(16.3%)	112	(20.7%)	112	(15.1%)	
Poverty Status									
Poor/Negative	278	(9.6%)	93	(9.0%)	90	(8.8%)	95	(11.1%)	
Near Poor	94	(4.0%)	24	(2.8%)	37	(5.0%)	33	(4.2%)	
Low Income	251	(11.3%)	84	(10.1%)	80	(12.6%)	87	(11.0%)	0.2468
Middle Income	627	(31.6%)	210	(31.2%)	208	(32.6%)	209	(30.7%)	
High Income	649	(43.5%)	220	(46.9%)	218	(41.1%)	211	(43.0%)	
Self-reported binge drinking status[b]									
No	563	(42.1%)	225	(39.5%)	148	(25.8%)	190	(29.3%)	<0.0001

Characteristics	Total Sample		Never tobacco user		Current exclusive smokeless user		Current exclusive cigarette smoker		P-value
	Unweighted 1,899 (100%)[a]		Unweighted 633 (%)[a]		Unweighted 633 (%)[a]		Unweighted 633 (%)[a]		
Yes									
	753	(57.9%)	153	(27.2%)	297	(49.5%)	303	(49.8%)	
Insurance type									
Uninsured	195	(8.6%)	63	(6.8%)	70	(10.9%)	62	(7.5%)	
Public	396	(15.3%)	128	(13.6%)	143	(15.8%)	125	(16.3%)	0.2840
Private	1308	(76.1%)	442	(79.5%)	420	(73.3%)	446	(76.1%)	

[a]Percentages were developed using survey weights to account for selection probabilities from the complex sampling design and to adjust for survey nonresponse

[b]Self-reported binge drinking status does not total 100% due to missing response.

Table 19:Descriptive Statistics of Study Sample, Among US Adults (≥ 18 years) by Health Status and Tobacco Use Status after propensity score matching, 2011–2017

Characteristics	Total Sample		Never tobacco user		Current exclusive smokeless user		Current exclusive cigarette smoker	
	Unweighted 1,899 (100%)[a]		Unweighted 633 (%)[a]		Unweighted 633 (%)[a]		Unweighted 633 (%)[a]	
Quality of life score (Short-Form 12)	**Mean (SE)**		**Mean (SE)**		**Mean (SE)**		**Mean (SE)***	
Physical Component Summary (0-100)	49.52 (0.35)		50.50 (0.66)		50.09 (0.70)		48.02 (0.54)*	
Mental Component Summary (0-100)	52.08 (0.32)		52.77 (0.47)		52.65 (0.59)		50.86 (0.53)*	
Mental illness score	**Mean (SE)**		**Mean (SE)**		**Mean (SE)**		**Mean (SE)**	
Kessler 6 Index	3.10 (0.13)		2.65 (0.16)		2.83 (0.24)		3.82 (0.22)	
Depression score					**n (%)[a]**			
Patient health Index								
0	1215	(71.8%)	438	(71.3%)	406	(65.8%)	371	(58%)
1	162	(10.3%)	52	(8.6%)	49	(8.0%)	61	(11.5%)
2	184	(9.7%)	55	(6.6%)	54	(7.3%)	75	(12.6%)
3	53	(2.4%)	-ˣ	(2.0%)	-ˣ	(1.9%)	22	(2.5%)
4	61	(3.0%)	-ˣ	(1.8%)	23	(2.6%)	27	(3.7%)
5	-ˣ	(0.8%)	-ˣ	(0.4%)	-ˣ	(0.3%)	-ˣ	(1.3%)
6	38	(2.1%)	-ˣ	(2.1%)	-ˣ	(1.3%)	-ˣ	(2.5%)

[a]Percentages were developed using survey weights to account for selection probabilities from the complex sampling design and to adjust for survey nonresponse

*Statistical significance difference between current exclusive cigarette smoker status and current exclusive smokeless user status and exclusive current cigarette smoker status and never tobacco user

ˣSample size for this cell is < 20 and cannot be disclose based on Agency for Healthcare Research and Quality (AHRQ) guidelines

SE= Standard error

Table 20:Descriptive Statistics of Study Sample, Among US Adults (>=18 years) by Comorbidities and Tobacco Use Status after propensity score matching, 2011–2017

Characteristics	Total Sample		Never tobacco user		Current exclusive smokeless user		Current exclusive cigarette smoker		P-Value
	Unweighted 1,899 (100%)[a]		Unweighted 633 (%)[a]		Unweighted 633 (%)[a]		Unweighted 633 (%)[a]		
Any Cancer Diagnosis									
Yes	159	(8.4%)	56	(8.5%)	46	(6.5%)	57	(10.4.%)	0.1424
No	1,737	(91.6%)	576	(91.5%)	586	(93.4%)	12,106	(89.6%)	
Cardiovascular Disease Diagnosis									
Angina									
Yes	53	(2.6%)	-[x]	(3.3%)	-[x]	(1.6%)	22	(3.2%)	0.0947
No	1,843	(97.4%)	614	(96.7%)	619	(98.3%)	610	(96.8%)	
Coronary heart disease									
Yes	99	(4.9%)	31	(4.3%)	27	(3.3%)	41	(7.4%)	0.0238
No	1797	(95.1%)	601	(95.7%)	605	(96.6%)	591	(92.6%)	
Myocardial Infarction									
Yes	83	(3.9%)	23	(3.3%)	25	(3.0%)	35	(5.4%)	0.1410

Characteristics	Total Sample		Never tobacco user		Current exclusive smokeless user		Current exclusive cigarette smoker		P-Value
	Unweighted 1,899 (100%)[a]		Unweighted 633 (%)[a]		Unweighted 633 (%)[a]		Unweighted 633 (%)[a]		
No	1813	(95.8%)	609	(96.7%)	607	(96.9%)	632	(94.6%)	
Stroke									
Yes	74	(3.4 %)	26	(3.9%)	21	(2.4%)	27	(4.1%)	0.2718
No	1822	(96.6%)	606	(96.1%)	611	(97.4%)	605	(95.9%)	
Emphysema Diagnosis									
Yes	65	(3.4%)	-[x]	(0.9%)	-[x]	(2.5%)	41	(6.9%)	<0.0001
No	1831	(96.6%)	624	(99.1%)	616	(97.4%)	591	(93.1%)	
Asthma Diagnosis									
Yes	171	(8.8%)	54	(8.0%)	64	(9.8%)	53	(8.3%)	0.6850
No	1726	(91.2%)	578	(92.0%)	568	(90.0%)	579	(91.7%)	
Arthritis Diagnosis									
Yes	482	(25.4%)	129	(20.7%)	109	(26.5%)	163	(28.8%)	0.0363
No	1413	(74.6%)	503	(79.3%)	442	(73.4%)	468	(71.2%)	

Characteristics	Total Sample		Never tobacco user		Current exclusive smokeless user		Current exclusive cigarette smoker		P-Value
	Unweighted 1,899 (100%)[a]		Unweighted 633 (%)[a]		Unweighted 633 (%)[a]		Unweighted 633 (%)[a]		
Diabetes Diagnosis									
Yes	230	(11.3%)	91	(12.2%)	81	(12.0%)	58	(9.5%)	0.4287
No	1666	(88.7%)	541	(87.8%)	551	(87.8%)	574	(90.5%)	
Lung cancer	**n=1,852,835**		**n=577,983**		**n= 537,005**		**n=737,847**		
Yes	-[×]	(2.7%)	-[×]	(1.7 %)	-[×]	0 (0.0%)	-[×]	(5.5%)	-
No	155	(97.3%)	55	(98.3%)	46	(100.0%)	54	(94.5%)	

[a]Percentages were developed using survey weights to account for selection probabilities from the complex sampling design and to adjust for survey nonresponse

[×]Sample size for this cell is < 20 and cannot be disclose based on Agency for Healthcare Research and Quality (AHRQ) guidelines

The next sections of this chapter will describe the results of Specific Aim 2: I-II and Specific Aim 1: V.

Health Care Utilization Results

Description of Health Care Utilization

Before Propensity Score matching

Table 21 shows the utilization rates and mean health care utilization for the 6 types of health care services by tobacco use status group examined in this study before matching. The utilization rates are presented as the percentages of participants that had at least 1 visit in the last 12 months for a given health care service. Mean health care utilization is based on individuals who have at least 1 visit in the last 12 months.

Current exclusive cigarette smokers had the highest ER utilization rate (21.8%) compared to 14.4 % and 13.3 % for never tobacco users and current exclusive smokeless tobacco users respectively. Of those who had an ER visit, never tobacco users had a mean of 1.44 ER visits, current exclusive smokeless users had a mean of 1.33 ER visits and current exclusive cigarette smokers had the highest mean value at 1.60 ER visits. This difference in mean ER visits was found to be statistically significant between current exclusive cigarette smokers vs never tobacco users (p-value <0.0001) and current exclusive cigarette smokers vs current exclusive smokeless users (p-value =0.0005). No significant difference was found between never tobacco users and current exclusive smokeless users.

Current exclusive smokeless users had the lowest office-based utilization rate (66.7%) compared to 79.6 % and 71.7 % for never tobacco users and current exclusive cigarette smokers respectively. Of those who had an office-based visit, never tobacco users had the highest mean

number of office-based visits at 9.32, smokeless users had a mean of 6.08 office -based visits and current exclusive cigarette smokers had a mean of 8.25 office- based visits. This difference in mean office-based visits was found to be statistically significant between current exclusive cigarette smokers vs never tobacco users (p-value <0.0001), current exclusive cigarette smokers vs current exclusive smokeless users (p-value =0.0044) and never tobacco users vs current exclusive smokeless users (p-value <0.0001).

Hospital outpatient utilization rates were lowest in current exclusive smokeless users (11.1%) and comparable for never tobacco users (20.9%) and current exclusive cigarette smokers (19.0%). In the past 12 months, current exclusive smokeless users had a mean of 2.41 hospital outpatient visits and never tobacco users and current exclusive cigarette smokers had a mean of 3.24 and 3.22 hospital outpatient visits respectively. There were statistically significant differences in mean hospital outpatient visits between current exclusive cigarette smokers vs current exclusive smokeless users (p-value =0.0237) and never tobacco users vs current exclusive smokeless users (p-value =0.0190).

Hospital inpatient utilization rates were lowest in current exclusive smokeless users (5.5%) and comparable for never tobacco users (8.6%)and current exclusive smokers (8.9%). In the past 12 months, current exclusive smokeless users had a mean of 5.42 hospital inpatient visits and never tobacco users and current exclusive cigarette smokers had a mean of 6.48 and 8.69 hospital inpatient visits respectively. There were statistically significant differences in mean hospital inpatient visits between current exclusive cigarette smokers vs current exclusive smokeless users (p-value =0.0315) and current exclusive cigarette smokers vs never tobacco users (p-value =0.0005).

Dental utilization rates were highest in never tobacco users (46.2%) and comparable for current exclusive cigarette smokers (29.5%)and current exclusive smokeless tobacco users (30.6%). In the past 12 months, the mean dental visits for never tobacco users, current exclusive cigarette smokers and current exclusive smokeless users were 2.28, 2.32 and 1.86 respectively. There were statistically significant differences in mean dental visits between current exclusive cigarette smokers vs current exclusive smokeless users (p-value <0.0001) and never tobacco users vs current exclusive smokeless users (p-value <0.0001).

Table 21:Health Care Utilization before Propensity Score Matching by Types of Health Care Service and Tobacco Use Status among US Adults, 2011–2017

Health Care Service	Total Sample N=68,866	Never Tobacco User N= 54,810	Current exclusive smokeless user N= 633	Current exclusive cigarette smoker N= 13,423
Hospital Outpatient visits (in the past 12 months)				
% with ≥ 1 visit	20.4	20.9	11.1ᵍ	19.0ϕ
Mean number of visits (se)	3.13 (0.06)	3.24(0.08)	2.41 (0.33)	3.22 (0.15)
Hospital Inpatient visits (in the past 12 months)				
% with ≥ 1 visit	8.9	8.6	5.5	10.1ϕº
Mean number of visits (se)	6.80 (0.15)	6.48 (0.19)	5.42 (1.38)	8.69 (0.55)
Emergency room (ER) visits (in the past 12 months)				
% with ≥ 1 visit	15.8	14.4	13.3ᵍ	21.8ϕ
Mean number of visits (se)	1.45 (0.01)	1.44 (0.01)	1.33 (0.07)	1.60 (0.03)
Office Based visits (in the past 12 months)				
% with ≥ 1 visit	77.9	79.6	66.7ᵍ	71.7*
Mean number of visits (se)	8.48 (0.08)	9.32 (0.11)	6.08 (0.73)	8.25(0.21)
Home Health care visits (in the past 12 months)				
% with ≥ 1 visit	4.2	4.3	2.4	3.8
Mean number of visits (se)	70.38 (2.57)	73.02 (3.34)	58.68 (12.33)	73.12 (6.45)
Dental visits (in the past 12 months)				
% with ≥ 1 visit	42.8	46.2	30.6ᵍ	29.5ϕ
Mean number of visits (se)	2.21 (0.01)	2.28 (0.02)	1.86 (0.08)	2.32 (0.05)

*Statistical significance difference between current exclusive cigarette smoker status and current exclusive smokeless user status and exclusive current cigarette smoker status and never tobacco user

ᵍ Statistical significance difference between never tobacco user and current exclusive smokeless user status ° Statistical significance difference between never tobacco user and current exclusive cigarette smoker status ϕ Statistical significance difference between current exclusive cigarette smoker status and current exclusive smokeless user status se= Standard error

After Propensity Score matching

Table ***22*** shows the utilization rates and mean health care utilization by health care services by tobacco use status after the tobacco status groups were balanced.

After matching, current exclusive cigarette smokers still had the highest ER utilization rate (16.4%) compared to 12.0 % and 13.3 % for never tobacco users and current exclusive smokeless tobacco users respectively. Of those who had an ER visit, never tobacco users had a mean of 1.25 ER visits, current exclusive smokeless users had a mean of 1.33 ER visits and current exclusive cigarette smokers had the highest mean value at 1.50 ER visits. A statistically significant difference was still found between current exclusive cigarette smokers vs never tobacco users (p-value =0.0009) and current exclusive cigarette smokers vs current exclusive smokeless users (p-value =0.0161), with no significant difference found between never tobacco users and current exclusive smokeless users.

Hospital outpatient utilization rates were still lowest in current exclusive smokeless users (11.1%) and comparable for never tobacco users (17.9%) and current exclusive cigarette smokers (14.1%) after matching. In the past 12 months, current exclusive smokeless users had a mean of 2.41 hospital outpatient visits and never tobacco users and current exclusive cigarette smokers had a mean of 3.29 and 3.32 hospital outpatient visits respectively. A statistically significant difference in mean hospital outpatient visits between current exclusive cigarette smokers vs current exclusive smokeless users (p-value =0.0051) was still observed. Never tobacco users vs current exclusive smokeless users was no longer statistically significantly different.

Hospital inpatient utilization rates were still lowest in current exclusive smokeless users (5.5%) and comparable for never tobacco users (6.9%)and current exclusive smokers (8.6%). In the past 12 months, current exclusive smokeless users had a mean of 5.42 hospital inpatient visits and never tobacco users and current exclusive cigarette smokers had a mean of 6.48 and 8.69 hospital inpatient visits respectively. A statistically significant differences in mean hospital inpatient visits between current exclusive cigarette smokers vs current exclusive smokeless users (p-value =0.0114) and current exclusive cigarette smokers vs never tobacco users (p-value =0.0085) was also still observed after matching.

After matching, dental utilization rates were still highest in never tobacco users (42.2%) and 28.2 % and 30.6% for current exclusive cigarette smokers and current exclusive smokeless tobacco users respectively. In the past 12 months, mean dental visits for never tobacco users, current exclusive cigarette smokers and current exclusive smokeless users were 2.11, 2.19 and 1.86 respectively. A statistically significant difference in mean dental visits was still found between current exclusive cigarette smokers vs current exclusive smokeless users (p-value =0.0365).

After matching, no statistically significant differences in mean office-based visit was found between current exclusive cigarette smokers vs never tobacco users, current exclusive cigarette smokers vs current exclusive smokeless users or never tobacco users vs current exclusive smokeless users. The results of home health care visit remained the same as before matching, were no statistically significant differences in mean home health care visit between tobacco use status group was observed.

Table 22:Health Care Utilization after Propensity Score Matching by Types of Health Care Service and Tobacco Use Status among US Adults, 2011–2017

Health Care Service	Total Sample N=1899	Never Tobacco User N= 633	Current exclusive smokeless user N= 633	Current exclusive cigarette smoker N= 633
Hospital Outpatient visits (in the past 12 months)				
% with ≥ 1 visit	14.2	17.9	11.1	14.1ϕ
Mean number of visits (se)	3.05 (0.21)	3.29 (0.47)	2.41 (0.21)	3.32 (0.18)
Hospital Inpatient visits (in the past 12 months)				
% with ≥ 1 visit	6.9	6.9	5.5	8.6ϕo
Mean number of visits (se)	6.11 (0.29)	5.34 (0.47)	5.42 (0.51)	7.22 (0.42)
Emergency room (ER) visits (in the past 12 months)				
% with ≥ 1 visit	13.9	12.0	13.3	16.4*
Mean number of visits (se)	1.37 (0.03)	1.25 (0.01)	1.33 (0.07)	1.50 (0.03)
Office Based visits (in the past 12 months)				
% with ≥ 1 visit	70.8	73.8	66.7	72.5
Mean number of visits (se)	6.40 (0.37)	6.93 (0.65)	6.08 (0.73)	6.23 (0.53)
Home Health Care visits (in the past 12 months)				
% with ≥ 1 visit	2.3	2.2	2.4	2.1
Mean number of visits (se)	61.94 (6.07)	55.69 (10.78)	58.68 (6.14)	72.32 (10.61)
Dental visits (in the past 12 months)				
% with ≥ 1 visit	33.4	42.2	30.6ϕ	28.2
Mean number of visits (se)	2.05 (0.08)	2.11 (0.16)	1.86 (0.07)	2.19 (0.13)

*Statistical significance difference between current exclusive cigarette smoker status and current exclusive smokeless user status and exclusive current cigarette smoker status and never tobacco user

ˢ Statistical significance difference between never tobacco user and current exclusive smokeless user status ᵒ Statistical significance difference between never tobacco user and current exclusive cigarette smoker status ϕ Statistical significance difference between current exclusive cigarette smoker status and current exclusive smokeless user status se= Standard error

Estimation of Annual Health Care Utilization by Tobacco Use Status

The final models were executed using the matched data set. Count-data models were fit and compared using a likelihood-based model-selection approach (i.e. identifying the model with the smallest AIC and BIC values) applying the full set of covariates (i.e. both sociodemographic and health status variables). The Hurdle count model, using a logit and a truncated Poisson regression was found to have the best fit for health care utilization data in this study.

Emergency room (ER) Visits

Table 23 shows the estimated results of the Hurdle model for ER visits including covariates; tobacco use status, age, gender, education, race/ethnicity, region, marital status, BMI, poverty status, binge drinking status, insurance type, physical component summary, mental component summary, Kessler index and Patient health index.

The logit indicated that there was no difference in the current exclusive smokeless tobacco user group or current exclusive cigarette smoker group compared to the never tobacco user group in their probability of having at least one ER visit (i.e. p-value =0.285 and p-value =0.092 respectively).

The Poisson model indicated for those who had at least one ER visit there was no statistically significant difference between current exclusive smokeless tobacco users or current exclusive cigarette smokers compared to never tobacco users at the 5% level (i.e. p-value =0.659 and p-value =0.268 respectively). Although not statistically significantly different, estimates for marginal effects on the conditional mean for the entire sample were calculated. Conditional on having at least one ER visit, current exclusive cigarette smokers averaged 0.16 visits more than

never tobacco users and current exclusive smokeless tobacco users on average have 0.07 more visits than never tobacco users.

However, no significant association between tobacco use status and ER visits was observed. Statistically significant associations were found for BMI ((p-value = 0.010, odds ratio (OR)=0.432), (p-value=0.074, OR= 0.587) for overweight and obesity respectively), self-reported binge status (p-value = 0.019, OR =0.378) and insurance type (p-value = 0.022, OR =3.130) for those who had at least one ER visit.

Table 23: Results for the Hurdle model for number of Emergency room visits, 2011-2017

Covariates	Emergency room visits (in the past 12 months)			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Tobacco Use Status				
Current exclusive cigarette smoker	0.367	0.092	0.428	0.268
Current exclusive smokeless tobacco user	0.251	0.285	0.209	0.659
Never Tobacco User	Reference		Reference	
Age	-0.014	0.074	0.007	0.515
Gender				
Male	Reference		Reference	
Female	0.308	0.259	0.578	0.121
Education				
College or Higher	0.096	0.679	-0.136	0.694
Some College	-0.023	0.932	-0.254	0.498

Covariates	Emergency room visits (in the past 12 months)			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
High School	-0.044	0.856	0.216	0.452
Less than High School	Reference		Reference	
Race/ethnicity				
White, non-Hispanic	Reference		Reference	
Black, non-Hispanic	0.065	0.790	-0.146	0.722
Asian, non-Hispanic	0.262	0.288	-0.390	0.339
Other or multiple race, non-Hispanic	0.180	0.698	-0.538	0.355
Hispanic	-0.104	0.744	-0.126	0.875
Region of residency				
Northeast	0.436	0.137	-0.228	0.605
Midwest	0.490	0.027	0.375	0.192
West	-0.179	0.520	-0.669	0.344
South	Reference		Reference	
Marital Status				
Married	-0.089	0.707	0.102	0.794
Living with Partner	0.325	0.287	-0.354	0.447
Widow/divorce/separated	0.125	0.636	-0.700	0.074
Never Married	Reference		Reference	
Body mass index (BMI)				

Covariates	Emergency room visits (in the past 12 months)			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Normal or Under Weight	Reference		Reference	
Overweight	-0.245	0.272	-0.839	0.010
Obesity	-0.110	0.597	-0.532	0.074
Poverty Status				
Poor/Negative	0.148	0.626	0.426	0.474
Near Poor	-0.210	0.638	-0.471	0.680
Low Income	-0.216	0.509	0.547	0.231
Middle Income	0.291	0.196	0.510	0.231
High Income	Reference		Reference	
Self-reported binge drinking status				
No	Reference		Reference	
Yes	-0.208	0.319	-0.974	0.019
Insurance type				
Uninsured	Reference		Reference	
Public	1.342	0.001	1.141	0.022
Private	0.519	0.138	0.737	0.197
Quality of life score (Short-Form 12)				
Physical Component Summary (0-100)	-0.016	0.068	0.011	0.488
Mental Component Summary (0-100)	0.001	0.958	0.017	0.395

Covariates	Emergency room visits (in the past 12 months)			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Mental illness score (Kessler 6 Index)	0.068	0.081	0.081	0.102
Depression score (Patient Health Index)				
0	Reference		Reference	
1	0.287	0.330	-0.755	0.260
2	-0.262	0.490	-0.160	0.775
3	-0.024	0.960	-0.540	0.398
4	-0.003	0.996	-0.310	0.585
5	-1.517	0.165	-13.811	0.0001
6	-0.235	0.745	-0.708	0.416

Statistical significance is at the 5% level

Office-Based Visits

Table 24 shows the estimated results of the Hurdle model for office-based visits including all covariates used in the ER model above.

The logit indicated that there was no difference in the current exclusive smokeless tobacco user group or current exclusive cigarette smoker group compared to the never tobacco user group in their probability of having at least one office- based visit (i.e. p-value =0.056 and p-value =0.074 respectively).

The Poisson model indicated for those who had at least one office-based visit there was no statistically significant difference between current exclusive smokeless tobacco users or current exclusive cigarette smokers compared to never tobacco users at the 5% level (i.e. p-value =0.310 and p-value =0.389 respectively). Although not statistically significantly different, estimates for marginal effects on the conditional mean for the entire sample were calculated. Current exclusive cigarette smokers averaged 0.63 visits less than never tobacco users and current exclusive smokeless tobacco users on average have 0.70 less visits than never tobacco users.

No significant association between tobacco use status and office-based visits was observed. However, statistically significant associations were found for self- reported binge status (p-value = 0.034, OR =0.785) , insurance type ((p-value =0.005 , OR = 2.243), (p-value = 0.007, OR = 2.10) public and private insurance respectively) for those who had at least one office-based visit. A participant is more likely to have at least one office-based visit with an unit increase in age (p-value = 0.008, OR =1.01) and less likely to have at least one office-based visit with an increase in health status-physical component score (p-value = 0.0001, OR =0.977)

Table 24: Results for the Hurdle model for number of Office-based visits, 2011-2017

Covariates	Office-based visits (in the past 12 months)			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Tobacco Use Status				
Current exclusive cigarette smoker	-0.295	0.074	-0.114	0.389
Current exclusive smokeless tobacco user	-0.350	0.056	-0.128	0.310
Never Tobacco User	Reference		Reference	
Age	0.17	0.005	0.011	0.008
Gender				
Male	Reference		Reference	
Female	1.047	0.001	-0.138	0.415
Education				
College or Higher	0.264	0.153	0.199	0.120
Some College	0.100	0.647	0.145	0.341
High School	0.053	0.783	-0.040	0.780
Less than High School	Reference		Reference	
Race/ethnicity				
White, non-Hispanic	Reference		Reference	
Black, non-Hispanic	-0.129	0.545	-0.144	0.387
Asian, non-Hispanic	-0.355	0.069	0.001	0.994
Other or multiple race, non-Hispanic	0.124	0.791	-0.236	0.407

Covariates	Office-based visits (in the past 12 months)			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Hispanic	-0.314	0.160	-0.082	0.553
Region of residency				
Northeast	0.389	0.121	0.117	0.461
Midwest	0.397	0.024	0.152	0.226
West	0.074	0.704	0.094	0.471
South	Reference		Reference	
Marital Status				
Married	-0.210	0.242	0.056	0.631
Living with Partner	-0.038	0.895	-0.215	0.429
Widow/divorce/separated	-0.124	0.583	0.069	0.680
Never Married	Reference		Reference	
Body mass index (BMI)				
Normal or Under Weight	Reference		Reference	
Overweight	0.325	0.061	0.037	0.821
Obesity	0.130	0.464	0.198	0.220
Poverty Status				
Poor/Negative	-0.590	0.040	0.067	0.679
Near Poor	-0.004	0.991	-0.021	0.905
Low Income	-0.215	0.393	-0.056	0.718

Covariates	Office-based visits (in the past 12 months)			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Middle Income	-0.321	0.052	0.063	0.608
High Income	Reference		Reference	
Self-reported binge drinking status				
No	Reference		Reference	
Yes	-0.197	0.186	-0.242	0.034
Insurance type				
Uninsured	Reference		Reference	
Public	1.899	0.0001	0.808	0.005
Private	1.776	0.0001	0.745	0.007
Quality of life score (Short-Form 12)				
Physical Component Summary (0-100)	-0.070	0.0001	-0.023	0.0001
Mental Component Summary (0-100)	-0.029	0.027	-0.015	0.094
Mental illness score (Kessler 6 Index)	-0.003	0.925	-0.010	0.665
Depression score (Patient Health Index)				
0	Reference		Reference	
1	0.067	0.784	0.067	0.676
2	-0.123	0.680	-0.032	0.843
3	0.933	0.147	-0.136	0.505
4	0.270	0.661	-0.172	0.414

Covariates	Office-based visits (in the past 12 months)			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
6	0.159	0.841	-0.114	0.700

Statistical significance is at the 5% level

Hospital Outpatient Visits

Table 25 shows the estimated results of the Hurdle model for hospital outpatient visits including all covariates previously mentioned for the other health care services. The logit indicated that there was no difference in current exclusive smokeless tobacco users or current exclusive cigarette smokers compared to never tobacco users in their probability of having at least one hospital outpatient visit (i.e. p-value =0.109 and p-value =0.832 respectively).

The Poisson model indicated for those who had at least one hospital outpatient visit there was no statistically significant difference between current exclusive smokeless tobacco users or current exclusive cigarette smokers compared to never tobacco users at the 5% level (i.e. p-value =0.130 and p-value =0.879 respectively). Although, not statistically significantly different, estimates for marginal effects on the conditional mean for the entire sample were calculated. Conditional on having at least one hospital outpatient visit, current exclusive cigarette smokers averaged 0.06 visits more than never tobacco users and current exclusive smokeless tobacco users on average have 0.79 less visits than never tobacco users.

No significant association between tobacco use status and hospital outpatient visits was observed. A statistically significant association was found between insurance type ((p-value =0.035 , OR = 0.390), (p-value = 0.033, OR = 0.365) public and private insurance respectively) and hospital outpatient visits.

Table 25:Results for the Hurdle model for number of Hospital Outpatient visits, 2011-2017

Covariates	**Hospital Outpatient visits (in the past 12 months)**			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Tobacco Use Status				
Current exclusive cigarette smoker	-0.048	0.832	0.033	0.879
Current exclusive smokeless tobacco user	-0.426	0.109	-0.629	0.130
Never Tobacco User	Reference		Reference	
Age	0.018	0.033	0.021	0.079
Gender				
Male	Reference		Reference	
Female	0.395	0.215	0.435	0.173
Education				
College or Higher	0.984	0.001	0.703	0.023
Some College	1.06	0.001	-0.287	0.415
High School	0.890	0.003	0.116	0.731
Less than High School	Reference		Reference	

Covariates	**Hospital Outpatient visits (in the past 12 months)**			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Race/ethnicity				
White, non-Hispanic	Reference		Reference	
Black, non-Hispanic	-0.409	0.169	0.003	0.997
Asian, non-Hispanic	0.068	0.809	0.309	0.423
Other or multiple race, non-Hispanic	-0.181	0.762	-1.019	0.365
Hispanic	-0.086	0.803	-0.178	0.680
Region of residency				
Northeast	0.630	0.049	0.530	0.246
Midwest	0.383	0.093	0.073	0.803
West	-0.504	0.134	0.848	0.044
South	Reference		Reference	
Marital Status				
Married	0.203	0.445	0.469	0.203
Living with Partner	0.562	0.148	0.787	0.064
Widow/divorce/separated	0.371	0.213	0.006	0.986
Never Married	Reference		Reference	
Body mass index (BMI)				
Normal or Under Weight	Reference		Reference	
Overweight	0.073	0.787	0.191	0.656

Covariates	**Hospital Outpatient visits (in the past 12 months)**			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Obesity	0.188	0.444	0.101	0.758
Poverty Status				
Poor/Negative	-0.495	0.187	-1.329	0.045
Near Poor	0.141	0.755	0.015	0.972
Low Income	-0.146	0.667	0.376	0.211
Middle Income	-0.166	0.488	-0.481	0.104
High Income	Reference		Reference	
Self-reported binge drinking status				
No	Reference		Reference	
Yes	-0.312	0.183	0.179	0.595
Insurance type				
Uninsured	Reference		Reference	
Public	0.853	0.079	-0.942	0.035
Private	1.030	0.019	-1.007	0.033
Quality of life score (Short-Form 12)				
Physical Component Summary (0-100)	-0.048	0.0001	-0.017	0.214
Mental Component Summary (0-100)	-0.002	0.878	0.010	0.625
Mental illness score (Kessler 6 Index)	0.052	0.209	0.044	0.311
Depression score (Patient Health Index)				

Covariates	**Hospital Outpatient visits (in the past 12 months)**			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
0	Reference		Reference	
1	0.325	0.291	0.894	0.028
2	-0.070	0.854	0.790	0.083
3	-0.431	0.409	0.678	0.253
4	-0.651	0.229	0.099	0.860
5	-1.145	0.370	0.608	0.448
6	-0.468	0.552	-0.717	0.448

Statistical significance is at the 5% level

Hospital Inpatient Visits

Table 26 shows the estimated results of the Hurdle model for hospital inpatient visits including all covariates previously mentioned for the other health care services. The logit indicated that there was no difference in current exclusive smokeless tobacco users or current exclusive cigarette smokers compared to never tobacco users in their probability of having at least one hospital inpatient visit (i.e. p-value =0.250 and p-value =0.304 respectively).

The Poisson model indicated for those who had at least one hospital inpatient visit there was no statistically significant difference between current exclusive smokeless tobacco users or current exclusive cigarette smokers compared to never tobacco users at the 5% level (i.e. p-value

=0.932and p-value =0.090 respectively). Although, not statistically significantly different, estimates for marginal effects on the conditional mean for the entire sample were calculated. Conditional on having at least one hospital inpatient visit, current exclusive cigarette smokers averaged 4.23 visits more than never tobacco users and current exclusive smokeless tobacco users on average have 0.11 less visits than never tobacco users.

No significant association between tobacco use status and hospital inpatient visits was observed. A statistically significant association was found between the Short -Form 12-mental component score (p-value = 0.0001, OR =0.941), were a participant is less likely to have at least one hospital inpatient visits with an increased score. Also, participants were less likely to have at least one hospital inpatient visit as the Kessler 6 index (mental illness score) decreases (p-value = 0.027, OR =0.924).

Table 26:Results for the Hurdle model for number of Inpatient visits, 2011-2017

Covariates	**Hospital Inpatient visits (in the past 12 months)**			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Tobacco Use Status				
Current exclusive cigarette smoker	0.353	0.250	0.643	0.090
Current exclusive smokeless tobacco user	-0.380	0.304	-0.026	0.934
Never Tobacco User	Ref		Ref	
Age	0.016	0.133	0.018	0.062
Gender				
Male	Ref		Ref	
Female	-0.125	0.761	0.384	0.118
Education				
College or Higher	-0.536	0.144	-0.106	0.780

Covariates	**Hospital Inpatient visits (in the past 12 months)**			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Some College	0.107	0.767	0.060	0.824
High School	-0.123	0.708	1.13	0.001
Less than High School	Ref		Ref	
Race/ethnicity				
White, non-Hispanic	Ref		Ref	
Black, non-Hispanic	0.036	0.920	0.773	0.011
Asian, non-Hispanic	0.207	0.581	-0.706	0.052
Other or multiple race, non-Hispanic	0.267	0.628	-0.769	0.069
Hispanic	0.048	0.926	1.18	0.016
Region of residency				
Northeast	-0.260	0.565	0.300	0.519
Midwest	-0.221	0.507	-0.616	0.065
West	-0.881	0.060	0.008	0.983
South	Ref		Ref	
Marital Status				
Married	0.177	0.649	-0.387	0.253
Living with Partner	0.395	0.476	0.862	0.121
Widow/divorce/separated	0.283	0.507	-0.508	0.174
Never Married	Ref		Ref	
Body mass index (BMI)				
Normal or Under Weight	Ref		Ref	
Overweight	0.507	0.199	0.402	0.191
Obesity	0.380	0.296	0.504	0.087
Poverty Status				

Covariates	**Hospital Inpatient visits (in the past 12 months)**			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Poor/Negative	-0.187	0.669	0.678	0.043
Near Poor	-0.265	0.663	-0.008	0.986
Low Income	-0.438	0.351	-0.763	0.125
Middle Income	-0.481	0.211	0.571	0.049
High Income	Ref		Ref	
Self-reported binge drinking status				
No	Ref		Ref	
Yes	-0.105	0.747	-0.336	0.371
Insurance type				
Uninsured	Ref		Ref	
Public	0.790	0.221	-0.368	0.466
Private	0.676	0.290	-0.568	0.325
Quality of life score (Short-Form 12)				
Physical Component Summary (0-100)	-0.050	0.0001	-0.006	0.646
Mental Component Summary (0-100)	-0.005	0.791	-0.061	0.0001
Mental illness score (Kessler 6 Index)	0.055	0.307	-0.079	0.027
Depression score (Patient Health Index)				
0	Ref		Ref	
1	-0.021	0.965	0.018	0.961
2	-0.844	0.181	0.766	0.045
3	-0.737	0.329	-0.510	0.393
4	0.171	0.796	0.003	0.995
5	-0.769	0.539	-2.160	0.005
6	-0.233	0.805	0.657	0.264

Dental Visits

Table *27* shows the estimated results of the Hurdle model for dental visits including all covariates previously mentioned for the other health care services. The logit indicated that both current exclusive smokeless tobacco users and current exclusive cigarette smokers have lower probability of having at least one dental visit compared to never tobacco users (i.e. p-value =0.004 and p-value =0.001 respectively).

The Poisson model indicated for those who had at least one dental visit there was no statistically significant difference between current exclusive smokeless tobacco users or current exclusive cigarette smokers compared to the never tobacco users at the 5% level (i.e. p-value =0.163 and p-value =0.940 respectively). Although, not statistically significantly different, estimates for marginal effects on the conditional mean for the entire sample were calculated. Conditional on having at least one dental visit, current exclusive cigarette smokers averaged 0.01 visits less than never tobacco users and current exclusive smokeless tobacco users on average have 0.23 less visits than never tobacco users.

No significant association between tobacco use status and dental visits was observed. A statistically significant association was found between BMI ((p-value =0.002 , OR = 1.657), (p-value = 0.038, OR = 1.383) overweight and obesity respectively) and dental visits.

Table 27:Results for the Hurdle model for number of Dental visits, 2011-2017

Covariates	Dental visits (in the past 12 months)			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Tobacco Use Status				
Current exclusive cigarette smoker	-0.531	0.001	-0.010	0.940
Current exclusive smokeless tobacco user	-0.499	0.004	-0.217	0.163
Never Tobacco User	Reference			
Age	-0.010	0.092	0.003	0.525
Gender				
Male	Reference		Reference	
Female	-0.040	0.885	-0.491	0.075
Education				
College or Higher	0.472	0.007	-0.077	0.654
Some College	0.257	0.230	-0.262	0.186
High School	0.147	0.449	-0.086	0.612
Less than High School	Reference		Reference	
Race/ethnicity				
White, non-Hispanic	Reference		Reference	
Black, non-Hispanic	-0.229	0.282	-0.162	0.458
Asian, non-Hispanic	-0.239	0.224	0.265	0.157

Covariates	Dental visits (in the past 12 months)			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Other or multiple race, non-Hispanic	-0.585	0.203	-0.490	0.356
Hispanic	-0.097	0.666	-0.014	0.948
Region of residency				
Northeast	0.502	0.030	0.050	0.816
Midwest	0.432	0.010	0.041	0.812
West	0.662	0.0001	0.243	0.164
South	Reference		Reference	
Marital Status				
Married	0.376	0.032	-0.105	0.549
Living with Partner	-0.070	0.803	-0.202	0.429
Widow/divorce/separated	0.213	0.321	0.005	0.976
Never Married	Reference		Reference	
Body mass index (BMI)				
Normal or Under Weight	Reference		Reference	
Overweight	0.062	0.714	0.505	0.002
Obesity	-0.041	0.813	0.324	0.038
Poverty Status				
Poor/Negative	-0.919	0.001	-0.276	0.327
Near Poor	-0.712	0.045	0.297	0.280

Covariates	Dental visits (in the past 12 months)			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Low Income	-1.077	0.0001	-0.081	0.776
Middle Income	-0.666	0.0001	0.027	0.853
High Income	Reference		Reference	
Self-reported binge drinking status				
No	Reference		Reference	
Yes	-0.111	0.465	-0.027	0.827
Insurance type				
Uninsured	Reference		Reference	
Public	1.314	0.0001	0.850	0.104
Private	1.086	0.001	0.395	0.426
Quality of life score (Short-Form 12)				
Physical Component Summary (0-100)	-0.002	0.802	-0.010	0.229
Mental Component Summary (0-100)	0.005	0.638	-0.009	0.522
Mental illness score (Kessler 6 Index)	0.019	0.544	-0.010	0.803
Depression score (Patient Health Index)				
0	Reference		Reference	
1	0.105	0.670	-0.049	0.805
2	0.284	0.297	-0.084	0.703
3	-0.493	0.319	-0.784	0.187

Covariates	Dental visits (in the past 12 months)			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
4	0.291	0.511	-0.433	0.262
5	-1.047	0.413	-11.797	0.0001
6	-0.868	0.184	-0.564	0.564

Statistical significance is at the 5% level

Home Health Care Visits

Table 28 shows the estimated results of the Hurdle model for home health care visits including tobacco use status and the sociodemographic variables: age, gender, Education, Race/ethnicity, region, marital status, BMI, poverty status, binge drinking status and insurance type. The logit indicated that there was no difference in current exclusive smokeless tobacco users or current exclusive cigarette smokers compared to never tobacco users in their probability of having at least one home health care visit (i.e. p-value =0.922 and p-value =0.286 respectively).

The Poisson model indicated for those who had at least one home health care visit there was no statistically significant difference between current exclusive smokeless tobacco users or current exclusive cigarette smokers compared to the never tobacco users at the 5% level (i.e. p-value =0.546 and p-value =0.127 respectively).

No significant association between tobacco use status and home health care visits was observed. Statistically significant associations were found for gender (p-value = 0.003, OR = 3.219) and education ((p-value =0.002 , OR = 0.291), (p-value = 0.0001, OR = 0.293) some college and High school respectively) with home health care visits.

Table 28:Results of the Hurdle model for number of Home Health Care visits, 2011-2017

Covariates	**Home Health Care visits (in the past 12 months)**			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Tobacco Use Status				
Current exclusive cigarette smoker	-0.412	0.286	-0.931	0.127
Current exclusive smokeless tobacco user	-0.033	0.922	-0.207	0.546
Never Tobacco User	Reference		Reference	
Age	0.039	0.002	-0.016	0.366
Gender				
Male	Reference		Reference	
Female	-0.184	0.622	1.169	0.003
Education				
College or Higher	-0.497	0.212	-1.236	0.114
Some College	-0.168	0.827	-1.233	0.002
High School	0.082	0.720	-1.211	0.0001
Less than High School	Reference		Reference	
Race/ethnicity				
White, non-Hispanic	Reference		Reference	
Black, non-Hispanic	0.059	0.880	-0.250	0.418
Asian, non-Hispanic	0.140	0.798	-1.396	0.050

Covariates	**Home Health Care visits (in the past 12 months)**			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Other or multiple race, non-Hispanic	0.483	0.416	-1.322	0.268
Hispanic	-0.310	0.628	0.567	0.254
Region of residency				
Northeast	0.216	0.652	-0.319	0.709
Midwest	-0.015	0.971	0.885	0.082
West	-1.934	0.069	2.082	0.007
South	Reference		Reference	
Marital Status				
Married	-0.556	0.250	-0.139	0.776
Living with Partner	-1.605	0.119	1.772	0.166
Widow/divorce/separated	-0.222	0.621	0.310	0.638
Never Married	Reference		Reference	
Body mass index (BMI)				
Normal or Under Weight	Reference		Reference	
Overweight	0.389	0.313	0.526	0.212
Obesity	0.137	0.717	-0.118	0.827
Poverty Status				
Poor/Negative	1.201	0.032	0.657	0.541
Near Poor	1.261	0.072	-0.156	0.884

Covariates	**Home Health Care visits (in the past 12 months)**			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Low Income	0.143	0.831	0.106	0.917
Middle Income	1.065	0.033	-0.349	0.716
High Income	Reference		Reference	
Self-reported binge drinking status				
No	Reference		Reference	
Yes	-0.436	0.311	-0.436	0.500
Insurance type				
Uninsured	Reference		Reference	
Public	2.190	0.035	0.869	0.107
Private	1.090	0.310	0.889	0.080

Statistical significance is at the 5% level

In summary, no statistically significant association was found between tobacco use status and the 6 health care services examined in this study using a Hurdle count model. Sociodemographic variables like age and insurance type were found to be associated with the health care services studied. Stata output for the final models for health care utilization for the 6 health care services can be found in Appendix B.

Sensitivity Analysis for Health Care Utilization

Sensitivity analysis was conducted using the original data set before propensity score matching. Hurdle count models were run for ER visits, office-based visits, hospital outpatient visits, dental visits and home health care visits adjusting for tobacco use status, sociodemographic and health status variables.

The Hurdle model results indicated that there was no statistically significant association between tobacco use status and ER visits or home health care visits. Statistically significant associations between tobacco use status and office-based visits, hospital outpatient visits, and dental visits (only for current smokeless tobacco users vs never tobacco users) were observed (see Table 29).

The Poisson model indicated for those who had at least one office-based visit there was a statistically significant difference between current exclusive smokeless tobacco users and current exclusive cigarette smokers compared to the never tobacco user group at the 5% level (i.e. p-value =0.039 and p-value =0.0001 respectively). Marginal effects on the conditional mean for the entire sample indicated that conditional on having at least one office-based visit, current exclusive cigarette smokers averaged 1.3 visits less than never tobacco users and current exclusive smokeless tobacco users on average have 1.6 less visits than never tobacco users.

For those who had at least one hospital outpatient visit there was a statistically significant difference between current exclusive smokeless tobacco users and current exclusive cigarette smokers compared to the never tobacco user group at the 5% level (i.e. p-value =0.003 and p-value =0.046 respectively). The marginal effects indicated that conditional on having at least one hospital outpatient visit, current exclusive cigarette smokers and current exclusive smokeless tobacco users on average have 0.16 and 0.44 less visits than never tobacco users respectively.

A statistically significant difference between current exclusive smokeless tobacco users and the never tobacco users at the 5% level for those having at least one dental visit , with marginal effects indicating that current exclusive smokeless users on average have 0.34 less visits than never tobacco users. The full results of the fitted models for this analysis are in Appendix A.

Table 29:Results for the Hurdle model for each of health service by tobacco use status

Covariate	**Emergency room visits (in the past 12 months)[a]**			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Tobacco Use Status				
Current exclusive cigarette smoker	0.265	0.0001	0.057	0.396
Current exclusive smokeless tobacco user	-0.043	0.836	-0.807	0.096
Never Tobacco User	Reference		Reference	
	Office based visits (in the past 12 months)[a]			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Current exclusive cigarette smoker	-0.407	0.0001	-0.180	0.0001
Current exclusive smokeless tobacco user	-0.375	0.005	-0.238	0.039
Never Tobacco User	Reference		Reference	
	Outpatient visits (in the past 12 months)[a]			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Current exclusive cigarette smoker	-0.184	0.0001	-0.229	0.046
Current exclusive smokeless tobacco user	-0.617	0.006	-1.161	0.003
Never Tobacco User	Reference		Reference	
	Inpatient visits (in the past 12 months)[a]			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Current exclusive cigarette smoker	0.005	0.938	0.121	0.046
Current exclusive smokeless tobacco user	-1.110	0.009	-0.412	0.526

Never Tobacco User	Reference		Reference	
	Dental visits (in the past 12 months)[a]			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Current exclusive cigarette smoker	-0.478	0.0001	0.032	0.318
Current exclusive smokeless tobacco user	-0.741	0.0001	-0.296	0.036
Never Tobacco User	Reference		Reference	
	Home Health care visits (in the past 12 months)[a]			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Current exclusive cigarette smoker	-0.017	0.862	-0.193	0.124
Current exclusive smokeless tobacco user	0.254	0.648	0.443	0.269
Never Tobacco User	Reference		Reference	

[a]All models control for tobacco use status, age, gender, education, race/ethnicity, region, marital status, BMI, poverty status, binge drinking status, insurance type, physical component summary, mental component summary, Kessler index and Patient health index.

Estimation of Annual Total Health Care Expenditure by Tobacco Use Status

The final models were executed using the matched data set. Two-part models were fit using different distribution and link functions (e.g. gaussian distribution with a log link) and compared using a likelihood-based model-selection approach (i.e. identifying the model with the smallest AIC and BIC values) applying the full set of covariates (i.e. both sociodemographic and health status variables). A generalized linear regression (GLM) with a Gamma distribution and a log link function was the best fit for the health expenditure data in this study.

Final Total Health Care Expenditure Model

Table 30 shows the estimated coefficients and linearized standard errors and associated p-values for tobacco use status and key variables that were found to be significant from the two-part model.

The logit indicated that there was no statistically significant difference in current exclusive smokeless tobacco users or current exclusive cigarette smokers compared to the never tobacco users in their probability of having at least some spending (i.e. p-value =0.439 and p-value =0.159 respectively). Among those who spend something, the GLM model indicates that there was no statistically significant difference between current exclusive smokeless tobacco users or current exclusive cigarette smokers compared to the never tobacco users at the 5% level (i.e. p-value =0.874 , OR= 1.029 and p-value =0.115, OR=1.306 respectively). Although not statistically significantly different, the marginal (or incremental) effects for the combined logit and GLM of the two-part model were estimated. The marginal effects for tobacco use status indicated that current exclusive cigarette smokers spend more than never tobacco users by about $1200 (standard error (se)=$902)and current exclusive smokeless tobacco user spend about $50 (se=$796)more than never tobacco users. The overall mean annual health care expenditure for US adults was $4868 (se=$436). Table 31 shows the estimated annual mean total health care expenditure by tobacco use status. Although no statistical association was found between total health care expenditure and tobacco use status, never tobacco user had the lowest annual mean total health care expenditure ($4426.89) followed by current exclusive smokeless users ($4478.33) and current exclusive cigarette smokers have the highest annual mean cost ($5627.64).

Covariates that were found to be statistically significantly associated with total health care expenditure, shown in Table 30, were age, insurance type and the quality of life scores (physical and mental health component scores).

The estimated coefficients for age were positive in both the logit and GLM and statistically significant at the 10% (p-value=0.069) and 5% levels (p-value=0.035, OR=1.013). Both the

probability of spending and the amount of spending conditional on any spending increased with age. The marginal effect of age averages $68.23 per year of age.

The estimated coefficients for insurance type were also statistically significant at the 5% level and positive in both the logit (p-value= 0.0001 and p-value=0.001 for public and private insurance respectively) and GLM((p-value =0.0001 , OR = 3.095), (p-value = 0.0001, OR = 3.669) for public and private insurance respectively). Those who have any type of health insurance (i.e. public or any private) are more likely than the uninsured to spend at least $1, and, conditional on spending any amount, they are more likely to spend more than the uninsured. The marginal effect of those with public or any private insurance spend more than the uninsured by about $4451.19 and $3608.88 respectively.

The two-part model results for physical and mental health component scores show adults who are in better health are both less likely to spend and to spend less when they do spend GLM((p-value =0.0001 , OR = 0.956), (p-value = 0.010, OR = 0.969) for PCS and MCS respectively).The physical and mental health component scores indicate that adults who are in better health spend significantly less than those in poorer health(i.e. about $250.97 and $154.71 less than those in poorer health). The model predicted that the overall total spending was about $4868 per person per year.

Given the statistically significant increase in spending with age, predicted values for age categories (i.e. decade of life) by tobacco use status were examined. Figure 9 shows that the predicted total health expenditures rise for all the tobacco use status groups with age. The predicted total health expenditures are highest for current exclusive cigarette smokers followed by current exclusive smokeless users and never tobacco users adults. The separation in spending for current exclusive smokeless users and never tobacco user increases with age. Just examining

the average marginal effect may mask the fact that the observed marginal effects may vary with age.

Based on the fact that the significant variables, like insurance type, in the model have a larger effect size as estimated by the odds ratio compared to the estimates of the non-significant tobacco use status variable with a small effect size, therefore the study should have enough power to detect a statistical significant difference.

The final model was also run with just the set of sociodemographic variables as a sensitivity analysis. Similar results for tobacco use status were observed. The logit indicated that there was no statistically significant difference in current exclusive smokeless tobacco users or current exclusive cigarette smokers compared to the never tobacco users in their probability of having at least some spending (i.e. p-value =0.615 and p-value =0.396 respectively). Among those who spend something, the GLM model indicates that there was no statistically significant difference between current exclusive smokeless tobacco users or current exclusive cigarette smokers compared to the never tobacco users at the 5% level (i.e. p-value =0.842 , OR= 1.036 and p-value =0.061, OR=1.363 respectively).

Table 30:Results of the Two-part model for Total Health Care Expenditure (2017 US $) per year, 2011-2017

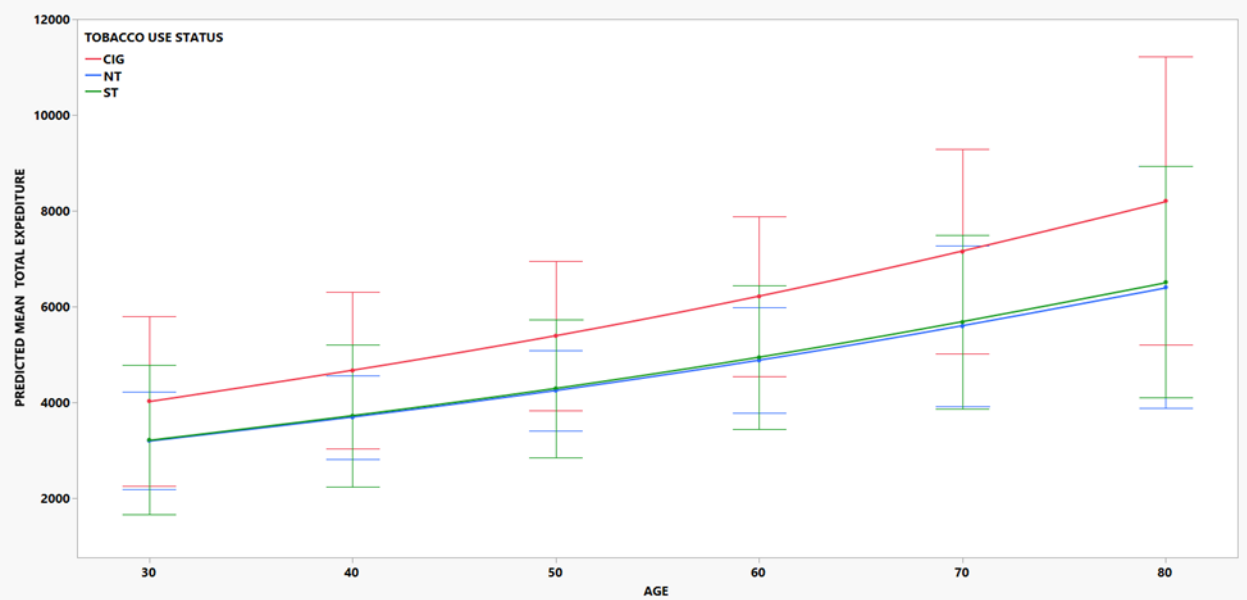
Covariates	**Total Health Care Expenditure**			
	Logit		GLM	
	Coefficient *(Linearized std. error)*	*P-value*	*Coefficient* *(Linearized std. error)*	*P-value*
Tobacco Use Status				

Covariates	Total Health Care Expenditure			
	Logit		GLM	
	Coefficient *(Linearized std. error)*	*P-value*	*Coefficient* *(Linearized std. error)*	*P-value*
Current exclusive cigarette smoker	-0.352 (0.248)	0.159	0.267 (0.169)	0.115
Current exclusive smokeless tobacco user	-0.229 (0.295)	0.439	0.029 (0.179)	0.874
Never Tobacco User	Ref		Ref	
Age	0.019 (0.010)	0.069	0.013 (0.005)	0.035
Insurance type				
Public	1.63 (0.326)	0.0001	1.13 (0.284)	0.0001
Private	1.50 (0.452)	0.001	1.30 (0.317)	0.0001
Uninsured	Ref		Ref	
Quality of life score (Short-Form 12)				
Physical Component Summary (0-100)	-0.086 (0.015)	0.0001	-0.045 (0.007)	0.0001
Mental Component Summary (0-100)	-0.010 (0.019)	0.595	-0.031 (0.012)	0.010

Table 31:Adjusted estimated mean Total Health Care Expenditure by Tobacco use status

Tobacco Use Status	*Mean health care[a] cost (US$,2017)*	*95% CI*
Current exclusive cigarette smoker	$5627.64	($4068.50, $7186.78)
Current exclusive smokeless tobacco user	$4478.33	($3035.05, $5921.62)
Never Tobacco User	$4426.89	($3514.19, $5339.59)

[a]Note model is adjusted for tobacco use status and sociodemographic characteristics



NT= never tobacco users ST= current exclusive smokeless user, CIG= current exclusive cigarette smoker
95% Confidence Interval

Figure 9: Conditional marginal effects of Age by Tobacco use status for Total Health Care Expenditure

Sensitivity Analysis-Age (≤ 65 years)

The analysis described above was rerun with restricting the data to adults between 18 and 65 years. Given that healthcare expenditure increases with adulthood, excluding older adults from the modeling analysis allows the assessment of total healthcare expenditure by tobacco use status in the general working population and removes older adults who may be sicker and have higher total healthcare spending than the average adult.

Table 32 show the estimated coefficients and linearized standard errors and associated p-values for tobacco use status and key variables that were found to be significant from the two-part model.

As previously found, the logit indicated that there was no statistically significant difference in current exclusive smokeless tobacco users or current exclusive cigarette smokers compared to never tobacco users in their probability of having at least some spending (i.e. p-value =0.358 and p-value =0.359 respectively). Among those who spend something, the GLM model indicates that there was no statistically significant difference between current exclusive smokeless tobacco users or current exclusive cigarette smokers compared to never tobacco users at the 5% level (i.e. p-value =0.100 and p-value =0.073respectively). At the 10% level a statistically significant difference would be observed between current exclusive cigarette smokers and never tobacco users. The marginal effects for the combined logit and GLM of the two-part model for tobacco use status indicated that current exclusive cigarette smokers spend more than never tobacco users by about $1130 (se=$717) and current exclusive smokeless tobacco users spend about $ 893 (se=$640)more than never tobacco users which is much higher than observed in the above model

($50). The overall mean annual health care expenditure for US adults between 18 and 65 was $3903 (se=$419). Table 33 shows the estimated annual mean total health care expenditure by tobacco use status. Although no statistically significant association was found between total health care expenditure and tobacco use status, never tobacco users had an annual mean total health care expenditure of $3196.20 followed by current exclusive smokeless users ($4089.01) and current exclusive cigarette smokers had annual mean cost ($4326.73).

Covariates that were found to be statistically significantly associated with total health care expenditure, shown in Table 32 included insurance type and quality of life scores (physical and mental health component scores). With the restriction of older adults, age no longer was a statistically significant covariate in the model.

The estimated coefficients for insurance type were also statistically significant at the 5% level and positive in both the logit (p-value= 0.002 and p-value=0.0001 for public and private insurance respectively) and GLM((p-value =0.0001 , OR = 6.959), (p-value = 0.0001, OR = 4.055) for public and private insurance respectively). Those who have any type of health insurance (i.e. public or any private) are more likely than the uninsured to spend at least $1, and, conditional on spending any amount, they are more likely to spend more than the uninsured. Those with public or any private insurance spend more than the uninsured by about $5519.58 and $3021.13 respectively.

The two-part model results for physical and mental health component scores show adults who are in better health are both less likely to spend and to spend less when they do spend GLM((p-value =0.0001 , OR = 0.939), (p-value = 0.004 OR = 0.960) for PCS and MCS respectively). The physical and mental health component scores indicate that adults who are in better health spend significantly less than those in poorer health (i.e. about $272.08 and $169.49 less than

those in poorer health). The model predicted that overall total spending was about $3902.96 per person per year which was $965.50 less than the results of the final model, indicating the impact of total health care spending in the older population.

Predicted values for age categories (i.e. 30 to 65 by 5 years) by tobacco use status were also examined. Figure 10 shows that the predicted total health expenditures rise for all tobacco use status groups with age. Predicted total health expenditures are highest for current exclusive cigarette smokers followed by current exclusive smokeless users and never tobacco users. A clear separation can be observed between the three groups and this difference is consistent overtime, which shows a different pattern than seen in the final model.

Table 32: Results of the Two-part model for Total Health Care Expenditure (2017 US $) per year (Age ≤65 years), 2011-2017

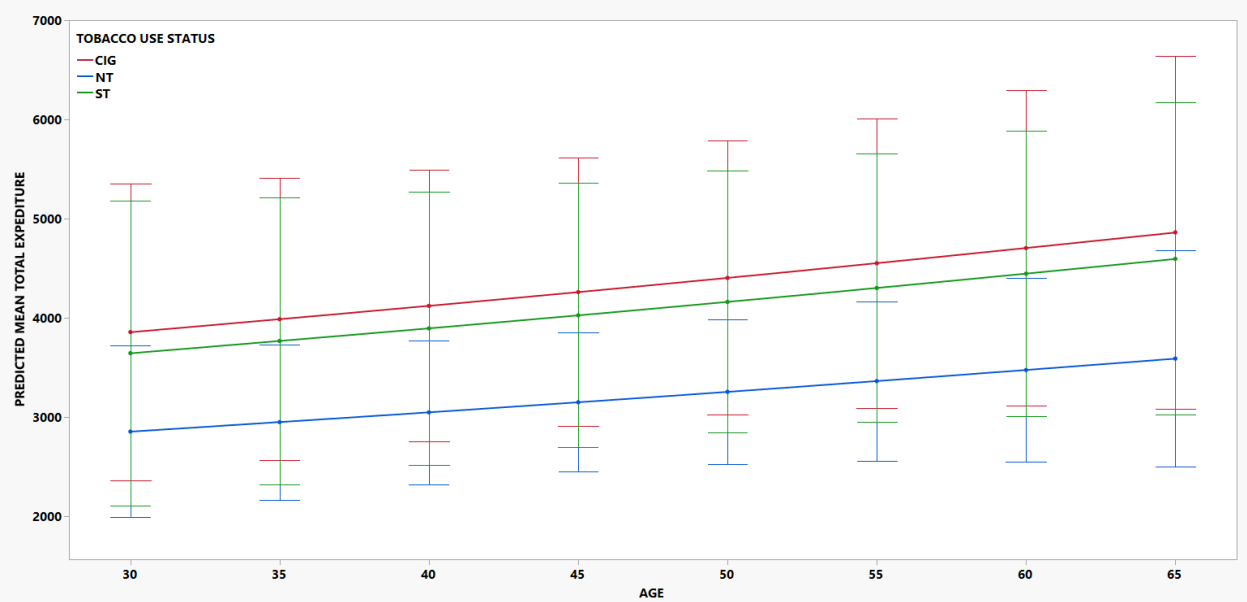
Covariates	Total Health Care Expenditure			
	Logit		GLM	
	Coefficient (Linearized std. error)	*P-value*	*Coefficient (Linearized std. error)*	*P-value*
Tobacco Use Status				
Current exclusive cigarette smoker	-0.219 (0.238)	0.359	0.323 (0.180)	0.073
Current exclusive smokeless tobacco user	-0.243 (0.264)	0.358	0.270 (0.163)	0.100
Never Tobacco User	Ref		Ref	
Insurance type				
Public	1.32 (0.427)	0.002	1.94 (0.266)	0.0001
Private	1.58 (0.318)	0.0001	1.40 (0.237)	0.0001

Covariates	Total Health Care Expenditure			
	Logit		GLM	
	Coefficient (Linearized std. error)	*P-value*	*Coefficient (Linearized std. error)*	*P-value*
Uninsured	Ref		Ref	
Quality of life score (Short-Form 12)				
Physical Component Summary (0-100)	-0.086 (0.015)	0.0001	-0.063 (0.008)	0.0001
Mental Component Summary (0-100)	-0.010 (0.019)	0.179	-0.041 (0.014)	0.004

Table 33: Adjusted Estimated mean Total Health Care Expenditure by Tobacco Use Status for Age (≤ 65 years), 2011-2017

Tobacco Use Status	*Mean health care[a] cost (US$,2017)*	*95% CI*
Current exclusive cigarette smoker	$4326.73	($2965.46, $5687.99)
Current exclusive smokeless tobacco user	$4089.01	($2759.49, $5418.54)
Never Tobacco User	$3196.20	($2476.35, $3916.04)

[a]Note model is adjusted for tobacco use status and sociodemographic characteristics



NT= never tobacco users ST= current exclusive smokeless user, CIG= current exclusive cigarette smoker

95% Confidence Interval

Figure 10: Conditional Marginal Effects of Age by Tobacco Use Status for Total Health Care Expenditure for Age (≤ 65 years)

Sensitivity Analysis-Age (≤ 65 years) and males only

This analysis restricts the data to adults equal to or less than 65 years and males only. Since smokeless tobacco products are predominately used by males in the U.S. population, females were removed to assess the impact on total health care expenditure. Both age and gender were evaluated in this sensitivity analysis.

The results of this model were similar to those in the previous sensitivity analysis. There was no statistically significant difference in current exclusive smokeless tobacco users or current exclusive cigarette smokers compared to never tobacco users in their probability of having at

least some spending (i.e. p-value =0.455 and p-value =0.366 respectively). Among those who spent something, there was no statistically significant difference between current exclusive smokeless tobacco users or current exclusive cigarette smokers compared to the never tobacco user group at the 5% level (i.e. p-value =0.162 and p-value =0.099 respectively). At the 10% level a statistically significant difference would be observed between current exclusive cigarette smokers and never tobacco users. The marginal effects for tobacco use status indicated that current exclusive cigarette smokers spend more than never tobacco users by about $1055 (se=$722) and current exclusive smokeless tobacco users spend about $796 (se=$655)more than never tobacco users which is not that different than the previous sensitivity analysis ($1130 and $892). The overall mean annual health care expenditure for US adults was $3668 (se=$417). Table 34 shows the estimated annual mean total health care expenditure by tobacco use status. Although no statistically significant association was found between total health care expenditure and tobacco use status, never tobacco users have the lowest annual mean total health care expenditures ($3053.73) followed by current exclusive smokeless users ($3796.06) and current exclusive cigarette smokers ($4073.95).

Two covariates were found to be statistically significantly associated with total health care expenditure - insurance type and the quality of life scores (physical and mental health component scores). The estimated coefficients for insurance type were statistically significant at the 5% level and positive in both the logit (p-value= 0.0001 and p-value=0.001 for public and private insurance respectively) and GLM((p-value =0.0001 , OR = 6.410), (p-value = 0.0001, OR = 3.559) for public and private insurance respectively). Those who had any type of health insurance (i.e. public or any private) were more likely than the uninsured to spend at least $1, and, conditional on spending any amount, they were more likely to spend more than the

uninsured. Those with public or any private insurance spend more than the uninsured by about $7391.55 and $5240.10 respectively.

The results for physical and mental health component scores also showed that adults who were in better health were both less likely to spend and to spend less when they do spend GLM((p-value =0.0001 , OR = 0.939), (p-value = 0.004 OR = 0.959) for PCS and MCS respectively). The physical and mental health component scores indicate that adults who were in better health spend significantly less than those in poorer health (i.e. about $256.20 and $160.86 less than those in poorer health). The model predicted that overall total spending was about $3667.98 per person per year which was close to the previous sensitivity analysis $3902.96.

Stata output for total healthcare expenditure models can be found in Appendix B.

Table 34: Adjusted estimated mean Total Health Care Expenditure by Tobacco use status for Age (≤ 65 years) and Males only

Tobacco Use Status	*Mean health care cost (US$,2017)*	*95% CI*
Current exclusive cigarette smoker	$4073.95	($2719.24, $5428.66)
Current exclusive smokeless tobacco user	$3796.06	($2496.46, $5095.67)
Never Tobacco User	$3053.73	($2334.92, $3772.54)

[a]Note model is adjusted for tobacco use status and sociodemographic characteristics

Chapter 5: Discussion

This final chapter discusses the results of the study, the study strengths and limitations, implications and ends with the study conclusion and future research.

Discussion

The study sample used in this analysis was U.S. adults age 18 years and older from the 2011-2017 linked MEPS/NHIS national surveys who were identified as current exclusive cigarette smokers, current exclusive smokeless tobacco users or never tobacco users. Approximately 35% of the individuals obtained from the MEPS/NHIS linked files met study eligibility. A slight increase in the prevalence of current exclusive smokeless tobacco use by year of MEPS data collection was observed across the seven years (2011 through 2017). A decrease was seen in current exclusive cigarette smoking over the same timeframe. This decrease is similar to what has been observed in the adult U.S. population over the same period of time.[2]

The average adult was 48 years of age for this study population. The majority of the population was female (55.2 %), non-Hispanic white (58.9%), reported being in middle- or high-income bracket (67.7%) and reported having some type of health care insurance (89.5%) . Similar socio-demographic trends were observed in Wang et al[1] whose study population data was obtained from NHIS 2012 through 2015. Majority of adults in the study population were never tobacco users (79.3%) followed by current exclusive cigarette smokers (19.5%) while current exclusive smokeless tobacco users had the smallest prevalence (1.2%).

The quality of life scores in both the PCS and MCS were slightly higher in never tobacco users indicating slightly better health status compared to the other two tobacco status groups in the study population. These mean differences in the quality of life scores were found to be statistically significant between current exclusive cigarette smokers vs current exclusive smokeless tobacco users and never tobacco users. No statistically significant difference was

observed between participants' mean scores in the never tobacco user group and current exclusive smokeless tobacco user group for either quality of life score.

Current exclusive cigarette smokers had a greater tendency towards mental disability as indicted by their higher Kessler index compared to never tobacco users and current exclusive smokeless tobacco users. These differences were found to be statistically significant between current exclusive cigarette smokers vs current exclusive smokeless tobacco and never tobacco user. No significant difference was observed between the Kessler index for never tobacco users and current exclusive smokeless tobacco users.

Of the ten selected comorbidities examined in this study, a statistically significant association was found between tobacco use status and having reported being diagnosed with emphysema, coronary heart disease and arthritis, with current exclusive cigarette smokers reporting the highest positive self-reported diagnosis rates. These findings are inline with the literature where cigarette usage has been found to be associated with increased emphysema and coronary heart disease due to the combustion produced during cigarette smoking.[2]

Health care utilization

Current exclusive cigarette smokers tended to be more likely to have one or more ER visits than current exclusive smokeless tobacco users and never tobacco users. Based on univariate analysis current exclusive cigarette smokers had a statistically significant higher mean number of ER and home health care visits than current exclusive smokeless tobacco users and never tobacco users.

The modeling results adjusting for sociodemographic and health status variables (multivariate analysis) indicated that current exclusive smokeless tobacco users and current exclusive cigarette smokers were not statistically significantly different than never tobacco users in their utilization of the health care services measured (i.e. with at least one ER visit, office-based visit, hospital

outpatient visit, hospital inpatient visit, home health care visit and dental visit in the past 12 months). Wang et al[1] found current smokeless tobacco users (i.e. an adult 18 years or older who now uses smokeless tobacco every day or some days) significantly differed from never tobacco users in ER visits in the past 12 months (p-value =0.043) but did not differ in the number of hospital nights (in the past 12 month), doctor visits (in the past 2 weeks) , and home care visits (in the past 2 weeks) using NHIS data from 2012-2015 and a Zero-Inflated Poisson regression model. This significant difference observed by Wang for ER visits could be due to the definition of current smokeless tobacco user which included current use of other tobacco products at the same time and former use of cigarettes.

Health Care Expenditure

The two-part model results indicated that no significant statistical association was found between total health care expenditure and tobacco use status. The marginal effects for tobacco use status indicated that current exclusive cigarette smokers spend more than never tobacco users by about $1200 and current exclusive smokeless tobacco users spend about $50 more than never tobacco users. The highest annual mean total health care expenditure in US 2017 dollars was observed in the current exclusive cigarette smoking group $5627.64 (95% CI = $4068.50, $7186.78) followed by current exclusive smokeless users $4478.33 (95% CI = $3035.05, $5921.62) and never tobacco users had the lowest annual mean cost $4426.89 (95% CI = $3514.19, $5339.59). A statistical association of age with total health care expenditure was also examined and showed that the predicted total health expenditures increased for all the tobacco use status groups with age. Swedler et al[13] examined the association between current smokers, former smokers and never smokers and medical expenditures. They found that current smokers had higher medical expenditure - $5244 in US 2015 dollar, (95%CI = $4707.9, $5580.3) - compared to never

smokers - $4360, (95%CI = $4154.3, $4566.3) - using 2011-2015 MEPS/NHIS linked data and former smokers had the highest medical expenditure ,$5590, (95%CI = $5267.4, $5913.5). Swedler et al[2] estimates for annual medical expenditures for all civilian non-institutionalized adults in the U.S. was $4830 in 2015 US dollars. The overall estimate for annual health care expenditure for U.S. adults in this study was $4869 in 2017 US dollars. The estimates for overall annual health care expenditure were close to estimates by Swedler et al and Mitchell and Machlin[13,73] estimate of $4978 for average total medical expenditures in 2015 using MEPS data.

Health care expenditure increases with aging, therefore a sensitivity analysis restricting the data to adults ≤ 65 years was run to remove the older adult population who may have higher total healthcare spending than the average adult. Although no statistical association was found between total health care expenditure and tobacco use status (at a 5% significant level), similar to the previous findings, at a 10% significance level a statistical difference was observed between current exclusive cigarette smokers and never tobacco users (p-value=0.073). The estimated annual mean total health care expenditure was highest in current exclusive cigarette smokers $4326.73 (95% CI = ($2965.46, $5687.99) followed by current exclusive smokeless users $4089.01 (95% CI = $2759.49, $5418.54) and never tobacco user has the lowest annual mean total health care expenditure $3196.20 (95% CI = $2476.35, $3916.04). Removing the older population which was about ~ 17% of the data. As expected, a decrease in total health care expenditure across all tobacco status groups was observed since the older population on average has higher health care expenditure compared to the younger population.

Another sensitivity analysis was conducted to evaluate the impact of total health care expenditure restricting the analysis to adults ≤ 65 years and males only, since smokeless tobacco products are predominately used by males(~95%) in the U.S. population.

Similar to the previous sensitivity analysis findings, no statistical association was found between total health care expenditure and tobacco use status (at a 5% significant level), but at a 10% significance level a statistical difference was observed between current exclusive cigarette smokers and never tobacco users (p-value=0.099). Never tobacco users still had the lowest annual mean total health care expenditures ($3053.73) followed by current exclusive smokeless users ($3796.06) and current exclusive cigarette smokers ($4073.95) which had the highest annual mean total health care expenditure. Excluding the small percentage of female participants (~5%) further reduced the annual mean total expenditures for all three groups which is expected given that females particularly in their childbearing age tend to have higher expenditure than males.[48,49]

Although the findings from the study showed no statistical association between total health care expenditure and tobacco use status, the highest annual mean expenditure was observed in current exclusive cigarette smokers. High medical expenditure estimates due to smoking are in line with other peer-reviewed work.[2,12,13,39] Unlike the study findings, a statistical significance difference was found between smokers and never/non smokers in terms of their total healthcare/ medical expenditure.[13,16] This difference in observing a statistical association between total health care expenditure and tobacco use status could be due to the definitions and classification of tobacco user groups and /or the study design and associated sample size.

Studies from the peer-viewed literature investigated only cigarette smoking status, where the smoking groups were defined as current smoker, former smoker and never/non smoker[2,12,13] or smoker (current and former combined) and non-smoker.[39] In studies where current smokers, former smokers and never/non-smokers were defined, former smokers tended to have the highest health care expenditure followed by current smokers and never/non-smokers. These

findings likely reflect quitting following the onset of an illness due to cigarette smoking. [2,4,5] In contrast this study's objective was to compare health care expenditures and health care utilization across different current exclusive tobacco product usage, therefore tobacco use status definitions were extended across two different tobacco products (cigarettes and smokeless tobacco use) and never/non users of these tobacco products and are further restricted to current exclusive use of the tobacco products at time of interview excluding former tobacco product users.

Secondly, this study is the first to my knowledge that uses propensity score matching to control for observable differences between the tobacco use status groups. This reduced the original sample size (i.e. 68,866 to 1,899) and therefore potentially increased the variability in the sample distribution. The current exclusive smokeless tobacco group (group1-'case') (N=633)was used to match observations in the other tobacco status groups (i.e. group 2; current cigarette smokers, group 3; never tobacco users-'controls'). The final matched analysis dataset on average had younger participants than observed in the original data set (also in both smokers and never smokers groups), ~83% of the sample was ≤ 65 years old and was predominately male (94%) since smokeless tobacco is mainly used by male in the U.S.[54] In comparison the peer-reviewed studies had larger sample sizes with less variability in their sample distributions and on average tended to have approximately similar ratio of males to females.

The sensitivity analysis excluding participants ≥ 65 years from the analysis resulted in a statistically significant difference between current exclusive cigarette user and never tobacco users at a 10% significance level.

Comparing estimates for annual expenditures from this study and other studies that used MEPS with estimates generated from using the National Health Expenditure Assessment (NHEA)

database, the study estimates are lower than similar estimates using NHEA data.[74,75] This is not unexpected given the fact that NHEA is more comprehensive than MEPS in capturing Medicaid costs covering institutionalized adults including those in nursing homes, active-duty military and foreign visitors to the USA.[75,76]

Strengths and Limitation

This is the first study to my knowledge that estimates and compares health care expenditures and utilization associated with current exclusive cigarette smoking, current exclusive smokeless tobacco use and never tobacco use, among U.S. adults aged 18 years and older. The rationale for selecting the two tobacco products (i.e. cigarettes and smokeless tobacco use) is because they are the most commonly used tobacco products in the U.S. and have been on the market for decades as compared to the more novel tobacco products, therefore providing the most comprehensive data. Also, these two products are on opposite ends of the continuum of risk for nicotine containing products; cigarettes on the highest level and smokeless tobacco on the lower end.

The analyses compared healthcare expenditure and use estimates across current exclusive use of tobacco products within the U.S. we are not aware of any other studies that have systematically quantified and compared the direct economic costs of adults who are current exclusive users of cigarettes, smokeless tobacco users or never users of tobacco products.

A common challenge in observational data analysis is addressing selection bias or confounding resulting from a lack of randomization. When the groups of interest are not randomized, there is the likelihood for the groups to differ in key variables (e.g. sociodemographics like age and gender) and not be comparable. To reduce the potential for bias, propensity score matching was used to balance the covariates across the three study groups. Covariates that were not balanced

through propensity score matching were controlled for in the models used in the multivariate analyses. As with most self-reported data, health care use may be subject to recall bias or underreporting. In the case of this study data, MEPS addresses nonresponse bias by imputing missing expenditure data instead of excluding these cases from the analysis dataset.[77] Also MEPS provides valuable data on characteristics of MEPS non respondents that reduces MEPS nonresponse bias through various checks and balances during the five rounds of MEPS data collects (e.g. variables like age are checked across a selected number of rounds for age verification).[78]

The model estimates from the study should be considered in the context of some limitations. Although two-part models are considered the best modeling approach for health expenditure data, the robustness of the estimates depends on the extent to which all the factors of healthcare spending are identified and considered.[13,68] As indicated in Swedler et al, while these types of models allocate costs to a specified risk factor or medical condition based on statistical estimations, the underlying cause of a person's medical event or cost is not known. Individuals who are more conscious of their health and seek out care may have more medical spending (e.g. more preventive care visits) than those who are involved in risky behavior.[13]

Findings from the study are only relevant to the non-institutionalized population. Due to this, NHIS and MEPS survey design may underestimate total health care expenditure given that cost information on institutionalized adults are excluded from this analysis. Another limitation is that there is a time lag of approximately a year that occurs between when individual characteristics are measured in the NHIS and when healthcare expenditures measured in the MEPS, so that the estimated health care costs associated with tobacco use status are accumulative.[42]

Smoking at least 100 cigarettes in one's lifetime (and smoked some days or every day at the time of the interview) was used to categorize smoking status to comply with the CDC's health surveillance definition and to be able to make comparisons with other studies. This definition is age dependent. Results from subgroup analysis by age group indicate that older adults with potentially longer smoking histories have substantially higher healthcare expenditure compared with their younger counterparts.

Conclusion and Future Research

In conclusion, cigarette smoking remains the leading preventable cause of death and disease in the U.S. and poses a major health hazard and public health issue, even though the results of this study found no statistical differences in health care expenditures and utilization among current exclusive cigarette smoking, current exclusive smokeless tobacco use and never tobacco use. This study is the first to compare medical expenditure and health care utilization associated with current exclusive cigarette smoking, current exclusive smokeless tobacco use and never tobacco use, in U.S. adults' over time. More data may be needed to capture a larger balanced dataset across current exclusive cigarette smoking, current exclusive smokeless tobacco use and never tobacco use. Given the relatively low prevalence of smokeless tobacco use compared to cigarette smoking and the stringent definitions used to define current exclusive tobacco product usage a large percentage of the data was not included in the study analysis.

Although not statistically significant, the mean annual healthcare expenditure for current exclusive cigarette smokers tended to be higher compared to the mean values for current smokeless tobacco users and never tobacco users. Study findings show the economic cost of tobacco products directionally reflects the continuum of risk in the Tobacco Harm Reduction Model, with cigarettes on one end and smokeless tobacco users on the other end. Indicating a possible reduce in healthcare expenditure by aiding cigarette smokers down the continuum of

risk of nicotine containing products to non-combustible tobacco products with less associated risk.

Future research should focus on examining other databases that capture both the institutionalized and non-institutionalized U.S. population using the same study design (e.g. use data from NHEA). Given the continuous changes to the tobacco use landscape, future studies where there is enough data should use a longitudinal design to estimate heath care expenditure of cigarette smokers who switched to a lower risk tobacco product (i.e. based on the Tobacco Harm Reduction model) and have a significant history of usage of the product to assess the potential reduction in healthcare expenditure. Results from the quality of life and depression analyses support this possibility, because in this study we observed both quality of life and tendencies towards mental disability were worst for current cigarette smokers compared to smokeless tobacco users.

Future research should also compare health care expenditure and use in the older adult U.S. population (i.e. age 65 and older), given that most of the tobacco related disease that lead to higher medical expenditures in tobacco users are typically observed at later stages in life. My current analysis dataset did not have sufficient sample size to explore this subset of the population. A better understanding of the comparison of healthcare expenditure and use across tobacco use status within the older population can help identify other underlying variables that may increase healthcare cost besides aging, insurance type and quality of life.

References

1. U.S. Department of Health and Human Services. The Health Consequences of Smoking: A Report of the Surgeon General; Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health: Atlanta, GA: U.S, 2004. [accessed November 2019].

2. U.S. Department of Health and Human Services. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. Atlanta, GA: U.S: Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2014.

3. Ostbye T, Taylor DH, Jung SH. A longitudinal study of the effects of tobacco smoking and other modifiable risk factors on ill health in middle-aged and old Americans: results from the Health and Retirement Study and Asset and Health Dynamics among the Oldest Old survey. Preventive Medicine 2002;34(3):334–45.

4. Johnson PB, Richter L. The relationship between smoking, drinking, and adolescents' self-perceived health and frequency of hospitalization: analyses from the 1997 National Household Survey on Drug Abuse. Journal of Adolescent Health 2002;30(3):175–83.

5. Kahende JW, Adhikari B, Maurice E, Rock V, Malarcher A. Disparities in health care utilization by smoking status—NHANES 1999–2004. International Journal of Environmental Research and Public Health 2009;6(3):1095–106.

6. Fishman PA, Khan ZM, Thompson EE, Curry SJ. Health care costs among smokers, former smokers, and never smokers in an HMO. Health Services Research. 2003;38(2):733–49.

7. Max W, Sung HY, Shi Y. Deaths from secondhand smoke exposure in the United States: economic implications. American Journal of Public Health. 2012;102(11):2173–80

8. Max W, Sung HY, Shi Y, et al. The cost of smoking in California. Nicotine Tob Res 2016; 18:1222–9.

9. Max W, Rice DP, Sung HY, et al. The economic burden of smoking in California. Tob Control 2004; 13:264–7.

10. Max W, Sung H-Y, Shi Y, et al. The cost of smoking in California, 2009. San Francisco, CA: Institute for Health & Aging, University of California, San Francisco, 2014.

11. Warner KE, Hodgson TA, Carroll CE. Medical costs of smoking in the United States: estimates, their validity, and their implications. Tobacco Control. 1999;8(3):290–300.

12. Xu X, Bishop EE, Kennedy SM, Simpson SA, Pechacek TF. Annual Healthcare Spending Attributable to Cigarette Smoking: An Update . American Journal of Preventive Medicine 2014;48(3):326–33

13. Swedler D, miller T, Ali B, et al. National medical expenditures by smoking status in American adults: an application of Manning's two-stage model to nationally

representative data. BMJ Open. 2019 Jul 16;9(7):e026592. doi: 10.1136/bmjopen-2018-026592

14. American Cancer Society. Cancer Facts & Figures 2017; American Cancer Society: Atlanta, GA, USA, 2017.

15. Global Initiative for Chronic Obstructive Pulmonary Disease. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease, 2011; ; accessed: August 6, 2013.

16. Congressional Budget Office. Raising the Excise Tax on Cigarettes: Effects on Health and the Federal Budget. Washington: Congressional Budget Office, 2012; accessed: March 8, 2021.

17. Levy DE, Newhouse JP. Assessing the effects of tobacco policy changes on smoking-related health expenditures. In: Bearman PS, Neckerman KM, L W, editors. After Tobacco: What Would Happen If Americans Stopped Smoking? New York: Columbia University Press, 2011:256–89.

18. U.S. Department of Health and Human Services. Smoking Cessation. A Report of the Surgeon General. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2020.

19. Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years' observations on male British doctors. BMJ 2004;328(7455):1519.

20. Jha P, Ramasundarahettige C, Landsman V, Rostron B, Thun M, Anderson RN, McAfee T, Peto R. 21st-cen tury hazards of smoking and benefits of cessation in the United States. New England Journal of Medicine 2013;368(4):341–50

21. CDC. Quitting smoking among adults—United States, 2001–2010. Morbidity and Mortality Weekly Report. 2011c;60(44):1513–1519.

22. Babb S, Malarcher A, Schauer G, Asman K, Jamal A. Quitting smoking among adults—United States, 2000–2015. Morbidity and Mortality Weekly Report 2017;65(52):1457–64.

23. Prochaska JJ, Benowitz NL. The past, present, and future of nicotine addiction therapy. Annual Review of Medicine 2016;67:467–86.

24. Schauer GL, Malarcher AM, Babb SD. Gradual reduction of cigarette consumption as a cessation strategy: prevalence, correlates, and relationship with quitting. Nicotine and Tobacco Research 2015;17(5):530–8.

25. Harm Reduction International. What is harm reduction? https://www.hri.global/what-is-harm-reduction

26. United Nations Office on Drugs and Crime. Reducing the Harm of Drug Use and Dependence and HIV Risk Reduction. https://www.unodc.org/ddt-training/treatment/VOLUME%20D/Topic%204/1.VolD_Topic4_Harm_Reduction.pdf

27. Moore CA. History and context of harm reduction in the united states. In: Richard Pates DR, editor. Harm reduction in substance use and high-risk behaviour. United States: John Wiley & Sons; 2012. p. 382–94.

28. des Jarlais DC. Harm reduction in the USA: the research perspective and an archive to David Purchase. Harm Reduct J. 2017 Jul 26;14(1):51. doi: 10.1186/s12954-017-0178-6. https://harmreductionjournal.biomedcentral.com/articles/10.1186/s12954-017-0178-6.

29. Langendam MW, van Brussel GH, Coutinho RA, van Ameijden EJ. The impact of harm-reduction-based methadone treatment on mortality among heroin users. Am J Public Health. 2001;91:774–80.

30. Burning Issues: Global State of Tobacco Harm Reduction 2020. London: Knowledge- -Action-Change, 2020.

31. Hatsukami, D.K.; Joseph, A.M.; Lesage, M.; Jensen, J.; Murphy, S.E.; Pentel, P.R.; Kotlyar, M.; Borgida, E.; Le, C.; Hecht, S.S. Developing the science base for reducing tobacco harm. Nicotine Tob. Res. 2007, 9, 537-553.

32. Zeller, M.; Hatsukami, D. The strategic dialogue on tobacco harm reduction: a vision and blueprint for action in the US. Tob. Control 2009, 18, 324-332.

33. Nutt, David J., Lawrence D. Phillips, David Balfour, H. Valerie Curran, Martin Dockrell, Jonathan Foulds, Karl Fagerstrom et al. 2014. “Estimating the Harms of Nicotine-Containing Products Using the MCDA Approach.” European Addiction Research 20, no. 5: 218–25. doi: 10.1159/000360220.

34. Scientific Committee on Emerging and Newly Identified Health Risks. 2008. "Scientific Opinion on the Health Effects of Smokeless Tobacco Products." February 6, 2008. ec.europa.eu/health/ph_risk/committees/04_scenihr/docs/scenihr_o_013.pdf.

35. Wang Y, Sung HY, Lightwood J, et al. Health Care Utilization and Expenditures Attributable to Smokeless Tobacco Use Among US Adults. Nicotine Tob Res. 2018 Sep 25; 20(11):1359-1368.

36. Wang Y, Sung HY, Yao T, Lightwood J, Max W. Health Care Utilization and Expenditures Attributable to Cigar Smoking Among US Adults, 2000-2015. Public Health Rep. 2018 May/Jun; 133(3):329-337.

37. Parsons VL, Moriarity C, Jonas K, et al. Design and estimation for the National Health Interview Survey, 2006–2015. National Center for Health Statistics. Vital Health Stat 2(165). 2014.

38. Chowdhury, S.R., Machlin, S.R., Gwet, K.L. Sample Designs of the Medical Expenditure Panel Survey Household Component, 1996–2006 and 2007–2016. Methodology Report #33. January 2019. Agency for Healthcare Research and Quality, Rockville, MD. https://meps.ahrq.gov/data_files/publications/mr33/mr33.pdf

39. An R. Health care expenses in relation to obesity and smoking among U.S. adults by gender, race/ethnicity, and age group: 1998- 2011. Public Health 2015;129:29–36

40. Agency for Healthcare Research and Quality. Medical Expenditure panel Survey. (2020, Dec. 20). MEPS NHIS Link: 1996 - 2018 MEPS Link Files to NHIS. Retrieved from Medical Expenditure Panel Survey Public Use File Details (ahrq.gov)

41. National Center for Health Statistics. National Health Interview Survey. 2020. https://www.cdc.gov/nchs/nhis/index.htm

42. Agency for Healthcare Research and Quality. Medical Expenditure Panel Survey. 2020. https://meps.ahrq.gov/mepsweb/

43. Dunn, A., Grosse, S.D., & Zuvekas S.H. (2018). Adjusting Health Expenditures for Inflation: A Review of Measures for Health Services Research in the United States. Health Services Research, 53(1), 175-196.

44. Ware, J.E., Kosinski, M., and Keller, S.D. (1996). A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. Medical Care 34:220.

45. Kessler, R.C., Andrews, G., Colpe, L.J., Hiripi, E., Mroczek, D.K., Normand, S.L., Walters, E.E., and Zaslavsky, A.M. (2002). Short screening scales to monitor population prevalence and trends in non-specific psychological distress. Psychological Medicine 32: 959-976. C-59 MEPS HC-192

46. Kroenke, K., Spitzer, R.L., and Williams, J.B. (2003). The Patient Health Questionnaire-2: Validity of a two-item depressive screener. Medical Care 41: 1284-1292.

47. Ware, Jr., J.E., Kosinski, M., Turner-Bowker, DM, and Gandek, B. How to Score Version 2 of the SF-12 (r) Health Survey. (October, 2002). QualityMetric, Inc., Lincoln, RI.

48. Mitchell, E.M. Concentration of Healthcare Expenditures and Selected Characteristics

of High Spenders, U.S. Civilian Noninstitutionalized Population, 2017. Statistical Brief #528. February 2020. Agency for Healthcare Research and Quality, Rockville, MD. http://www.meps.ahrq.gov/mepsweb/data_files/publications/st528/stat528.pdf

49. Kaiser Family Foundation analysis of Medical Expenditure Panel Survey, Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services.

50. Sengupta N, Nichol MB, Wu J, et al. Mapping the SF-12 to the HUI3 and VAS in a managed care population. Med Care 2004;42:927–37.

51. Arroll B, Goodyear-Smith F, Crengle S, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. Ann Fam Med 2010;8:348–53.

52. Kessler RC, Barker PR, Colpe LJ, et al. Screening for serious mental illness in the general population. Arch Gen Psychiatry 2003; 60:184–9.

53. Elixhauser A, Steiner C, Harris DR, et al. Comorbidity measures for use with administrative data. Med Care 1998; 36:8–27.

54. Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. Trends in Smokeless Tobacco Use and Initiation: 2002 to 2012; The National Survey on Drug Use and Health (NSDUH): Rockville, MD, USA, 2014.

55. MEPS HC-036: 1996-2018 Pooled Linkage File for Common Variance Structure. July 2020. Agency for Healthcare Research and Quality, Rockville, Md. https://www.meps.ahrq.gov/data_stats/download_data/pufs/h36/h36u18doc.pdf

56. SAS Institute Inc. 2016. SAS/STAT® 14.2 User's Guide. Cary, NC: SAS Institute Inc.

57. Rosenbaum P.R. and Rubin D.B. "The Central Role of the Propensity Score in Observational Studies for Causal Effects", Biometrika, 1983, 70, 41-55.

58. Rubin, D. B. (2001). "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation." Health Services and Outcomes Research Methodology 2:169–188.

59. Stuart, E. A. (2010). "Matching Methods for Causal Inference: A Review and a Look Forward." Statistical Science 25:1–21

60. Normand, S.-L. T., Landrum, M. B., Guadagnoli, E., Ayanian, J. Z., Ryan, T. J., Cleary, P. D., and McNeil, B. J. (2001). "Validating Recommendations for Coronary Angiography Following Acute Myocardial Infarction in the Elderly: A Matched Analysis Using Propensity Scores." Journal of Clinical Epidemiology 54:387–398.

61. Mamdani, M., Sykora, K., Li, P., Normand, S. L., Streiner, D. L., Austin, P. C., Rochon, P. A., and Anderson, G. M. (2005). "Reader's Guide to Critical Appraisal of Cohort Studies: 2. Assessing Potential for Confounding." BMJ 330:960–962.

62. Austin, P. C. (2009). "Balance Diagnostics for Comparing the Distribution of Baseline Covariates between Treatment Groups in Propensity-Score Matched Samples." Statistics in Medicine 28:3083–3107.

63. Manning WG, Basu A, Mullahy J. Generalized modeling approaches to risk adjustment of skewed outcomes data. J Health Econ.2005;24:465–88.

64. Deb P, Norton EC, Manning WG. 2017. Health Econometrics Using Stata. College Station, TX:Stata Press

65. Duan N, Manning WG, Morris CN, Newhouse JP. A comparison of alternative models for the demand for medical care. J Bus Econ Stat 1983;1(2):115e26.

66. Belotti F, Deb P, Manning WG, et al. Twopm: Two-Part Models. Stata J 2015;15:3–20.

67. StataCorp. 2017. Stata: Release 15. Statistical Software. College Station, TX: StataCorp LLC.

68. Deb Partha and Norton Edward C. Modeling Health Care Expenditures and Use Annual Review of Public Health 2018 39:1, 489-505

69. Stewart AL, Greenfield S, Hays RD, et al. Functional status and well-being of patients with chronic conditions. Results from the Medical Outcomes Study. JAMA. 1989;262:907-913.

70. Warkentin LM, Majumdar SR, Johnson JA, et al. Weight loss required by the severely obese to achieve clinically important differences in health-related quality of life: two-year prospective cohort study. BMC Med. 2014;12:175.

71. Wyrwich KW, Tierney WM, Babu AN, Kroenke K, Wolinsky FD. A comparison of clinically important differences in health-related quality of life for patients with chronic lung disease, asthma, or heart disease. Health Serv Res. 2005;40:577-591.

72. H.P. Kale, N.V. Carroll. Self-reported financial burden of cancer care and its effect on physical and mental health-related quality of life among US cancer survivors. Cancer, 122 (2016), pp. 283-289

73. Mitchell E, Machlin S. Concentration of health expenditures and selected characteristics of high spenders, U.S. civilian noninstitutionalized population, 2015. Statistical Brief #506. Rockville, MD: Agency for Healthcare Research and Quality, 2017.

74. Lassman D, Hartman M, Washington B, et al. US health spending trends by age and gender: selected years 2002-10. Health Aff 2014; 33:815–22.

75. Finkelstein EA, Trogdon JG, Cohen JW, et al. Annual medical spending attributable to obesity: payer-and service-specific estimates. Health Aff 2009;28: w822–w831.

76. Sing M, Banthin JS, Selden TM, et al. Reconciling medical expenditure estimates from the MEPS and NHEA, 2002. Health Care Financ Rev 2006;28:25–40.

77. Machlin, S. R. and Dougherty, D. D. Overview of Methodology for Imputing Missing Expenditure Data in the Medical Expenditure Panel Survey. Method-ology Report No. 19. March 2007. Agency for Healthcare Research and Quality, Rockville, Md. http://www.meps.ahrq.gov/mepsweb/data_files/publications/mr19/mr19.shtml

78. Cohen SB. Design strategies and innovations in the Medical Expenditure Panel Survey. Med Care. 2003;41(7):(suppl) iii5-III1212865722

Appendix A

Section 1 : Sensitivity Analysis- Healthcare Utilization Models before Propensity Score Matching

Table 35: Results of the Two-part model for Total Health Care Expenditure (2017 US $) per year, 2011-2017 before Propensity Score Matching

Covariates	Total Health Care Expenditure			
	Logit		GLM	
	Coefficient (Linearized std. error)	*P-value*	*Coefficient (Linearized std. error)*	*P-value*
Tobacco Use Status				
Current exclusive cigarette smoker	-0.441 (0.058)	0.0001	-0.018 (0.049)	0.720
Current exclusive ST user	-0.616 (0.217)	0.005	-0.389 (0.234)	0.099
Never Tobacco User	Ref		Ref	
Age	0.032 (0.002)	0.0001	0.015 (0.001)	0.0001
Gender				
Female	1.046 (0.055)	0.0001	0.147 (0.041)	0.0001
Male	Ref		Ref	
Poverty Status				
Poor/Negative	-0.601 (0.085)	0.0001	-0.268 (0.064)	0.0001
Near Poor	-0.516 (0.113)	0.0001	-0.230 (0.090)	0.012
Low Income	-0.596 (0.078)	0.0001	-0.270 (0.059)	0.0001

Covariates	Total Health Care Expenditure			
	Logit		GLM	
	Coefficient (Linearized std. error)	*P-value*	*Coefficient (Linearized std. error)*	*P-value*
Middle Income	-0.417 (0.062)	0.0001	-0.160 (0.048)	0.001
High Income	Ref		Ref	
Insurance type				
Public	1.47 (0.062)	0.0001	1.13 (0.284)	0.0001
Private	1.19 (0.084)	0.0001	1.30 (0.317)	0.0001
Uninsured	Ref		Ref	
Quality of life score (Short-Form 12)				
Physical Component Summary (0-100)	-0.056 (0.004)	0.0001	-0.045 (0.007)	0.0001
Mental Component Summary (0-100)	-0.021 (0.004)	0.0001	-0.031 (0.012)	0.010

Table 36:Results of the Two-part model for Total Health Care Expenditure (2017 US $) per year (Age ≥ 65 years), 2011-2017

Covariates	Total Health Care Expenditure			
	Logit		GLM	
	Coefficient (Linearized std. error)	*P-value*	*Coefficient (Linearized std. error)*	*P-value*
Tobacco Use Status				
Current exclusive cigarette smoker	-0.405 (0.058)	0.0001	-0.033 (0.056)	0.559
Current exclusive ST user	-0.593 (0.215)	0.006	-0.340 (0.249)	0.173
Never Tobacco User	Ref		Ref	
Age	0.027 (0.003)	0.0001	0.015 (0.002)	0.0001
Gender				

Covariates	Total Health Care Expenditure			
	Logit		GLM	
	Coefficient (Linearized std. error)	*P-value*	*Coefficient (Linearized std. error)*	*P-value*
Female	1.056 (0.056)	0.0001	0.198 (0.045)	0.0001
Male	Ref		Ref	
Poverty Status				
Poor/Negative	-0.571 (0.089)	0.0001	-0.305 (0.071)	0.0001
Near Poor	-0.532 (0.116)	0.0001	-0.233 (0.106)	0.029
Low Income	-0.579 (0.079)	0.0001	-0.297 (0.067)	0.0001
Middle Income	-0.409 (0.066)	0.0001	-0.160 (0.053)	0.004
High Income	Ref		Ref	
Insurance type				
Public	1.46 (0.064)	0.0001	0.609 (0.109)	0.0001
Private	1.13 (0.084)	0.0001	0.685 (0.112)	0.0001
Uninsured	Ref		Ref	
Quality of life score (Short-Form 12)				
Physical Component Summary (0-100)	-0.055 (0.004)	0.0001	-0.045 (0.007)	0.0001
Mental Component Summary (0-100)	-0.022 (0.004)	0.0001	-0.031 (0.012)	0.010

Section 2 : Sensitivity Analysis- Healthcare Expenditure Models before Propensity Score Matching

Table 37: Results for Hurdle model for number of Emergency room visits, 2011-2017 before Propensity Score Matching

Covariates	Emergency room visits (in the past 12 months)			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Tobacco Use Status				
Current exclusive cigarette smoker	0.265	0.0001	0.057	0.396
Current exclusive ST user	-0.043	0.836	-0.807	0.096
Never Tobacco User	Ref		Ref	
Age	-0.007	0.0001	-0.006	0.008
Gender				
Male	Ref		Ref	
Female	0.278	0.0001	0.072	0.244
Education				
College or Higher	0.041	0.309	0.065	0.393
Some College	0.089	0.098	0.248	0.003
High School	0.084	0.102	0.124	0.144
Less than High School	Ref		Ref	
Race/ethnicity				
White, non-Hispanic	Ref		Ref	
Black, non-Hispanic	0.239	0.0001	-0.029	0.743
Asian, non-Hispanic	-0.221	0.0001	-0.157	0.111
Other or multiple race, non-Hispanic	0.178	0.111	-0.184	0.344
Hispanic	-0.078	0.138	-0.179	0.032
Region of residency				
Northeast	0.116	0.028	0.131	0.123
Midwest	0.073	0.123	-0.074	0.358
West	-0.131	0.007	-0.224	0.009
South	Ref		Ref	

Marital Status				
Married	Ref		Ref	
Living with Partner	0.162	0.019	0.186	0.101
Widow/divorce/separated	0.164	0.001	0.061	0.472
Never Married	0.128	0.007	0.127	0.160
Body mass index (BMI)				
Normal or Under Weight	Ref		Ref	
Overweight	0.018	0.698	-0.101	0.198
Obesity	0.126	0.005	-0.042	0.572
Poverty Status				
Poor/Negative	0.265	0.0001	0.228	0.043
Near Poor	0.233	0.005	0.152	0.237
Low Income	0.143	0.017	0.128	0.233
Middle Income	0.062	0.203	0.067	0.494
High Income	Ref		Ref	
Self-reported binge drinking status				
No	Ref		Ref	
Yes	-0.006	0.886	-0.020	0.759
Insurance type				
Uninsured	Ref		Ref	
Public	0.544	0.0001	0.314	0.001
Private	0.232	0.0001	0.178	0.092
Quality of life score (Short-Form 12)				
Physical Component Summary (0-100)	-0.041	0.0001	-0.026	0.0001
Mental Component Summary (0-100)	-0.014	0.0001	-0.006	0.124
Mental illness score (Kessler 6 Index)	0.012	0.117	0.023	0.018
Depression score (Personal health Index)				

0	Ref		Ref	
1	-0.022	0.724	0.054	0.602
2	0.057	0.393	-0.041	0.712
3	0.144	0.185	-0.041	0.795
4	0.098	0.390	-0.014	0.934
5	-0.261	0.135	-0.096	0.671
6	0.209	0.165	-0.344	0.137

Table 38: Results for Hurdle model for number of Office-based visits, 2011-2017 before Propensity Score Matching

Covariates	**Office based visits (in the past 12 months)**			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Tobacco Use Status				
Current exclusive cigarette smoker	-0.407	0.0001	-0.180	0.0001
Current exclusive ST user	-0.375	0.005	-0.238	0.039
Never Tobacco User	Ref		Ref	
Age	0.313	0.0001	0.008	0.0001
Gender				
Male	Ref		Ref	
Female	0.954	0.0001	0.240	0.0001
Education				
College or Higher	0.228	0.0001	0.169	0.0001
Some College	0.150	0.001	0.121	0.0001
High School	-0.038	0.391	0.051	0.120
Less than High School	Ref		Ref	

Race/ethnicity				
White, non-Hispanic	Ref		Ref	
Black, non-Hispanic	-0.380	0.0001	-0.211	0.0001
Asian, non-Hispanic	-0.141	0.003	-0.215	0.0001
Other or multiple race, non-Hispanic	-0.014	0.892	-0.067	0.286
Hispanic	-0.402	0.0001	-0.207	0.0001
Region of residency				
Northeast	0.113	0.017	0.079	0.0001
Midwest	0.209	0.0001	0.228	0.005
West	-0.304	0.451	0.157	0.0001
South	Ref		Ref	
Marital Status				
Married	Ref		Ref	
Living with Partner	-0.400	0.493	0.050	0.188
Widow/divorce/separated	0.032	0.472	0.059	0.022
Never Married	0.116	0.004	0.107	0.0001
Body mass index (BMI)				
Normal or Under Weight	Ref		Ref	
Overweight	-0.061	0.103	-0.051	0.052
Obesity	-0.014	0.724	-0.041	0.137
Poverty Status				
Poor/Negative	-0.543	0.0001	-0.214	0.0001
Near Poor	-0.500	0.0001	-0.195	0.002
Low Income	-0.443	0.0001	-0.283	0.0001
Middle Income	-0.346	0.0001	-0.219	0.0001
High Income	Ref		Ref	
Self-reported binge drinking status				

No	Ref		Ref	
Yes	0.085	0.011	-0.032	0.196
Insurance type				
Uninsured	Ref		Ref	
Public	1.242	0.0001	0.532	0.0001
Private	1.312	0.0001	0.522	0.0001
Quality of life score (Short-Form 12)				
Physical Component Summary (0-100)	-0.048	0.0001	-0.029	0.0001
Mental Component Summary (0-100)	-0.027	0.0001	-0.015	0.0001
Mental illness score (Kessler 6 Index)	0.021	0.008	0.004	0.329
Depression score (Personal health Index)				
0	Ref		Ref	
1	0.125	0.030	-0.005	0.882
2	-0.026	0.684	-0.037	0.393
3	-0.140	0.249	0.025	0.726
4	-0.237	0.059	-0.062	0.397
5	-0.339	0.119	-0.156	0.150
6	-0.239	0.194	-0.106	0.234

Table 39: Results for Hurdle model for number of Hospital Outpatient visits, 2011-2017 before Propensity Score Matching

Covariates	**Hospital Outpatient visits (in the past 12 months)**			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Tobacco Use Status				
Current exclusive cigarette smoker	-0.184	0.0001	-0.229	0.046

Current exclusive ST user	-0.617	0.006	-1.161	0.003
Never Tobacco User	Ref		Ref	
Age	0.023	0.0001	0.004	0.091
Gender				
Male	Ref		Ref	
Female	0.524	0.0001	-0.212	0.008
Education				
College or Higher	0.129	0.004	0.001	0.995
Some College	0.107	0.046	-0.077	0.547
High School	0.093	0.086	-0.019	0.885
Less than High School	Ref		Ref	
Race/ethnicity				
White, non-Hispanic	Ref		Ref	
Black, non-Hispanic	-0.244	0.0001	0.394	0.001
Asian, non-Hispanic	-0.265	0.0001	0.082	0.414
Other or multiple race, non-Hispanic	0.0005	0.997	0.467	0.108
Hispanic	-0.427	0.0001	0.040	0.743
Region of residency				
Northeast	0.565	0.0001	0.171	0.124
Midwest	0.562	0.0001	0.119	0.274
West	-0.170	0.001	0.286	0.025
South	Ref		Ref	
Marital Status				
Married	Ref		Ref	
Living with Partner	-0.101	0.164	0.001	0.990
Widow/divorce/separated	-0.026	0.549	0.001	0.966
Never Married	-0.157	0.001	0.127	0.0001

Body mass index (BMI)				
Normal or Under Weight	Ref		Ref	
Overweight	-0.011	0.797	0.067	0.475
Obesity	0.135	0.002	0.082	0.341
Poverty Status				
Poor/Negative	-0.168	0.009	-0.008	0.957
Near Poor	-0.285	0.002	0.399	0.106
Low Income	-0.227	0.0001	-0.153	0.166
Middle Income	-0.188	0.0001	-0.068	0.400
High Income	Ref		Ref	
Self-reported binge drinking status				
No	Ref		Ref	
Yes	0.039	0.327	0.053	0.636
Insurance type				
Uninsured	Ref		Ref	
Public	0.834	0.0001	0.506	0.026
Private	0.924	0.0001	0.337	0.003
Quality of life score (Short-Form 12)				
Physical Component Summary (0-100)	-0.038	0.0001	-0.027	0.0001
Mental Component Summary (0-100)	-0.011	0.0001	-0.016	0.003
Mental illness score (Kessler 6 Index)	0.006	0.411	-0.003	0.852
Depression score (Personal health Index)				
0	Ref		Ref	
1	0.112	0.056	0.013	0.939
2	0.040	0.556	-0.215	0.101
3	0.177	0.125	-0.270	0.154
4	0.071	0.552	-0.101	0.705

5	0.086	0.630	-0.494	0.094
6	0.051	0.750	-0.705	0.008

Table 40:Results for Hurdle model for number of Dental visits, 2011-2017 before Propensity Score Matching

Covariates	Dental visits (in the past 12 months)			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Tobacco Use Status				
Current exclusive cigarette smoker	-0.478	0.0001	0.032	0.318
Current exclusive ST user	-0.741	0.0001	-0.296	0.036
Never Tobacco User	Ref		Ref	
Age	0.4119	0.0001	0.010	0.0001
Gender				
Male	Ref		Ref	
Female	0.412	0.0001	0.034	0.139
Education				
College or Higher	0.263	0.0001	0.038	0.198
Some College	-0.033	0.429	0.003	0.935
High School	-0.038	0.340	0.061	0.082
Less than High School	Ref		Ref	
Race/ethnicity				
White, non-Hispanic	Ref		Ref	
Black, non-Hispanic	-0.407	0.0001	-0.189	0.0001

Asian, non-Hispanic	-0.030	0.444	-0.016	0.589
Other or multiple race, non-Hispanic	-0.279	0.002	0.131	0.071
Hispanic	-0.412	0.0001	-0.014	0.697
Region of residency				
Northeast	0.274	0.0001	0.126	0.0001
Midwest	0.400	0.0001	0.094	0.001
West	0.275	0.0001	0.138	0.0001
South	Ref		Ref	
Marital Status				
Married	Ref		Ref	
Living with Partner	-0.136	0.013	0.001	0.990
Widow/divorce/separated	-0.108	0.002	0.001	0.966
Never Married	0.063	0.076	0.127	0.0001
Body mass index (BMI)				
Normal or Under Weight	Ref		Ref	
Overweight	-0.108	0.001	-0.048	0.060
Obesity	-0.203	0.0001	0.009	0.738
Poverty Status				
Poor/Negative	-0.753	0.0001	-0.041	0.427
Near Poor	-0.772	0.0001	-0.169	0.019
Low Income	-0.883	0.0001	-0.113	0.008
Middle Income	-0.479	0.0001	-0.091	0.0001
High Income	Ref		Ref	
Self-reported binge drinking status				
No	Ref		Ref	
Yes	-0.018	0.554	-0.059	0.024
Insurance type				

Uninsured	Ref		Ref	
Public	0.661	0.0001	0.107	0.097
Private	1.078	0.0001	0.154	0.008
Quality of life score (Short-Form 12)				
Physical Component Summary (0-100)	0.005	0.007	-0.005	0.0001
Mental Component Summary (0-100)	-0.001	0.551	-0.006	0.005
Mental illness score (Kessler 6 Index)	0.018	0.005	-0.008	0.174
Depression score (Personal health Index)				
0	Ref		Ref	
1	0.095	0.045	-0.021	0.584
2	-0.042	0.436	0.005	0.917
3	-0.126	0.211	-0.025	0.827
4	-0.229	0.030	-0.030	0.747
5	-0.395	0.020	0.222	0.119
6	-0.247	0.081	0.110	0.381

Table 41: Results for Hurdle model for number of Home Health care visits, 2011-2017 before Propensity Matching

Covariates	**Home Health care visits (in the past 12 months)**			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Tobacco Use Status				
Current exclusive cigarette smoker	-0.017	0.862	-0.193	0.124
Current exclusive ST user	0.254	0.648	0.443	0.269
Never Tobacco User	Ref		Ref	

Age	0.046	0.0001	0.012	0.006
Gender				
Male	Ref		Ref	
Female	0.124	0.164	-0.269	0.017
Education				
College or Higher	0.157	0.140	0.102	0.445
Some College	0.041	0.743	-0.109	0.439
High School	0.062	0.603	-0.350	0.015
Less than High School	Ref		Ref	
Race/ethnicity				
White, non-Hispanic	Ref		Ref	
Black, non-Hispanic	0.349	0.002	0.116	0.404
Asian, non-Hispanic	-0.190	0.178	0.186	0.298
Other or multiple race, non-Hispanic	0.474	0.041	0.019	0.940
Hispanic	-0.124	0.351	0.268	0.084
Region of residency				
Northeast	0.369	0.001	-0.092	0.599
Midwest	0.156	0.172	-0.135	0.388
West	0.187	0.097	0.190	0.147
South	Ref		Ref	
Marital Status				
Married	Ref		Ref	
Living with Partner	0.060	0.793	0.228	0.380
Widow/divorce/separated	0.452	0.0001	0.455	0.002
Never Married	0.651	0.0001	0.662	0.0001
Body mass index (BMI)				
Normal or Under Weight	Ref		Ref	

Overweight	-0.215	0.048	-0.307	0.027
Obesity	-0.028	0.792	-0.254	0.072
Poverty Status				
Poor/Negative	0.375	0.008	0.592	0.004
Near Poor	0.010	0.959	0.348	0.146
Low Income	0.073	0.607	0.445	0.030
Middle Income	-0.024	0.841	0.020	0.927
High Income	Ref		Ref	
Self-reported binge drinking status				
No	Ref		Ref	
Yes	-0.131	0.229	-0.162	0.216
Insurance type				
Uninsured	Ref		Ref	
Public	1.352	0.0001	1.039	0.013
Private	1.085	0.0001	0.843	0.046
Quality of life score (Short-Form 12)				
Physical Component Summary (0-100)	-0.081	0.0001	-0.016	0.001
Mental Component Summary (0-100)	-0.031	0.0001	-0.004	0.631
Mental illness score (Kessler 6 Index)	0.021	0.204	0.030	0.125
Depression score (Personal health Index)				
0	Ref		Ref	
1	-0.010	0.944	-0.074	0.641
2	-0.237	0.098	-0.132	0.479
3	-0.183	0.384	0.197	0.441
4	-0.269	0.231	-0.269	0.318
5	-0.522	0.099	-0.161	0.652
6	-0.237	0.403	-0.429	0.181

Table 42: Results for Hurdle model for number of Hospital Inpatient visits, 2011-2017 before Propensity Matching

Covariates	**Hospital Inpatient visits (in the past 12 months)**			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Tobacco Use Status				
Current exclusive cigarette smoker	0.005	0.938	0.121	0.195
Current exclusive smokeless tobacco user	-1.110	0.009	-0.412	0.526
Never Tobacco User	Ref		Ref	
Age	0.002	0.287	0.002	0.562
Gender				
Male	Ref		Ref	
Female	0.395	0.0001	-0.298	0.001
Education				
College or Higher	0.084	0.192	-0.150	0.099
Some College	0.132	0.076	0.187	0.101
High School	0.058	0.423	0.290	0.022
Less than High School	Ref		Ref	
Race/ethnicity				
White, non-Hispanic	Ref		Ref	
Black, non-Hispanic	0.026	0.712	0.133	0.193
Asian, non-Hispanic	-0.027	0.728	0.180	0.238

Covariates	**Hospital Inpatient visits (in the past 12 months)**			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Other or multiple race, non-Hispanic	0.157	0.311	0.208	0.340
Hispanic	-0.031	0.666	-0.194	0.072
Region of residency				
Northeast	-0.024	0.743	-0.030	0.788
Midwest	0.050	0.441	-0.203	0.030
West	-0.163	0.015	-0.147	0.253
South	Ref		Ref	
Marital Status				
Married	Ref		Ref	
Living with Partner	0.083	0.391	0.026	0.847
Widow/divorce/separated	-0.120	0.056	0.173	0.116
Never Married	-0.204	0.003	0.079	0.469
Body mass index (BMI)				
Normal or Under Weight	Ref		Ref	
Overweight	-0.205	0.001	-0.056	0.416
Obesity	-0.077	0.206	-0.078	0.619
Poverty Status				
Poor/Negative	0.370	0.0001	-0.004	0.979
Near Poor	0.253	0.027	0.165	0.318
Low Income	0.067	0.431	0.162	0.232
Middle Income	0.059	0.371	-0.012	0.901
High Income	Ref		Ref	
Self-reported binge drinking status				
No	Ref		Ref	

Covariates	**Hospital Inpatient visits (in the past 12 months)**			
	Logit		Truncated Poisson	
	Coefficient	*P-value*	*Coefficient*	*P-value*
Yes	0.032	0.580	0.050	0.541
Insurance type				
Uninsured	Ref		Ref	
Public	1.129	0.0001	0.251	0.154
Private	0.899	0.0001	0.061	0.718
Quality of life score (Short-Form 12)				
Physical Component Summary (0-100)	-0.059	0.0001	-0.025	0.0001
Mental Component Summary (0-100)	-0.017	0.0001	-0.019	0.002
Mental illness score (Kessler 6 Index)	-0.002	0.882	-0.016	0.236
Depression score (Patient Health Index)				
0	Ref		Ref	
1	0.010	0.903	-0.035	0.810
2	-0.134	0.142	-0.122	0.517
3	-0.100	0.494	-0.116	0.601
4	0.028	0.851	-0.073	0.759
5	-0.119	0.607	-0.005	0.988
6	-0.0004	0.998	0.233	0.466

Appendix B

Outputs for Healthcare Expenditure and Healthcare Utilization Models

Healthcare Expenditure Models

Full Expenditure Model

```
* Two-part model, with logit first part and GLM second part all variables (i.e
sociodemographic and comorbidities)
svy:twopm totexpi c.agelast ib3.tobs2 i.sex ib3.educat3 ib3.marcat2 i.bmicat2 ib2.racethx
ib3.regcat2 i.binge2 ib5.povcat ib3.inscov c.pcs42 c.mcs42 c.k6sum42 i.phq242  ,
firstpart(logit) secondpart(glm, family(gamma) link(log))
(running twopm on estimation sample)

Survey data analysis

Number of strata   =       162                Number of obs     =       1,298
Number of PSUs     =       297                Population size   =  15,368,643
                                              Design df         =         135
                                              F(  34,    102)   =        6.66
                                              Prob > F          =      0.0000
------------------------------------------------------------------------------
                     |             Linearized
             totexpi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
logit                |
             agelast |   .0185455   .0101043     1.84   0.069    -.0014376    .0385287
               tobs2 |
                   1 |   -.351833   .2483546    -1.42   0.159    -.8430021     .139336
                   2 |   -.229222   .2953677    -0.78   0.439    -.8133684    .3549245
                     |
                 2.sex |   .8284252   .4388348     1.89   0.061    -.0394549    1.696305
             educat3 |
COLLEGE GRAD OR HIGH |  -.1590572   .2743212    -0.58   0.563      -.70158    .3834657
                  HS |  -.0906717   .2749685    -0.33   0.742    -.6344749    .4531314
        SOME COLLEGE |  -.0778594    .335173    -0.23   0.817    -.7407285    .5850098
             marcat2 |
                   1 |   -.423125   .2783135    -1.52   0.131    -.9735435    .1272936
                   2 |  -.1540523   .3726572    -0.41   0.680    -.8910535     .582949
                   4 |   .0719318   .4346279     0.17   0.869    -.7876286    .9314921
             bmicat2 |
                   2 |   .0576359   .2466911     0.23   0.816    -.4302431     .545515
                   3 |  -.0489334   .2540454    -0.19   0.848     -.551357    .4534902
```

```
                racethx |
                     1  |  -.4531627   .3224679    -1.41   0.162    -1.090905    .1845795
                     3  |   .0270999   .3087996     0.09   0.930    -.5836107    .6378104
                     4  |  -.2905011   .2493947    -1.16   0.246    -.7837271    .2027249
                     5  |  -.3448759   .6868663    -0.50   0.616    -1.703286    1.013534
                regcat2 |
                     1  |   .4420181   .3618344     1.22   0.224     -.273579    1.157615
                     2  |   .5771603   .2449452     2.36   0.020      .092734    1.061587
                     4  |   .4965339   .3367915     1.47   0.143    -.1695362    1.162604
               1.binge2 |  -.4568187   .2639217    -1.73   0.086    -.9787745    .0651371
                povcat |
                     1  |  -.9555049   .4380112    -2.18   0.031    -1.821756   -.0892536
                     2  |  -.2478917   .5249494    -0.47   0.638     -1.28608    .7902967
                     3  |  -.9535605   .4322168    -2.21   0.029    -1.808352   -.0987687
                     4  |  -.8089066   .2314143    -3.50   0.001    -1.266573   -.3512404
                inscov |
                     1  |   1.630839   .3263201     5.00   0.000     .9854779    2.276199
                     2  |   1.500713   .4523487     3.32   0.001     .6061066     2.39532
                 pcs42  |   -.085994   .0150781    -5.70   0.000    -.1158139   -.0561741
                 mcs42  |  -.0098869   .0185662    -0.53   0.595    -.0466051    .0268313
               k6sum42  |   .0647854   .0512515     1.26   0.208    -.0365743    .1661452
                phq242  |
                     1  |   .1059072   .3142879     0.34   0.737    -.5156576     .727472
                     2  |   .2047224    .450697     0.45   0.650    -.6866175    1.096062
                     3  |   1.212215   .8330678     1.46   0.148    -.4353362    2.859767
                     4  |   .0925209    .712875     0.13   0.897    -1.317326    1.502368
                     5  |          0  (empty)
                     6  |  -.0389743   1.327388    -0.03   0.977     -2.66414    2.586191
                        |
                  _cons |   5.105925   1.752638     2.91   0.004     1.639747    8.572104
---------------------+------------------------------------------------------------------
glm                  |
               agelast |   .0125686   .0058949     2.13   0.035     .0009102    .0242269
                 tobs2 |
                     1  |   .2671609   .1685367     1.59   0.115    -.0661528    .6004747
                     2  |   .0285081   .1787777     0.16   0.874    -.3250591    .3820754
                2.sex   |  -.1815554   .2220142    -0.82   0.415    -.6206312    .2575204
               educat3 |
COLLEGE GRAD OR HIGH   |   .2363892   .1719946     1.37   0.172    -.1037632    .5765417
                   HS  |   .0569295    .189573     0.30   0.764    -.3179876    .4318466
        SOME COLLEGE   |   .0817419   .2368567     0.35   0.731    -.3866878    .5501715
               marcat2 |
                     1  |   .1089897   .1717587     0.63   0.527    -.2306962    .4486756
                     2  |   .0700352   .1980271     0.35   0.724    -.3216014    .4616719
                     4  |   .0278552   .3016614     0.09   0.927    -.5687381    .6244486
               bmicat2 |
                     2  |  -.0686064   .1534513    -0.45   0.656    -.3720859    .2348732
                     3  |     .06135   .1668256     0.37   0.714    -.2685798    .3912797
               racethx |
                     1  |  -.0907519   .2457461    -0.37   0.712    -.5767621    .3952583
                     3  |   .1496835   .1946592     0.77   0.443    -.2352925    .5346595
                     4  |   .0384385    .160984     0.24   0.812    -.2799384    .3568154
                     5  |  -.4601158    .228316    -2.02   0.046    -.9116546    -.008577
               regcat2 |
                     1  |   .1782905   .2442561     0.73   0.467    -.3047729    .6613539
                     2  |  -.0170598   .1654098    -0.10   0.918    -.3441895      .31007
                     4  |    .137903   .2588982     0.53   0.595     -.374118     .649924
               1.binge2 |  -.4771534   .1379921    -3.46   0.001    -.7500593   -.2042475
                povcat |
                     1  |  -.0072448   .2064153    -0.04   0.972    -.4154708    .4009811
                     2  |  -.0920637   .3563864    -0.26   0.797    -.7968863    .6127589
```

```
                   3  |  -.3963214   .1910549    -2.07   0.040    -.7741691   -.0184736
                   4  |   .1602088   .1758436     0.91   0.364    -.1875557    .5079734
                inscov |
                   1  |   1.134259   .2838527     4.00   0.000     .5728853    1.695632
                   2  |   1.304568   .3170781     4.11   0.000      .677485    1.931651
                 pcs42 |  -.0446406   .0067848    -6.58   0.000    -.0580588   -.0312225
                 mcs42 |  -.0310079    .011916    -2.60   0.010     -.054574   -.0074417
               k6sum42 |  -.0217276   .0221661    -0.98   0.329    -.0655654    .0221102
                phq242 |
                   1  |    -.09249    .199679    -0.46   0.644    -.4873936    .3024136
                   2  |  -.1284248   .2948409    -0.44   0.664    -.7115293    .4546797
                   3  |  -.2207303   .3444264    -0.64   0.523    -.9018998    .4604391
                   4  |  -.2283418   .3151703    -0.72   0.470    -.8516517     .394968
                   5  |  -1.232874   .3842957    -3.21   0.002    -1.992893   -.4728554
                   6  |  -.2327141   .4774268    -0.49   0.627    -1.176917    .7114891
                _cons |   10.41554   .9596208    10.85   0.000     8.517709    12.31338
------------------------------------------------------------------------------
Note: Strata with single sampling unit centered at overall mean.
*Overall conditional mean
 margins
Predictive margins

Number of strata   =       162                Number of obs     =       1,298
Number of PSUs     =       297                Population size   =  15,368,643
Model VCE    : Linearized                     Design df         =         135

Expression   : twopm combined expected values, predict()

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   4868.459   436.0241    11.17   0.000     4006.137    5730.781
------------------------------------------------------------------------------

 * Conditional mean by Tobacco Status
margins tobs2 /* change in Tob_stat2*/
Predictive margins
Number of strata   =       162                Number of obs     =       1,298
Number of PSUs     =       297                Population size   =  15,368,643
Model VCE    : Linearized                     Design df         =         135

Expression   : twopm combined expected values, predict()

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       tobs2 |
          1  |   5627.641    788.362     7.14   0.000     4068.503    7186.778
          2  |   4478.333   729.7839     6.14   0.000     3035.045    5921.621
          3  |   4426.889   461.4979     9.59   0.000     3514.189     5339.59
------------------------------------------------------------------------------

* Marginal effects, averaged over the sample
margins, dydx(*)
Average marginal effects
Number of strata   =       162                Number of obs     =       1,298
Number of PSUs     =       297                Population size   =  15,368,643
Model VCE    : Linearized                     Design df         =         135
```

```
Expression   : twopm combined expected values, predict()
dy/dx w.r.t. : agelast 1.tobs2 2.tobs2 2.sex 1.educat3 2.educat3 4.educat3 1.marcat2 2.marcat2
4.marcat2 2.bmicat2 3.bmicat2 1.racethx 3.racethx 4.racethx 5.racethx 1.regcat2 2.regcat2
4.regcat2 1.binge2 1.povcat 2.povcat3.povcat 4.povcat 1.inscov 2.inscov pcs42 mcs42 k6sum42
1.phq242 2.phq242 3.phq242 4.phq242 5.phq242 6.phq242

------------------------------------------------------------------------------
                     |            Delta-method
                     |      dy/dx   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
             agelast |    68.2292   28.13612     2.42   0.017      12.5846    123.8738
               tobs2 |
                   1 |   1200.751   902.1839     1.33   0.185    -583.4911    2984.993
                   2 |   51.44363   795.7262     0.06   0.949    -1522.258    1625.145
              2.sex  |  -612.1792    1027.63    -0.60   0.552    -2644.515    1420.156
             educat3 |
COLLEGE GRAD OR HIGH |   1107.783   852.9509     1.30   0.196    -579.0916    2794.657
                  HS |   226.6126   860.8732     0.26   0.793     -1475.93    1929.155
        SOME COLLEGE |   347.7521   1104.229     0.31   0.753    -1836.073    2531.577
             marcat2 |
                   1 |   365.2657   832.3003     0.44   0.661    -1280.768      2011.3
                   2 |   280.9105   942.2005     0.30   0.766    -1582.472    2144.293
                   4 |   152.2877   1442.717     0.11   0.916    -2700.963    3005.539
             bmicat2 |
                   2 |  -299.4418     748.79    -0.40   0.690    -1780.318    1181.434
                   3 |   284.8546    840.115     0.34   0.735    -1376.634    1946.344
             racethx |
                   1 |  -595.5511   1065.522    -0.56   0.577    -2702.827    1511.724
                   3 |   801.0806   1091.076     0.73   0.464    -1356.733    2958.894
                   4 |   72.64053   803.8245     0.09   0.928    -1517.077    1662.358
                   5 |  -1891.176   757.6234    -2.50   0.014    -3389.522   -392.8305
             regcat2 |
                   1 |   1103.589    1336.25     0.83   0.410    -1539.102    3746.279
                   2 |   127.8672   782.4741     0.16   0.870    -1419.626     1675.36
                   4 |   896.8828    1427.97     0.63   0.531    -1927.203    3720.969
            1.binge2 |  -2264.394   654.7739    -3.46   0.001    -3559.336   -969.4531
              povcat |
                   1 |   -399.185   1033.218    -0.39   0.700    -2442.572    1644.202
                   2 |  -513.5474   1659.374    -0.31   0.757    -3795.279    2768.184
                   3 |  -1900.979   777.7645    -2.44   0.016    -3439.158   -362.7999
                   4 |   532.6112   951.6994     0.56   0.577    -1349.557     2414.78
              inscov |
                   1 |   3608.875   562.8657     6.41   0.000       2495.7    4722.051
                   2 |   4451.189   1015.397     4.38   0.000     2443.046    6459.332
               pcs42 |  -249.9732   48.10175    -5.20   0.000    -345.1036   -154.8427
               mcs42 |  -154.7134   63.87207    -2.42   0.017    -281.0327   -28.39411
             k6sum42 |  -81.18823   109.0209    -0.74   0.458     -296.798    134.4216
              phq242 |
                   1 |  -422.3858   1012.692    -0.42   0.677     -2425.18    1580.408
                   2 |  -559.4209   1420.242    -0.39   0.694    -3368.222     2249.38
                   3 |  -762.9956   1563.337    -0.49   0.626    -3854.795    2328.804
                   4 |  -1034.501   1429.481    -0.72   0.471    -3861.574    1792.572
                   5 |           .  (not estimable)
                   6 |  -1094.172   2028.011    -0.54   0.590    -5104.952    2916.609
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.

* Two-part model, with logit first part and GLM second part all variables
```

```
svy: twopm totexpi c.agelast ib3.tobs2 i.sex ib3.educat3 ib3.marcat2 i.bmicat2 ib2.racethx
ib3.regcat2 i.binge2 ib5.povcat ib3.inscov c.pcs42 c.mcs42 c.k6sum42 i.phq242 ,
firstpart(logit) secondpart(glm, family(gamma) link(log))
(running twopm on estimation sample)

Survey data analysis
Number of strata   =       162                Number of obs     =       1,298
Number of PSUs     =       297                Population size   =  15,368,643
                                              Design df         =         135
                                              F(  34,    102)   =        6.66
                                              Prob > F          =      0.0000
------------------------------------------------------------------------------
                     |             Linearized
             totexpi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
logit                |
            agelast  |   .0185455   .0101043     1.84   0.069    -.0014376    .0385287
              tobs2  |
                 1   |   -.351833   .2483546    -1.42   0.159    -.8430021     .139336
                 2   |   -.229222   .2953677    -0.78   0.439    -.8133684    .3549245
              2.sex  |   .8284252   .4388348     1.89   0.061    -.0394549    1.696305
            educat3  |
COLLEGE GRAD OR HIGH |  -.1590572   .2743212    -0.58   0.563      -.70158    .3834657
                 HS  |  -.0906717   .2749685    -0.33   0.742    -.6344749    .4531314
        SOME COLLEGE |  -.0778594    .335173    -0.23   0.817    -.7407285    .5850098
            marcat2  |
                 1   |   -.423125   .2783135    -1.52   0.131    -.9735435    .1272936
                 2   |  -.1540523   .3726572    -0.41   0.680    -.8910535     .582949
                 4   |   .0719318   .4346279     0.17   0.869    -.7876286    .9314921
            bmicat2  |
                 2   |   .0576359   .2466911     0.23   0.816    -.4302431     .545515
                 3   |  -.0489334   .2540454    -0.19   0.848     -.551357    .4534902
            racethx  |
                 1   |  -.4531627   .3224679    -1.41   0.162    -1.090905    .1845795
                 3   |   .0270999   .3087996     0.09   0.930    -.5836107    .6378104
                 4   |  -.2905011   .2493947    -1.16   0.246    -.7837271    .2027249
                 5   |  -.3448759   .6868663    -0.50   0.616    -1.703286    1.013534
            regcat2  |
                 1   |   .4420181   .3618344     1.22   0.224     -.273579    1.157615
                 2   |   .5771603   .2449452     2.36   0.020      .092734    1.061587
                 4   |   .4965339   .3367915     1.47   0.143    -.1695362    1.162604
           1.binge2  |  -.4568187   .2639217    -1.73   0.086    -.9787745    .0651371
             povcat  |
                 1   |  -.9555049   .4380112    -2.18   0.031    -1.821756   -.0892536
                 2   |  -.2478917   .5249494    -0.47   0.638     -1.28608    .7902967
                 3   |  -.9535605   .4322168    -2.21   0.029    -1.808352   -.0987687
                 4   |  -.8089066   .2314143    -3.50   0.001    -1.266573   -.3512404
            inscov   |
                 1   |   1.630839   .3263201     5.00   0.000     .9854779    2.276199
                 2   |   1.500713   .4523487     3.32   0.001     .6061066     2.39532
              pcs42  |   -.085994   .0150781    -5.70   0.000    -.1158139   -.0561741
              mcs42  |  -.0098869   .0185662    -0.53   0.595    -.0466051    .0268313
            k6sum42  |   .0647854   .0512515     1.26   0.208    -.0365743    .1661452
             phq242  |
                 1   |   .1059072   .3142879     0.34   0.737    -.5156576     .727472
                 2   |   .2047224    .450697     0.45   0.650    -.6866175    1.096062
                 3   |   1.212215   .8330678     1.46   0.148    -.4353362    2.859767
                 4   |   .0925209    .712875     0.13   0.897    -1.317326    1.502368
                 5   |          0  (empty)
                 6   |  -.0389743   1.327388    -0.03   0.977     -2.66414    2.586191
              _cons  |   5.105925   1.752638     2.91   0.004     1.639747    8.572104
```

```
---------------------+------------------------------------------------------------------
glm                  |
            agelast  |   .0125686   .0058949     2.13   0.035     .0009102    .0242269
              tobs2  |
                1    |   .2671609   .1685367     1.59   0.115    -.0661528    .6004747
                2    |   .0285081   .1787777     0.16   0.874    -.3250591    .3820754
             2.sex   |  -.1815554   .2220142    -0.82   0.415    -.6206312    .2575204
            educat3  |
COLLEGE GRAD OR HIGH |   .2363892   .1719946     1.37   0.172    -.1037632    .5765417
                HS   |   .0569295    .189573     0.30   0.764    -.3179876    .4318466
        SOME COLLEGE |   .0817419   .2368567     0.35   0.731    -.3866878    .5501715
            marcat2  |
                1    |   .1089897   .1717587     0.63   0.527    -.2306962    .4486756
                2    |   .0700352   .1980271     0.35   0.724    -.3216014    .4616719
                4    |   .0278552   .3016614     0.09   0.927    -.5687381    .6244486
            bmicat2  |
                2    |  -.0686064   .1534513    -0.45   0.656    -.3720859    .2348732
                3    |     .06135   .1668256     0.37   0.714    -.2685798    .3912797
            racethx  |
                1    |  -.0907519   .2457461    -0.37   0.712    -.5767621    .3952583
                3    |   .1496835   .1946592     0.77   0.443    -.2352925    .5346595
                4    |   .0384385    .160984     0.24   0.812    -.2799384    .3568154
                5    |  -.4601158    .228316    -2.02   0.046    -.9116546    -.008577
            regcat2  |
                1    |   .1782905   .2442561     0.73   0.467    -.3047729    .6613539
                2    |  -.0170598   .1654098    -0.10   0.918    -.3441895      .31007
                4    |    .137903   .2588982     0.53   0.595     -.374118     .649924
           1.binge2  |  -.4771534   .1379921    -3.46   0.001    -.7500593   -.2042475
             povcat  |
                1    |  -.0072448   .2064153    -0.04   0.972    -.4154708    .4009811
                2    |  -.0920637   .3563864    -0.26   0.797    -.7968863    .6127589
                3    |  -.3963214   .1910549    -2.07   0.040    -.7741691   -.0184736
                4    |   .1602088   .1758436     0.91   0.364    -.1875557    .5079734
             inscov  |
                1    |   1.134259   .2838527     4.00   0.000     .5728853    1.695632
                2    |   1.304568   .3170781     4.11   0.000      .677485    1.931651
              pcs42  |  -.0446406   .0067848    -6.58   0.000    -.0580588   -.0312225
              mcs42  |  -.0310079    .011916    -2.60   0.010     -.054574   -.0074417
            k6sum42  |  -.0217276   .0221661    -0.98   0.329    -.0655654    .0221102
             phq242  |
                1    |    -.09249    .199679    -0.46   0.644    -.4873936    .3024136
                2    |  -.1284248   .2948409    -0.44   0.664    -.7115293    .4546797
                3    |  -.2207303   .3444264    -0.64   0.523    -.9018998    .4604391
                4    |  -.2283418   .3151703    -0.72   0.470    -.8516517     .394968
                5    |  -1.232874   .3842957    -3.21   0.002    -1.992893   -.4728554
                6    |  -.2327141   .4774268    -0.49   0.627    -1.176917    .7114891
              _cons  |   10.41554   .9596208    10.85   0.000     8.517709    12.31338
------------------------------------------------------------------------------
Note: Strata with single sampling unit centered at overall mean.

 * Margin plots for Age by Tobacco Status
 margins, at (agelast=(30(10)80) tobs2 = (1,2,3))

Predictive margins
Number of strata   =       162                Number of obs     =       1,298
Number of PSUs     =       297                Population size   =  15,368,643
Model VCE    : Linearized                     Design df         =         135

Expression   : twopm combined expected values, predict()
1._at        : agelast         =          30
               tobs2           =           1
```

```
2._at        : agelast         =          30
               tobs2           =           2
3._at        : agelast         =          30
               tobs2           =           3
4._at        : agelast         =          40
               tobs2           =           1
5._at        : agelast         =          40
               tobs2           =           2
6._at        : agelast         =          40
               tobs2           =           3
7._at        : agelast         =          50
               tobs2           =           1
8._at        : agelast         =          50
               tobs2           =           2
9._at        : agelast         =          50
               tobs2           =           3
10._at       : agelast         =          60
               tobs2           =           1
11._at       : agelast         =          60
               tobs2           =           2
12._at       : agelast         =          60
               tobs2           =           3
13._at       : agelast         =          70
               tobs2           =           1
14._at       : agelast         =          70
               tobs2           =           2
15._at       : agelast         =          70
               tobs2           =           3
16._at       : agelast         =          80
               tobs2           =           1
17._at       : agelast         =          80
               tobs2           =           2
18._at       : agelast         =          80
               tobs2           =           3
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   4034.126   896.6801     4.50   0.000     2260.768    5807.483
          2  |   3222.614    789.643     4.08   0.000     1660.943    4784.285
          3  |   3206.349   513.6461     6.24   0.000     2190.515    4222.183
          4  |   4670.742   828.1915     5.64   0.000     3032.834     6308.65
          5  |   3725.209   750.9914     4.96   0.000     2239.979    5210.439
          6  |   3696.127   441.3434     8.37   0.000     2823.286    4568.969
          7  |   5394.925   784.6905     6.88   0.000     3843.049    6946.802
          8  |   4296.429   726.4633     5.91   0.000     2859.708     5733.15
          9  |   4252.043   426.9187     9.96   0.000     3407.729    5096.357
         10  |   6217.656   841.6065     7.39   0.000     4553.218    7882.095
         11  |   4944.927   761.5421     6.49   0.000     3438.831    6451.023
         12  |   4882.529   556.5951     8.77   0.000     3781.755    5983.303
         13  |   7151.399   1079.167     6.63   0.000     5017.138     9285.66
         14  |   5680.536   914.3955     6.21   0.000     3872.143    7488.929
         15  |   5597.177    847.267     6.61   0.000     3921.544    7272.811
         16  |    8210.32   1522.536     5.39   0.000     5199.213    11221.43
         17  |   6514.439   1220.495     5.34   0.000     4100.676    8928.203
         18  |   6406.903   1277.926     5.01   0.000      3879.56    8934.247
------------------------------------------------------------------------------
```

Age (≤ 65 years) Expenditure Model

```
svy:twopm totexpi c.agelast ib3.tobs2 i.sex ib3.educat3 ib3.marcat2 i.bmicat2 ib2.racethx
ib3.regcat2 i.binge2 ib5.povcat ib3.inscov c.pcs42 c.mcs42 c.k6sum42 i.phq242 ,
firstpart(logit) secondpart(glm, family(gamma) link(log))
(running twopm on estimation sample)

Survey data analysis
Number of strata   =       160                Number of obs     =       1,088
Number of PSUs     =       294                Population size   =  13,025,624
                                              Design df         =         134
                                              F(  34,    101)   =        2.54
                                              Prob > F          =      0.0002

------------------------------------------------------------------------------
                             |             Linearized
                     totexpi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------------------+------------------------------------------------------------------
logit                        |
                    agelast |     .00393   .0112094     0.35   0.726    -.0182403    .0261003
                      tobs2 |
                          1 |  -.2188141   .2379397    -0.92   0.359    -.6894175    .2517892
                          2 |  -.2437136   .2640158    -0.92   0.358     -.765891    .2784637
                      2.sex |   1.108227   .5715452     1.94   0.055    -.0221896    2.238644
                    educat3 |
COLLEGE GRAD OR HIGH        |  -.1043773   .3039787    -0.34   0.732    -.7055941    .4968396
                         HS |  -.3348051   .2666419    -1.26   0.211    -.8621762    .1925661
               SOME COLLEGE |  -.1129375   .3056098    -0.37   0.712    -.7173804    .4915055
                    marcat2 |
                          1 |  -.0064504   .2725879    -0.02   0.981    -.5455818    .5326809
                          2 |   .2499429   .3056257     0.82   0.415    -.3545314    .8544173
                          4 |  -.3818723   .4019774    -0.95   0.344    -1.176914    .4131689
                    bmicat2 |
                          2 |   .3070919    .211206     1.45   0.148    -.1106368    .7248206
                          3 |   .2515883   .2377777     1.06   0.292    -.2186945    .7218712
                    racethx |
                          1 |  -.4755246   .3508007    -1.36   0.178    -1.169347     .218298
                          3 |  -.4836295   .2819257    -1.72   0.089    -1.041229    .0739704
                          4 |   -.418038   .2574672    -1.62   0.107    -.9272632    .0911873
                          5 |  -.3818102   .7598424    -0.50   0.616    -1.884646    1.121026
                    regcat2 |
                          1 |   .5886035   .4053385     1.45   0.149    -.2130855    1.390293
                          2 |   .1429476   .2810094     0.51   0.612      -.41284    .6987351
                          4 |  -.1080479   .3419419    -0.32   0.753    -.7843494    .5682537
                   1.binge2 |  -.1723804   .2335171    -0.74   0.462    -.6342365    .2894757
                     povcat |
                          1 |  -.8897641    .413883    -2.15   0.033    -1.708353   -.0711756
                          2 |  -.4704084   .5089496    -0.92   0.357    -1.477022    .5362052
                          3 |  -1.302972   .3734933    -3.49   0.001    -2.041676    -.564267
                          4 |  -.9470192   .2285889    -4.14   0.000    -1.399128   -.4949102
                    inscov |
                          1 |   1.575004   .3183678     4.95   0.000     .9453281     2.20468
                          2 |   1.317395   .4266214     3.09   0.002     .4736121    2.161178
                     pcs42 |  -.0652491   .0161115    -4.05   0.000    -.0971148   -.0333834
                     mcs42 |   -.023464   .0173705    -1.35   0.179    -.0578199    .0108919
                   k6sum42 |   .0044122    .041712     0.11   0.916    -.0780868    .0869112
                    phq242 |
                          1 |  -.2151516   .3291931    -0.65   0.515    -.8662382     .435935
                          2 |   .2597069   .4146509     0.63   0.532    -.5604004    1.079814
                          3 |          0  (empty)
                          4 |   .8194732   .7345623     1.12   0.267    -.6333631    2.272309
                          5 |    .704313   1.061803     0.66   0.508    -1.395749    2.804375
```

```
                   6  |   1.526835   1.179469     1.29   0.198    -.8059493     3.85962
                _cons |   5.363542   1.609042     3.33   0.001     2.181137    8.545947
---------------------+------------------------------------------------------------------
glm                  |
              agelast |    .006204   .0059218     1.05   0.297    -.0055083    .0179163
                tobs2 |
                   1  |   .3239129   .1795632     1.80   0.073    -.0312318    .6790577
                   2  |   .2700106   .1630855     1.66   0.100    -.0525441    .5925654
               2.sex  |   .3181315   .2716639     1.17   0.244    -.2191723    .8554354
              educat3 |
COLLEGE GRAD OR HIGH  |  -.2095454   .1664307    -1.26   0.210    -.5387163    .1196255
                  HS  |   -.293531   .2397692    -1.22   0.223    -.7677528    .1806908
        SOME COLLEGE  |   .1479551   .2379525     0.62   0.535    -.3226735    .6185837
              marcat2 |
                   1  |    .323562   .1623109     1.99   0.048     .0025393    .6445846
                   2  |   .3208715   .2275727     1.41   0.161    -.1292277    .7709706
                   4  |   .2056018    .312568     0.66   0.512    -.4126033    .8238069
              bmicat2 |
                   2  |   .0781474   .1644558     0.48   0.635    -.2471175    .4034122
                   3  |   .1435625   .1523446     0.94   0.348    -.1577485    .4448735
              racethx |
                   1  |   .3866123   .3028786     1.28   0.204    -.2124287    .9856534
                   3  |  -.2794542   .2139657    -1.31   0.194    -.7026411    .1437327
                   4  |   .0192359   .1817564     0.11   0.916    -.3402466    .3787183
                   5  |  -.7125657   .3024568    -2.36   0.020    -1.310773   -.1143588
              regcat2 |
                   1  |  -.0858684   .2473916    -0.35   0.729    -.5751659    .4034291
                   2  |   .1635595    .167265     0.98   0.330    -.1672616    .4943806
                   4  |   .0820755   .2584243     0.32   0.751    -.4290428    .5931938
             1.binge2 |  -.2930343    .142511    -2.06   0.042    -.5748963   -.0111723
               povcat |
                   1  |  -.1323469   .2585875    -0.51   0.610    -.6437879    .3790941
                   2  |  -.2809101   .3636527    -0.77   0.441    -1.000152    .4383317
                   3  |    -.60652   .2109812    -2.87   0.005    -1.023804    -.189236
                   4  |   -.020115   .1671071    -0.12   0.904    -.3506238    .3103937
                inscov |
                   1  |   1.401576   .2370262     5.91   0.000     .9327799    1.870373
                   2  |   1.944569   .2663946     7.30   0.000     1.417687    2.471451
                pcs42 |  -.0632391   .0078288    -8.08   0.000    -.0787231   -.0477551
                mcs42 |  -.0410952   .0140024    -2.93   0.004    -.0687895   -.0134008
              k6sum42 |  -.0210613   .0292644    -0.72   0.473    -.0789411    .0368185
               phq242 |
                   1  |  -.3784691   .2122905    -1.78   0.077    -.7983428    .0414045
                   2  |  -.4123546   .2714674    -1.52   0.131    -.9492699    .1245607
                   3  |  -.8206237   .2794668    -2.94   0.004     -1.37336    -.267887
                   4  |  -1.255511   .4025839    -3.12   0.002    -2.051752   -.4592699
                   5  |  -1.775673   .3801353    -4.67   0.000    -2.527515   -1.023832
                   6  |  -1.099219   .4895141    -2.25   0.026    -2.067393   -.1310459
                _cons |   11.79766   1.147556    10.28   0.000     9.527998    14.06733
------------------------------------------------------------------------------------------
Note: Strata with single sampling unit centered at overall mean.
*Overall conditional mean
margins
Predictive margins
Number of strata   =       160                Number of obs     =       1,085
Number of PSUs     =       294                Population size   =  13,025,624
Model VCE    : Linearized                     Design df         =         134

Expression   : twopm combined expected values, predict()
------------------------------------------------------------------------------
             |            Delta-method
```

```
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   3902.958   418.9036     9.32   0.000      3074.44    4731.476
------------------------------------------------------------------------------

* Conditional mean by Tobacco Status
margins tobs2 /* change in Tob_stat2*/
Predictive margins
Number of strata   =       160                Number of obs     =       1,085
Number of PSUs     =       294                Population size   =  13,025,624
Model VCE    : Linearized                     Design df         =         134
Expression   : twopm combined expected values, predict()
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       tobs2 |
          1  |   4326.727   688.2648     6.29   0.000     2965.459    5687.995
          2  |   4089.012   672.2159     6.08   0.000     2759.486    5418.538
          3  |   3196.196   363.9581     8.78   0.000     2476.351    3916.042
------------------------------------------------------------------------------


 * Marginal effects, averaged over the sample
 margins, dydx(*)

Average marginal effects
Number of strata   =       160                Number of obs     =       1,085
Number of PSUs     =       294                Population size   =  13,025,624
Model VCE    : Linearized                     Design df         =         134
Expression   : twopm combined expected values, predict()
dy/dx w.r.t. : agelast 1.tobs2 2.tobs2 2.sex 1.educat3 2.educat3 4.educat3 1.marcat2 2.marcat2
4.marcat2 2.bmicat2 3.bmicat2 1.racethx 3.racethx 4.racethx 5.racethx 1.regcat2 2.regcat2
4.regcat2 1.binge2 1.povcat 2.povcat 3.povcat 4.povcat 1.inscov 2.inscov pcs42 mcs42 k6sum42
1.phq242 2.phq242 3.phq242 4.phq242 5.phq242 6.phq242

------------------------------------------------------------------------------------------
                     |            Delta-method
                     |      dy/dx   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------------
             agelast |   25.73528    23.2201     1.11   0.270    -20.19003    71.66059
               tobs2 |
                  1  |   1130.531   716.8714     1.58   0.117    -287.3158    2548.378
                  2  |    892.816   639.5516     1.40   0.165    -372.1056    2157.738
              2.sex  |   1846.117   1454.739     1.27   0.207    -1031.103    4723.338
             educat3 |
COLLEGE GRAD OR HIGH |  -861.4615   691.3877    -1.25   0.215    -2228.906    505.9829
                  HS |  -1222.768    918.612    -1.33   0.185    -3039.623    594.0861
        SOME COLLEGE |   642.6883   1128.927     0.57   0.570    -1590.133     2875.51
             marcat2 |
                  1  |   1183.402   631.4255     1.87   0.063    -65.44806    2432.251
                  2  |   1275.903   895.8394     1.42   0.157    -495.9112    3047.717
                  4  |   543.1766   1156.556     0.47   0.639     -1744.29    2830.643
             bmicat2 |
                  2  |   407.8242   605.6497     0.67   0.502    -790.0453    1605.694
                  3  |    650.924   587.3619     1.11   0.270    -510.7754    1812.623
             racethx |
                  1  |   1599.391   1668.447     0.96   0.339    -1700.507    4899.289
                  3  |  -1120.218   700.7717    -1.60   0.112    -2506.222    265.7866
                  4  |  -91.75282   729.2547    -0.13   0.900    -1534.091    1350.586
                  5  |  -2105.651   691.6886    -3.04   0.003    -3473.691   -737.6118
```

```
             regcat2 |
                  1  |  -130.0111   877.2487    -0.15   0.882    -1865.056    1605.034
                  2  |     719.16   693.6563     1.04   0.302    -652.7714    2091.091
                  4  |   269.7005   1013.346     0.27   0.791    -1734.522    2273.923
            1.binge2 |   -1173.71   575.1396    -2.04   0.043    -2311.235   -36.18368
              povcat |
                  1  |  -886.4802   1063.305    -0.83   0.406    -2989.513    1216.553
                  2  |  -1256.043   1302.424    -0.96   0.337    -3832.011    1319.925
                  3  |  -2427.784   628.1044    -3.87   0.000    -3670.065   -1185.502
                  4  |  -472.2329   730.4016    -0.65   0.519     -1916.84    972.3743
               inscov |
                  1  |   3021.128    421.644     7.17   0.000      2187.19    3855.067
                  2  |   5519.578   1124.153     4.91   0.000     3296.199    7742.958
               pcs42 |  -272.0785   51.65142    -5.27   0.000     -374.236    -169.921
               mcs42 |   -169.476   61.79692    -2.74   0.007    -291.6996   -47.25247
             k6sum42 |  -80.49336   116.9578    -0.69   0.493    -311.8156    150.8288
              phq242 |
                  1  |  -1817.027   1006.609    -1.81   0.073    -3807.924    173.8703
                  2  |  -1790.362   1281.457    -1.40   0.165     -4324.86    744.1352
                  3  |           .  (not estimable)
                  4  |  -3863.893   1307.663    -2.95   0.004    -6450.223   -1277.563
                  5  |  -4543.219   1170.308    -3.88   0.000    -6857.884   -2228.553
                  6  |   -3533.42   1583.011    -2.23   0.027    -6664.339   -402.5006
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.

* Margin plots for Age by Tobacco Status
margins, at (agelast=(30(5)65) tobs2 = (1,2,3))
Predictive margins
Number of strata   =       160                Number of obs     =       1,085
Number of PSUs     =       294                Population size   =  13,025,624
Model VCE    : Linearized                     Design df         =         134

Expression   : twopm combined expected values, predict()
1._at        : agelast         =          30
               tobs2           =           1
2._at        : agelast         =          30
               tobs2           =           2
3._at        : agelast         =          30
               tobs2           =           3
4._at        : agelast         =          35
               tobs2           =           1
5._at        : agelast         =          35
               tobs2           =           2
6._at        : agelast         =          35
               tobs2           =           3
7._at        : agelast         =          40
               tobs2           =           1
8._at        : agelast         =          40
               tobs2           =           2
9._at        : agelast         =          40
               tobs2           =           3
10._at       : agelast         =          45
               tobs2           =           1
11._at       : agelast         =          45
               tobs2           =           2
12._at       : agelast         =          45
               tobs2           =           3
13._at       : agelast         =          50
               tobs2           =           1
14._at       : agelast         =          50
```

```
              tobs2           =           2
15._at        : agelast       =          50
              tobs2           =           3
16._at        : agelast       =          55
              tobs2           =           1
17._at        : agelast       =          55
              tobs2           =           2
18._at        : agelast       =          55
              tobs2           =           3
19._at        : agelast       =          60
              tobs2           =           1
20._at        : agelast       =          60
              tobs2           =           2
21._at        : agelast       =          60
              tobs2           =           3
22._at        : agelast       =          65
              tobs2           =           1
23._at        : agelast       =          65
              tobs2           =           2
24._at        : agelast       =          65
              tobs2           =           3
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   3855.589   756.6631     5.10   0.000     2359.041    5352.136
          2  |   3643.078   777.9289     4.68   0.000     2104.471    5181.686
          3  |   2852.506    436.862     6.53   0.000     1988.469    3716.543
          4  |   3985.705   717.8111     5.55   0.000         2566    5405.411
          5  |   3766.162   733.2432     5.14   0.000     2315.935     5216.39
          6  |   2947.867   396.2475     7.44   0.000     2164.159    3731.575
          7  |   4120.094   691.6778     5.96   0.000     2752.076    5488.112
          8  |   3893.291   696.9289     5.59   0.000     2514.887    5271.695
          9  |   3046.337   366.7051     8.31   0.000     2321.058    3771.616
         10  |   4258.892   683.4774     6.23   0.000     2907.093    5610.691
         11  |   4024.595   673.5335     5.98   0.000     2692.463    5356.727
         12  |   3148.016    354.994     8.87   0.000       2445.9    3850.132
         13  |   4402.241   698.0393     6.31   0.000     3021.641    5782.841
         14  |   4160.207   668.1067     6.23   0.000     2838.809    5481.606
         15  |   3253.006   367.2089     8.86   0.000     2526.731    3979.281
         16  |   4550.287   738.5258     6.16   0.000     3089.612    6010.963
         17  |   4300.267    685.183     6.28   0.000     2945.095     5655.44
         18  |   3361.413   405.6644     8.29   0.000     2559.079    4163.747
         19  |   4703.181   805.7177     5.84   0.000     3109.612    6296.751
         20  |   4444.917   727.5737     6.11   0.000     3005.903    5883.931
         21  |   3473.346   468.2953     7.42   0.000      2547.14    4399.553
         22  |   4861.079   898.3889     5.41   0.000     3084.222    6637.936
         23  |   4594.305   795.7926     5.77   0.000     3020.366    6168.244
         24  |   3588.919   551.1012     6.51   0.000     2498.937    4678.901
------------------------------------------------------------------------------
```

Age (≤ 65 years) and Male Expenditure Model

```
* Two-part model, with logit first part and GLM second part all variables (i.e
sociodemographic and comorbidities)
svy: twopm totexpi c.agelast ib3.tobs2 i.sex ib3.educat3 ib3.marcat2 i.bmicat2 ib2.racethx
ib3.regcat2 i.binge2 ib5.povcat ib3.inscov c.pcs42 c.mcs42 c.k6sum42 i.phq242
,firstpart(logit) secondpart(glm, family(gamma) link(log))
(running twopm on estimation sample)
```

```
Survey data analysis
Number of strata   =       160                Number of obs     =       1,023
Number of PSUs     =       288                Population size   =  12,527,088
                                              Design df         =         128
                                              F(  33,     96)   =        2.64
                                              Prob > F          =      0.0001
------------------------------------------------------------------------------
                     |             Linearized
             totexpi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
logit                |
            agelast  |   .0006949   .0116623     0.06   0.953    -.0223809    .0237708
              tobs2  |
                 1   |  -.2209336   .2435647    -0.91   0.366     -.702868    .2610007
                 2   |  -.2007821   .2677302    -0.75   0.455     -.730532    .3289678
            educat3  |
COLLEGE GRAD OR HIGH |  -.0748593   .3100052    -0.24   0.810    -.6882576     .538539
                  HS |  -.3019874   .2739054    -1.10   0.272    -.8439561    .2399812
        SOME COLLEGE |  -.0521699   .3134091    -0.17   0.868    -.6723034    .5679636
            marcat2  |
                 1   |   .0548619   .2842743     0.19   0.847    -.5076234    .6173472
                 2   |    .220344   .3143677     0.70   0.485    -.4016861    .8423742
                 4   |  -.4382878   .3981575    -1.10   0.273    -1.226111    .3495349
            bmicat2  |
                 2   |   .3356067   .2205691     1.52   0.131    -.1008269    .7720403
                 3   |   .2610218   .2464633     1.06   0.292     -.226648    .7486915
            racethx  |
                 1   |   -.472339   .3650725    -1.29   0.198    -1.194697    .2500193
                 3   |  -.5582216   .2880107    -1.94   0.055      -1.1281    .0116568
                 4   |  -.4053915   .2597167    -1.56   0.121    -.9192854    .1085024
                 5   |   .0554819   .7443055     0.07   0.941    -1.417254    1.528218
            regcat2  |
                 1   |   .6704057   .4051286     1.65   0.100    -.1312104    1.472022
                 2   |   .1994403   .2843247     0.70   0.484    -.3631447    .7620253
                 4   |   .0256272   .3390304     0.08   0.940    -.6452025    .6964568
           1.binge2  |  -.2183974   .2382927    -0.92   0.361    -.6899002    .2531055
             povcat  |
                 1   |  -.7697931   .4323271    -1.78   0.077    -1.625226    .0856399
                 2   |  -.5472831    .513587    -1.07   0.289    -1.563503    .4689366
                 3   |   -1.40861   .3968032    -3.55   0.001    -2.193752   -.6234666
                 4   |  -.9827213    .229652    -4.28   0.000    -1.437127   -.5283157
              inscov |
                 1   |   1.581559   .3341463     4.73   0.000     .9203932    2.242724
                 2   |   1.561297    .474469     3.29   0.001      .622479    2.500115
               pcs42 |  -.0713578   .0175271    -4.07   0.000    -.1060381   -.0366775
               mcs42 |  -.0225321   .0185876    -1.21   0.228    -.0593109    .0142468
             k6sum42 |   .0225179     .04473     0.50   0.616    -.0659881     .111024
              phq242 |
                 1   |  -.2340909   .3410201    -0.69   0.494    -.9088575    .4406756
                 2   |   .2063572   .4382907     0.47   0.639    -.6608758     1.07359
                 4   |   1.578631   .8059011     1.96   0.052     -.015982    3.173244
                 5   |   .4480438   1.081358     0.41   0.679    -1.691607    2.587695
                 6   |   1.116415   1.198713     0.93   0.353    -1.255443    3.488273
               _cons |   5.630388    1.71755     3.28   0.001     2.231922    9.028854
---------------------+--------------------------------------------------------------
glm                  |
            agelast  |   .0053067   .0061233     0.87   0.388    -.0068093    .0174226
              tobs2  |
                 1   |   .3100775   .1866656     1.66   0.099    -.0592723    .6794272
                 2   |   .2373006   .1685764     1.41   0.162    -.0962565    .5708578
            educat3  |
```

```
COLLEGE GRAD OR HIGH  |  -.1806012   .1767689    -1.02   0.309    -.5303687    .1691663
                  HS  |  -.3110565   .2490626    -1.25   0.214    -.8038694    .1817565
        SOME COLLEGE  |   .1994781   .2565327     0.78   0.438    -.3081157    .7070718
              marcat2 |
                   1  |   .3146846   .1698144     1.85   0.066    -.0213222    .6506915
                   2  |    .323048    .243754     1.33   0.187    -.1592609     .805357
                   4  |   .2171749   .3195987     0.68   0.498    -.4152059    .8495556
              bmicat2 |
                   2  |    .132129   .1709914     0.77   0.441    -.2062068    .4704648
                   3  |   .1805135   .1663864     1.08   0.280    -.1487103    .5097374
              racethx |
                   1  |   .4533077   .3067169     1.48   0.142    -.1535841    1.060199
                   3  |  -.3759723   .2182129    -1.72   0.087    -.8077439    .0557993
                   4  |   .0437133   .1900559     0.23   0.818    -.3323448    .4197714
                   5  |   -.697845   .3180671    -2.19   0.030    -1.327195    -.068495
              regcat2 |
                   1  |  -.0638655   .2506461    -0.25   0.799    -.5598116    .4320805
                   2  |   .1640336   .1681374     0.98   0.331     -.168655    .4967222
                   4  |   .0920951   .2605941     0.35   0.724    -.4235348     .607725
             1.binge2 |  -.2987401   .1447225    -2.06   0.041    -.5850983    -.012382
               povcat |
                   1  |  -.1071281   .2678212    -0.40   0.690    -.6370582     .422802
                   2  |  -.2480645   .3842643    -0.65   0.520    -1.008397     .512268
                   3  |  -.6809079   .2163951    -3.15   0.002    -1.109083   -.2527332
                   4  |  -.0329724   .1693833    -0.19   0.846    -.3681262    .3021813
               inscov |
                   1  |   1.269363   .2645949     4.80   0.000     .7458167    1.792909
                   2  |   1.857952   .2867521     6.48   0.000     1.290564     2.42534
                pcs42 |  -.0626617   .0077215    -8.12   0.000      -.07794   -.0473834
                mcs42 |  -.0415874   .0143726    -2.89   0.004    -.0700261   -.0131487
              k6sum42 |  -.0223108   .0295279    -0.76   0.451    -.0807367    .0361152
               phq242 |
                   1  |  -.3987987   .2239127    -1.78   0.077    -.8418482    .0442507
                   2  |  -.3663879   .2732496    -1.34   0.182    -.9070589    .1742831
                   3  |  -.8870842   .2948263    -3.01   0.003    -1.470448   -.3037201
                   4  |  -1.258579   .4169911    -3.02   0.003    -2.083668   -.4334913
                   5  |  -1.659381   .3904153    -4.25   0.000    -2.431885   -.8868779
                   6  |  -1.199127   .4915472    -2.44   0.016    -2.171737   -.2265172
                _cons |   11.92157   1.155468    10.32   0.000     9.635279    14.20786
------------------------------------------------------------------------------
Note: Strata with single sampling unit centered at overall mean.
*Overall conditional mean
margins
Predictive margins
Number of strata   =       160                Number of obs     =       1,020
Number of PSUs     =       288                Population size   =  12,527,088
Model VCE    : Linearized                     Design df         =         128
Expression   : twopm combined expected values, predict()
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   3667.983   416.8423     8.80   0.000     2843.189    4492.777
------------------------------------------------------------------------------
* Conditional mean by Tobacco Status
margins tobs2 /* change in Tob_stat2*/
Predictive margins
Number of strata   =       160                Number of obs     =       1,020
Number of PSUs     =       288                Population size   =  12,527,088
Model VCE    : Linearized                     Design df         =         128
Expression   : twopm combined expected values, predict()
```

```
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       tobs2 |
          1  |   4073.945   684.6569     5.95   0.000     2719.235    5428.656
          2  |   3796.062   656.8069     5.78   0.000     2496.457    5095.666
          3  |    3053.73    363.281     8.41   0.000     2334.917    3772.544
------------------------------------------------------------------------------
 Marginal effects, averaged over the sample
 margins, dydx(*)
Average marginal effects
Number of strata   =       160                Number of obs     =       1,020
Number of PSUs     =       288                Population size   =  12,527,088
Model VCE    : Linearized                     Design df         =         128
Expression   : twopm combined expected values, predict()
dy/dx w.r.t. : agelast 1.tobs2 2.tobs2 1.sex 1.educat3 2.educat3 4.educat3 1.marcat2 2.marcat2
4.marcat2 2.bmicat2 3.bmicat2 1.racethx 3.racethx 4.racethx 5.racethx 1.regcat2 2.regcat2
4.regcat2 1.binge2 1.povcat 2.povcat 3.povcat 4.povcat 1.inscov 2.inscov pcs42 mcs42 k6sum42
1.phq242 2.phq242 3.phq242 4.phq242 5.phq242 6.phq242
------------------------------------------------------------------------------------------
                     |            Delta-method
                     |      dy/dx   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------------
             agelast |   19.72142     22.643     0.87   0.385    -25.08163    64.52447
               tobs2 |
                  1  |   1055.764   721.9839     1.46   0.146    -372.8045    2484.333
                  2  |   796.2622   654.6393     1.22   0.226    -499.0535    2091.578
             educat3 |
COLLEGE GRAD OR HIGH |   -690.089   664.8362    -1.04   0.301    -2005.581     625.403
                  HS |  -1252.479   952.3155    -1.32   0.191    -3136.798    631.8397
        SOME COLLEGE |   712.4149   939.2606     0.76   0.450    -1146.073    2570.903
             marcat2 |
                  1  |   1174.519   677.0094     1.73   0.085    -165.0594    2514.098
                  2  |   1266.312   908.5017     1.39   0.166    -531.3142    3063.937
                  4  |   634.7261    1198.99     0.53   0.597     -1737.68    3007.132
             bmicat2 |
                  2  |   608.5925   610.7085     1.00   0.321    -599.7986    1816.984
                  3  |   758.5206   610.2654     1.24   0.216    -448.9938    1966.035
             racethx |
                  1  |   1488.282   1143.814     1.30   0.196    -774.9492    3751.512
                  3  |  -1585.221   888.1245    -1.78   0.077    -3342.528    172.0848
                  4  |   10.62121   712.7059     0.01   0.988    -1399.589    1420.832
                  5  |  -2539.193   1253.192    -2.03   0.045    -5018.848   -59.53782
             regcat2 |
                  1  |   13.33518    908.936     0.01   0.988     -1785.15     1811.82
                  2  |   675.3293   618.5639     1.09   0.277    -548.6051    1899.264
                  4  |   347.2678    964.155     0.36   0.719    -1560.478    2255.013
            1.binge2 |  -1176.432   585.2533    -2.01   0.047    -2334.455   -18.40826
              povcat |
                  1  |  -677.2424   1019.389    -0.66   0.508    -2694.277    1339.792
                  2  |  -1112.018   1429.205    -0.78   0.438    -3939.944    1715.909
                  3  |  -3017.783   850.5411    -3.55   0.001    -4700.724   -1334.843
                  4  |   -483.879   618.5462    -0.78   0.435    -1707.778    740.0203
               inscov |
                  1  |   5240.099   1219.284     4.30   0.000     2827.538    7652.661
                  2  |   7391.552   1484.194     4.98   0.000     4454.821    10328.28
               pcs42 |  -256.1957   49.86388    -5.14   0.000    -354.8599   -157.5315
               mcs42 |  -160.8633    59.1569    -2.72   0.007    -277.9153   -43.81125
             k6sum42 |  -73.51918   111.1418    -0.66   0.509    -293.4322    146.3938
              phq242 |
```

```
                   1  |  -1549.241   864.3641    -1.79   0.075    -3259.533    161.0512
                   2  |  -1267.693    1044.68    -1.21   0.227     -3334.77    799.3844
                   3  |   -3253.81   1316.707    -2.47   0.015    -5859.139   -648.4806
                   4  |  -4033.431    1714.15    -2.35   0.020     -7425.17   -641.6919
                   5  |  -5921.112   1800.511    -3.29   0.001     -9483.73   -2358.494
                   6  |  -3986.066   1986.156    -2.01   0.047    -7916.016   -56.11649
------------------------------------------------------------------------------
```

Healthcare Utilization Models

Emergency room visits (ER)

```
* Hurdle:Two-part model, with logit first part and Piosson second part for all variables (i.e
sociodemographic and comorbidities)
logit any_off c.agelast ib3.tobs2 i.sex ib3.educat3 ib3.marcat2 i.bmicat2 ib2.racethx
ib3.regcat2 i.binge2 ib5.povcat ib3.inscov c.pcs42 c.mcs42 c.k6sum42 i.phq242 , vce(robust)
Iteration 0:   log pseudolikelihood = -527.65512
Iteration 1:   log pseudolikelihood = -488.65173
Iteration 2:   log pseudolikelihood = -483.40995
Iteration 3:   log pseudolikelihood = -483.39281
Iteration 4:   log pseudolikelihood = -483.39281

Logistic regression                             Number of obs     =      1,308
                                                Wald chi2(35)     =      85.57
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -483.39281               Pseudo R2         =     0.0839
------------------------------------------------------------------------------
                     |               Robust
            any_off |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------------+---------------------------------------------------------------
            agelast |  -.0135145   .0075628    -1.79   0.074    -.0283375    .0013084
              tobs2 |
                 1  |   .3668492   .2174766     1.69   0.092    -.0593971    .7930955
                 2  |   .2505223   .2343581     1.07   0.285    -.2088113    .7098558
              2.sex |   .3077268   .2726864     1.13   0.259    -.2267287    .8421824
            educat3 |
COLLEGE GRAD OR HIGH |   .0961242    .232606     0.41   0.679    -.3597753    .5520236
                HS  |  -.0442821    .243313    -0.18   0.856    -.5211667    .4326026
       SOME COLLEGE |   -.023248   .2707336    -0.09   0.932    -.5538762    .5073801
            marcat2 |
                 1  |  -.0888831   .2364731    -0.38   0.707    -.5523618    .3745957
                 2  |   .1250408   .2643815     0.47   0.636    -.3931374    .6432189
                 4  |   .3252572   .3054655     1.06   0.287    -.2734441    .9239585
            bmicat2 |
```

```
                   2  |  -.2453346   .2231541    -1.10   0.272    -.6827085    .1920394
                   3  |  -.1099313   .2077931    -0.53   0.597    -.5171983    .2973357
             racethx  |
                   1  |  -.1042317   .3197723    -0.33   0.744    -.7309739    .5225105
                   3  |   .0647847   .2435401     0.27   0.790    -.4125452    .5421146
                   4  |   .2619719   .2466642     1.06   0.288    -.2214811     .745425
                   5  |    .179846   .4641951     0.39   0.698    -.7299596    1.089652
             regcat2  |
                   1  |   .4361185   .2933942     1.49   0.137    -.1389235     1.01116
                   2  |   .4896813   .2219864     2.21   0.027      .054596    .9247667
                   4  |  -.1794305   .2791778    -0.64   0.520     -.726609    .3677479
            1.binge2  |  -.2080421   .2089706    -1.00   0.319     -.617617    .2015328
              povcat  |
                   1  |   .1479691   .3038153     0.49   0.626    -.4474979    .7434361
                   2  |   -.210096   .4468939    -0.47   0.638    -1.085992    .6657999
                   3  |   -.215715   .3262807    -0.66   0.509    -.8552133    .4237834
                   4  |   .2911892   .2250969     1.29   0.196    -.1499925     .732371
             inscov   |
                   1  |   .5192446   .3501232     1.48   0.138    -.1669842    1.205473
                   2  |    1.34212   .3886189     3.45   0.001     .5804411    2.103799
              pcs42   |  -.0164959   .0090514    -1.82   0.068    -.0342362    .0012445
              mcs42   |   .0007247   .0136023     0.05   0.958    -.0259352    .0273847
            k6sum42   |    .067565   .0386872     1.75   0.081    -.0082606    .1433906
             phq242   |
                   1  |   .2873864   .2948171     0.97   0.330    -.2904445    .8652173
                   2  |  -.2621818    .379441    -0.69   0.490    -1.005873     .481509
                   3  |  -.0240553   .4827771    -0.05   0.960    -.9702809    .9221703
                   4  |  -.0028663   .5691528    -0.01   0.996    -1.118385    1.112653
                   5  |   -1.51684   1.091314    -1.39   0.165    -3.655777    .6220965
                   6  |  -.2345369   .7225263    -0.32   0.745    -1.650662    1.181589
               _cons  |  -1.737642   1.196008    -1.45   0.146    -4.081774    .6064906
------------------------------------------------------------------------------
*Hurdle Piosson Model estimates
tpoisson ertot c.agelast ib3.tobs2 i.sex ib3.educat3 ib3.marcat2 i.bmicat2 ib2.racethx
ib3.regcat2 i.binge2 ib5.povcat ib3.inscov c.pcs42 c.mcs42 c.k6sum42 i.phq242  if ertot>0 ,
ll(0) vce(robust)
Iteration 0:   log pseudolikelihood = -175.48536
Iteration 1:   log pseudolikelihood = -139.49958
Iteration 2:   log pseudolikelihood = -138.24711
Iteration 3:   log pseudolikelihood = -138.20419
Iteration 4:   log pseudolikelihood = -138.19524
Iteration 5:   log pseudolikelihood = -138.19312
Iteration 6:   log pseudolikelihood = -138.19264
Iteration 7:   log pseudolikelihood = -138.19253
Iteration 8:   log pseudolikelihood = -138.19251
Iteration 9:   log pseudolikelihood = -138.19251
Truncated Poisson regression                    Number of obs     =        182
Limits:        lower =          0               Wald chi2(35)     =     268.59
               upper =       +inf               Prob > chi2       =     0.0000
Log pseudolikelihood = -138.19251               Pseudo R2         =     0.2779
------------------------------------------------------------------------------
                     |               Robust
               ertot |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
             agelast |   .0065554   .0100686     0.65   0.515    -.0131788    .0262895
               tobs2 |
                   1 |   .4280123   .3860392     1.11   0.268    -.3286106    1.184635
                   2 |   .2088302   .4732322     0.44   0.659    -.7186878    1.136348
               2.sex |   .5784416   .3733406     1.55   0.121    -.1532926    1.310176
             educat3 |
COLLEGE GRAD OR HIGH |  -.1361085   .3454611    -0.39   0.694    -.8131999    .5409829
```

```
              HS |   .2163224     .28767     0.75   0.452    -.3475004    .7801452
    SOME COLLEGE |  -.2538544   .3745386    -0.68   0.498    -.9879366    .4802279
         marcat2 |
              1 |   .1018368   .3895788     0.26   0.794    -.6617237    .8653973
              2 |  -.7000288   .3914136    -1.79   0.074    -1.467185    .0671279
              4 |  -.3542364   .4662407    -0.76   0.447    -1.268051    .5595786
         bmicat2 |
              2 |  -.8394927   .3280143    -2.56   0.010    -1.482389   -.1965964
              3 |  -.5318744   .2981098    -1.78   0.074    -1.116159    .0524101
         racethx |
              1 |  -.1261414   .8017612    -0.16   0.875    -1.697564    1.445282
              3 |  -.1460761   .4100365    -0.36   0.722    -.9497329    .6575807
              4 |   -.389784   .4077349    -0.96   0.339     -1.18893    .4093617
              5 |  -.5380496   .5820267    -0.92   0.355    -1.678801    .6027017
         regcat2 |
              1 |  -.2283217    .441643    -0.52   0.605    -1.093926    .6372828
              2 |   .3745224   .2873538     1.30   0.192    -.1886807    .9377256
              4 |  -.6689206   .7073545    -0.95   0.344     -2.05531    .7174688
        1.binge2 |   -.973808   .4158016    -2.34   0.019    -1.788764   -.1588518
          povcat |
              1 |   .4258721    .594184     0.72   0.474    -.7387072    1.590451
              2 |  -.4708315   1.139644    -0.41   0.680    -2.704492    1.762829
              3 |   .5473565   .4569473     1.20   0.231    -.3482438    1.442957
              4 |   .5097922    .425224     1.20   0.231    -.3236315    1.343216
          inscov |
              1 |   .7365302   .5477095     1.34   0.179    -.3369607    1.810021
              2 |   1.141021   .4980706     2.29   0.022     .1648203    2.117221
           pcs42 |   .0108142   .0155894     0.69   0.488    -.0197405    .0413688
           mcs42 |   .0168534   .0198135     0.85   0.395    -.0219803    .0556871
         k6sum42 |   .0807462   .0493186     1.64   0.102    -.0159165    .1774089
          phq242 |
              1 |  -.7549338   .6702038    -1.13   0.260    -2.068509    .5586416
              2 |   -.159994   .5608412    -0.29   0.775    -1.259223    .9392346
              3 |  -.5396092   .6379838    -0.85   0.398    -1.790034     .710816
              4 |   -.309715   .5665842    -0.55   0.585      -1.4202    .8007696
              5 |  -13.81117   1.524454    -9.06   0.000    -16.79904   -10.82329
              6 |  -.7084642    .871547    -0.81   0.416    -2.416665    .9997365
           _cons |  -2.906615   1.617434    -1.80   0.072    -6.076728    .2634979-------
------------------------------------------------------------------------------
```

Office-Based visits

```
*Hurdle Piosson Model 1st part estimates
logit any_off c.agelast ib3.tobs2 i.sex ib3.educat3 ib3.marcat2 i.bmicat2 ib2.racethx
ib3.regcat2 i.binge2 ib5.povcat ib3.inscov c.pcs42 c.mcs42 c.k6sum42 i.phq242 , vce(robust)
Iteration 0:   log pseudolikelihood = -816.39687
Iteration 1:   log pseudolikelihood = -681.31435
Iteration 2:   log pseudolikelihood = -672.42965
Iteration 3:   log pseudolikelihood = -672.26699
Iteration 4:   log pseudolikelihood = -672.26674
Iteration 5:   log pseudolikelihood = -672.26674
Logistic regression                             Number of obs     =      1,298
                                                Wald chi2(34)     =     192.25
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -672.26674               Pseudo R2         =     0.1765
------------------------------------------------------------------------------
                   |               Robust
          any_off |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
          agelast |   .0168368   .0059994     2.81   0.005     .0050782    .0285954
            tobs2 |
              1 |   -.295484   .1654275    -1.79   0.074    -.6197158    .0287479
              2 |  -.3503746   .1833178    -1.91   0.056    -.7096708    .0089217
```

```
                2.sex |   1.047489   .3154674     3.32   0.001     .4291845    1.665794
              educat3 |
COLLEGE GRAD OR HIGH  |   .2637928   .1845999     1.43   0.153    -.0980164     .625602
                  HS  |   .0530896   .1926303     0.28   0.783    -.3244588     .430638
        SOME COLLEGE  |   .1003371   .2189855     0.46   0.647    -.3288665    .5295407
              marcat2 |
                   1  |  -.2095862   .1791778    -1.17   0.242    -.5607683    .1415959
                   2  |  -.1237406   .2253885    -0.55   0.583    -.5654939    .3180126
                   4  |  -.0376856   .2862484    -0.13   0.895    -.5987222     .523351
              bmicat2 |
                   2  |   .3248639   .1732316     1.88   0.061    -.0146638    .6643915
                   3  |   .1304622   .1779765     0.73   0.464    -.2183652    .4792897
              racethx |
                   1  |  -.3135539   .2230608    -1.41   0.160    -.7507451    .1236373
                   3  |  -.1288438   .2128717    -0.61   0.545    -.5460647    .2883772
                   4  |  -.3550016   .1951329    -1.82   0.069    -.7374551    .0274518
                   5  |   .1237247   .4661889     0.27   0.791    -.7899886    1.037438
              regcat2 |
                   1  |   .3878437    .250147     1.55   0.121    -.1024354    .8781227
                   2  |   .3965552   .1758603     2.25   0.024     .0518754    .7412349
                   4  |   .0739696   .1944397     0.38   0.704    -.3071252    .4550644
            1.binge2  |   -.196506   .1485125    -1.32   0.186     -.487585    .0945731
               povcat |
                   1  |  -.5897605   .2865432    -2.06   0.040    -1.151375   -.0281461
                   2  |  -.0041755   .3701905    -0.01   0.991    -.7297356    .7213845
                   3  |  -.2145813    .251385    -0.85   0.393    -.7072868    .2781242
                   4  |  -.3213465   .1656436    -1.94   0.052     -.646002    .0033089
               inscov |
                   1  |   1.899153   .2474467     7.67   0.000     1.414166    2.384139
                   2  |   1.775936   .3098596     5.73   0.000     1.168623     2.38325
                pcs42 |  -.0704207   .0113077    -6.23   0.000    -.0925833    -.048258
                mcs42 |  -.0292244   .0131746    -2.22   0.027     -.055046   -.0034027
              k6sum42 |  -.0033955   .0358442    -0.09   0.925    -.0736489    .0668578
               phq242 |
                   1  |   .0671737   .2448048     0.27   0.784    -.4126348    .5469822
                   2  |  -.1226567   .2973064    -0.41   0.680    -.7053665    .4600532
                   3  |   .9335344   .6439804     1.45   0.147     -.328644    2.195713
                   4  |   .2702809   .6161677     0.44   0.661    -.9373856    1.477947
                   5  |          0  (empty)
                   6  |   .1590792   .7941821     0.20   0.841    -1.397489    1.715648
                _cons |   3.700347   1.239762     2.98   0.003     1.270459    6.130236
------------------------------------------------------------------------------
```

```
*Hurdle Piosson Model estimates
tpoisson obtotv c.agelast ib3.tobs2 i.sex ib3.educat3 ib3.marcat2 i.bmicat2 ib2.racethx
ib3.regcat2 i.binge2 ib5.povcat ib3.inscov c.pcs42 c.mcs42 c.k6sum42 i.phq242  if obtotv>0 ,
ll(0) vce(robust)
Iteration 0:   log pseudolikelihood = -4195.4307
Iteration 1:   log pseudolikelihood = -4193.9898
Iteration 2:   log pseudolikelihood = -4193.9872
Iteration 3:   log pseudolikelihood = -4193.9872
Truncated Poisson regression                    Number of obs     =        889
Limits:        lower =          0               Wald chi2(35)     =     204.48
               upper =       +inf               Prob > chi2       =     0.0000
Log pseudolikelihood = -4193.9872               Pseudo R2         =     0.1566
------------------------------------------------------------------------------
                     |               Robust
              obtotv |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
```

```
                agelast |   .0113009   .0042423     2.66   0.008     .0029862    .0196156
                  tobs2 |
                      1 |  -.1143025   .1328096    -0.86   0.389    -.3746045    .1459996
                      2 |  -.1283631   .1264181    -1.02   0.310    -.3761379    .1194118
                  2.sex |  -.1380887   .1693459    -0.82   0.415    -.4700006    .1938231
                educat3 |
COLLEGE GRAD OR HIGH    |   .1985675   .1277037     1.55   0.120    -.0517272    .4488621
                     HS |  -.0400095   .1431778    -0.28   0.780    -.3206328    .2406138
          SOME COLLEGE  |   .1448467   .1522355     0.95   0.341    -.1535294    .4432228
                marcat2 |
                      1 |   .0563614    .117405     0.48   0.631    -.1737482     .286471
                      2 |   .0689678   .1671322     0.41   0.680    -.2586053    .3965408
                      4 |  -.2146906    .271524    -0.79   0.429     -.746868    .3174867
                bmicat2 |
                      2 |   .0372111   .1643063     0.23   0.821    -.2848233    .3592455
                      3 |   .1977419   .1612058     1.23   0.220    -.1182157    .5136996
                racethx |
                      1 |  -.0818116   .1379132    -0.59   0.553    -.3521166    .1884934
                      3 |   -.143568   .1661066    -0.86   0.387     -.469131     .181995
                      4 |   .0012652   .1590829     0.01   0.994    -.3105315    .3130619
                      5 |  -.2359867     .28442    -0.83   0.407    -.7934396    .3214662
                regcat2 |
                      1 |   .1165981   .1581629     0.74   0.461    -.1933955    .4265918
                      2 |   .1519154   .1254316     1.21   0.226     -.093926    .3977568
                      4 |   .0935053   .1297415     0.72   0.471    -.1607834    .3477939
              1.binge2  |  -.2417031   .1137872    -2.12   0.034    -.4647218   -.0186843
                povcat  |
                      1 |    .066964   .1618873     0.41   0.679    -.2503292    .3842572
                      2 |  -.0214503   .1802957    -0.12   0.905    -.3748234    .3319228
                      3 |  -.0558711   .1549167    -0.36   0.718    -.3595022     .24776
                      4 |    .063079   .1229607     0.51   0.608    -.1779195    .3040774
                 inscov |
                      1 |   .7448672   .2780018     2.68   0.007     .1999937    1.289741
                      2 |   .8076762   .2894717     2.79   0.005     .2403221     1.37503
                  pcs42 |  -.0234392   .0054743    -4.28   0.000    -.0341687   -.0127098
                  mcs42 |  -.0153196   .0091351    -1.68   0.094    -.0332241    .0025849
                k6sum42 |  -.0097936   .0225821    -0.43   0.665    -.0540537    .0344666
                 phq242 |
                      1 |   .0676756    .162153     0.42   0.676    -.2501386    .3854897
                      2 |  -.0320265   .1619046    -0.20   0.843    -.3493537    .2853007
                      3 |   -.136403   .2045272    -0.67   0.505    -.5372689    .2644629
                      4 |  -.1720725   .2104385    -0.82   0.414    -.5845245    .2403794
                      5 |      .1653   .3558211     0.46   0.642    -.5320965    .8626964
                      6 |  -.1144187   .2974172    -0.38   0.700    -.6973457    .4685084
                  _cons |   2.295728   .8768698     2.62   0.009      .577095    4.014361
------------------------------------------------------------------------------
```

Hospital Outpatient visits

```
*Hurdle Piosson Model 1st part estimates
logit any_off c.agelast ib3.tobs2 i.sex ib3.educat3 ib3.marcat2 i.bmicat2 ib2.racethx
ib3.regcat2 i.binge2 ib5.povcat ib3.inscov c.pcs42 c.mcs42 c.k6sum42 i.phq242 , vce(robust)
Iteration 0:   log pseudolikelihood = -485.96161
Iteration 1:   log pseudolikelihood = -427.48217
Iteration 2:   log pseudolikelihood = -415.46802
Iteration 3:   log pseudolikelihood = -415.32979
Iteration 4:   log pseudolikelihood =  -415.3296
Iteration 5:   log pseudolikelihood =  -415.3296
```

```
Logistic regression                             Number of obs     =      1,308
                                                Wald chi2(35)     =     141.26
                                                Prob > chi2       =     0.0000
Log pseudolikelihood =  -415.3296               Pseudo R2         =     0.1453
------------------------------------------------------------------------------
                     |               Robust
            any_off |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------------+---------------------------------------------------------------
            agelast |   .0175505   .0082517     2.13   0.033     .0013774    .0337235
              tobs2 |
                 1  |  -.0481112   .2271333    -0.21   0.832    -.4932844    .3970619
                 2  |  -.4257712   .2655933    -1.60   0.109    -.9463246    .0947821
             2.sex  |   .3947112   .3182892     1.24   0.215    -.2291242    1.018547
            educat3 |
COLLEGE GRAD OR HIGH|   .9842729   .2899006     3.40   0.001     .4160781    1.552468
                 HS |   .8902138    .295547     3.01   0.003     .3109523    1.469475
       SOME COLLEGE |   1.056447   .3189116     3.31   0.001     .4313922    1.681503
            marcat2 |
                 1  |   .2032592   .2661672     0.76   0.445    -.3184189    .7249374
                 2  |   .3711417   .2980781     1.25   0.213    -.2130807     .955364
                 4  |   .5618671   .3881295     1.45   0.148    -.1988527    1.322587
            bmicat2 |
                 2  |   .0725509   .2685909     0.27   0.787    -.4538775    .5989794
                 3  |    .187736   .2452924     0.77   0.444    -.2930282    .6685003
            racethx |
                 1  |  -.0858986   .3445966    -0.25   0.803    -.7612956    .5894984
                 3  |  -.4093614   .2977689    -1.37   0.169    -.9929777    .1742548
                 4  |   .0677415   .2795668     0.24   0.809    -.4801994    .6156825
                 5  |  -.1813286   .5985208    -0.30   0.762    -1.354408    .9917505
            regcat2 |
                 1  |   .6303421   .3208834     1.96   0.049     .0014222    1.259262
                 2  |   .3829294   .2276408     1.68   0.093    -.0632383    .8290971
                 4  |  -.5039126    .336653    -1.50   0.134     -1.16374    .1559152
           1.binge2 |  -.3117728   .2341094    -1.33   0.183    -.7706188    .1470732
             povcat |
                 1  |  -.4949694   .3749276    -1.32   0.187    -1.229814    .2398751
                 2  |   .1409888   .4513187     0.31   0.755    -.7435796    1.025557
                 3  |  -.1462253   .3399149    -0.43   0.667    -.8124462    .5199956
                 4  |  -.1660053    .239161    -0.69   0.488    -.6347522    .3027415
             inscov |
                 1  |   1.029891   .4385278     2.35   0.019     .1703924     1.88939
                 2  |   .8534226   .4859753     1.76   0.079    -.0990716    1.805917
              pcs42 |  -.0479879   .0098784    -4.86   0.000    -.0673493   -.0286266
              mcs42 |  -.0023717   .0153972    -0.15   0.878    -.0325496    .0278063
            k6sum42 |    .051732   .0411445     1.26   0.209    -.0289097    .1323736
             phq242 |
                 1  |   .3251578   .3082234     1.05   0.291    -.2789488    .9292645
                 2  |  -.0701067   .3800023    -0.18   0.854    -.8148974     .674684
                 3  |  -.4311004    .521605    -0.83   0.409    -1.453427    .5912265
                 4  |   -.650799    .541252    -1.20   0.229    -1.711633    .4100353
                 5  |  -1.145806   1.279351    -0.90   0.370    -3.653289    1.361676
                 6  |  -.4678984   .7865148    -0.59   0.552    -2.009439    1.073642
              _cons |  -2.363821   1.440814    -1.64   0.101    -5.187764    .4601215
------------------------------------------------------------------------------


.
*Hurdle Piosson Model estimates
tpoisson optotv c.agelast ib3.tobs2 i.sex ib3.educat3 ib3.marcat2 i.bmicat2 ib2.racethx
ib3.regcat2 i.binge2 ib5.povcat ib3.inscov c.pcs42 c.mcs42 c.k6sum42 i.phq242  if optotv>0 ,
ll(0) vce(robust)
```

```
Iteration 0:   log pseudolikelihood = -321.96632
Iteration 1:   log pseudolikelihood = -308.66006
Iteration 2:   log pseudolikelihood = -308.51122
Iteration 3:   log pseudolikelihood = -308.50979
Iteration 4:   log pseudolikelihood = -308.50979
Truncated Poisson regression                    Number of obs     =        160
Limits:        lower =          0               Wald chi2(34)     =          .
               upper =       +inf               Prob > chi2       =          .
Log pseudolikelihood = -308.50979               Pseudo R2         =     0.2746
------------------------------------------------------------------------------
                     |               Robust
              optotv |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
            agelast |   .0206494   .0117688     1.75   0.079    -.0024171    .0437158
             tobs2 |
                1  |   .0328895   .2166145     0.15   0.879     -.391667     .457446
                2  |  -.6288453   .4155325    -1.51   0.130    -1.443274    .1855836
             2.sex |   .4347539   .3188746     1.36   0.173    -.1902287    1.059737
            educat3 |
COLLEGE GRAD OR HIGH  |   .7025447   .3084982     2.28   0.023     .0978993     1.30719
                HS  |   .1157596   .3365218     0.34   0.731    -.5438111    .7753303
       SOME COLLEGE  |  -.2871887   .3524441    -0.81   0.415    -.9779663     .403589
            marcat2 |
                1  |   .4690977   .3683503     1.27   0.203    -.2528556    1.191051
                2  |   .0058586   .3399006     0.02   0.986    -.6603343    .6720516
                4  |     .78712   .4242371     1.86   0.064    -.0443694    1.618609
            bmicat2 |
                2  |   .1907473   .4279312     0.45   0.656    -.6479825    1.029477
                3  |   .1014148   .3291321     0.31   0.758    -.5436723    .7465018
            racethx |
                1  |  -.1778024   .4316192    -0.41   0.680     -1.02376    .6681556
                3  |   .0025738   .6401453     0.00   0.997    -1.252088    1.257236
                4  |   .3087647   .3854124     0.80   0.423    -.4466297    1.064159
                5  |  -1.019079   1.124085    -0.91   0.365    -3.222245    1.184086
            regcat2 |
                1  |   .5299183   .4563484     1.16   0.246    -.3645081    1.424345
                2  |   .0730268   .2932798     0.25   0.803     -.501791    .6478446
                4  |   .8476742   .4201825     2.02   0.044     .0241316    1.671217
           1.binge2 |   .1793379    .337259     0.53   0.595    -.4816775    .8403533
             povcat |
                1  |  -1.329007     .66186    -2.01   0.045    -2.626229   -.0317857
                2  |   .0146332   .4177433     0.04   0.972    -.8041286    .8333951
                3  |   .3756531   .3005981     1.25   0.211    -.2135084    .9648145
                4  |  -.4807965   .2953234    -1.63   0.104     -1.05962    .0980268
             inscov |
                1  |  -1.007375   .4729389    -2.13   0.033    -1.934319   -.0804322
                2  |   -.942767   .4475977    -2.11   0.035    -1.820042   -.0654915
              pcs42 |  -.0166281   .0133852    -1.24   0.214    -.0428626    .0096063
              mcs42 |   .0100274   .0205089     0.49   0.625    -.0301693    .0502241
            k6sum42 |   .0443209   .0437465     1.01   0.311    -.0414207    .1300624
             phq242 |
                1  |    .894428   .4070109     2.20   0.028     .0967013    1.692155
                2  |   .7900444   .4550466     1.74   0.083    -.1018307    1.681919
                3  |   .6776741   .5933243     1.14   0.253    -.4852202    1.840568
                4  |   .0999072   .5684551     0.18   0.860    -1.014244    1.214059
                5  |   .6075025   .8002736     0.76   0.448     -.961005     2.17601
                6  |  -.7170258   .9449706    -0.76   0.448    -2.569134    1.135083
              _cons |   -.408934   1.705857    -0.24   0.811    -3.752352    2.934484
------------------------------------------------------------------------------
```

Hospital Inpatient visits

```
*Hurdle Piosson Model 1st part estimates
logit any_off c.agelast ib3.tobs2 i.sex ib3.educat3 ib3.marcat2 i.bmicat2 ib2.racethx
ib3.regcat2 i.binge2 ib5.povcat ib3.inscov c.pcs42 c.mcs42 c.k6sum42 i.phq242 , vce(robust)
Iteration 0:   log pseudolikelihood = -298.30249
Iteration 1:   log pseudolikelihood = -280.47312
Iteration 2:   log pseudolikelihood = -248.85825
Iteration 3:   log pseudolikelihood =  -248.1469
Iteration 4:   log pseudolikelihood = -248.14543
Iteration 5:   log pseudolikelihood = -248.14543

Logistic regression                             Number of obs     =      1,308
                                                Wald chi2(35)     =     114.70
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -248.14543               Pseudo R2         =     0.1681

------------------------------------------------------------------------------
                     |               Robust
            any_off |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
            agelast |   .0161383   .0107411     1.50   0.133    -.0049138    .0371903
              tobs2 |
                 1  |   .3538343   .3075652     1.15   0.250    -.2489823     .956651
                 2  |  -.3799129   .3696586    -1.03   0.304     -1.10443    .3446047
              2.sex |  -.1249233   .4101811    -0.30   0.761    -.9288634    .6790168
            educat3 |
COLLEGE GRAD OR HIGH |  -.5361302   .3672294    -1.46   0.144    -1.255887    .1836262
                 HS  |  -.1226658   .3280373    -0.37   0.708     -.765607    .5202754
        SOME COLLEGE |   .1074975   .3634546     0.30   0.767    -.6048605    .8198555
            marcat2 |
                 1  |   .1766569   .3877062     0.46   0.649    -.5832332     .936547
                 2  |   .2833022   .4267593     0.66   0.507    -.5531306    1.119735
                 4  |   .3954837   .5543624     0.71   0.476    -.6910465    1.482014
            bmicat2 |
                 2  |   .5066758   .3941332     1.29   0.199    -.2658112    1.279163
                 3  |   .3807568   .3645736     1.04   0.296    -.3337942    1.095308
            racethx |
                 1  |   .0483068   .5175404     0.09   0.926    -.9660537    1.062667
                 3  |   .0356478   .3561874     0.10   0.920    -.6624667    .7337624
                 4  |   .2072959   .3756397     0.55   0.581    -.5289443    .9435361
                 5  |   .2674564   .5525373     0.48   0.628    -.8154967     1.35041
            regcat2 |
                 1  |  -.2600751   .4516127    -0.58   0.565     -1.14522    .6250696
                 2  |  -.2209548   .3331348    -0.66   0.507    -.8738871    .4319774
                 4  |  -.8819413   .4697311    -1.88   0.060    -1.802597    .0387146
           1.binge2 |  -.1045359   .3243102    -0.32   0.747    -.7401721    .5311003
             povcat |
                 1  |  -.1867277   .4363721    -0.43   0.669    -1.042001    .6685459
                 2  |  -.2653157    .608982    -0.44   0.663    -1.458898    .9282671
                 3  |  -.4378296   .4693251    -0.93   0.351     -1.35769    .4820307
                 4  |  -.4810704   .3849748    -1.25   0.211    -1.235607    .2734664
              inscov |
                 1  |   .6764091    .639147     1.06   0.290     -.576296    1.929114
                 2  |   .7901079   .6451614     1.22   0.221    -.4743853    2.054601
              pcs42 |   -.050687   .0115256    -4.40   0.000    -.0732767   -.0280972
              mcs42 |  -.0053079   .0200527    -0.26   0.791    -.0446104    .0339947
            k6sum42 |   .0547161    .053548     1.02   0.307    -.0502361    .1596684
             phq242 |
                 1  |  -.0206853    .472317    -0.04   0.965    -.9464095     .905039
```

```
                   2  |  -.8440632   .6312365    -1.34   0.181    -2.081264    .3931376
                   3  |  -.7373763   .7551436    -0.98   0.329    -2.217431     .742678
                   4  |    .170513   .6584609     0.26   0.796    -1.120047    1.461073
                   5  |  -.7686657   1.252487    -0.61   0.539    -3.223496    1.686164
                   6  |  -.2333355   .9442539    -0.25   0.805    -2.084039    1.617368
                 _cons |  -1.803927   1.785307    -1.01   0.312    -5.303065    1.695211
------------------------------------------------------------------------------------------

 *Hurdle Piosson Model estimates
tpoisson ipngtd c.agelast ib3.tobs2 i.sex ib3.educat3 ib3.marcat2 i.bmicat2 ib2.racethx
ib3.regcat2 i.binge2 ib5.povcat ib3.inscov c.pcs42 c.mcs42 c.k6sum42 i.phq242  if ipngtd>0 ,
ll(0) vce(robust)

Iteration 0:   log pseudolikelihood =  -895.2041
Iteration 1:   log pseudolikelihood = -464.83757
Iteration 2:   log pseudolikelihood = -309.17592
Iteration 3:   log pseudolikelihood = -192.58375
Iteration 4:   log pseudolikelihood = -191.13686
Iteration 5:   log pseudolikelihood = -191.12919
Iteration 6:   log pseudolikelihood = -191.12919

------------------------------------------------------------------------------------------
                       |               Robust
                ipngtd |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------------+------------------------------------------------------------------
              agelast  |   .0180873   .0097017     1.86   0.062    -.0009277    .0371023
                tobs2  |
                   1   |   .6434621   .3795601     1.70   0.090     -.100462    1.387386
                   2   |  -.0262953   .3153238    -0.08   0.934    -.6443186     .591728
                2.sex  |   .3836765   .2453334     1.56   0.118    -.0971681     .864521
              educat3  |
COLLEGE GRAD OR HIGH   |  -.1058688   .3788531    -0.28   0.780    -.8484072    .6366695
                   HS  |   1.129212   .3476693     3.25   0.001     .4477925    1.810631
        SOME COLLEGE   |   .0596198   .2674494     0.22   0.824    -.4645715     .583811
              marcat2  |
                   1   |   -.386643   .3382997    -1.14   0.253    -1.049698    .2764123
                   2   |  -.5084107   .3744118    -1.36   0.174    -1.242244    .2254229
                   4   |   .8624364   .5561576     1.55   0.121    -.2276124    1.952485
              bmicat2  |
                   2   |   .4016985    .307116     1.31   0.191    -.2002378    1.003635
                   3   |   .5039153   .2945032     1.71   0.087    -.0733004    1.081131
              racethx  |
                   1   |   1.177186   .4885521     2.41   0.016     .2196418    2.134731
                   3   |   .7728373   .3053966     2.53   0.011     .1742709    1.371404
                   4   |  -.7064896   .3628239    -1.95   0.052    -1.417612    .0046322
                   5   |  -.7690285   .4226368    -1.82   0.069    -1.597381    .0593244
              regcat2  |
                   1   |    .299896    .465468     0.64   0.519    -.6124044    1.212197
                   2   |  -.6155872   .3330922    -1.85   0.065    -1.268436    .0372615
                   4   |   .0082274   .3759822     0.02   0.983    -.7286842    .7451389
             1.binge2  |  -.3361988   .3757412    -0.89   0.371    -1.072638    .4002404
               povcat  |
                   1   |   .6774472   .3353876     2.02   0.043     .0200995    1.334795
                   2   |   -.008223   .4617167    -0.02   0.986    -.9131711    .8967251
                   3   |   -.763004   .4973438    -1.53   0.125     -1.73778    .2117719
                   4   |   .5714829   .2903333     1.97   0.049     .0024401    1.140526
               inscov  |
                   1   |  -.5677419   .5772532    -0.98   0.325    -1.699137    .5636536
                   2   |  -.3685124   .5060062    -0.73   0.466    -1.360266    .6232416
                pcs42  |  -.0061151   .0133202    -0.46   0.646    -.0322221     .019992
                mcs42  |  -.0613642   .0143415    -4.28   0.000    -.0894731   -.0332553
```

```
             k6sum42 |  -.0795008   .0359464    -2.21   0.027    -.1499545   -.0090471
              phq242 |
                  1  |   .0178828   .3683339     0.05   0.961    -.7040385     .739804
                  2  |   .7663659   .3828313     2.00   0.045     .0160303    1.516702
                  3  |  -.5101972   .5969091    -0.85   0.393    -1.680118    .6597232
                  4  |   .0027136   .4633434     0.01   0.995    -.9054229      .91085
                  5  |  -2.159661   .7731497    -2.79   0.005    -3.675007   -.6443157
                  6  |   .6569241   .5886077     1.12   0.264    -.4967258    1.810574
                     |
               _cons |   3.423672   1.375536     2.49   0.013     .7276713    6.119673
------------------------------------------------------------------------------
```

Dental visits

```
*Hurdle Piosson Model 1st part estimates
logit any_off c.agelast ib3.tobs2 i.sex ib3.educat3 ib3.marcat2 i.bmicat2 ib2.racethx
ib3.regcat2 i.binge2 ib5.povcat ib3.inscov c.pcs42 c.mcs42 c.k6sum42 i.phq242 , vce(robust)
Iteration 0:   log pseudolikelihood = -793.51301
Iteration 1:   log pseudolikelihood =  -719.7301
Iteration 2:   log pseudolikelihood = -716.58987
Iteration 3:   log pseudolikelihood = -716.56532
Iteration 4:   log pseudolikelihood = -716.56532
Logistic regression                               Number of obs     =      1,308
                                                  Wald chi2(35)     =     123.14
                                                  Prob > chi2       =     0.0000
Log pseudolikelihood = -716.56532                 Pseudo R2         =     0.0970
------------------------------------------------------------------------------
                     |               Robust
             any_off |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
             agelast |  -.0098785   .0058577    -1.69   0.092    -.0213594    .0016025
               tobs2 |
                  1  |  -.5306906   .1611925    -3.29   0.001     -.846622   -.2147592
                  2  |  -.4992461   .1732044    -2.88   0.004    -.8387204   -.1597718
              2.sex  |  -.0402554   .2786983    -0.14   0.885     -.586494    .5059832
             educat3 |
COLLEGE GRAD OR HIGH |   .4721171   .1754653     2.69   0.007     .1282114    .8160229
                  HS |   .1470452   .1940867     0.76   0.449    -.2333577    .5274481
        SOME COLLEGE |    .256672   .2138055     1.20   0.230    -.1623791     .675723
             marcat2 |
                  1  |   .376459   .1752615     2.15   0.032     .0329527    .7199653
                  2  |  .2126135   .2144062     0.99   0.321     -.207615     .632842
                  4  |  -.0699466   .2802357    -0.25   0.803    -.6191985    .4793053
             bmicat2 |
                  2  |   .0622733   .1700683     0.37   0.714    -.2710543     .395601
                  3  |  -.0413113   .1746216    -0.24   0.813    -.3835634    .3009408
             racethx |
                  1  |  -.0970693   .2249226    -0.43   0.666    -.5379095    .3437709
                  3  |  -.2286144   .2125018    -1.08   0.282    -.6451103    .1878816
                  4  |  -.2393281   .1968749    -1.22   0.224    -.6251958    .1465397
                  5  |  -.5851657   .4601301    -1.27   0.203    -1.487004    .3166728
             regcat2 |
                  1  |   .5026442   .2311992     2.17   0.030     .0495021    .9557863
                  2  |    .432802   .1669266     2.59   0.010     .1056319    .7599722
                  4  |   .6621034   .1895773     3.49   0.000     .2905386    1.033668
            1.binge2 |  -.1109451   .1519592    -0.73   0.465    -.4087798    .1868895
              povcat |
                  1  |  -.9193867   .2653548    -3.46   0.001    -1.439473   -.3993009
                  2  |  -.7118437   .3550522    -2.00   0.045    -1.407733   -.0159543
```

```
                   3  |  -1.076715   .2619222    -4.11   0.000    -1.590073   -.5633573
                   4  |  -.6659233   .1563328    -4.26   0.000    -.9723299   -.3595166
              inscov  |
                   1  |   1.086058   .3270659     3.32   0.001     .4450203    1.727095
                   2  |   1.314026   .3544898     3.71   0.000     .6192389    2.008813
                pcs42 |  -.0019713   .0078739    -0.25   0.802    -.0174038    .0134612
                mcs42 |   .0050473   .0107243     0.47   0.638    -.0159719    .0260665
              k6sum42 |   .0186635   .0307805     0.61   0.544    -.0416652    .0789922
               phq242 |
                   1  |   .1045737   .2450067     0.43   0.670    -.3756306    .5847779
                   2  |    .284065   .2721872     1.04   0.297    -.2494121    .8175422
                   3  |  -.4932559   .4949637    -1.00   0.319    -1.463367    .4768551
                   4  |   .2911963   .4430826     0.66   0.511    -.5772296    1.159622
                   5  |  -1.047299    1.28057    -0.82   0.413    -3.557171    1.462572
                   6  |  -.8675374   .6528331    -1.33   0.184    -2.147067    .4119919
               _cons  |  -1.505468    .986796    -1.53   0.127    -3.439552     .428617
------------------------------------------------------------------------------
```

```
*Hurdle Piosson Model estimates
tpoisson dvtot c.agelast ib3.tobs2 i.sex ib3.educat3 ib3.marcat2 i.bmicat2 ib2.racethx
ib3.regcat2 i.binge2 ib5.povcat ib3.inscov c.pcs42 c.mcs42 c.k6sum42 i.phq242  if dvtot>0 ,
ll(0) vce(robust)
Iteration 0:   log pseudolikelihood = -548.38111
Iteration 1:   log pseudolikelihood = -529.50306
Iteration 2:   log pseudolikelihood = -529.15182
Iteration 3:   log pseudolikelihood = -529.10636
Iteration 4:   log pseudolikelihood = -529.09648
Iteration 5:   log pseudolikelihood = -529.09438
Iteration 6:   log pseudolikelihood = -529.09391
Iteration 7:   log pseudolikelihood = -529.09379
Iteration 8:   log pseudolikelihood = -529.09377
Truncated Poisson regression                    Number of obs     =        386
Limits:        lower =          0               Wald chi2(35)     =     204.85
               upper =       +inf               Prob > chi2       =     0.0000
Log pseudolikelihood = -529.09377               Pseudo R2         =     0.0772
------------------------------------------------------------------------------
                     |               Robust
               dvtot |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
             agelast |   .0034422   .0054142     0.64   0.525    -.0071695     .014054
               tobs2 |
                   1 |  -.0104245     .13819    -0.08   0.940     -.281272     .260423
                   2 |  -.2172335   .1557458    -1.39   0.163    -.5224896    .0880226
               2.sex |  -.4912071   .2754773    -1.78   0.075    -1.031133    .0487184
             educat3 |
COLLEGE GRAD OR HIGH |  -.0771707   .1721864    -0.45   0.654    -.4146499    .2603084
                  HS |  -.0856638   .1688669    -0.51   0.612    -.4166369    .2453092
        SOME COLLEGE |  -.2629496   .1986369    -1.32   0.186    -.6522707    .1263715
             marcat2 |
                   1 |  -.1054379   .1761199    -0.60   0.549    -.4506265    .2397507
                   2 |   .0051965   .1697501     0.03   0.976    -.3275077    .3379007
                   4 |  -.2023416   .2555857    -0.79   0.429    -.7032804    .2985972
             bmicat2 |
                   2 |   .5047509   .1653514     3.05   0.002     .1806682    .8288336
                   3 |   .3242999   .1563735     2.07   0.038     .0178135    .6307863
             racethx |
                   1 |  -.0137411   .2094031    -0.07   0.948    -.4241637    .3966815
                   3 |  -.1617194   .2177928    -0.74   0.458    -.5885855    .2651468
```

```
                   4 |   .2656673   .1877859     1.41   0.157    -.1023862    .6337209
                   5 |  -.4896896   .5300754    -0.92   0.356    -1.528618     .549239
             regcat2 |
                   1 |   .0502691    .215536     0.23   0.816    -.3721738     .472712
                   2 |   .0409246   .1720089     0.24   0.812    -.2962066    .3780558
                   4 |   .2434774   .1748587     1.39   0.164    -.0992393    .5861942
            1.binge2 |  -.0272248   .1245548    -0.22   0.827    -.2713478    .2168982
              povcat |
                   1 |  -.2760822   .2814241    -0.98   0.327    -.8276633     .275499
                   2 |    .296516   .2745461     1.08   0.280    -.2415846    .8346165
                   3 |  -.0811298   .2855806    -0.28   0.776    -.6408574    .4785978
                   4 |    .026816   .1449073     0.19   0.853    -.2571972    .3108292
             inscov |
                   1 |   .3951923   .4969009     0.80   0.426    -.5787155      1.3691
                   2 |   .8502336   .5226723     1.63   0.104    -.1741852    1.874652
               pcs42 |   -.009872   .0082064    -1.20   0.229    -.0259563    .0062123
               mcs42 |  -.0089227   .0139343    -0.64   0.522    -.0362335    .0183881
             k6sum42 |  -.0100694   .0402612    -0.25   0.803    -.0889798     .068841
              phq242 |
                   1 |  -.0492447   .1996165    -0.25   0.805    -.4404857    .3419964
                   2 |  -.0840059   .2205595    -0.38   0.703    -.5162944    .3482827
                   3 |  -.7843748   .5939681    -1.32   0.187    -1.948531    .3797813
                   4 |  -.4332627   .3859984    -1.12   0.262    -1.189806    .3232802
                   5 |  -11.79689   1.099792   -10.73   0.000    -13.95244   -9.641335
                   6 |  -.5644977   .9787853    -0.58   0.564    -2.482882    1.353886
               _cons |   .7088899   1.056085     0.67   0.502    -1.360999    2.778778
------------------------------------------------------------------------------
```

Home Health Care visits

```
*Hurdle Piosson Model 1st part estimates
 generate any_off = hhtotdy>0
logit any_off c.agelast ib3.tobs2 i.sex ib3.educat3 ib3.marcat2 i.bmicat2 ib2.racethx
ib3.regcat2 i.binge2 ib5.povcat ib3.inscov, vce(robust)
Iteration 0:   log pseudolikelihood =   -241.085
Iteration 1:   log pseudolikelihood = -199.90625
Iteration 2:   log pseudolikelihood = -186.08095
Iteration 3:   log pseudolikelihood = -185.54356
Iteration 4:   log pseudolikelihood = -185.53592
Iteration 5:   log pseudolikelihood = -185.53591
Logistic regression                             Number of obs     =      1,753
                                                Wald chi2(26)     =     125.19
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -185.53591               Pseudo R2         =     0.2304
------------------------------------------------------------------------------
                     |               Robust
            any_off |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------------+---------------------------------------------------------------
            agelast |   .0392873   .0125304     3.14   0.002     .0147281    .0638465
              tobs2 |
                1   |  -.4117874   .3856154    -1.07   0.286     -1.16758     .344005
                2   |  -.0330975   .3393711    -0.10   0.922    -.6982526    .6320576
              2.sex |  -.1838346   .3724577    -0.49   0.622    -.9138384    .5461691
            educat3 |
COLLEGE GRAD OR HIGH|  -.4973059   .3988767    -1.25   0.212     -1.27909     .284478
                 HS |   .0824611   .3779785     0.22   0.827    -.6583631    .8232853
       SOME COLLEGE |  -.1680935    .468466    -0.36   0.720     -1.08627     .750083
            marcat2 |
                1   |  -.5557302   .4833085    -1.15   0.250    -1.502997     .391537
```

```
                   2  |  -.2219835   .4489408    -0.49   0.621    -1.101891    .6579244
                   4  |    -1.6054   1.028722    -1.56   0.119    -3.621658    .4108585
             bmicat2  |
                   2  |   .3896975   .3861904     1.01   0.313    -.3672219    1.146617
                   3  |   .1373177    .379374     0.36   0.717    -.6062417    .8808772
             racethx  |
                   1  |  -.3100632    .640145    -0.48   0.628    -1.564724    .9445979
                   3  |   .0591062   .3913985     0.15   0.880    -.7080207    .8262331
                   4  |   .1404872   .5476556     0.26   0.798     -.932898    1.213872
                   5  |   .4830411   .5934985     0.81   0.416    -.6801946    1.646277
             regcat2  |
                   1  |   .2164733   .4793731     0.45   0.652    -.7230808    1.156027
                   2  |  -.0148554    .402865    -0.04   0.971    -.8044564    .7747455
                   4  |  -1.933773   1.062189    -1.82   0.069    -4.015626    .1480788
            1.binge2  |  -.4355634   .4302853    -1.01   0.311    -1.278907    .4077803
              povcat  |
                   1  |   1.201492   .5592529     2.15   0.032     .1053766    2.297608
                   2  |   1.261699   .7011311     1.80   0.072    -.1124925    2.635891
                   3  |   .1430912   .6722072     0.21   0.831    -1.174411    1.460593
                   4  |   1.065295   .4997887     2.13   0.033     .0857276    2.044863
              inscov  |
                   1  |   1.089894   1.072962     1.02   0.310    -1.013073     3.19286
                   2  |   2.189757   1.038439     2.11   0.035      .154455     4.22506
                _cons |  -7.266709   1.335786    -5.44   0.000    -9.884802   -4.648616
------------------------------------------------------------------------------
```

```
*Hurdle Piosson Model estimates
tpoisson hhtotdy c.agelast ib3.tobs2 i.sex ib3.educat3 ib3.marcat2 i.bmicat2 ib2.racethx
ib3.regcat2 i.binge2 ib5.povcat ib3.inscov  if hhtotdy>0 , ll(0) vce(robust)
Iteration 0:   log pseudolikelihood = -740.93191
Iteration 1:   log pseudolikelihood = -732.23292
Iteration 2:   log pseudolikelihood = -732.13868
Iteration 3:   log pseudolikelihood = -732.13846
Iteration 4:   log pseudolikelihood = -732.13846
Truncated Poisson regression                    Number of obs     =         47
Limits:        lower =          0               Wald chi2(23)     =          .
               upper =       +inf               Prob > chi2       =          .
Log pseudolikelihood = -732.13846               Pseudo R2         =     0.6867
------------------------------------------------------------------------------
                     |               Robust
             hhtotdy |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
             agelast |  -.0161357   .0178378    -0.90   0.366     -.051097    .0188257
               tobs2 |
                   1 |  -.9307224   .6100287    -1.53   0.127    -2.126357    .2649119
                   2 |  -.2067102   .3422505    -0.60   0.546    -.8775088    .4640884
                 2.sex |   1.169077   .3930684     2.97   0.003     .3986772    1.939477
             educat3 |
COLLEGE GRAD OR HIGH |  -1.235567    .782363    -1.58   0.114     -2.76897    .2978368
                  HS |  -1.210934    .260059    -4.66   0.000     -1.72064   -.7012275
        SOME COLLEGE |  -1.233186   .4045504    -3.05   0.002     -2.02609   -.4402817
             marcat2 |
                   1 |  -.1393737   .4907958    -0.28   0.776    -1.101316    .8225684
                   2 |   .3095278   .6581212     0.47   0.638    -.9803661    1.599422
                   4 |   1.771605   1.278658     1.39   0.166    -.7345191    4.277729
```

bmicat2						
2	.5265402	.4215924	1.25	0.212	-.2997657	1.352846
3	-.1182179	.5414623	-0.22	0.827	-1.179464	.9430287
racethx						
1	.5674279	.4973629	1.14	0.254	-.4073855	1.542241
3	-.2495987	.3079433	-0.81	0.418	-.8531565	.3539592
4	-1.395812	.7107053	-1.96	0.050	-2.788769	-.0028549
5	-1.322271	1.193616	-1.11	0.268	-3.661716	1.017175
regcat2						
1	-.3187938	.8531722	-0.37	0.709	-1.990981	1.353393
2	.8849095	.5080861	1.74	0.082	-.1109209	1.88074
4	2.082465	.7657232	2.72	0.007	.5816756	3.583255
1.binge2	-.4362296	.6474849	-0.67	0.500	-1.705277	.8328175
povcat						
1	.6568208	1.074768	0.61	0.541	-1.449687	2.763328
2	-.1562196	1.070479	-0.15	0.884	-2.25432	1.94188
3	.1061492	1.014628	0.10	0.917	-1.882485	2.094783
4	-.3489774	.9604565	-0.36	0.716	-2.231438	1.533483
inscov						
1	.8689721	.5387951	1.61	0.107	-.1870469	1.924991
2	.889071	.5077466	1.75	0.080	-.106094	1.884236
_cons	4.240403	1.295451	3.27	0.001	1.701365	6.779441

Vita

RAHEEMA MUHAMMAD-KAH, MSPH

8518 Foster Ridge Ter
Moseley, VA 23120
Work: (804)-335-2493
Cell: (404)-323-1823
Email: rsmuham@gmail.com

SUMMARY

Experienced research professional with proven abilities in statistical planning, analysis and reporting for scientific research including clinical, product, and consumer research studies. Substantial experience in population modeling and epidemiological, reduced harm, risk assessment and product support studies. Results oriented with the ability to work individually and on a team in a fast-paced Fortune 200 environment.

PROFESSIONAL PROFILE

- 16 years' experience as a biostatistician working on multiple types of studies and analysis.
- Experienced in Population Modeling and Decision analysis
- Possess a broad knowledge of statistical modeling and analysis with emphasis on statistical applications and methodologies in both the Clinical/Public health research and industrial settings.
- Demonstrated strong analytical, qualitative and quantitative abilities.
- Possess a strong knowledge of public health issues and demonstrated experience in Epidemiology.
- Experienced in report writing, manuscript writing and poster presentations. Published work includes 18 manuscripts and 21 scientific abstracts.
- Proven record of team and individual performance, project/vendor management, communication and presenting to diverse audiences.

EDUCATION

School of Pharmacy, Virginia Commonwealth University, Richmond, VA
Doctor of Philosophy (PhD), Pharmacoeconomics and Health Outcomes
Expected May 2021

Rollins School of Public Health, Emory University, Atlanta, GA
Master of Science in Public Health, Biostatistics **May 2004**
Thesis: A Comparison of Missing Data Methods for Evaluating a Drug Treatment for Depression (Longitudinal data analysis). **Charles C. Shepard Award finalist** for the best Thesis in the Rollins School of Public Health.

Saint Mary's University, Halifax, Canada
Bachelor of Science, Mathematics, Magna cum laude **October 2000**

Dean's List throughout my undergraduate program

PROFESSIONAL EXPERIENCE

Principal Scientist II, Altria Client Services, Richmond, VA 2020-present
Population Science, Lead the Population Assessment and Statistics Team

Lead the development and execution of population assessment strategies, monitoring of surveillance systems and provide advance statistical expertise to determine the population impact related to product standards and reduced risk products.

Principal Scientist I, Altria Client Services, Richmond, VA 2017-present
Population Science, Lead the Population Health Impact Team

Provide population modeling and statistical expertise to determine the population health impact related to product standards and reduced risk products.

Some specific Accomplishments:

- Lead the development and validation of two population health models; a Cohort based and Agent-based model to assess the overall health impact of reduce risk products on the U.S. population.
- Designed experiments and conducted various statistical analyses using an array of statistical methodologies to support regulatory science.

Senior Research Scientist (Senior Biostatistician), Altria Client Services, Richmond, VA 2010-2017
Modeling & Simulation and Health Sciences

Provide statistical expertise to accelerate product development, facilitate cost savings and assess tobacco use behavior, exposure, and population health effects.

Some specific Accomplishments:

- Lead the efforts in the development of dynamic population health effect models.
- Performed various statistical analyses, designed experiments for various projects and model development.
- Investigate and analyze data from publicly available surveys.
- Applied innovative Missing Data techniques to evaluate the effect of menthol and nicotine under different sensations (Psychophysical data analysis).
- Analyze data to answer questions posed by Federal Regulatory Agencies.
- Evaluate the reliability and reproducibility of analytical lab data.

Research Scientist (Biostatistician), Altria Client Services, Richmond, VA 2005- 2010
Clinical Evaluation

Provided statistical expertise, lead and /or collaborate with Contract Research Organizations to facilitate the statistical analysis of clinical studies, including the review, revision and editing of study protocols, case report forms, data management plans, statistical analysis plans, statistical reports, clinical study reports and other study related documents.

Some specific Accomplishments:

- Performed multiple in-house analyses including linear and non-linear mixed models, general linear models, analysis of variance and covariance, regression analysis, logistic regression, etc. in lieu of hiring outside statistical firm, resulting in cost-savings to company.
- A lead statistician for a 5000 participant clinical study with over 700 variables, including demographics, Biomarkers of exposure, Biomarkers of Potential Harm, behavior questions, etc.
- Chosen as subject matter expert for extensive Survey Data Analysis (182 item Questionnaire) - including demographics, behavior questions, factor analysis and trend analysis, enabling the completion of the project and publication of results ahead of schedule.
- Designated statistical expert on numerous clinical research studies which evaluated potential reduced risk products and biomarkers of potential harm.
- Managed Contract Research Organizations, supervising seven (7) projects through inception to final report, writing statements of work, budgeting, and presenting results ahead of schedule and at a cost-savings to company.
- Statistical report writing, manuscript writing, poster presentations and PowerPoint presentations to various groups within the company and external audiences.
- Introduced statistical techniques (Bland-Altman method) which enabled clinical evaluation to determine that spot-urine collection was comparable to 24hr urine collection which would, when implemented, result in cost-savings to the company.
- Aided in the development of standard operating procedures (SOPs) for the Biostatistics and Data Management group and the Clinical Evaluation department.

Biostatistician – Kelly Scientific Services, 2004 – 2005,
Contractor, Clinical Evaluation

Took initiative on assignments and was offered full-time permanent position within 3 months of working in Clinical Evaluation group.

- Performed a wide range of statistical analyses on data from clinical studies-e.g. descriptive statistics, general linear models, linear and non-linear mixed models, regression analysis, analysis of variance and covariance, etc.
- Wrote statistical reports, tabulated results and produced graphs of findings with MS Excel and SAS graph.

Research Assistant –Rollins School of Public Health, Biostatistics Consulting Center, Emory University, Atlanta, GA 2003–2004

Provided statistical and data management support for various studies, including a Suicide Prevention study, Birth control survey, Depression study, and a Parkinson Disease study.

- Performed statistical analyses e.g. Paired analysis, logistic regressions, summary statistics, Friedman's test, Chi-square test, Fisher test etc.
- Performed statistical analyses for a Birth Control Survey involving 502 individuals; looking at 27 variables, including demographic and behavioral knowledge of women ranging from ages 25 to 45.
- Created tables and figures to facilitate communications among professional peers and to give a comprehensive summary of the findings of the studies.
- SAS programming created SAS datasets and MS Excel spreadsheets. Collected, entered and cleaned data.

COMPUTER SKILLS

Proficient with SAS software, JMP, SPLUS, SPSS, R, Minitab, nQuery Advisor, Epi Info, experience with MATLAB and MARS software.

SELECTED CONTINUING EDUCATION

- **Meta-Analysis: Combining Results of Multiple Studies**, August 2015
- **Advanced Decision Modeling for Health Economic Evaluations**, May 2015
- **Applications in using Large Databases**, May 2015
- **Cost-Effectiveness Analysis alongside Clinical Trials**, May 2015
- **Introduction to Modeling methods,** May 2014
- **Agent-Based Modeling for Economic Evaluations**, June 2014
- **SAS® Procedures for Analyzing Survey Data,** August 2013
- **Patient-Reported Outcomes- Item Response Theory,** May 2013
- **Bayesian Analysis- Overview and Applications,** May 2013
- **Applying Mixed methods to Establish content validity of Patient-Reported, Clinician-Reported and Observer-Reported Outcome Assessment Instruments,** May 2013
- **Successful Data Mining in Practice,** May 2013
- **Kepner-Tregoe Training**: Problem solving and Decision making, November, 2012
- **Data Mining, Knowledge Modeling and Causal Analysis with Bayesian Belief Networks,** October, 2011
- **Introduction to Mars: Predictive Modeling with Nonlinear Automated Regression Tools**, August, 2007
- **Advances in Data mining: Jerome Friedmans's TreeNet/MART and Leo Breiman's Random Forests**, August, 2007

ADDITIONAL INFORMATION

- **Member of the American Statistical Association** - Member since 2003
- **Member of International Society For Pharmacoeconomics and Outcomes Research** - Member since 2013
- **Research Abstract Reviewer**- ISPOR 20th and 21st Annual International Meeting

- **Marquis Who's Who in America 2012 (Sixty-sixth Edition)**
- **Finalist for the Charles C. Shepard's Award for Best Thesis 2004**, Emory University, Rollins School of Public Health

PUBLICATIONS

Wei, L., **Muhammad-Kah**, R.S., Hannel, T. *et al.* (2020). The impact of cigarette and e-cigarette use history on transition patterns: a longitudinal analysis of the population assessment of tobacco and health (PATH) study, 2013–2015. *Harm Reduct J* **17,** 45 https://doi.org/10.1186/s12954-020-00386-z

Peter N Lee, David Abrams, Annette Bachand, Gizelle Baker, Ryan Black, Oscar Camacho, Geoffrey Curtin, Smilja Djurdjevic, Andrew Hill, David Mendez, **Raheema S Muhammad-Kah**, Jose Luis Murillo, Raymond Niaura, et al. (2020). Estimating the Population Health Impact of Recently Introduced Modified Risk Tobacco Products: A Comparison of Different Approaches, *Nicotine & Tobacco Research*, https://doi.org/10.1093/ntr/ntaa102

Muhammad-Kah R, Pithawalla YB, Boone EL, Wei L, Jones M, Black R, Bryan T, Sarkar M (2019). A Computational Model for Assessing the Population Health Impact of Introducing a Modified Risk Claim on an Existing Smokeless Tobacco Product. *International Journal of Environmental Research and Public Health*, 16(7): 1264.

Saxena, K.; Liang, Q; **Muhammad-Kah, R**.; Sarkar, M. (2016) "Evaluating the relationship between biomarkers of potential harm and biomarkers of tobacco exposure among current, past, and nonsmokers: data from the National Health and Nutrition Examination Survey 2007-2012". *Biomarker*, 22(3):1-10.

Frost-Pineda, K.; **Muhammad-Kah, R**.; Rimmer, L.; Liang, Q. (2014) "Predictors, indicators, and validated measures of dependence in menthol smokers". *Journal of Addictive Diseases*, 33(2): 94-113.

Sarkar, M., **Muhammad-Kah, R.,** Liang, Q., Kapur, S., Feng, S., Roethig, H. (2013). Evaluation of Spot Urine as an Alternative to 24 hour Urine Collection for Determination of Biomarkers of Exposure to Cigarette Smoke in Adult Smokers. Environmental Toxicology and Pharmacology. 36(1):108-14.

Fisher MT, Bennett CB, Hayes A, Kargalioglu Y, Knox BL, Xu D, **Muhammad-Kah R,** Gaworski CL.(2012). Sources of and technical approaches for the abatement of tobacco specific nitrosamine formation in moist smokeless tobacco products. Food Chem Toxicol. Mar;50(3-4):942-8.

Muhammad-Kah R, Liang Q, Frost-Pineda K, Mendes PE, Roethig HJ, Sarkar M. (2011) Factors affecting exposure to nicotine and carbon monoxide in adult cigarette smokers. Regul Toxicol Pharmacol. Oct; 61(1):129-36.

Muhammad-Kah RS, Hayden AD, Liang Q, Frost-Pineda K, Sarkar M (2011). The relationship between nicotine dependence scores and biomarkers of exposure in adult cigarette smokers. Regulatory Toxicology and Pharmacology. Jun 1; 60(1):79-83.

Muhammad-Kah RS, Mendes P, Rimmer L, Liang Q, Serafin R, Roethig HJ, Sarkar M (2011) Exposure to Cigarette Smoke Constituents in a Population of Adult Cigarette Smokers in the U.S. Who Spontaneously Switched to Cigarettes with Lower or Higher Machine Measured 'Tar' Yield. Beitr. Tabakforsch. Int. 24(4):66-173.

Liu J, Liang Q, Frost-Pineda K, **Muhammad-Kah R**, Rimmer L, Roethig H J, Mendes P, Sarkar M (2011). Relationship between Biomarkers of Cigarette Smoke Exposure and Biomarkers of Inflammation, Oxidative Stress, and Platelet Activation in Adult Cigarette Smokers. Cancer Epidemiol Biomarkers Prev. August 20:1760-1769.

Scherer G, Urban M, Hagedorn HW, Serafin R, Feng S, Kapur S, **Muhammad-Kah R**, Jin Y, Sarkar M, Roethig HJ (2010). Determination of methyl-, 2-hydroxyethyl- and 2-cyanoethylmercapturic acids as biomarkers of exposure to alkylating agents in cigarette smoke. Journal of Chromatography B. October; 878(27):2520-8.

Roethig, HJ., Koval T., **Muhammad-Kah R.,** Jin Y., Mendes P., Unverdorben M. (2010). Short term effects of reduced exposure to cigarette smoke on white blood cells, platelets and red blood cells in adult cigarette smokers. Regulatory Toxicology and Pharmacology. July-August; 57(2-3):333-7.

Wang J, Roethig HJ, Appleton S, Werley M, **Muhammad-Kah R**, Mendes P.(2010). The effect of menthol containing cigarettes on adult smokers' exposure to nicotine and carbon monoxide. Regulatory Toxicology and Pharmacology. June; 57(1):24-30.

Munjal S, Koval T**, Muhammad R**, Jin Y, Demmel V, Roethig HJ, Mendes P, Unverdorben M. (2009). Heart rate variability increases with reductions in cigarette smoke exposure after 3 days. Journal of Cardiovascular Pharmacology and Therapeutics. September; 14(3):192-8.

Urban, M., Scherer, G., Kavvadias, D., Hagedorn, H.W., Feng, S., Serafin, R., Kapur, S., **Muhammad, R**., Jin, Y., Mendes, P. and Roethig H.J. (2009). Quantitation of *N'*-nitrosonornicotine (NNN) in smokers' urine by liquid chromatography-tandem mass spectrometry. Journal of Analytical Toxicology. June; 33(5):260-5.

Feng, S., Kapur, S., Sarkar, M., K., **Muhammad, R.,** Mendes, P., Newland, K., and Roethig, H. J. (2007). Respiratory retention of nicotine and urinary excretion of nicotine and its five major metabolites in adult male smokers. Toxicology Letters, September 10; Vol. 173 (2), 101–106.

Feng, S., Plunkett, S. E., Lam, K., Kapur, S., **Muhammad, R**., Jin, Y., Zimmerman, M., Mendes, P., Kinser, R., and Roethig, H. J. (2007). A new method for estimating the retention of selected smoke constituents in the respiratory tract of smokers during cigarette smoking. Inhalation Toxicology**.** Feb; 19(2):169-79**.**

SELECTED PRESENTATIONS

Muhammad-Kah, R.; Hannel, T.; Cheng, H.; Sarkar, M., Assessing the Potential Population Health Impact of a Market Authorization of an Oral Nicotine Pouch Product in the U.S. Poster presentation at Society for Research on Nicotine and Tobacco (SRNT) 26th Annual Meeting, San Antonio, Texas, February 24 - 27, 2021

Muhammad-Kah R., Hannel T., Jones M., Wei L., Black R., Pithawalla Y. B., Bryan T., Sarkar M., Evaluating the Population Impact from Introduction of a Novel Oral Tobacco-Derived Nicotine Product in the US. Poster presentation at Society for Research on Nicotine and Tobacco (SRNT) 24th Annual Meeting, San Francisco, CA, February 20 - 23, 2019

Muhammad-Kah, R., Pithawalla,Y.B., Hannel, T., Wei, L., Black, R., Bryan, T., Gogova, M., Assessing the potential population health impact of authorizing the marketing of e-cigarettes in the US. Poster presentation at 2018 CORESTA Congress, Kunming, China, October 22 - 26, 2018

Muhammad-Kah, R., Pithawalla,Y.B., Jones, M., Wei, L., Bryan, T., Black, R., Boone, E.L., Sarkar, M. Assessing the population health impact of a marketing authorization of a smokeless tobacco product with a proposed modified risk claim. Oral presentation at 2018 CORESTA Congress, Kunming, China, October 22 - 26, 2018

Muhammad-Kah, R., Pithawalla,Y.B., Jones, M., Wei, L., Bryan, T., Black, R., Boone, E.L., Sarkar, M. Estimating the population health impact of authorizing the marketing of a smokeless tobacco product with a proposed modified risk claim. Society for Research on Nicotine and Tobacco (SRNT) 24th Annual Meeting, Baltimore, MD, February 21 - 24, 2018

Muhammad-Kah, R., Hannel, T., Wei, L., Black, R., Bryan, T., Gogova, M., Pithawalla,Y.B. Estimating the potential population health impact of authorizing the marketing of e-cigarettes in the US. Society for Research on Nicotine and Tobacco (SRNT) 24th Annual Meeting, Baltimore, MD, February 21 - 24, 2018

Muhammad-Kah, R.; Pithawalla, Y.; Wei, L.; Hannel, T.; Gogova, M.; Boone, E. An agent based modeling approach for tobacco product risk assessments. Joint Statistical Meeting (JSM), Chicago, IL, July 30 - August 4, 2016

Boone, E.L.; **Muhammad-Kah, R.S**.; Pithawalla, Y.B.; Wei, L.; Frost-Pineda, K.; Gogova, M. Combining statistical and compartmental models for use in tobacco product risk assessments American Statistical Association Conference on Statistical Practice, February 18-20, 2016

Saxena,K; Liang,Q; **Muhammad-Kah, R**; Sarkar, M. Investigation of relationships between biomarkers of potential harm and cigarette smoking measures among current, past, and non-smokers based on national health and nutrition examination survey 2007-2012. International Society for Pharmacoeconomics and Outcomes Research (ISPOR), May 16-20, 2015

Sarkar M, Frost-Pineda K, **Muhammad-Kah R**, Rimmer L, Liang Q, Liu J. Characterization of selected biomarkers of potential harm in adult smokers and nonsmokers. Society for Research on Nicotine and Tobacco(SRNT),18th Annual Meeting March 16, 2012, Houston, Texas. http://www.srnt.org/conferences/abstracts/index.cfm

Muhammad-Kah R, Liang Q, Rimmer L, Frost-Pineda K. The effect of mentholated cigarettes on measures of nicotine dependence. Society for Research on Nicotine and Tobacco (SRNT) 16th Annual Meeting, February 27, 2010, Baltimore, MD. http://www.srnt.org/conferences/abstracts/index.cfm

Frost-Pineda K, **Muhammad-Kah R**, Liang Q, Rimmer L. The effect of mentholated cigarettes on biomarkers of potential harm. Society for Research on Nicotine and Tobacco (SRNT) 16th Annual Meeting, February 27, 2010, Baltimore, MD. http://www.srnt.org/conferences/abstracts/index.cfm

Munjal, S., **Muhammad, R.,** Jin, Y., Mendes, P., Demmel, V., McDonald, S., Unverdorben, M. "Heart Rate Variability in Adult Smokers Increases with Reductions in Cigarette Smoke Exposure and No Smoking After 3 Days Each" Poster presented at the 37th Annual Meeting of the American College of Clinical Pharmacology, Philadelphia, PA, September 14- 16, 2008.

Koval, T., **Muhammad, R.,** Munjal, S., Mendes, P., Jin, Y., and Roethig, H.J. "White blood cells decline significantly 3 days after switching from conventional cigarettes to reduced tobacco exposure or not-smoking" Poster presented at the 37th Annual Meeting of the American College of Clinical Pharmacology, Philadelphia, PA, September 14-16, 2008.

References available upon request