Summer 8-12-2016

Essays on Smoking, Drinking and Obesity: Evidence from a Randomized Experiment

Benjamin David Thomas Johannes Ukert

ABSTRACT

ESSAYS ON SMOKING, DRINKING, AND OBESITY: EVIDENCE FROM A RANDOMIZED EXPERIMENT

BY

BENJAMIN DAVID THOMAS JOHANNES UKERT

August 2016

Committee Chair: Dr. Rusty Tchernis

Major Department: Economics

This dissertation consists of three chapters analyzing risky health behaviors utilizing data from the Lung Health Study (LHS), a randomized smoking cessation program. The first two chapters of this dissertation analyze the effects of smoking on alcohol consumption and BMI, respectively. The third chapter studies whether and how much the objective smoking information, which is defined by clinicians, may be misreported.

The first chapter examines the effect of smoking on alcoholic beverage consumption. The epidemiology literature suggests that both behaviors affect similar brain regions and are commonly consumed together. So far, the economics literature has presented inconclusive causal evidence on the relationship. Building on the theory of rational addiction, I estimate the relationship between smoking and alcohol consumption using several different smoking measures. I report four salient findings. First, self-reported and clinically verified smoking variables suggest that quitting smoking lowers alcoholic beverages consumption by 11.5%. Second, cigarette consumption dating back 12 months affects alcohol consumption, and those with the highest past 12 months average cigarette consumption see the largest increase in alcohol consumption. Third, I find that the length of quitting affects future alcohol consumption as well. Continuously abstaining from smoking for 12 months reduces alcoholic beverage consumption by 27.5% per week. Fourth, non-smoking for 12 months also reduces the probability of drinking any alcoholic beverages by 31%.

The second chapter aims to identify the causal effect of smoking on body mass index (BMI). Since nicotine is a metabolic stimulant and appetite suppressant, quitting or reducing smoking could

lead to weight gain. Using randomized treatment assignment to instrument for smoking, we estimate that quitting smoking leads to an average long-run weight gain of 1.8-1.9 BMI units, or 11-12 pounds at the average height. These results imply that the drop in smoking in recent decades explains 14% of the concurrent rise in obesity. Semi-parametric models provide evidence of a diminishing marginal effect of smoking on BMI, while subsample regressions show that the impact is largest for younger individuals, females, those with no college degree, and those with healthy baseline BMI levels.

The third chapter analyzes and compares self-reported and clinically verified smoking information. Descriptive statistics show that about 8% of clinically verified smokers self-report that they do not smoke (under-report participation), and that smoking cessation treatment group participants misreport smoking participation 2 to 1 relative to control group participants. In our first methodological approach we regard the objectively verified smoking measure as the gold standard. We estimate linear probability models and find that being male and married increases the probability of misreporting by 10 percentage points. Additionally, older participants are more likely to misreport smoking status, while those using nicotine gum and with a higher BMI are less likely to misreport. However, all variables can only explain a small fraction of the variation that explains misreporting. Our second methodological approach takes an agnostic view on whether the clinically verified smoking information is accurate. We utilize BMI, Carbon Monoxide (CO), and Cotinine level information to inform whether a person is a smoker. We estimate a Bayesian mixture model to account for the heterogeneity in BMI, CO and Cotinine levels after a substantial decrease in post treatment smoking participation. All of our models show that smokers are more likely assigned to the low BMI, high CO and high Cotinine level distributions. Among those classified as misreporters, we find that 30% have a very high probability of being part of the non-smoking distributions. As a result, we believe that objectively- verified smoking measure may not be better than the self-reported measure.

ESSAYS ON SMOKING, DRINKING, AND OBESITY: EVIDENCE FROM A RANDOMIZED EXPERIMENT

BY

BENJAMIN DAVID THOMAS JOHANNES UKERT

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree
of
Doctor of Philosophy
in the
Andrew Young School of Policy Studies
of
Georgia State University

GEORGIA STATE UNIVERSITY

2016

ACCEPTANCE

This dissertation was prepared under the direction of the candidate's Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics in the Andrew Young School of Policy Studies of Georgia State University.

Dissertation Chair:	Dr. Rusty Tchernis
Committee:	Dr. Charles Courtemanche
	Dr. Ian McCarthy
	Dr. Tom Mroz

Electronic Version Approved:

Mary Beth Walker, Dean
Andrew Young School of Policy Studies
Georgia State University
August, 2016

ACKNOWLEDGMENTS

My academic achievements would have been impossible without the constant support of family and friends. Specifically, the support and encouragement of my parents played a pivotal part in me pursuing graduate studies.

I am also very grateful to the community at the Andrew Young School for all of its support. First and foremost I want to give special gratitude to my adviser Rusty Tchernis. He recruited me after taking the Bayesian Econometrics class and without his constant support and sharp mind my dissertation would not have been possible. I am also grateful to him for consistently challenging me to improve my writing and economic thought. Your kindness, open ear, and availability for a fun chat during stressful times will be missed.

I am also grateful to Charles Courtemanche for his advice and support. I have had the opportunity to interact with him during class and in research. He is exceptionally great at writing clear and concise papers, and I can only hope to reach your writing skills in some distant future. His health economics class and STATA class reveal how to think about research and connect theory to application. His wit and intuition are unique and he personifies everything that make him great applied economist.

I am thankful to Tom Mroz for his constant support, his advice, and his incisive questions to improve my research and my thinking about economics and econometrics. Tom Mroz is one of few economists that I see as a complete economist. He has an unprecedented insight into econometrics that allows him to provide help in almost any circumstance related to applied research.

I also owe thanks to Ian McCarthy, I value your calmness and your helpful comments.

Additionally, I want to thank Bess Blyler and her colleagues who make everything work so smoothly around the department.

Finally, I am grateful to my wife Elena for her continuous support, her constant admiration and love.

Table of Contents

List of Tables

List of Figures

Introduction

During the early 20th century, smoking was a conventional habit across social classes in the United States. Tobacco consumption was considered a regular good. For example, R.J.Reynolds advertised with this slogan: "More doctors smoke Camels than any other cigarette" (Gardner & Brandt 2006). As a consequence, by the mid 1960s smoking participation peaked at 42% among adults. However, smoking participation has decreased every year since 1965. What caused the change?

One possible explanation is that smoking was not associated with health risks until the 60s. The 1964 surgeon general report is the first official statement summarizing the health risks associated with smoking. Today, smoking is regarded as a health risk. According to the 2014 surgeon general report, smoking accounts for 90% of all lung cancer deaths and 80% of all chronic obstructive pulmonary disease (COPD) deaths. Moreover, the risk of dying from smoking increased within the last 50 years among smokers (Thun et al. 2013).

The recognition of health risks associated with smoking prompted federal, state and local governments to impose smoking restrictions. Some of the steps taken by policymakers to reduce smoking include increasing cigarette taxes, introducing workplace smoking bans and requiring tobacco free zones. Subsequently smoking prevalence decreased by roughly 20 percentage points from the mid 70s.

Unfortunately, imposing smoking restrictions can result in severe unintended consequences. Evidence suggests that higher cigarette prices lead to a higher BMI and obesity rate (Chou et al. 2004). Therefore, the overall health benefit from quitting smoking can be overstated if it does not include the health risks associated with a potentially higher BMI. However, recent evidence (Courtemanche 2009, Wehby et al. 2012) suggests that higher cigarette costs lower BMI. Thus, the precise impact of cigarette cost on BMI remains controversial.

In addition, smoking restrictions may also affect alcoholic beverage consumption. Cigarette consumption and alcohol consumption are highly correlated (Bobo et al. 1987, Grant et al. 2004), but the literature is inconclusive on the causal relationship. If alcohol consumption and cigarette consumption are complements, smoking restrictions would decrease alcohol consumption as well.

Therefore, the overall health benefits of quitting smoking are understated, as the benefits of lower alcohol consumption are ignored.

My dissertation uses a randomized smoking cessation study, Lung Health Study (LHS), to investigate the effect of smoking on BMI and alcoholic beverage consumption. My first chapter analyzes the causal relationship between smoking and alcoholic beverage consumption. I provide evidence that smoking and alcohol consumption are positively correlated. My paper differs from the previous literature in several ways. First, I estimate the relationship along the intensive and extensive margin of smoking. Second, the LHS, a five-year panel study, allows me to look at the effect of smoking in the short- and long-run. Third, I estimate a causal relationship at the individual level between cigarette consumption and alcoholic beverage consumption over-time. My main results suggest that quitting smoking lowers the alcoholic beverage consumption by half a drink per week. The effect is stronger the higher the cigarette consumption in the past, and the effects persist over five years.

My second chapter evaluates the relationship between smoking and BMI. The chapter is the result of co-authored work with Dr. Charles Courtemanche and Dr. Rusty Tchernis. We estimate the effects along the intensive and extensive margin of smoking. We further provide evidence that Eisenberg and Quinn (2006), who apply the same data, overestimate the effect of quitting smoking on BMI. Lastly, we present evidence that quitting smoking not only reduces BMI, but that the effect of smoking on BMI depends mostly on the intensity of smoking.

My third chapter has compares self-reported to objectively clinically verified smoking information provided in the LHS, by generating a misreporting measure captured difference between both smoking variables. The first methodological approach regard the objectively verified smoking measure as the gold standard, and we estimate linear probability models and find that being male and married increases the probability of misreporting. However, all variables can only explain a small fraction of the variation that explains misreporting. The second methodological approach takes an agnostic view on whether the clinically verified smoking information is accurate. We utilize BMI, Carbon Monoxide (CO), and Cotinine level information to inform whether a person is a

smoker. All of our models show that smokers have a lower BMI, high CO and high Cotinine level. However, all models also show that many misreporters are similar to non-smokers in characteristics affected by smoking, allowing us to conclude that the objectively verified smoking measure may not clearly identify smokers from-non-smokers.

I The Effect of Smoking on Alcohol Consumption

Introduction

Smoking cigarettes is the leading cause of preventable deaths in the U.S., causing roughly 480,000 deaths per year (Center for Disease Control (CDC) 2012). It is well known that smoking causes diseases such as lung cancer, stroke, coronary heart disease, and chronic obstructive pulmonary disease (COPD).[1] Despite a 50% decrease in smoking prevalence since the 1960s, an estimated 42 million U.S. adults smoke, generating roughly $289 billion per year in smoking related costs (CDC news release 2014). Similarly, short- and long-term risks of excessive alcohol use, defined as 15 or more drinks per week or 5 or more drinks per session for men, the third largest modifiable risk factor in the U.S., include injuries, violence, dementia, stroke, and liver disease. Currently, economic costs of excessive alcohol consumption account for about $249 billion per year (Sacks et al. 2015).

Given the severe consequences of both risky behaviors, understanding how they influence each other is an important policy question. A large health literature suggests that smoking and drinking are highly correlated (Bobo et al. 1987, Bobo et al 2000, Bien et al. 1990). Increasing cigarette taxes or expanding smoking ban regulation may induce smoking cessation, but could increase drinking if they are substitutes. Therefore, the net health benefits of quitting smoking would be overstated, as the costs of higher alcohol consumption are ignored. On the other hand, if quitting smoking reduces drinking, the benefits of quitting smoking would be understated.

In this paper I exploit a randomized smoking cessation study, the Lung Health Study (LHS), to investigate the relationship between cigarette smoking and alcohol consumption. My empirical strategy uses the randomized intervention assignment as an instrumental variable for smoking. This allows me to estimate a causal effect and analyze the effect of smoking on alcohol consumption along the intensive and extensive margin. Additionally, I also analyze the effect by gender, marital status, and among excessive alcohol users.

[1] Surgeon General Report 2015.

This paper contributes to two strands of economic literature: First, it provides new evidence on the causal relationship between smoking and drinking. The typical approach to identify complementarity or substitutability is to regress consumption levels of one good on the price of the other good. However, the current literature report inconclusive results on the relationship between smoking and drinking.[2] Cameron and Williams (2001) and Bask and Melkersson (2004) find complementarity between smoking and drinking in both alcohol and cigarette demand equations, others, such as Goel and Morey (1995) and Picone et al. (2004), find that smoking and drinking are substitutes. Additionally, Decker and Schwartz (2000) find complementarity in a cigarette demand equation and substitutability in an alcohol demand equation.[3] I find that quitting smoking reduces alcohol consumption, implying both goods can be complements in consumption.

Second, I contribute to a growing literature on habit formation that tests how past cigarette consumption, also known as the smoking habit stock, affects current consumption of alcohol or other addictive goods (Goel and Morey 1995, Pacula 1997, 1998, Kenkel et al. 2001). Becker and Murphy's (1988) rational addiction model provides a formal theoretical framework that Bask and Melkersson (2004) extend to two addictive goods and then show that, depending on the interaction between both addictive goods, quitting one addictive good can increase or decrease the overall consumption of another addictive good in the future. To date, there is mixed and limited empirical evidence on the effect of past smoking on current alcohol consumption (see Bask & Melkersson 2004, Picone et al. 2004, and Pierani et al. 2009). I provide evidence that a smoking habit stock, measured by up to 5 years of historic smoking decisions, increases current alcohol consumption. I also show that the effect of the smoking habit stock on alcohol consumption is larger among heavy smokers. Lastly, I show that those who permanently abstain from smoking longer consume

[2]For a general overview on the literature see Pierani et al. (2009).

[3]The application of different data and identification strategies may be an explanation for the empirical different results among similar demand equations. For example, Bask and Melkersson (2004) use aggregate time series sales volume data from Sweden and estimate demand equations based on the rational addiction model. Picone et al. (2004) rely on individual level data from the Health and Retirement Survey (HRS) and apply an instrumental variable approach, using prices and lagged consumption as instruments for current consumption. Whereas Decker and Schwarz (2000) concentrate on the estimation of the relationship between cigarette and alcohol prices on alcohol and cigarette consumption levels, respectively. Another reason for the diverse results could be that the price variation is not exogenous.

considerable less alcohol.

Overall, I conclude that cigarettes and alcohol are positively correlated. Quitting smoking leads to a reduction of alcoholic beverage consumption by 11.5% per week.[4] Restricting the sample to unmarried smokers shows that quitters reduce alcohol consumption by 16%. Similarly, when looking at the sample of excessive alcohol users, defined as consuming 15 or more drinks per week or 5 or more drinks per session, quitting smoking reduces alcohol consumption by roughly 11% per week.[5] Excessive male drinkers reduce alcohol consumption by 14% drinks per week after quitting smoking, while excessive female drinkers do not reduce alcohol consumption after quitting smoking.[6]

Results also suggest that the habit stock of cigarettes affects current alcohol consumption significantly. Both the intensity and length of smoking significantly affect current alcohol consumption. Consuming on average 20 cigarettes per day in the past year increases current alcohol consumption by 28% from the mean baseline alcoholic beverage consumption level. Similarly, those who quit smoking 12 months ago, whether successful or not, decrease current alcoholic beverage consumption by 20% per week while also decreasing the probability of drinking any alcohol by 20%. Abstaining from smoking continuously for 12 months leads to 27.5% fewer consumed drinks per week and decreases the probability of any drinking by 30%. The effects of permanently quitting for 12 months are particularly pronounced for women who reduce alcohol consumption by roughly 53% vs. 23% for men.

The rest of the paper is organized as follows. The next section provides a summary of the conceptual framework. Section 3 discusses the data. Section 4 outlines the estimation strategy. Section 5 presents the results. Section 6 concludes.

[4] 11.5% equal about 0.5 drinks per week on an average pre-treatment consumption of 4.35 drinks per week.

[5] 11% translates into 1.6 drinks per week.

[6] On average Excessive male drinkers consume 15 drinks per week pre-treatment. Males reduce alcohol consumption by 2.1 drinks per week.

Conceptual Framework

This section discusses a simple conceptual framework on the relationship between smoking and alcohol consumption. We formulate the framework by taking advantage of the medical literature documenting in many cases a positive correlation between smoking and alcohol consumption (Carmelli et al. 1993, Enoch et al. 2001, Swan et al. 1997). The Supplement outlines in detail the theoretical implications for a model of addiction with two addictive goods. The outlined framework considers a utility maximizing agent, and builds its foundation on the rational addiction model by Becker and Murphy (1988), and Bask and Melkersson (2004). Consider an individual making two simultaneous decisions: whether to consume alcohol and cigarettes, and the quantity of each good. For simplification the individual does not maximize his lifetime utility, i.e. the individual is myopic, and we assume complementarity between smoking and drinking. Opposite conclusions will be drawn for the case of substitutability.

The Supplement presents two conclusions. First, the choice of quitting smoking immediately decreases the consumption of alcoholic beverages. Complementary assumes that consuming both at the same time also raises the marginal utility of both good. Thus, choosing not to consume cigarettes will reduce the marginal utility of consuming alcoholic beverages. Formally, marginal utility must equal marginal cost, and a reduction in the marginal utility of alcoholic beverages requires a decrease in consumption to satisfy the equality, ceteris paribus. Therefore, quitting smoking will reduce alcohol consumption.

Second, quitting smoking will reduce alcohol consumption over-time. The habit stock, a function of past smoking and beverage consumption, increases the marginal utility of smoking and also increases the marginal utility of alcoholic beverage consumption. The epidemiology literature presents evidence supporting the aforementioned response to a change in the habit stock, as smoking increases the nicotine intake affects the limbic system. Barrett et al. (2006) show that nicotine increases dopamine output and higher dopamine levels lead to an increased craving for alcoholic beverage consumption. Similarly, the consumption of cigarettes increases a person's familiarity with its psychological and physical benefits and contributes to habit formation, resulting in an in-

creased desire to consume additional cigarettes and can spill over to alcohol consumption as well. The influence of past physical and psychological experience on today's behavior is summarized in the form of the habit stock.

Similarly, a large literature on youth addictive behavior shows that the consumption of legal addictive goods leads to future consumption of other legal and illegal addictive goods. The experience of ones addictive good consumption generates a benefit that can increase over time, and affect the enjoyment of other addictive goods (See for example Pacula 1997 & 1998, Kenkel et al. 2001). Put differently, today's consumption depends on the habit stock. The longer someone smokes, the higher the habit stock, while quitting smoking depreciates the habit stock over-time. The level of the habit stock directly affects the marginal utility of alcohol consumption.[7]

In summary, the framework outlines the relationship between past and current smoking behavior on current alcohol consumption. Quitting smoking will cause two distinct changes in the marginal utility of alcoholic beverage consumption. First, quitting smoking decreases immediately the marginal utility of alcohol consumption, leading to lower alcohol demand. Second, quitting smoking decreases the marginal utility of alcohol consumption over time through a depreciation of the habit stock. Again, leading to a reduction in alcoholic beverage consumption.

Data

I utilize data from the Lung Health Study (LHS), a randomized smoking cessation trial implemented to measure the effect of smoking cessation on lung function. O'Hara et al. (1993, 1998) present a comprehensive analysis of the LHS recruitment and implementation. I briefly summarize the most important aspects relevant to the paper. Recruitment took place between 1986 and 1989 within a wide region of the 10 hospitals (9 in U.S. and 1 in Canada) participating in the study.[8] Participation

[7] In a broader economic context, the habit stock of smoking has similar properties to the concept of human capital stock in labor economics (Ben-Porath 1967). Education, training, and work experience increase the total human capital stock, which improves productivity and consequently wages. Similarly, the human capital stock depreciates over time, mostly due to age. Depreciation can be stopped or at least minimized by consistent investments into education or training. The habit stock of cigarettes can be thought of in a similar fashion, replacing education, training and work with cigarette consumption and productivity with smoking appreciation.

[8] Participants should live no more than 75 miles away from the hospital and should have no intentions to move away from the area. This requirement was implemented to minimize attrition. The list of hospitals participating can

required applicants to be 35 to 59 years old, show signs of mild lung function impairment, have no history of certain medications, consume less than 25 alcoholic beverages per week.[9], have no severe illnesses, and have no chronic medical conditions.[10] After 3 screening interviews the final study samples included 5887 participants.[11]

After recruitment, participants were randomly assigned into three different groups, two treatment groups that include smoking cessation treatment with an inhaled placebo (SI-P) and smoking cessation treatment with an inhaled bronchodilator (SI-A) and a control group receiving no treatment (UC). The only difference in treatment between the SI-P and SI-A group were that the SI-A group received an inhaler with ipratropium bromide to treat early COPD symptoms, while the SI-P group received a placebo inhaler. Both treat treatment groups, SI-A and SI-P, also received nicotine gum prescriptions, an intensive 12 session quit week accompanied by frequent contact with support personal, and invitations to bring a spouse or relative to the meetings. The usual care group received no treatment. Most of the treatment was completed within the first 4 months of the study and there was no differential effect on smoking cessation for the SI-P and SI-A treatment group members.

All participants were interviewed individually at a medical clinic near the residence of the participant to collect information about average alcoholic beverage consumption, weight, age, gender, employment status, smoking behavior, type of tobacco use and family smoking habits. Concerns about measurement error in cigarette consumption as well as smoking status can be minimized given that smoking status is annually verified by clinicians through carbon monoxide and cotinine level tests at each of the annual follow-up visits. For the empirical analysis I use the objectively verified binary smoking measure, the continuous carbon monoxide measure, and monthly information on self-reported cigarette consumption per day. The results section presents mostly regression outputs for the annual verified smoking status and the monthly self-reported smoking variables. In

be found online http://www.biostat.umn.edu/lhs/centers.html

[9]The alcohol limit was only implemented for the first third of the recruitment process and deleted afterwards.

[10]History of certain medication use includes medicine for tuberculosis, theophylline or other xanthines, beta-blockers, insulin, any corticosteriods, antipsychotic drugs, nitroglyercine, digitalis, anticoagulants and antiarrhythmics. Alcoholics were also excluded.

[11]The clinical trial ended in 1994.

any case, I draw similar empirical conclusions when I use the carbon monoxide measure.

The sample includes 6 years of data: one pre-treatment year and five post-treatment years. Table 1 presents descriptive statistics at the randomization meeting. At that time, the average alcoholic beverage consumption was close to 4.3 drinks per week. Among the portion of the sample that drinks alcohol, the average alcoholic beverage consumption was 6.3 drinks per week. The average age of smokers was 48 and the average cigarette consumption per day was roughly 30 cigarettes per day. Additionally, The survey does not collect information regarding income, but a high average education level, at 13.50 completed years, allows me to assume that the average income level should be higher than in the general population as well.[12] The data does not have information on race, but O'Hara et al. (1998) mention that the sample is 97% white.

Table 1: Summary Statistics at Randomized Treatment Assignment Meeting

Variable	SI-A	SI-P	UC
Drinks per week	4.30	4.39	4.36
	(5.57)	(5.56)	(5.47)
Age	48.41	48.55	48.43
	(6.84)	(6.83)	(6.84)
Marital Status	0.65	0.67	0.65
	(0.47)	(0.47)	(0.47)
Education	13.57	13.56	13.68
	(2.83)	(2.84)	(2.79)
Male	0.61	0.64	0.64
	(0.48)	(0.48)	(0.48)
Cigarettes per Day	29.59	29.49	29.53
	(14.08)	(13.64)	(14.11)
Carbon Monoxide	25.97	26.70	25.98
	(13.47)	(12.67)	(12.66)
Observations	1961	1962	1964

Standard deviation in parenthesis. SI-A, SI-P and UC refer to the three treatment groups. Data collected at time of randomized assignment to intervention groups. Cigarettes per day represents the self-reported average daily cigarette consumption of each participant. Education is measured in years of highest grade completed.

[12]The 1990 NHIS suggest an average level 12.50 years of completed education.

The summary characteristics in Table 1 are balanced across treatment groups and signify a successful randomization. T-tests and F-tests confirm the hypothesis that the variable means are not different across treatment group. Besides the reported variables presented in Table 1,participants are also asked about their average levels of cigarette consumption per day in every of the past 12 months. Appendix Table A.1 presents the monthly smoking status and average cigarette consumption per day surveyed at the first annual follow-up.

The LHS has two benefits relative to observational data. First, I have an objectively verified smoking variable, reducing potentially the problem of mismeasurement relative to self-reported information. Second, the data is randomized, allowing for causal inference. Due to the efforts to retain contact with the participants, attrition rates were low. By the fifth annual follow-up visit at the local clinic, 95 percent of men and 96 percent of women attended.[13]

Some limitations remain that are common across randomized control trials. The data consists of smokers who responded to recruitment efforts to participate in a smoking cessation study. The self-selection indicates that the sample may not be representative of the general smoking population and may result in estimates that have questionable external validity. With that said, the estimates may in fact not be valid because the latest Surgeon General Report documents that almost 70% of current smokers indicate that they would be willing to quit smoking. Thus, generally smokers are willing to quit, but may be unable to quit without any support group or intervention (2014 Surgeon General Report). Therefore, any inference may be at least attributable to 70% of the smoking population.

Estimation Strategy

The empirical strategy estimates the effect of smoking on alcoholic beverage consumption, conducts sub-sample analyses, and tests for differences in short- and long run effects. I identify the relationship between smoking and drinking with two different models. First, I estimate whether smoking and alcohol are complements in consumption. Second, I test how smoking in the past, the habit stock of smoking, affects current alcohol consumption.

[13] See O'Hara et al. (1998) for more details on attrition.

The first model estimates cross-sectional regression between current smoking status and current alcohol consumption at the first annual follow-up visit. The cross-sectional data from the first follow-up visit should identify the short-term effect of quitting smoking on drinking. I begin my analysis with OLS:

$$Drinks_{it} = \gamma_0 + \gamma_1 Smoke_{it} + X_i\gamma + \mu_{it} \tag{1}$$

Where the dependent variable, $Drinks_{it}$, represents alcoholic beverage consumption per week for person i at the first follow-up visit, $Smoke_{it}$ is equal to an indicator for whether a person is a current smoker and zero otherwise, a continuous cigarettes per day variable, or a measure of carbon monoxide level, and X_i is a vector of demographic characteristics at baseline including age, gender, education level, and marital status.

To generate exogenous variation in smoking, I utilize an instrumental variables approach, or a two stage least square (2SLS) approach. The first follow-up visit allows me to apply the randomized smoking cessation treatment as an instrumental variable, because treatment was received earlier in the year. The approach allows me to analyze the causal relationship between both choices.[14] The appropriate first and second stages are presented here:

$$Smoke_{it} = \beta_0 + Treatment_i\beta + X_i\beta + \epsilon_{it}$$

Second Stage:

$$Drinks_{it} = \gamma_0 + \gamma_1 \hat{Smoke}_{it} + X_i\gamma + \eta_{it}$$

Where $Treatment_i$ represents a vector of two dummy variables equal to one for participants in each of the smoking cessation programs and zero otherwise. The underlying assumption on the treatment suggests that it is uncorrelated with the error term η_{it}, and highly correlated with smoking status.

[14] I am unable to merge cigarette price data, because I have no state identifiers. This is the best I can do to understand the contemporaneous relationship between cigarettes and alcohol.

The random assignment to treatment allows for such a strong assumption. It is conceivable that the treatment has a direct effect on drinking and violating the assumption of an uncorrelated error term in the second stage. I test and reject that the treatment has a direct effect on drinking in the results section. Therefore, the instruments should allow a2SLS estimation strategy to recover coefficients that have a causal interpretation.

The second identification strategy estimates cross-sectional regressions between historic smoking measures and current alcohol consumption. Such regressions should identify the habit stock effect of smoking on alcohol consumption. Based on the addiction theory and the literature, I include the previous period's smoking status.[15] Moreover, the benefit of the LHS is that it includes detailed monthly self-reported historic smoking information. Thus, it allows me to estimate the effect of historic smoking measures that are a function of monthly information on alcohol consumption. Those measures include dummy variables measuring the average intensity of cigarette consumption in the last year, and also cumulative 12 months cigarette consumption variables with and without some form of depreciation over-time. The general setup of the regression follows equation (1) with the only difference being the inclusion of lagged smoking measures:

$$Drinks_{it} = \gamma_0 + \gamma_1 Smoke_{it-1} + \gamma_2 X_{it} + \mu_{it} \tag{2}$$

All variables are exactly the same as defined in equation (11), except that smoking is some measure of historic 12 months cigarette consumption. Where applicable, I use a 2SLS estimation strategy to control for endogeneity. The instrument should yield be valid because treatment group assignment was random and occurred at least 13 months prior to the information collected on cigarette consumption and alcoholic beverages. In both cases, I use the random assignment to smoking cessation groups as a instrument for current and lagged smoking status.

I also estimate the effects of smoking on the probability of drinking, where I replace the left-hand side variable in (1)-(2) with a binary indicator equal to one if the person drinks alcohol and zero otherwise. As a robustness check, I also estimate the relationship between smoking and drinking

[15] See the Supplement that explains the theoretical background of the habit stock effect.

with Negative Binomial (NB) models, because the dependent variable is a count variable with an over-dispersed variance. Regular NB estimates will result in biased coefficients. I apply a control function approach to account for endogeneity. NB control function approach estimates are abbreviated with CF-NB. The approach includes the predicted residuals from the IV's first stage as an independent variable in the second stage. This is a commonly used approach for non-linear correction for endogeneity.[16] Lastly, I also estimate equation (11) with a log-transformed dependent variable, where I generate log-drinks per week variable by adding a constant of one to each observation before taking logs.

The results section presents estimates from OLS, 2SLS, NB, NB-CF, log, and log-2SLS models. The 2SLS and control function estimates have a 'causal' interpretation under the assumption that the instrumental variables have no direct effect on drinking. They estimate a local average treatment effect (LATE). In other words, the causal effect estimates how a shock in cigarette consumption induced by the intervention changes the level of alcohol consumption. Equation (1) estimates the instantaneous effect of quitting smoking on alcohol consumption for those who received the treatment earlier in the year. Therefore, equation (1) implies a temporal relationship by definition of the LATE, which I explore deeper with equation (2) where I estimate how past smoking decisions affect current alcohol consumption. The latter equation is a first step at getting at the effect of the smoking habit stock, because past quitters are less likely to smoke today.

Results

The Impact of Smoking on Alcohol Consumption

O'Hara et al. (1993) show that the smoking cessation treatment was effective in reducing smoking. I verify their results by presenting estimates for the first stage of the 2SLS approach. Table 2 presents marginal effects for the treatment variables and shows that the treatment was effective in reducing smoking. Stock et al. (2002) provide evidence that large F-statistics indicate a strong

[16]See Wooldridge, J. Econometric Analysis of Cross Section and Panel Data, 2nd Edition or Cameron, C. and Travedi, P. Microeconmetrics Methods and Application

instrument. In this case F-statistics are well above 10 in both regressions with and without demographic variables.

Table 2: First Stage Estimates between Smoking Cessation Treatment and Smoking at the First Annual Follow-up Visit

First Stage IV Regressions

	Verified Smoking Status		Self-Reported Smoking Status	
Dep. Var.	Smoke	Carbon Monoxide	Smoke	Cig. per day
Treatment (SI-A)	-0.272***	-8.014***	-0.322***	-11.731***
	(0.013)	(0.464)	(0.014)	(0.476)
Treatment (SI-P)	-0.267***	-7.782***	-0.311***	-11.108***
	(0.013)	(0.470)	(0.014)	(0.482)
Demographics	Yes	Yes	Yes	Yes
F- Stat	**115.21**	**95.08**	**144.86**	**125.98**
Observations	5,584	5,269	5,584	5,584
R-squared	0.0867	0.092	0.109	0.134

Robust standard errors are in parenthesis. ***,**, and * indicate significance at the 1, 5, and 10 percent level. All first stages include demographic characteristics. Every row entry represents a different regression on the dependent smoking status in year 1.

A concern remains regarding the validity of estimated coefficients from the 2SLS models. Specifically, Courtemanche et al. (2015) show that including a dummy variable for smoking status is invalid in a 2SLS setup under the condition that the treatment affects smoking intensity, but does not lead to smoking cessation. In other words, I have to test whether the instrument affects drinking among those who do not change their smoking habits. A significant association suggests that those who reduce cigarette consumption, but do not quit smoking, reduce alcohol consumption. This will lead to a biased estimator, because I would be unable to capture the reduction in the intensive margin of smoking with the objectively verified smoking measure. I test this hypothesis by running cross-sectional regressions and limiting the sample to subsets of smokers. First, I generate five post treatment sub-samples limiting the sample to objectively verified smokers in each year. Second, in separate regressions, I limit the sample to individuals satisfying a non-sustained quitter measure. The non-sustained quitter measure, which is only available for follow-up years 2 through 5, only includes individuals who smoke in at least one of the follow-up waves. All sub-

samples have very similar demographic characteristics to those presented in Table 1, with the only difference that those participants did not quit smoking permanently. All tests should identify if the treatment has an effect on those who temporarily change smoking status. Across both restrictions and all regressions, I find no significant effects of the treatment on drinking.[17] I conclude that there is no evidence that the exclusion restriction is violated in the 2SLS model and that the estimator should not be biased the objectively verified smoking dummy measure.

Next, I concentrate on the estimates on the effect of smoking on alcohol consumption from equation (1). Table 3 provides marginal effects for OLS, 2SLS, NB, control function approach NB, log, and log-2SLS estimates. Each coefficient in Table 3 represents estimates from a different regression of equation (1). For comparison purposes, Table 3 presents the effect of smoking on alcohol consumption for the objectively verified and self-reported smoking variables. Additionally, I also generate a quasi objectively-verified cigarette consumption measure and estimates yield similar average effects of quitting smoking on drinking.[18]

The OLS and NB estimates suggest that those who quit smoking at the first annual visit consume less alcohol than smokers. All results are significant at least at the 5% percent level. The OLS results for the binary smoking variable reveal that smokers consume roughly half a drink more than quitters, implying an increase of 11.5% from the mean alcoholic beverage consumption level at randomization. Similarly, OLS results for the cigarettes per day variable present the same average effect for an average smoker consuming 20 cigarettes per day. However, all 2SLS estimates are insignificant, but present similar point estimates to the OLS results. There is good reason for the insignificant but similar 2SLS estimates. In any 2SLS estimation standard errors increase from OLS estimates. There is also a good reason why the coefficient does not change. The highly effective smoking cessation program generates many non-smokers in the treatment group, but only few non-smokers in the control group. Therefore, the variation in the OLS is very similar to the variation in

[17]The reduced form results are not included. Results will be made available upon request.

[18]I generate the quasi-objectively verified cigarette per day variable measure by comparing the verified smoking status variable to the self-reported cigarette consumption per day. If the self-reported cigarette per day variable and objective smoking dummy variable do not suggest the same smoking status, i.e. dummy variable indicates smoker (equal to one), but self reported cigarette consumption measures equals zero. In all cases of mismatch I drop those observations, leaving me with a quasi-objective cigarette measure.

the 2SLS estimates. In other words, there is very little endogeneity.

Table 3: The Effect of Smoking on weekly Alcoholic Beverage Consumption

Obj. Verified Smoking Status:			Negative Binomial		Log Drinks per Week	
	OLS	2SLS	NB	CF-NB	OLS†	2SLS†
Smoke	0.447**	0.483	0.438**	0.524	0.038	0.099
	(0.185)	(0.741)	(0.182)	(0.705)	(0.029)	(0.104)
Carbon Monoxide	0.020*	0.022	0.015*	0.023	0.0004	0.003
	(0.010)	(0.026)	(0.008)	(0.024)	(0.001)	(0.003)
Observations	5584	5584	5584	5584	5584	5584

Self-Reported Smoking Status:			Negative Binomial		Log Drinks per Week	
	OLS	2SLS	NB	CF-NB	OLS†	2SLS†
Smoke	0.633***	0.420	0.585***	0.476	0.071***	0.085
	(0.180)	(0.630)	(0.186)	(0.581)	(0.027)	(0.088)
Cigs per Day	0.024***	0.012	0.022***	0.014	0.0002**	0.002
	(0.006)	(0.017)	(0.005)	(0.016)	(0.0008)	(0.002)
Observations	5584	5584	5584	5584	5584	5584

† Coefficients in Columns 5-6 represent percentage change when multiplied by 100. On average a drinker consumes 6.40 drinks per week.
Robust standard errors are in parenthesis. ***,**, and * indicate significance at the 1, 5, and 10 percent level. All models include demographic characteristics. Every row entry represents a different regression on the dependent variable, drinks per week, at the first annual follow-up visit. Log Drinks per Week refers to regressions where the dependent variable is transformed into logs.

I also apply two approaches to identify if the OLS coefficient is biased. First, I test if non-smokers differ from smokers in observable characteristics. If they do not differ in observable characteristics, it is less likely that they only differ in unobservable characteristics. There is no evidence that smokers differ from non-smokers in observable characteristics. Second, I test for difference in coefficients between OLS and 2SLS with the Hausman test. However, the Hausman test is not valid for an endogenous dummy variable because the errors are not distributed normally. As a result, I only apply the Hausman test for the non-reported cigarettes per day and carbon monoxide variables. The hypothesis of the Hausman test is that OLS and 2SLS coefficients are equal to each other cannot be rejected at the 10 percent level for all estimates. In other words, the endogeneity of the OLS estimate is not strong enough to justify the use of 2SLS estimate with increased standard

errors.

Separately, I also run Probit and IV-Probit to estimate to understand if there are any extensive margin effects on drinking. Unreported Probit and IV-Probit estimates present no statistically significant effect of quitting smoking on alcohol cessation.

Overall, I conclude that smokers consume more alcohol, implying that cigarettes and alcohol are complements in consumption. However, significantly different effects of smoking on alcohol consumption for gender and socioeconomic status may persist. Previous studies support this notion (See Pierani et al 2009, Picone et al. 2004) and the next subsection discusses sub-sample differences.

Heterogeneous Effects

Smoking can have different effects by gender, age, and socioeconomic background. For example, women are more easily addicted to cigarettes than men (See 2001 Surgeon General Report - Women and Smoking). Genetic differences also cause women to absorb more alcohol and take longer to metabolize alcohol. Therefore, this leaves women with higher alcohol levels in their blood than men, conditional on drinking the same amount of alcohol. As a consequence, women are more susceptible to alcohol's long-term negative health effects than men (Ashley et al. 1977). Married participants are exposed to a different household environment than singles that reduces the likelihood of smoking and that reduces how often they report being in poor health (Schoenborn 2004).

Separately, sub-sample analysis among drinkers may be an important sample to analyze because drinkers may be affected differently by the treatment cessation program than non-drinkers.[19] Current drinkers and smokers may experience complementarity in consumption while non-drinkers may experience substitutability. Therefore, quitting smoking may strongly reduce alcohol consumption among the sample consuming both, a result that may not be observable in the full sam-

[19] Among the sample of drinkers at randomization, I tested for difference in demographic characteristics by treatment groups and did not find any significant differences.

ple.[20] Initially, I tested the hypothesis that all coefficient are equal across gender and marital status in Table 3, which was rejected at the 1 percent level. Moreover, the hypothesis that the coefficients of smoking are equal across gender was rejected at the 10 percent level. While I was unable to reject the null hypothesis of no difference in smoking across age, I include results for a sample including people between 45 to 55 years old. The 45 to 55 age range may represent the longest addicted smokers that are still receptive to quitting smoking while allowing for spillover effect in alcohol consumption.

Table 4 presents sub-sample estimates for individuals who drink alcohol at randomization, by gender, by marital status, and only for those between the ages of 45-55. The results are presented in the same format as Table 3. All OLS results in column 1 present statistically significant results, except for the sub-sample of individuals between the ages of 45-55. Given that the objectively verified smoking variable does not suffer from mis-measurement and present similar average effects as the continuous smoking measure, I concentrate my discussion on those results. The effect of smoking on drinking is larger in the sub-sample estimates of smokers who drink at the time of randomization than for the full sample estimates. The effect is also larger for males than females and non-married participants relative to married participants. For example, among those who drink at randomization, quitting smoking reduces alcohol consumption by 0.6 drinks per week. Moreover, male smokers consume roughly 0.5 drinks more than non-smokers, and non-married smokers consume 0.7 drinks more than non-smokers. The coefficients in the female and married sub-sample estimates are roughly half the magnitude than in the their counterpart male and non-married sub-sample regressions. Again, all 2SLS estimates are insignificant at the 10 percent level, but Hausman tests suggest that OLS estimates can be utilized for all groups except for the age 45-55 sample.[21]

Overall, the OLS marginal effects present significant changes in average alcohol consumption per week for several groups. For example, the mean baseline alcohol consumption among drinkers

[20] I suspect that the full-sample results which includes some individuals who don't drink at randomization, may result in attenuated results towards zero.

[21] I find that some sub-sample Probit estimates suggest that smoking decreases the probability of drinking by about 4-6%. IV-Probit estimates indicate no significant effect and present significantly different marginal effects. I conclude that there is no effect on the extensive margin of drinking.

Table 4: The Effect of Smoking on Weekly Alcoholic Beverage Consumption Among Sub-Samples (Verified Smoking Variable Only)

Drinks at Baseline > 0			Negative Binomial		Log Drinks per Week	
	OLS	2SLS	NB	CF-NB	OLS†	2SLS†
Smoke	0.607***	0.211	0.595***	0.321	0.060**	0.139
	(0.235)	(0.889)	(0.230)	(0.876)	(0.030)	(0.108)
Observations	3878	3878	3878	3878	3878	3878
Males			Negative Binomial		Log Drinks per Week	
	OLS	2SLS	NB	CF-NB	OLS†	2SLS†
Smoke	0.520**	0.386	0.506*	0.345	0.034	-0.007
	(0.261)	(1.028)	(0.261)	(1.035)	(0.039)	(0.132)
Observations	3491	3491	3491	3491	3491	3491
Females			Negative Binomial		Log Drinks per Week	
	OLS	2SLS	NB	CF-NB	OLS†	2SLS†
Smoke	0.355***	0.837	0.351	0.768	0.048	0.319*
	(0.227)	(0.837)	(0.222)	(0.814)	(0.042)	(0.169)
Observations	2093	2093	2093	2093	2093	2093
Married			Negative Binomial		Log Drinks per Week	
	OLS	2SLS	NB	CF-NB	OLS†	2SLS†
Smoke	0.343	0.578	0.338	0.572	0.028	0.061
	(0.218)	(0.770)	(0.216)	(0.735)	(0.035)	(0.115)
Observations	3749	3749	3749	3749	3749	3749
Not Married			Negative Binomial		Log Drinks per Week	
	OLS	2SLS	NB	CF-NB	OLS†	2SLS†
Smoke	0.694**	0.334	0.661**	0.534	0.060	0.215
	(0.346)	(1.785)	(0.334)	(1.671)	(0.052)	(0.223)
Observations	1835	1835	1835	1835	1835	1835
Age: Between 45-55			Negative Binomial		Log Drinks per Week	
	OLS	2SLS	NB	CF-NB	OLS†	2SLS†
Smoke	0.239	1.971*	0.252	1.865**	0.0264	0.274*
	(0.271)	(1.014)	(0.263)	(0.894)	(0.043)	(0.148)
Observations	2532	2532	2532	2532	2532	2532

† Coefficients in Columns 5-6 represent percentage change when multiplied by 100.
Robust standard errors are in parenthesis. ***,**, and * indicate significance at the 1, 5, and 10 percent level. All models include demographic characteristics. Every row entry represents a different regression of the binary verified smoking variable on drinks per week in year 1. Log Drinks per Week refers to regression estimates with a log transformed dependent.

of 6.3 drinks per week suggests that quitting smoking reduces alcohol consumption by 10% for the average drinker at baseline. The average level of alcohol consumption per week among male and non-married smokers can be reduced by 11% and 16% if they quit smoking. Besides the demographic sub-samples, it is important to analyze the effect of quitting smoking among the excessive drinking population alone. One might worry that the presented results are less important because they concern moderate drinking levels. However, even at moderate levels of drinking the risk of mortality increases significantly (Bouchery et al. 2006). Therefore, reducing moderate drinking levels can significantly reduce the death toll. Since the cost of excessive drinking is extremely large the next subsection concentrates on heavy and binge drinkers.

The Relationship Between Smoking and Drinking Among Excessive Drinkers

The Center for Disease Control (CDC) defines excessive drinking if someone is a heavy drinker, defined as consuming 15 or more drinks per week for men, or binge drinker, defined as consuming 5 or more drinks per session per day for men. That group represent at least half the death toll and three quarters of all costs associated with alcohol consumption alone (CDC 2016).[22] Thus, identifying the relationship between smoking and drinking is especially important among the high-risk population. Table 5 presents estimates for 4 different samples – a sample including heavy drinkers consuming more than 14 drinks per week, a sample of heavy and binge drinkers defined by the male standard, and samples of excessive drinkers by gender.

The estimates for the heavy drinking sub-sample do not suggest that quitting smoking reduces alcohol consumption. However, the estimates are imprecise because of the small sample size. The sub-sample including all excessive drinkers suggests that quitting smoking reduces alcohol consumption by 1.68 drinks per week. The OLS coefficient is statistically significant at the 10 % level, while the 2SLS estimates present insignificant results. Hausman tests do however indicate that OLS estimates are preferred to 2SLS. Among the sample of excessive male drinkers, the OLS estimates suggest that quitting smoking reduces alcohol consumption by 2 drinks per week. 2SLS

[22] Heavy drinking for females is defined as consuming more than 7 drinks per week and binge drinking as consuming more than 3 drinks per occasion.

Table 5: The Effect of Smoking on Weekly Alcoholic Beverage Consumption among Excessive Drinkers (Verified Smoking Variable Only)

Heavy Drinkers>14			Negative Binomial		Log Drinks per Week	
	OLS	2SLS	NB	CF-NB	OLS†	2SLS†
Smoke	1.106	-1.932	1.100	-1.867	0.069	-0.097
	(1.170)	(5.622)	(1.150)	(5.814)	(0.100)	(0.416)
Observations	347	347	347	347	347	347
Excessive Drinkers			Negative Binomial		Log Drinks per Week	
	OLS	2SLS	NB	CF-NB	OLS†	2SLS†
Smoke	1.68*	-0.500	1.513*	-1.501	0.132	-0.254
	(0.885)	(4.114)	(0.877)	(4.658)	(0.084)	(0.348)
Observations	591	591	591	591	591	591
Male Excessive Drinkers			Negative Binomial		Log Drinks per Week	
	OLS	2SLS	NB	CF-NB	OLS†	2SLS†
Smoke	2.124**	2.117	1.982*	1.818	0.192*	-0.041
	(1.023)	(4.535)	(1.027)	(4.611)	(0.098)	(0.372)
Observations	472	472	472	472	472	472
Female Excessive Drinkers			Negative Binomial		Log Drinks per Week	
	OLS	2SLS	NB	CF-NB	OLS†	2SLS†
Smoke	-0.095	-2.158	0.039	-1.965	-0.021	-0.067
	(0.695)	(2.991)	(0.687)	(3.767)	(0.077)	(0.381)
Observations	416	416	416	416	416	416

† Coefficients in Columns 5-6 represent percentage change when multiplied by 100.
Robust standard errors are in parenthesis. ***,**, and * indicate significance at the 1, 5, and 10 percent level. All models include demographic characteristics. Every row entry represents a different regression on the dependent variable Drinks per Week in year 1. Log Drinks per Week refers to regressions where the dependent variable is transformed into logs.

estimates have the same point estimate, but are again insignificant. Among the sample of excessive female drinkers, I find insignificant estimates across all specifications, but the 2SLS estimates have economic significance implying a 3 drinks per week increase for quitters. I speculate that gender specific difference in addiction may contribute to this result. For example, if females are more heavily addicted to drinking, then quitting smoking may not result in an immediate response in alcohol consumption. Instead, changes in alcohol consumption may occur more gradually over time. [23] On the other hand, they may suffer from oral-fixation and there increase alcoholic intake.

In summary, I conclude that quitting smoking reduces alcohol consumption by roughly 1.6 drinks or 11% per week among excessive drinkers. Also, quitting smoking reduces alcohol consumption by 2 drinks per week for males and does not statistically affect consumption levels for females. The results have important policy implications, because individuals consuming alcohol at excessive levels burden large costs onto society. Therefore, policymakers need to implement strategies that effectively reduce alcohol consumption to below excessive levels. On average, excessive drinkers consume roughly 14 drinks per week. Subsequently, quitting smoking reduces alcoholic beverage consumption below the excessive threshold. It also suggests that those people consume about 6.5 drinks per month less than smokers. These effects can have significant income effects and health benefits. Among the sub-sample of excessive male drinkers, the reduction of 2 drinks per week translates into a decrease in consumption of 15% per week. Again quitting smoking will lead on average to an alcoholic beverage consumption below the CDC threshold of excessive drinking.[24]

The Impact of Past Cigarette Consumption on Current Alcohol Consumption

The preceding results only investigate the contemporaneous effect of quitting on alcohol consumption without taking the habit stock effect into account. This analysis has three motivations. First,

[23] Among the sub-samples, I find statistically significant effects on the extensive margin of drinking in Probit models by gender, suggesting that smoking reduces the probability of drinking by 6% and 9% for males and females. While IV-Probit present different insignificant effects.

[24] Initial consumption for heavy and binge drinking men prior to the smoking cessation program is roughly 15 drinks per week.

the empirical evidence and the rational addiction model suggest that the smoking habit stock affects current alcohol consumption. Second, the previous section compares the instantaneous relationship between smoking and drinking, but it can be difficult to understand how one consumption choice affects another consumption choice immediately. It is possible that a smoking cessation treatment changes smoking behavior and affects alcohol consumption over time rather than at the same point. For example, health consciousness and changes towards healthy behaviors are a process that can be initiated by quitting smoking, but it takes time to change other risky behaviors. Third, relating past smoking status on current alcohol consumption can be understood as a long-term effect.

I begin by regressing weekly alcoholic beverage consumption on different historic smoking variables. Specifically, I regress last year's smoking status on today's alcohol consumption. However, including a smoking status dummy variable excludes the variation among individuals along the intensity of historic smoking. A clear feature of the habit effect in the rational addiction model is that the intensity of smoking conveys information on the level of addiction. Someone smoking in the past, on average, one cigarette per day should see a different response in alcohol consumption than someone consuming, on average, 20 cigarettes per day. The availability of monthly cigarette consumption in all of the last twelve months allows me generate a more comprehensive measure of the habit stock that affects today's alcohol consumption. Thus, I utilize all 12 months of smoking information collected at the first annual follow-up visit to generate a measure of the habit effect.

My preferred specification includes dummy variables for different average levels of past 12 months cigarette consumption.[25] This measure differentiates between smoking status as well as in the intensity of past smoking. For simplification, I generate dummy variable bins in steps of 5 cigarettes s consumed per day in the last year. For example, a participants is in the very low average smoking bin if the person smoked on average 1-5 cigarettes per day in each month of the last year. The low smoking bin includes participants with an average cigarettes consumption of 6-10 cigarettes per day in the last year of each month. The heaviest smokers includes highest bin

[25]I have also generate a cumulative 12 months cigarette consumption variable. For the cumulative variable I also included depreciation rates from 0-100% per year. I have also modeled non-linear polynomials relaxing the linearity assumption of the cumulative smoking measure. In all cases I find similar results

includes those who consume on average 36 or more cigarettes per day. Such a categorization leads to an estimation of 8 dummy variables in the regression. I have also used different bin ranges and draw to similar conclusions. I estimate the average long-term smoking status as the average of the five annual smoking information. The model for the short-term average smoking level is presented here:

$$Drinks_{it} = \gamma_0 + \sum_{j=1}^{N} \gamma_j Cigs_{i,j} + \gamma X_{it} + \mu_{it} \tag{3}$$

Where $Cigs_{i,j}$ is a dummy variable equal to one in the jth bin for individual i if his 12 months average cigarette consumption falls in that bin range and equals zero otherwise. The omitted category includes participants who are non-smokers in the last 12 months.

Endogeneity in the OLS estimates may be one concern that leads to biased estimates. However, as presented earlier, OLS did not seem to suffer from endogeneity and most of the variation in the extensive margin of smoking occurs among participants in the treatment group. Additionally, the changes on the intensive margin of smoking are also very strong among the treatment group participants with only small reductions in the average consumption of cigarettes among the control group.[26] As a result, the coefficients should not be significantly affected by bias.

Table 6 presents marginal effects for eight dummy variables representing different average levels of cigarette consumption in the last 12 months in column 1. Columns 2 and 3 present subsample estimates by sex. The estimates follow the theoretical predictions that higher past consumption affects alcohol consumption more strongly. Participants with above average levels of cigarette consumption consume significantly more alcohol today than those with below average cigarette consumption levels. Specifically, column 1 shows that participants averaging between 1-5 cigarettes per day do not consume more alcohol than non-smokers. The low consumption of cigarettes suggests that it may represent "stress" smokers or people who quit smoking for several months. Someone consuming 6 to 10 cigarettes per day also consumes roughly 1 drink per week more than non-smokers, a significant increase in alcohol consumption. Estimates also show that

[26]The average consumption of cigarettes per day at randomization is 29.5 cigarettes. At the first annual follow-up visit the average cigarette consumption among the treated is about 12 cigarettes per day and 24 cigarettes among the control group.

those consuming on average 31 to 35 cigarettes per day do not consume more than non-smokers. This finding may be surprising, but may have to do with the fact of the small population in that category (181 observations). Column 2-3 show that the effect of past smoking on alcohol consumption is larger for females and that past smoking increases alcohol consumption significantly.[27]

Table 6: The Effect of the 12 Months Average Cigarette Consumption per Day History on Drinking At the First Follow-up Meeting

	Dep. Var. Drinks per week		
Average 12 Months Consumption:	OLS	Men	Female
1-5 Cigarettes per Day $\equiv \mathbb{1}$	0.571	0.420	.923***
	(0.353)	(0.516)	(0.313)
6-10 Cigarettes per Day $\equiv \mathbb{1}$	0.973***	0.738	1.489***
	(0.375)	(0.553)	(0.3253)
11-15 Cigarettes per Day $\equiv \mathbb{1}$	1.285***	1.187**	1.625***
	(0.377)	(0.548)	(0.370)
16-20 Cigarettes per Day $\equiv \mathbb{1}$	1.184***	1.137*	1.468***
	(0.405)	(0.615)	(0.356)
21-25 Cigarettes per Day $\equiv \mathbb{1}$	1.418***	1.202*	1.955***
	(0.462)	(0.676)	(0.440)
26-30 Cigarettes per Day $\equiv \mathbb{1}$	1.660***	1.278**	2.493***
	(0.408)	(0.575)	(0.454)
31-35 Cigarettes per Day $\equiv \mathbb{1}$	0.754	0.196	1.794***
	(0.501)	(0.718)	(0.535)
36+ Cigarettes per Day $\equiv \mathbb{1}$	1.731***	1.616***	2.007***
	(0.438)	(0 .596)	(.512)
Observations	5579	3488	2091

Robust standard errors are in parenthesis. ***,**, and * indicate significance at the 1, 5, and 10 percent level. All models include demographic characteristics. Each column present results from a single regression. The Bins variable indicates dummy variables measuring average cigarette consumption in that specific range.

In unreported regressions I also estimate the relationship between a binary smoking variable measuring smoking status 12 months ago. All estimates are similar in magnitude and are significant at least at the 5% level. Suggesting that not smoking 12 months ago decreases current alcohol

[27] As a robustness check I also regressed drinks per week on one continuous average 12 months cigarette per day consumption variable. The results present significantly smaller effects for the average smoker. Including square and cube terms of average cigarette consumption generates similar results to those presented in table 6.

consumption by roughly 1 drink per week or 23%. 2SLS estimates are insignificant, but a Hausman test indicates that OLS estimates are preferred to 2SLS estimates. Probit estimates suggest that quitting smoking 12 months ago reduces the probability of drinking by 20%, but I find no statistically significant IV-Probit estimates.[28]

In summary, the smoking historic influences current alcohol consumption significantly. I find large heterogeneous effects from the intensity of past smoking decisions on alcohol consumption. Low levels of smoking tend to have no effect on alcohol consumption, but consuming one pack per day increases alcohol consumption by about 1.2 drinks per week. Suggesting a significant increase in overall alcohol consumption of about 28% from the baseline mean. The results, however, have some caveats., because I only measure an average consumption across the last 12 months. The fact that some people may quit smoking in some months and relapse in others may misrepresent the average cigarette consumption per day. Therefore, the next section tries to identify the effect of those participants who continuously abstain from smoking on alcoholic beverage consumption.

Identifying the Time-Effect of Permanently Quitting Smoking on Alcohol Consumption

Table 6 presents estimates on the effect of 12 months historic average cigarette consumption on current alcohol consumption. The 12 months average cigarette consumption variable allows for the possibility that some people change the level of cigarette consumption (or quit smoking) across month, i.e. they might consume in one month 30 cigarettes and do not smoke in the next month, which would result in an average consumption of 15 cigarettes per day. Therefore, t this section tries to understand whether permanently quitting smoking affects alcohol consumption differently. Additionally, I investigate how the length or timing of quitting smoking affects alcohol consumption.

The addiction theory implies that the level of the habit stock matters as well as the timing of quitting smoking matters on future alcohol consumption. For example, consider two people with the same level of the habit stock, where one of the smokers quits smoking and the other continuous

[28]Results are not reported, but are available upon request. I ran Linear Probability Models (LPM), and Hausman tests suggest that OLS estimates can be used over 2SLS estimates.

smoking. The habit stock of the quitter depreciates over time, while the continuous smoker does not see a change in the habit stock. As a result, at any future point in time alcohol consumption of the quitter should be differently affected due to the lower habit stock.[29] The following paragraphs explain how I try to measures the timing and the differential habit stock effect on alcohol consumption.

Specifically, I estimate the timing effect by generating a sample of beginning non-smokers measured by the months of quitting from the first annual follow-up visit. I specifically exclude temporary non-smokers (those who relapse in later months). The empirical strategy includes a dummy variable equal to one if a participant quits smoking permanently. For example, to measure the effect of 12 months non-smoking on alcohol consumption, the dummy variable equals one if a person quits smoking 12 months prior to the first annual follow-up visit and does not smoke in the following 11 months up to the first annual follow-up visit. The dummy variable equals zero for those participants who are 12 months continuous smokers and part of the control group. I utilize smokers in the control group as a counterfactual because they have the least differences in demographic characteristics from the treatment group.[30] This gives me a measure of how much alcohol smokers would have consumed had they not quit. In other words, the counterfactual thought experiment suggests that what would have happened if the person did not quit smoking for twelve months, but instead smoked the full year.

The LHS includes self-reported information on smoking for all 12 months prior to the first annual follow-up year, allowing me to run 12 regression between beginning non-smokers in a given month and alcohol consumption at the first annual follow-up visit. To clarify, each regression, therefore, includes a different mutually exclusive treatment group from month to month. The treatment group changes from month to month as it only includes beginning sustained non-smokers, while always including the continuous smokers of the control group. By limiting the sample, I identify the actual changes in alcohol consumption due to permanent smoking cessation and im-

[29] A detailed explanation of the theoretical background is in the Supplement part 3

[30] The estimation strategy can be thought of as a classical difference in differences model, where quitting smoking is the exogenous treatment for the smokers.

plicitly measure changes in the habit stock of cigarettes. The regression equation is similar to equation (1) but includes the here-discussed definition of the smoking variable. For example, for the 12 month sustained non-smoker the new smoking variable equals one if the following equation holds, and zero for the continuous smokers in the control group:

$$Non - Smoker_{12month} = 1, iff \big[\sum_{m=1}^{12} smoke_{im} \big] = 0$$

$$Non - Smoker_{12month} = 0, iff \big[\sum_{m=1}^{12} smoke_{im} \big] = 12$$

Similarly, $Smoke_{11month}$ equals one for smoking in $m = 1$ and non-smoking in months $m = 2, 3, ..., 12$, and zero as defined above. To reiterate, each regression only includes one of the 12 generated smoking variables and treatment group is limited to those who smoked in month $t - 1$ but quit from month t until the first follow-up meeting.

2SLS estimates will identify the effect of sustained non-smokers in the treatment group relative to continuous smokers in the control group.[31] I impose one additional strong assumption to estimate valid 2SLS models. I assume that the random assignment to the treatment group, which happened at different times following the third screening survey also results in a random date of receiving treatment. The first date of possibly receiving treatment would be 12 months prior to the first annual follow-up visit. Therefore, I assume that those who permanently quit for 12 months are only 12 months non-smokers because they received the treatment earlier in the year than people who quit smoking for 11 months.

The estimates include a different sample than the estimates presented in Table 6. The interpretation of quitting smoking in month 12 suggests that the participant did not smoke in all of the following months up to the first annual follow-up visit. The previous section did not make any restrictions on month-to-month smoking habits. To distinguish the results from the previous section I refer to a smoker quitting in month 12 who remain a non-smoker in all of the following months

[31] Non-parametric estimation strategy should yield similar results, where I could compare the difference in alcohol consumption between permanent quitters in the treatment group relative to continuous smokers in the control.

as a "12 month sustained non-smoker". The sample of sustained non-smokers may be different from the overall sample and suffer from sample selection. I tested for differences in any of the five observable characteristics (gender, education level, marital status, age, and cigarette consumption at baseline) for the 12-8 months sustained non-smoker from the control group participants. I only find a statistically significant difference in marital status for 12, 10 and 9 months sustained quitters and gender for 12 months sustained quitters. In total I conducted 20 tests and had to reject in 5 cases and only once at the 1% level. I conclude that there may be some evidence for possible differences by marital status between treatment and control group.

Figure 1 graphs the OLS and 2SLS marginal effects on the effect of becoming a sustained non-smokers for 12 to 8 months on alcohol consumption.[32] I find statistically significant results for 12 months sustained non-smokers. 2SLS results indicate that 12 months sustained non-smokers consume 1.2 drinks per week or 27.5% less than continuous smokers. 11 through 9 months sustained non-smokers consume about 0.6 drinks more per week than a 12 month sustained non-smoker. Nevertheless, sustained quitters beginning in month 11 through 9 consume 0.6 drinks, or 14%, less that continuous smokers.[33] Overall, the marginal effects follow the expected inverse relationship between quitting smoking and alcohol consumption. Quitting smoking later in the year and closer to the first annual follow-up visit results in a higher habit stock of smoking, which increases alcohol consumption.

Similarly, I also find large extensive margin effects of quitting smoking permanently on drinking. Table 7 includes average marginal effects for probit, and CF-probit estimates. 12 months sustained non-smokers are 31% less likely to drink than continuous smokers. 11 months sustained non-smokers are roughly 5% less likely to drink than continuous smokers.[34] The extensive margin effects follow the results from Figure 1. The longer one permanently abstains from smoking the

[32]I present only months 12 through 8, because limiting the sample has its cost on observations and power. Beginning in month 7 and any later month sustained quitters represent only a small fraction of the sample, questioning any internal validity. For example, in June only 36 people become sustained quitters for the rest of the year. Beginning in August, the average month sees 11 sustained quitters. Therefore, I am not confident that the estimates from month 7 and onward represent representative average effects.

[33]Appendix Table A.2& A.3 present regression and Negative Binomial marginal effects.

[34]Logit results presents similar average marginal effects. Table not included.

Figure 1: Average Alcoholic Beverage Consumption for Beginning Sustained Non-Smokers in Different Months Prior to the First Annual Follow-up Visit

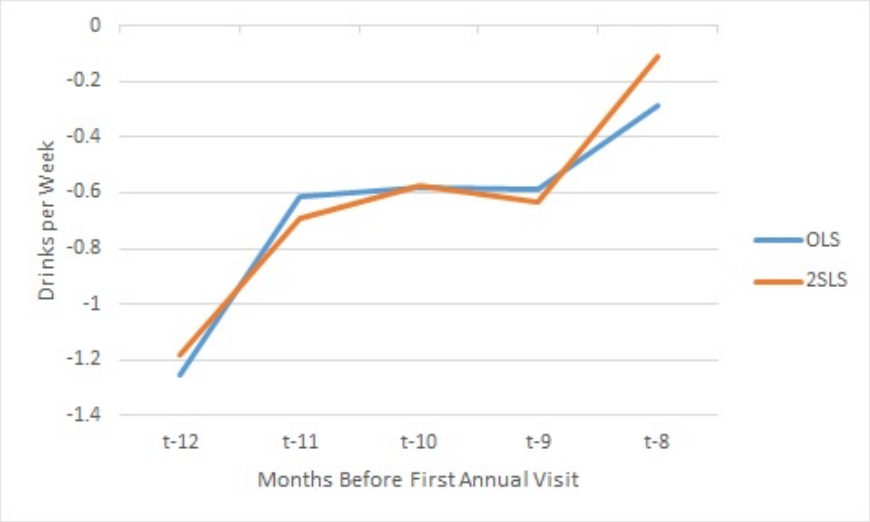


Figure 1 plots OLS and 2SLS marginal effects from Table A.2. Results present the effect of becoming a sustained non-smoker in months 12,11,10... prior to the first annual follow-up visit on alcohol consumption.

larger the effects on alcohol consumption.

Taking all results, I conclude that the length of quitting smoking affects alcohol consumption differently. Moreover, since the average cigarette consumption as baseline for those who permanently quit 12 months is about 30 cigarettes per day, I conclude that the estimates are similar compared to Table 6, and therefore providing evidence that the effect is robust to specification.[35] Additionally, the theoretical model supports the prediction that a longer period of abstaining from smoking decreases the habit stock through depreciation and results in a larger drop in alcohol consumption. Next, I present results by gender, as Pierani et al. (2009) also suggest that those differ.

Gender Differences on the Effect of the Habit Stock on Alcohol Consumption

Pierani et al. (2009) present significantly different results for the habit stock by gender. This section estimates the gender difference in the effect of permanently quitting smoking on alcohol consumption. Table 8 presents OLS and 2SLS estimates for all beginning sustained non-smokers

[35]On average a 12 through 8 months sustained quitter consumed close to 30 cigarettes per day which leads to very similar point estimate for the 12 month sustained quitter.

Table 7: The Effect of the Length of Abstaining from Smoking on the Probability of Drinking at the First Annual Follow-Up Visit

	Dep. Var. Any Drinks	
Length of Non-Smoking	Probit	CF-Probit
12 Months Sustained Non-Smoker	-0.39***	-0.31***
	(0.02)	(0.03)
11 Months Sustained Non-Smoker	-0.04***	-0.05
	(0.03)	(0.03)
10 Months Sustained Non-Smoker	-0.01	-0.01
	(0.026)	(0.03)
9 Months Sustained Non-Smoker	-0.016	-0.02
	(0.03)	(0.027)
8 Months Sustained Non-Smoker	-0.004	-0.001
	(0.04)	(0.04)
7 Months Sustained Non-Smoker	-0.003	-0.026
	(0.06)	(0.08)
6 Months Sustained Non-Smoker	-0.021	-0.099
	(0.08)	(0.10)
5 Months Sustained Non-Smoker	-0.10	-0.021
	(0.08)	(0.12)
4 Months Sustained Non-Smoker	-0.021	-0.10
	(0.09)	(0.12)
3 Months Sustained Non-Smoker	0.057	-0.30
	(0.09)	(0.01)
2 Months Sustained Non-Smoker	0.017	-0.12
	(0.09)	(0.14)
1 Month Sustained Non-Smoker	0.04	-0.03
	(0.08)	(0.13)

Robust standard errors are in parenthesis. ***,**, and * indicate significance at the 1, 5, and 10 percent level. Column 1 presents Probit estimates between a binary drinking indicator, equal to one if the person quit smoking continuously beginning that months and zero otherwise, and alcoholic beverage consumption per week. All estimates present average marginal effects.

beginning in month 12 through 1 before the first follow-up visit. Figure 2 graphs 2SLS marginal effects by gender for months 12 to 8 beginning sustained non-smokers. The figure shows that the shorter the period of abstaining from smoking the higher the alcohol consumption. Across the length of quitting the trend for females are more linear than males. However, among the male sub-sample I only find statistically significant OLS results for 12 months non-smoking. I find statistically significant OLS and 2SLS marginal effects for 12 and 11 months female sustained non-smokers. Specifically, I find that female 12 month sustained non-smokers consume 1.8 drinks per week less, while 12 months male non-smokers consume 1 drink per week less than continuous smokers.

Figure 2: Effects of Becoming a Sustained Non-Smokers on Alcohol Consumption by Gender

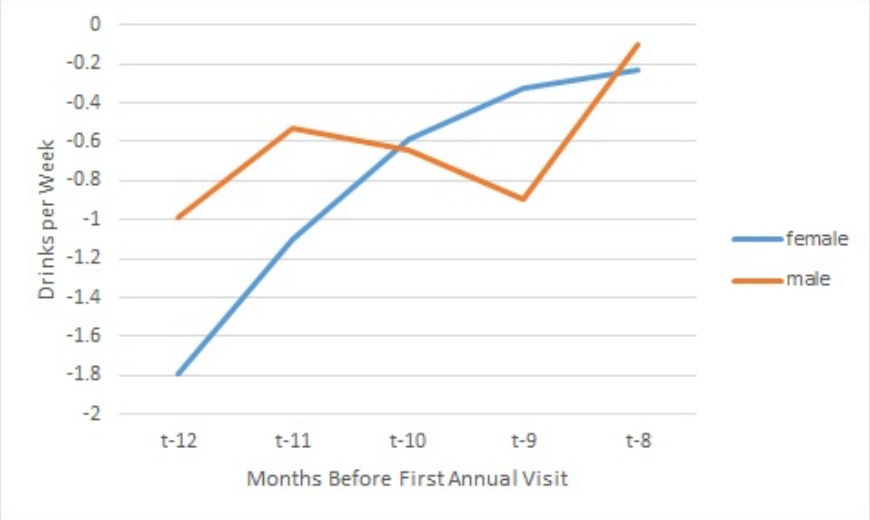


Figure 3 presents 2SLS marginal effects for Table 8. Results present the effect of becoming a sustained non-smoker in months 12,11,10... prior to the first annual follow-up visit on alcohol consumption.

Similarly, I also estimate probit and CF-profit models by gender. Table 9 presents probit, and CF-probit average marginal effects. I find that 12 months sustained male and female non-smokers are 30% and 32% less likely to drink, respectively. Both estimates are statistically significant at the 1% level. Additionally, 11 months female sustained non-smokers are roughly 10% less likely to drink than continuous female smokers. Interestingly, women are slightly more likely to quit

Table 8: The Effect of the Length of Abstaining from Smoking on Alcohol Consumption at the First Annual-Follow-up Visit

	Female		Male	
Length of Non-Smoking	OLS	2SLS	OLS	2SLS
12 Months Sustained Non-Smoker	-1.744***	-1.791***	-1.111*	-0.986
	(0.365)	(0.379)	(.596)	(0.633)
11 Months Sustained Non-Smoker	-1.014*	-1.096**	-0.445	-0.527
	(0.540)	(0.548)	(0.618)	(0.620)
10 Months Sustained Non-Smoker	-0.422	-0.583	-0.750	-0.647
	(0.445)	(0.449)	(0.500)	(0.509)
9 Months Sustained Non-Smoker	-0.215	-0.328	-0.891*	-0.893*
	(0.485)	(0.511)	(0.496)	(0.502)
8 Months Sustained Non-Smoker	-0.040	-0.230	-0.428	-0.099
	(0.713)	(0.736)	(0.735)	(0.789)
7 Months Sustained Non-Smoker	0.572	-2.377	-1.144	-1.915**
	(1.838)	(0.677)	(1.084)	(0.823)
6 Months Sustained Non-Smoker	-1.408	-0.701	0.254	-0.993
	(0.937)	(1.486)	(1.392)	(1.584)
5 Months Sustained Non-Smoker	-0.947	2.917***	1.149	1.959
	(1.073)	(0.399)	(1.940)	(2.488)
4 Months Sustained Non-Smoker	0.033	0.460	-0.009	-0.464
	(1.272)	(2.177)	(1.717)	(1.671)
3 Months Sustained Non-Smoker	-0.071	-0.444	-3.284***	-1.033
	(1.272)	(2.179)	(1.006)	(1.233)
2 Months Sustained Non-Smoker	-2.158***	2.551***	-1.678	-3.185***
	(0.362)	(0.418)	(1.240)	(1.087)
1 Month Sustained Non-Smoker	-1.417	-1.249	0.811	0.869
	(0.929)	(1.414)	(1.223)	(1.802)

Robust standard errors are in parenthesis. ***,**, and * indicate significance at the 1, 5, and 10 percent level. All models include demographic characteristics. Every row entry represents a different regression and different sample on the dependent variable Drinks per Week in year 1. Table 8 presents OLS and 2SLS marginal effects for a binary indicator equal to one for a beginning non-smoker for the rest of the year and zero for a continuous smoker.

drinking than men. This is counterintuitive to the current evidence that women tend to become addicted more easily and should have a more difficult time quitting abruptly.

Finally, the evidence shows that women reduce alcoholic beverage consumption and are more likely to quit drinking if they become a sustained non-smoker for 12 or 11 months. 12 month non-smoking females reduces alcohol consumption by 53% per week and 11 months non-smoking females reduce alcohol consumption by 29% relative to baseline consumption levels at the randomization meeting. Among men, those who become sustained non-smokers for 12 months decrease alcohol consumption by 23%. The evidence suggests that the length of quitting affects alcohol consumption significantly. Moreover, I present evidence that implies that the habit stock affects females more than males, because the change in alcoholic beverage consumption among females is twice as large as the change for males. In a different light, the estimates suggest that there are long-term benefits from quitting smoking via a reduction in alcoholic beverage consumption.

Table 9: Sub-sample Gender Effects of the Length of Abstaining from Smoking on the Probability of Drinking Any Alcoholic Beverages at the First Annual Follow-up Visit

	Female		Male	
Length of Non-Smoking	Probit	CF-Probit	Probit	CF-Probit
12 Months Sustained Non-Smoker	-0.37***	-0.32***	-0.40***	-0.30***
	(0.04)	(0.05)	(.03)	(0.03)
11 Months Sustained Non-Smoker	-0.10*	-0.12*	-0.01	-0.01
	(0.054)	(0.06)	(0.03)	(0.04)
10 Months Sustained Non-Smoker	-0.02	-0.03	-0.006	-0.01
	(0.04)	(0.005)	(0.03)	(0.03)
9 Months Sustained Non-Smoker	0.01	0.01	0.01	0.02
	(0.04)	(0.05)	(0.03)	(0.034)
8 Months Sustained Non-Smoker	0.05	0.05	-0.03	-0.02
	(0.07)	(0.07)	(0.04)	(0.05)

Robust standard errors are in parenthesis. ***,**, and * indicate significance at the 1, 5, and 10 percent level. Table 8 presents average Probit marginal effects and Probit Control Function Approach marginal effects for a binary variable equal to one if the person is a beginning sustained non-smoker for the rest of the year and zero for a continuous smoker.

The Effect of Quitting Smoking on Alcohol Consumption after Five Years

All previous results present short-term average effects on the effect of smoking on alcohol consumption. This section estimates the long-term effects and compares them with the short-term effects. It is possible that people who quit smoking abruptly decrease alcohol consumption by the first follow-up visit, but increase consumption in the long-term. To this end, I present OLS and 2SLS long-run estimates for two different smoking variables on alcohol consumption by the 5th annual follow-up visit. I present long-term effects for the objectively-verified smoking variable and the 12 months sustained quitters variable from the previous sections.

Table 10: Long-Run Effects of Quitting Smoking

Objective Smoking Status:	Short-Run		Long-Run	
	OLS	2SLS	OLS	2SLS
Smoking status at first follow-up visit	0.447**	0.482	0.497**	0.542
	(0.185)	(0.741)	(0.195)	(0.771)
12 months sustained non-smokers	-1.251***	-1.183***	-1.112***	-1.115***
	(0.403)	(0.428)	(0.401)	(0.422)

Robust standard errors are in parenthesis. ***,**, and * indicate significance at the 1, 5, and 10 percent level. All models include demographic characteristics. Every row entry represents a different regression on the dependent variable Drinks per Week in year 1 for the Short-Run and year 5 for the Long-Run.

Table 10 presents the main results and shows no significant changes in coefficient size after five years. I test for differences between the short-run OLS and long-run OLS coefficients but do not find statistically significant differences.[36] The key take away is that the short-run effect of lower alcohol consumption persists after five years.

Conclusion

This paper presents strong evidence that smoking and alcoholic beverage consumption is positively correlated. On average, quitting smoking reduces alcohol consumption by roughly 0.5 drinks per week or 11.5%. Among excessive alcohol users, quitting smoking reduces alcohol consumption

[36] A Table presenting those results can be made available upon request.

by about 1.7 drinks or 15% per week. The estimates also reveal heterogeneous smoking habit stock effects on alcohol consumption. I show that participants consuming on average between 1-5 cigarettes per day in the last 12 months prior to the first annual follow-up meeting do not increase current alcohol consumption relative to non-smokers. However, consuming on average 20 cigarettes per day increases current alcohol consumption by 1.2 drinks per week. . Additionally, I show that the longer the participant abstains from smoking the stronger the negative effect on alcohol consumption

The study suggests that smoking interventions present a powerful policy approach that minimize cigarette and alcohol demand. Smoking interventions have a dynamic second-order effect on alcohol demand, suggesting that policymakers could have underestimated the total benefits of smoking interventions. I hypothesize that the second-order effects from a smoking intervention can also be partially achieved in the general smoking population, if they are combined with an effective nicotine prescription program. However, the consistently shrinking smoking population lessens the total second order benefits, requiring broader actions targeting non-smoking drinkers.

A separate literature points out that quitting smoking increases BMI, but the mechanisms are still debated.[37] Alcoholic beverage consumption significantly increases caloric and could be one mechanism through which recent quitters compensate or ameliorate the painful detoxification process to stay away from cigarettes. This paper provides evidence that quitters do not substitute their physical and psychological cigarette dependence by increasing alcoholic beverage consumption. To the contrary, especially for sustained non-smokers, those who stop smoking tend to drink less alcoholic beverages, cutting additional calories out of their diet.

Lastly, a back of the envelope calculation estimating the costs savings from the intervention that would otherwise have occurred due to excessive alcohol consumption suggests annual saving of about 645,000 dollars.[38] The recent passage of the Affordable Care Act also requires insurers to cover smoking cessation treatment costs of the insured and could generate cost savings similar as

[37] See Courtemanche 2009, Chou et al. 2004.

[38] 918 LHS participants were classified as continuous non-smokers by the 5th year. 129 were classified as excessive drinkers at the baseline meeting. Based on Sack et al. 2015 a conservative estimate suggests that the cost of excessive drinking per excessive drinker is at 5000 dollars.

the LHS per person. For example, taking current estimates from Sachs et al. (2015), a 1 percentage point drop in smoking participation can save around 300 million dollars annually that would have otherwise been incurred from excessive drinking.[39]

II The Effect of Smoking on Obesity: Evidence from a Randomized Trial

Introduction

In the last 40 years obesity[40] rates have steadily increased in the United States (US), rising from 13% in the early 1960s to 35% in 2011-2012 (Flegal et al., 1998, Ogden et al., 2014). This rise in obesity has contributed significantly to increasing rates of diabetes, heart disease, and stroke (Mokdad et al. 2001, Manson et al. 1990, Rexrode et al. 1997), with Flegal et al. (2005) finding that obesity-related diseases lead to 112,000 deaths per year. Wang et al. (2011) project that by 2030 the number of obese adults in the US will grow by another 65 million. Cawley and Meyerhoefer (2012) estimate that obesity leads to $190 billion per year in medical expenses, while Wang et al. (2011) project that this number will increase by $48 to $66 billion by 2030.

During the same time frame, the percentage of adults who use tobacco in the US declined from 42% to 19% (National Center for Health Statistics, 2011). The 1964 Surgeon General's Report concluded that smoking leads to adverse health conditions such as lung cancer and heart disease and increases mortality risk (US Department of Health and Human Services, 1964). Subsequently, federal and state governments launched an aggressive tobacco control campaign featuring advertising restrictions, warning labels, information-spreading programs, cigarette taxes, and smoking bans in public places. Despite the success of these efforts in reducing smoking, tobacco is still responsible for one out of every five deaths in the United States and at least $130 billion per year

[39] CDC and the LHS suggest that 15 percent of smokers consume at excessive alcohol levels

[40] Obesity is defined as having a body mass index (BMI) greater than 30, where BMI is equal to weight in kilograms divided by height in meters squared.

in medical expenses (US Department of Health and Human Services, 2014).

The inverse trends in smoking and obesity raise the question of whether they are causally related, in which case tobacco control policies may had the unintended consequence of contributing to the rise in obesity. Quitting or reducing smoking could increase body weight since nicotine can act as an appetite suppressant and metabolic stimulant (Pinkowish, 1999). In standard economic models of body weight (e.g. Philipson and Posner, 1999), nicotine's appetite-suppressing properties could be seen as decreasing the marginal utility of food consumption, leading to less eating and therefore lower body weight. Stimulating the metabolism would mean more calories burned holding physical activity constant, again reducing weight. On the other hand, smoking reduces lung capacity (Hedenstrom et al., 1986), which could lead to weight gain by increasing the marginal disutility from exercise.

A large public health literature documents that individuals tend to gain weight following smoking cessation. A meta-analysis of 15 studies found that individuals who quit smoking gain an average of four pounds more than a comparison group of continued smokers (U.S. Department of Health and Human Services, 1990). Another meta-analysis, which included a larger number of studies (62) but no comparison group, found that weight gain steadily increases in the year after smoking cessation, ultimately reaching 4.67 kg (10.3 lbs) (Aubin et al., 2012). Evidence regarding longer-run effects is mixed. Some studies have found that much of the weight gain after quitting smoking is temporary (Chen et al., 1993; Mizoue et al., 1998), but others conclude that the effect remains sizeable five to ten years after cessation (Flegal et al., 1995; Travier et al., 2012; Williamson et al., 1991). These associational estimates could be susceptible to bias from unobservable characteristics, such as time preference and level of interest in one's health, that likely influence both the probability of smoking cessation and weight trajectory. Additionally, studies that simply track changes in weight without utilizing a comparison group of continued smokers are susceptible to additional confounding from the tendency to gain weight with age.

The economics literature has attempted to move closer toward causality by examining the effects of plausibly exogenous sources of variation in economic factors that influence smoking on

BMI. The results from this literature are mixed. Chou et al. (2004), Rashad et al. (2006), and Baum (2009) estimate positive relationships between cigarette costs and BMI. Since higher cigarette prices have been shown to reduce smoking, these results are consistent with reduced smoking leading to weight gain. However, Gruber and Frakes (2006), Courtemanche (2009), and Wehby and Courtemanche (2012) estimate the effect of cigarette costs on BMI to actually be negative, while Nonnemaker et al. (2009) and Courtemanche et al. (forthcoming) find little evidence of an effect in either direction. The discrepancies in observed results in the literature hinge on methodological issues such as whether cigarette prices or tax rates are used as the measure of cigarette costs, whether time is modeled using a quadratic trend or time period dummies, and whether the difference between short-run and long-run effects is considered. A particularly controversial issue is whether cigarette costs can actually be considered exogenous. Cigarette prices may depend on the demand for cigarettes, while high cigarette taxes may be more politically palatable in states where a relatively small percentage of the population smokes. Fletcher (2014) considers a different tobacco-control policy – workplace smoking bans – and finds evidence that smoking cessation induced by these bans increases BMI.

To our knowledge, the only paper that uses a randomized intervention to estimate the causal effect of smoking on weight is Eisenberg and Quinn (2006; hereafter EQ). EQ use the Lung Health Study (LHS), which randomly assigned smokers to a comprehensive smoking cessation program and then tracked their health for five years. EQ do not actually use the LHS microdata, but instead take advantage of the fact that O'Hara et al. (1998) report differences between the treatment and control groups' average changes in weight and smoking status to compute a Wald instrumental variables (IV) estimate of the weight gain from quitting smoking. EQ find that quitting smoking leads to a very large average weight gain of 9.7 kg (21.4 pounds), about two to five times the magnitude typically found in the associational literature.

Despite their use of randomization, there are reasons to suspect that EQ's estimate is overstated. They use assignment into the smoking cessation program as an instrument for "sustained quitting", which is defined as being a medically verified quitter in all five LHS follow-up waves. The IV

strategy therefore requires the strong assumption that the program only affects weight through its effect on sustained quitting. However, in addition to helping some smokers quit immediately and permanently, the program may help others quit smoking in some but not all follow-up periods, and still others by reducing their number of cigarettes smoked per day even though they never quit entirely. To the extent that such partial quitting exists and influences BMI, EQ's estimated average effect of quitting smoking will be biased upwards.

We contribute to the literature on the effect of smoking on weight in several ways. First, we provide, in our view, the most reliable estimates to date of the average causal effect of quitting smoking on weight. We use the LHS microdata to exploit the randomized nature of the study while also constructing detailed smoking measures – such as cigarettes smoked per day and average carbon monoxide (CO) level over the entire five-year period of the study – that account for delayed or temporary quitting as well as smoking intensity. Our preferred estimates imply that quitting smoking leads to an average weight gain of 1.5-1.7 BMI units (10-11 pounds at the average height) at the end of the first year of the study period. The effect persists over time, reaching 1.7-1.9 BMI units (11-12 pounds) by the end of the fifth year – a magnitude that implies that the fall in smoking explains around 14% of the rise in BMI in recent decades. Our estimated effects are toward the high end of the range of results from the associational literature, but substantially smaller than EQ's estimate that uses randomization but relies on the "sustained quitting" measure.

Our paper also contributes by providing new information related to the heterogeneity of the effect across the smoking and weight distributions as well as by demographic characteristics. We estimate a semi-parametric instrumental variables model that allows the data to determine the functional form of the relationship between smoking and BMI. The results suggest a diminishing marginal effect, with additional smoking having little long-run impact on BMI beyond about a pack of cigarettes per day or a CO level of about 20 parts per million (ppm). We also conduct subsample analyses by age, gender, education, and baseline BMI and find that on average younger individuals, females, those with no college degree, and those with healthy baseline BMI levels gain the most weight in response to smoking cessation.

Data

This section provides a brief introduction to the LHS, with an emphasis on the information most relevant for our paper. O'Hara et al. (1993, 1998) provide a more detailed discussion of the LHS, and further information is also available online at https://www.clinicaltrials.gov.

The purpose of the LHS was to observe changes in the severity of chronic obstructive pulmonary disease (COPD) among smokers. The study consisted of 5887 smokers with initial ages between 35 and 59. Recruitment started in 1986 and ended in 1989. The clinical trial ended in 1994. To be eligible for selection, potential participants had to show signs of mild lung function impairment, have no history of certain medications, consume less than 25 drinks per week, and have no severe illnesses or chronic medical conditions. Each year all participants were extensively interviewed individually at a medical clinic near the residence of the participant (no more than 75 miles away from the participant's permanent residence). The data therefore consist of the baseline period (1989) plus five annual follow-up periods (1990 through 1994). Attrition was relatively low, as 5,297 individuals remained in the sample in the final wave. The attriters included 315 participants who died during the study period.

Participants were randomly assigned into three different groups: two treatment groups and one control group. Both treatment groups received a special intervention (SI) consisting of free nicotine gum, an intensive quit week, and frequent contact with support personnel with invitations to bring a spouse or relative to the meetings. The only difference between the two treatment groups is that, in addition to the SI, one group received an inhaled bronchodilator (SI-A) while the other received an inhaled placebo (SI-P). Most of the intensive intervention treatments were completed within the first 4 months of the study. The control group referred to as the usual care (UC) group received no intervention and members continued to use their own private sources for medical care.

The LHS collected information about weight, height, smoking behavior, family smoking habits, health status, and demographic characteristics. Weight and height were measured by medical staff at the participants' clinic visits, so our BMI measure is not susceptible to the concern about measure-

Table 11: Summary Statistics

(Means, with Standard Deviations in Parentheses)

Variable	SI-A	SI-P	UC
Age	48.41	48.55	48.43
	(6.84)	(6.83)	(6.80)
Cigarettes per Day	29.59	29.49	29.53
	(14.08)	(13.60)	(14.11)
Carbon Monoxide	25.97	26.70	25.98
	(13.47)	(12.67)	(12.66)
BMI	25.42	25.67	25.55
	(3.91)	(3.92)	(3.92)
High School Degree	0.29	0.30	0.29
	(0.45)	(0.46)	(0.45)
Some College	0.34	0.34	0.35
	(0.47)	(0.47)	(0.47)
College Degree	0.22	0.22	0.24
	(0.41)	(0.42)	(0.42)
Male	0.60	0.64	0.64
	(0.49)	(0.48)	(0.48)
Married	0.65	0.67	0.65
	(0.47)	(0.47)	(0.47)
Observations	1961	1962	1964

ment error that is common in the economics of obesity literature.[41] The data contain self-reported smoking information as well as CO test results. We consider three different measures of smoking: a dummy variable for whether the respondent currently smokes (clinically measured through the CO test), number of cigarettes typically smoked per day (self reported), and CO level in ppm. We also utilize the LHS' information on education (dummies for high school graduate, some college, and college graduate), gender (dummy for male), age (years), and marital status (dummy for married) as controls. Note that we do not control for race/ethnicity because 97% of LHS participants were white. Our sample is therefore not representative of the overall population of US smokers along this dimension.

Table 11 presents descriptive statistics for the three groups at the time of randomization. Average cigarette consumption was roughly 30 cigarettes per day, average CO level was about 26, and

[41] See Courtemanche et al. (2015) for an overview of the challenges involved with using self-reported weight and height.

the average respondent was just slightly overweight. The summary statistics for all variables are very similar across the three groups, indicating the randomization was successful.

Figure 3: Changes Over Time in BMI and Smoking for Treatment and Control Groups

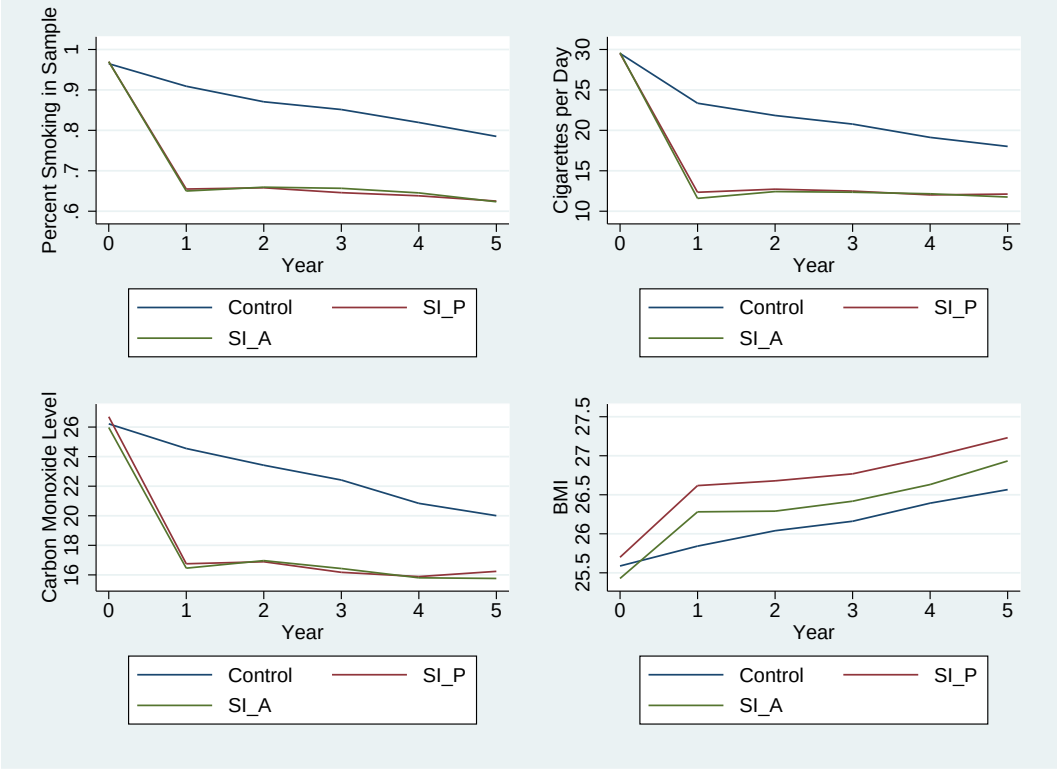


Figure 3 displays changes throughout the sample period in the average number of cigarettes smoked per day, objectively-verified smoking status, CO level, and BMI for each group. Sharp decreases in cigarette smoking, smoking status and CO level are evident for both treatment groups in the first year after the intervention. The decrease in smoking for the control group is much more moderate. Average BMI is trending upward for all three groups, but the two treatment groups experience much sharper increases in BMI than the control group in the first year. The graph therefore suggests both that the intervention was effective in reducing smoking and that smoking reduces BMI. We next use econometric methods to estimate the magnitude of these effects.

Econometric Analyses

Our econometric objectives are to identify 1) the average short- and long-run causal effects of quitting smoking on weight gain, 2) how the effect of smoking on weight changes across the smoking distribution, and 3) how the effect of smoking on weight varies by demographic characteristics and baseline BMI. We begin by using parametric regressions to answer the fist question and running falsification tests to evaluate the validity of the models. We then conduct semiparametric estimation allowing for a flexible relationship between smoking and weight to address the second question. Next, we answer the third question with subsample analyses. Finally, we address issues related to the generalizability of the results and show why our results differ from those of EQ.

Average Effects

Short Run

We begin by aiming to identify the average causal effect of quitting smoking on weight gain with a series of parametric regressions. Economists typically consider body weight to be a capital stock. Individuals start with an exogenous endowment of weight that changes over time due to depreciation as well as "investments" that take the form of caloric intake or expenditure. In the LHS, weight at the end of the first follow-up year can therefore be modeled as a function of weight at baseline and investments – such as smoking – in year one. This leads to the model

$$bmi_{i1} = \beta_0 + \beta_1 bmi_{i0} + \beta_2 S_{i1} + \beta_{3t}\mathbf{X}_i + \varepsilon_{i1} \tag{4}$$

where bmi_{i1} is individual i's BMI at the end of year 1, bmi_{i0} is BMI at the beginning of the study (year 0), S_{i1} is smoking in year 1, $\mathbf{X}_i$ is a vector of demographic controls that are assumed to be constant over time since they are only available for the baseline wave, and ε_{i1} is period 1's error term. β_2, the coefficient of interest, gives the short-run association between smoking and weight. We estimate the OLS model given by (4) as well as an IV model that uses the randomized treatment

assignment to instrument for S_i. The first stage of the IV model is given by

$$S_{i1} = \gamma_0 + \gamma_1 w_{i0} + \gamma_2 si\_a_i + \gamma_3 si\_p_i + \gamma_4 \mathbf{X}_i + \mu_{i1} \tag{5}$$

where si_a_i and si_p_i reflect whether the individual was assigned into the SI-A or SI-P treatment group, respectively. The second stage of the IV model is identical to (4) except it replaces S_{i1} with the predicted value generated by (5). In the IV model, β_2 can be interpreted as the short-run local average treatment effect (LATE) of intervention-induced changes in smoking on BMI. We estimate linear models in both stages due to their relative ease of interpretation, their ability to produce reliable average effects (e.g. Angrist and Pischke, 2009), and the inherent difficulties with non-linear IV estimation (e.g. Terza et al., 2008). We define S_{i1} three different ways: a dummy for smoking cessation, number of cigarettes smoked per day, and CO level. We next discuss these three smoking variables.

The first smoking measure is a dummy equal to one if and only if individual i was a medically validated non-smoker at the end of year one. We consider this to be a naive measure of smoking because it ignores variation in smoking intensity among smokers. This could lead to an overstatement of the average weight gain from quitting smoking estimated by IV models. When a quit dummy is used as the smoking measure, the IV estimator effectively scales the difference in BMI between the treatment and control groups by the difference in smoking cessation rates between the two groups. The validity of this estimator therefore hinges on the assumption that the randomized intervention only affected the BMIs of people who fully quit smoking. To the extent that the intervention also affected the BMIs of those who cut back on smoking but did not quit entirely, the difference in BMI will be scaled by too small a number and the resulting IV estimate will be too large. (This point is central to our critique of EQ, and we will discuss it in more detail in Section 3.5.)

Our second smoking measure is therefore self-reported number of cigarettes smoked per day, with a value of zero assigned to those who reported quitting. This measure incorporates both reducing smoking and quitting entirely and therefore is not susceptible to the above criticism. In

order to make the results using cigarettes per day comparable to $\widehat{\beta}_2$ from the regressions using the smoking cessation dummy, we need to compute an implied average weight gain from quitting smoking. Since the best way to do this is not immediately obvious, we consider several possibilities. First, we use the average weight that would be gained if all individuals in the sample switch from their baseline number of cigarettes to none. We do this by multiplying the coefficient estimate on the cigarettes smoked per day variable by each individual's number of cigarettes smoked at baseline, and then taking the average across all individuals. Formally, this means we compute $\left(\sum_{i=1}^{N}\beta_2 cigday_{i0}\right)/N$, where $cigday$ is cigarettes smoked per day and i indexes the N observations. Second, we use the same formula but average over only those individuals who actually quit smoking. Third, we only average over quitters from the treatment groups. Fourth, we compute the effect of quitting smoking from the sample mean baseline smoking level, i.e. $\beta_2\overline{cigday}_{i0}$. Fifth and sixth, we again compute $\beta_2\overline{cigday}_{i0}$ but using only quitters, and quitters from the treatment group, respectively. The estimated average weight gain from quitting smoking is similar using all six approaches, so we only report the results using the first method: averaging the predicted effects across all individuals. Results from the other approaches are available upon request.

A key limitation with cigarettes per day is its self-reported nature. At issue for the validity of our IV estimates is not whether cigarettes per day are reported with error, but whether this error is correlated with treatment status. It is not obvious that this is the case, but it is possible that, for instance, being assigned into the treatment group creates more pressure to report progress toward smoking cessation, leading to differentially large reporting error among the treatment group. Alternatively, perhaps reporting error simply rises with number of cigarettes smoked per day, in which case we would expect the amount of error in the follow-up periods to be highest among the control group.

We therefore also utilize a third smoking variable that is both clinically measured and incorporates both the intensive and extensive margins of smoking: CO level from a test conducted during the follow-up interview. Using the CO regression estimates to compute the average weight gain from quitting smoking is somewhat more complicated than using cigarettes smoked per day since

even non-smokers generally have a positive CO level. We therefore compute each individual's predicted effect of quitting smoking as the effect of switching from her baseline CO level to the mean CO level for non-smokers, rather than to a CO level of zero. For the mean CO level of non-smokers, we use Deveci et al.'s (2004) estimate of 3.61 ppm; this is similar to the mean CO level of verified non-smokers in the follow-up waves of the LHS. The average effect of quitting smoking on weight across the entire sample is therefore given by $\left(\sum_{i=1}^{N}\beta_2(CO_{i0}-3.61)\right)/N$ where CO is CO level in ppm. Note that CO levels are only available at baseline for 922 individuals, so our average effect is computed using only this portion of the sample (though our regressions still utilize the full sample). We doubt that this limitation is of consequence since reported numbers of cigarettes smoked per day at baseline are virtually identical for those with missing baseline CO levels and those with non-missing levels. We have also considered analogs of the other five approaches to computing the average effects of quitting smoking discussed above and verified that, as with cigarettes per day, the results are robust.

While using CO levels solves the probability of reporting error, it should be noted that it is not immune to all sources of measurement error. In particular, it only reflects smoking in the past couple of days. Therefore, for some people self-reported number of cigarettes smoked per day could actually be more indicative of typical smoking behavior than clinically measured CO. Consequently, we take an agnostic view about which measure is preferred and present the results for both alongside each other throughout the paper.[42]

Long Run

The above specifications estimate the short-run causal effect of smoking on BMI. We also aim to identify the long-run effect by asking how smoking across all five follow-up waves affects BMI at the end of the study (year five). Comparing the short- and long-run effects is important since, as

[42] If we regress CO level in the first follow-up year on cigarettes smoked per day as well as the interaction of cigarettes with the two treatment dummies, the R-squared is 0.48. This suggests that, while cigarettes smoked per day and CO are highly correlated, they do convey different information. Additionally, the coefficients on the interaction terms are positive and significant, though small. In other words, measurement error does appear to be slightly correlated with treatment status. Both of these results underscore the importance of verifying that the results are similar using the two different measures.

discussed in Section 1, the evidence from the associational literature is mixed as to whether at least some of the weight gained after quitting smoking is temporary. Ideally, we would like to estimate

$$bmi_{i5} = \beta_0 + \beta_1 bmi_{i0} + \sum_{t=1}^{5} \beta_{2t} S_{it} + \beta_3 \mathbf{X}_i + \varepsilon_{i5} \tag{6}$$

where bmi_{i5} is individual i's BMI at the end of year 5 and S_{it} is smoking in year t. However, the need to utilize IV estimation prevents us from allowing separate coefficients for each of the five smoking variables, as this would require five instruments. In other words, in order to operationalize an IV model we need to compress the five years of smoking information into a single variable S_i. The easiest way to do this is to take a simple average across the five years:

$$S_i = \frac{S_{i1} + S_{i2} + S_{i3} + S_{i4} + S_{i5}}{5}. \tag{7}$$

However, this approach assumes that smoking in each of the five periods has the same effect on weight. To the extent that weight is a depreciating capital stock, we might expect smoking in more recent years to have a larger effect on BMI than smoking in more distant years. We therefore also estimate models defining S as a weighted rather than simple average of quit status in the five follow-up years:

$$S_i = \frac{S_{i1} + (1-\delta)S_{i2} + (1-\delta)^2 S_{i3} + (1-\delta)^3 S_{i4} + (1-\delta)^4 S_{i5}}{1 + (1-\delta) + (1-\delta)^2 + (1-\delta)^3 + (1-\delta)^4}. \tag{8}$$

Since we do not have a sufficient number of instruments to credibly estimate the depreciation rate δ, we simply try several plausible values: 0.05, 0.1, 0.15, 0.2, 0.25, and 0.3. In all our regressions, the coefficient estimate on the baseline BMI variable will barely be below one, so we consider it probable that the "true" value of δ is toward the low end of this range; i.e. there is little reason to consider values of δ above 0.3.

We estimate equation (6) using both OLS and IV, with si_a_i and si_p_i again serving as the instruments. We again use the three different measures of S_{it}: smoking cessation, cigarettes smoked

per day, and CO level. For the cigarettes per day and CO regressions, we compute implied average effects of quitting smoking in the same manner as the short-run specifications.

Results

Table 12 reports the results of interest from the parametric regressions. Panel A presents the OLS and IV estimates of the effects of the different smoking measures on BMI. Panel B shows the estimated effects of the treatment dummies on the smoking variables from the first stage of the IV models, along with the F-statistic from a test of their joint significance. The first three columns show the effect of year 1 smoking on year 1 BMI (short-run effect), with the first column using the binary quitting variable as the smoking measure, the second using cigarettes smoked per day, and the third using CO. The last three columns present the effects of the simple averages of these three smoking measures across years 1-5 on BMI in year 5 (long-run effect). The results using the weighted averages, available in Appendix Table A1, are similar to those obtained using simple averages.

Coefficient estimates are shown in the table, with heteroskedasticity-robust standard errors in parentheses. The stars represent 0.1%, 1%, and 5% significance levels. For the regressions with the non-binary smoking measures cigarettes per day and CO, the implied average effects of quitting smoking on BMI are presented in brackets. In other words, the numbers in brackets from the cigarettes and CO regressions are comparable to the coefficient estimates from the quit status regressions. The row labeled "Hausman" gives the p-values from Hausman tests of the consistency of the OLS estimator compared to IV. The sample sizes, provided in the row labeled "N", vary somewhat across specifications due to differing amounts of missing information. In unreported regressions (available upon request), we re-estimated the models using only observations with no missing smoking information and verified that any meaningful differences between the results cannot simply be attributed to the difference in samples.

The first column presents the short-run estimates using the quit dummy. The OLS regression estimates that quitting smoking increases BMI by 1.295 units, or 8.2 pounds at the US average

height of 66.55 inches.[43] This is well within the range of estimates from the associational literature discussed in Section 1. The IV estimate is a larger 2.202 BMI units, and the Hausman test strongly rejects the consistency of OLS. This IV estimate equates to 13.9 pounds, which is larger than most estimates of the average short-run weight gain from quitting smoking from the associational literature.

The next two columns use the smoking measures that incorporate intensity: cigarettes per day and CO. In the IV specifications, we estimate that in the short run an additional cigarette smoked per day reduces BMI by 0.052 units while an additional ppm of CO reduces BMI by 0.077 units. The average effects of quitting smoking implied by these two regressions are 1.52 and 1.71 BMI units, which translate to 9.6 and 10.8 pounds at the mean height. These estimates are 31% and 22% smaller than the 13.9 pounds we obtained using the quit dummy. This is consistent with our prediction that neglecting to account for smoking intensity leads to an exaggerated IV estimate of the average weight gain from quitting smoking.

The last three columns turn to the long-run estimates. The key result is that the long-run effects are slightly stronger than the short-run effects. This is an important result, as the issue of whether the effect diminishes over time has been a point of contention in the associational literature, as discussed in Section 1. In the IV specification using average quit status, quitting for all five follow-up years is estimated to increase BMI by 2.646 units, or 16.7 pounds. An additional cigarette smoked per day over the five years reduces BMI by 0.065 units, while an additional ppm of average CO reduces BMI by 0.082 units. These latter two estimates imply average weight gains from quitting smoking of 1.91 and 1.81 units of BMI, or 12.0 and 11.4 pounds. As with the short-run estimates, these results suggest that incorporating smoking intensity is necessary to avoid overstating the magnitude of the weight gain from smoking cessation.

We can use these long-run results to estimate the percentage of the rise in BMI that can be attributed to falling smoking, under the admittedly strong assumption that the results generalize.

[43] Average height is computed by taking a simple average of the male and female heights given by http://www.cdc.gov/nchs/fastats/body-measurements.htm.

Table 12: Parametric Regression Results

	Short Run (BMI Year 1)			Long Run (BMI Year 5)		
	Quit	Cigarettes	CO	Average Quit	Average Cigarettes	Average CO
Panel A: Effects of Smoking Measures on BMI						
OLS	1.295***	-0.031***	-0.032***	2.299***	-0.052***	-0.060***
	(0.047)	(0.001)	(0.002)	(0.077)	(0.002)	(0.003)
		[0.91]	[0.72]		[1.52]	[1.33]
IV	2.202***	-0.052***	-0.077***	2.646***	-0.065***	-0.082***
	(0.148)	(0.004)	(0.006)	(0.285)	(0.007)	(0.011)
		[1.52]	[1.71]		[1.91]	[1.81]
Hausman	0.000	0.000	0.000	0.211	0.047	0.030
Panel B: Effects of Treatment Dummies on Smoking Measures (First Stage of IV)						
SI-A	0.277***	-11.952***	-8.025***	0.206***	-9.092***	-6.634***
	(0.013)	(0.489)	(0.464)	(0.012)	(0.431)	(0.416)
SI-P	0.272***	-11.390***	-7.728***	0.211***	-8.874***	-6.380***
	(0.013)	(0.495)	(0.472)	(0.012)	(0.437)	(0.417)
F Statistic	320.6	371.1	190.3	227.8	284.3	163.6
N	5345	5344	5274	5446	4966	4517

Notes: Heteroskedasticity-robust standard errors are in parentheses. ***,** and * indicate significance at the 0.1, 1, and 5 percent levels. For the non-binary smoking measures, the implied average effect of quitting smoking is in brackets. The controls for education, gender, marital status, age, and baseline BMI are included in all regressions.

This percentage is given by $\frac{dbmi}{dcigday}\frac{\Delta\overline{cigday}}{\Delta\overline{bmi}} * 100\%$. For $\frac{dbmi}{dcigday}$, we use the long-run IV estimate for cigarettes smoked per day: -0.065. $\Delta\overline{cigday}$ and $\Delta\overline{bmi}$ are the changes in the population means of cigarettes smoked per day and BMI among those at least eighteen years old. We compute these using the oldest and newest waves of the National Health and Nutrition Examination Survey (NHANES) that contain data on both smoking (self-reported) and BMI (medically measured): 1971-1974 (NHANES I) and 2011-2012. This period spans the entirety of the sharp rise in obesity, which did not begin until the late 1970s. During this time frame, average cigarettes smoked per day fell from 9.165 to 2.188, so $\Delta\overline{cigday} = 6.977$. Average BMI rose from 25.425 to 28.617, so $\Delta\overline{bmi} = 3.192$. Plugging in these numbers suggests that the drop in smoking explains 14.2% of the rise in BMI. We view this as a relatively substantial contribution to the trend. Courtemanche et al. (2015) examine the extent to which 27 different economic factors contributed to the rise in BMI, finding that the increased prevalence of big box grocers and restaurants explain 17% and 12%, respectively, while no other factors explain more than 4%.

Finally, the first-stage estimates in Panel B of Table 12 show that the treatment was effective in reducing smoking. In the short run, being assigned into the SI-A or SI-P groups increased the probability of quitting by 27-28 percentage points while decreasing cigarettes smoked per day by 11-12 and CO level by 8 ppm. In the long run, SI-A or SI-P assignment increased the fraction of the five follow-up years quit by 0.21 while decreasing average cigarettes per day by 9 and average CO by 6-7 ppm. The treatment variables are all highly significant in the first stage and the F-statistics from the test of their joint significance are easily large enough to conclude that they are sufficiently strong instruments. Also noteworthy is the fact that there is essentially no difference in the coefficient estimates for the two treatment variables; in other words, the inhaled bronchodilator given to the SI-A group did not influence smoking. This also means that, though our IV model is technically overidentified, the instruments are not sufficiently distinct to make an overidentification test informative or to consider instrumenting for two endogenous variables.

Falsification Tests

We next conduct falsification tests to support our contention that the IV results using cigarettes per day and CO are more credible than those using the quit dummy. In our IV models, the identifying assumption is that the randomized treatment only influences BMI via the smoking variable. Our falsification tests evaluate this assumption by asking whether the instruments influence the BMIs of individuals who did not change their smoking habits during the sample period, according to each smoking measure. A significant association would provide evidence that the randomized intervention influenced BMI through pathways other than the particular smoking measure, invalidating the causal interpretation of the corresponding IV estimate. For the quit measure, we restrict the sample to those with values of 0, meaning those who smoked in the first follow-up wave for the short-run analysis and those who smoked in all five follow-up waves for the long-run analysis. For the continuous smoking measures, there are obviously very few individuals with literally no change in smoking across the sample period (i.e. the exact same number of cigarettes smoked per day/CO level in the baseline period as in the follow-up waves), so a judgment call is required as to what magnitude change in smoking should be considered "meaningful". We report results restricting the sample to those whose post-treatment level of smoking is within 25% of their baseline level; results using neighboring cutoffs are similar. Using these subsamples, we estimate the reduced-form version of the short- and long-run IV models; i.e. we regress BMI on the two treatment variables plus the controls.

Ex ante, our prediction is that the binary quit measure will perform the worst in the falsification tests since it leaves people in the sample who did actually experience a meaningful change in smoking but did not quit entirely. Note that it is not obvious that the falsification tests will produce null results even for the smoking measures that incorporate intensity, though, since it is conceivable that the treatment could affect BMI through pathways other than smoking. For instance, perhaps being exposed to an intensive health-related intervention might increase some people's level of general health consciousness, which could lead to improved health behaviors along other dimensions besides smoking. The falsification tests are therefore important in assessing whether even our most

conservative estimates of the effect of smoking on weight can be given a causal interpretation.

Table 13 reports the results. The left half of the table presents the results from the short-run falsification tests ("effects" of the treatment dummies on year 1 BMI for those with unchanged smoking status) while the right half shows the long-run results (year 5 BMI). For comparison purposes, the first column of each half of the table presents the reduced-form results for the full sample. The remaining three columns of each half include those with no meaningful changes in the quit, cigarettes per day, and CO variables, respectively. The sample sizes in the tests based on cigarettes per day are smaller than those using quit status simply because more individuals are excluded as the measure of smoking becomes more comprehensive. The sample sizes in the two columns using CO are very small because, as discussed previously, much of the sample is missing baseline CO information, preventing the calculation of the percentage change. The falsification tests using cigarettes per day are therefore much more highly powered – and consequently more informative – than those using CO.

The columns labeled "full sample" show that, in both the short and long run, the reduced-form effects of the two treatment variables on BMI are between 0.54 and 0.61 before excluding any observations. Dropping those who quit smoking reduces the magnitude of these effects by about half, but significant effects of 0.21-0.28 remain. There is therefore clear evidence that the intervention affected BMI through a pathway besides quitting smoking, implying that the IV estimates using quitting smoking are too large. The falsification test results are much more favorable if we also exclude those with meaningful (>25%) changes in smoking intensity. In the two regressions that use cigarettes per day, the coefficient estimates for the treatment variables are small (between -0.014 and 0.091) and highly statistically insignificant. There is therefore no evidence that the exclusion restriction in the IV model is violated if $cigday$ is used as the smoking measure. Excluding on the basis of changes in CO also leads to highly insignificant effects, with three being negative (the opposite direction of the full-sample relationship) and one positive. The estimates are imprecise due to the small sample size, so these results are not as compelling as those using $cigday$, but the lack of a clear pattern is at least somewhat reassuring. To summarize, the results in this section suggest

Table 13: Falsification Test Results

	Short Run (BMI Year 1)				Long Run (BMI Year 5)			
	Full Sample	Quit=0	<25% Change in Cigs.	<25% Change in CO	Full Sample	Average Quit=0	<25% Change in Avg. Cigs.	<25% Change in Avg. CO
SI-A	0.596***	0.278***	0.091	-0.253	0.541***	0.210*	0.086	-0.368
	(0.048)	(0.051)	(0.130)	(0.365)	(0.075)	(0.082)	(0.121)	(0.342)
SI-P	0.614***	0.214***	0.049	-0.127	0.561***	0.252**	-0.014	0.204
	(0.049)	(0.052)	(0.127)	(0.378)	(0.074)	(0.084)	(0.118)	(0.367)
N	5345	3812	1703	231	5446	2958	1379	211

See notes for Table 12.

that while the IV results using the quitting indicator are contaminated by an alternative effect, the results using both number of cigarettes and CO levels are more reliable.

Semi-Parametric Estimation

An issue with the parametric regressions for $cigday$ and CO is that they assume that smoking intensity affects BMI linearly. This is a strong assumption, as it seems likely that there is either a non-linear dose-response effect of nicotine on metabolism/appetite or a non-linear effect of metabolism/appetite on weight-related behaviors. While it is not clear that this will bias estimates of the average weight gain from quitting smoking, such a restrictive functional form is likely to lead to systematically inappropriate predictions for at least some individuals. Moreover, given the complicated chain of biological and behavioral pathways through which smoking influences BMI, the nature of the non-linearity is not clear *ex ante*. In other words, it is not obvious that the non-linearity could be captured through simple approaches such as logarithmic or quadratic specifications. We therefore next estimate a semi-parametric model that allows the data to determine the functional form of the relationship between smoking and BMI. Specifically, we implement Robinson's (1988) semi-parametric double residual estimator with local smoothing. This approach allows us to model the expectation of the dependent variable at every point on the distribution of the independent variable, thereby enabling the prediction of the weight gained (or lost) from switching from any level

of smoking to any other level.[44]

Semi-parametric IV models can be estimated using a control function approach (Blundell and Powell, 2004; Lee, 2007). The first stage takes the same form as equation (5). The second stage differs from equation (4) in two ways. First, it does not specify the functional form for the smoking measure. Second, rather than using the predicted value of the smoking variable from the first-stage regression, the second stage includes the residual from the first stage as a regressor.[45] The second stage short-run regression can therefore be expressed as

$$bmi_{i1} = \beta_0 + \beta_1 bmi_{i0} + f(S_i) + \beta_2 \mathbf{X}_i + \beta_3 \widehat{\mu}_i + \varepsilon_i \tag{9}$$

where S is either $cigday$ or CO and $\widehat{\mu}$ is the first-stage residual. The second stage long-run regression is similar but replaces bmi_{i1} with bmi_{i5} and S_i with the average smoking measures discussed previously.

The estimation was conducted using the Stata program "semipar" by Deparsy and Verardi (2012). The first step is to estimate E($bmi|S$), E($\mu|S$) and E($\mathbf{X}|S$), which are approximated by the predicted values $\widehat{bmi}$, $\widehat{\mu}$, and $\widehat{\mathbf{X}}$ by a kernel weighted local polynomial regression. The second step is to form the residuals $\widehat{\mu}_1 = bmi - \widehat{bmi}$, $\widehat{\mu}_2 = \mathbf{X} - \widehat{\mathbf{X}}$, $\widehat{\mu}_3 = \mu - \widehat{\mu}$. Then the coefficients $\widehat{\beta}_0$, $\widehat{\beta}_1$, $\widehat{\beta}_{\mathbf{2}}$, and $\widehat{\beta}_3$, representing the relationships between the independent variables and BMI, are estimated by regressing $\widehat{\mu}_1$ on $\widehat{\mu}_2$ and $\widehat{\mu}_3$. Thus, all parameters in equation (9) are identified except the relationship between cigarette consumption and BMI. The last step is, therefore, to identify this relationship with a non-parametric regression of cigarette consumption on the predicted BMI residual, $\hat{bmi}_{i1} - \widehat{\beta}_0 - \widehat{\beta}_1 bmi_{i0} - \widehat{\beta}_2 \mathbf{X}_i - \widehat{\beta}_3 \widehat{\mu}_i$. This relationship is estimated at every level of cigarette smoking, allowing independent marginal effects. The idea behind this strategy is to estimate the non-parametric cigarette function by the residual variation that is unrelated to the parametric independent variables.

[44] For simplicity, we round smoking values to the nearest integer; e.g. if someone averaged 21.2 cigarettes per day over the five follow-up waves we round this to 21.

[45] For an overview of the control function approach to dealing with endogeneity, see Heckman (1979) and Heckman and Robb (1986).

We calculate the average effect of quitting smoking on BMI using the semi-parametric estimates as follows. When using the cigarettes smoked per day variable, we first calculate the change in predicted weight from switching from the number of cigarettes smoked at baseline to zero. We then take the average of these predicted changes across all individuals in the sample. The process for the CO variable is similar; the only difference is that we compute the predicted effect of switching to the average CO level for non-smokers of 3.61 ppm, as opposed to zero.

In semi-parametric estimation, the confidence interval becomes very wide at extreme values where there are very few observations. We therefore drop the top 1% of the smoking distribution, which means those who smoke more than 50 cigarettes per day on average across the five follow up years and those with average CO levels of over 50 ppm. We doubt that this restriction is consequential, since if we drop the same individuals in the parametric regressions the results (available upon request) remain similar.

Figure 4: Estimated Short-Run Effect of Cigarettes Per Day on BMI from Semi-Parametric Model

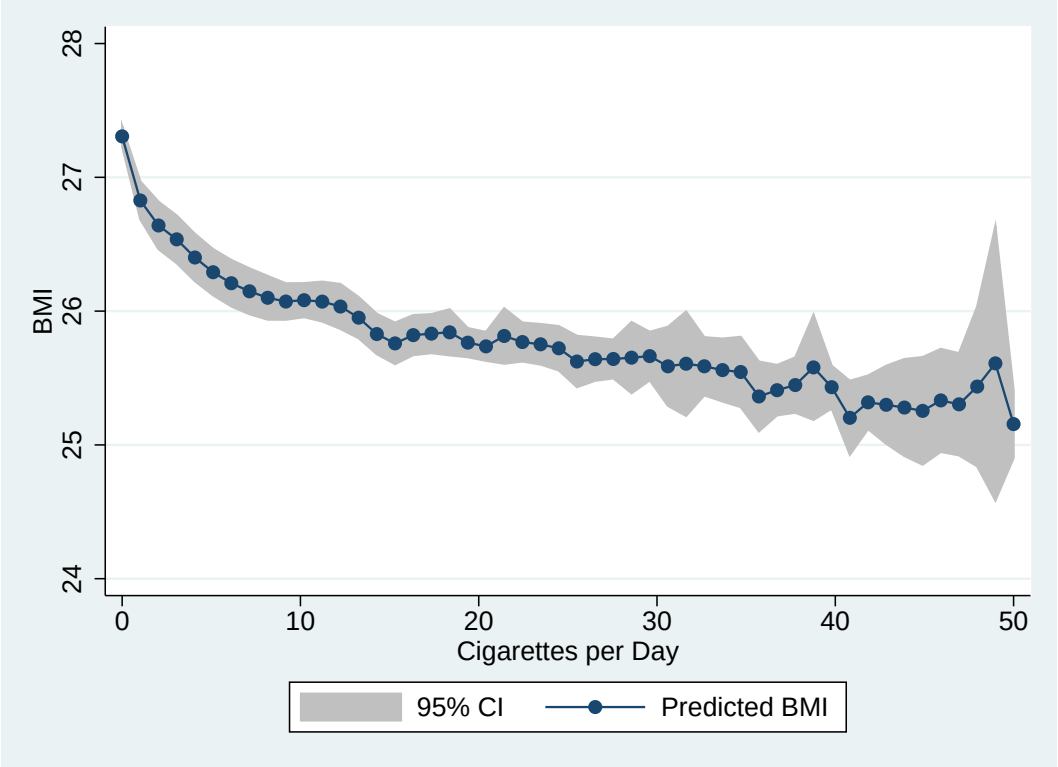
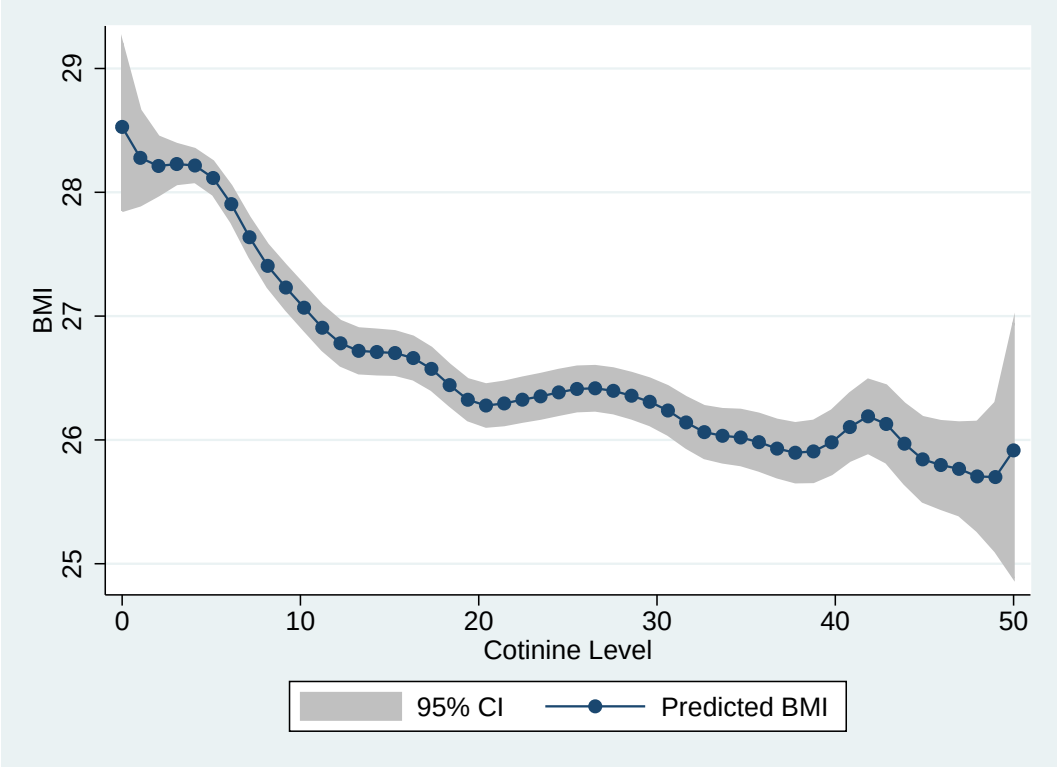


Figure 5: Estimated Short-Run Effect of CO Level on BMI from Semi-Parametric Model

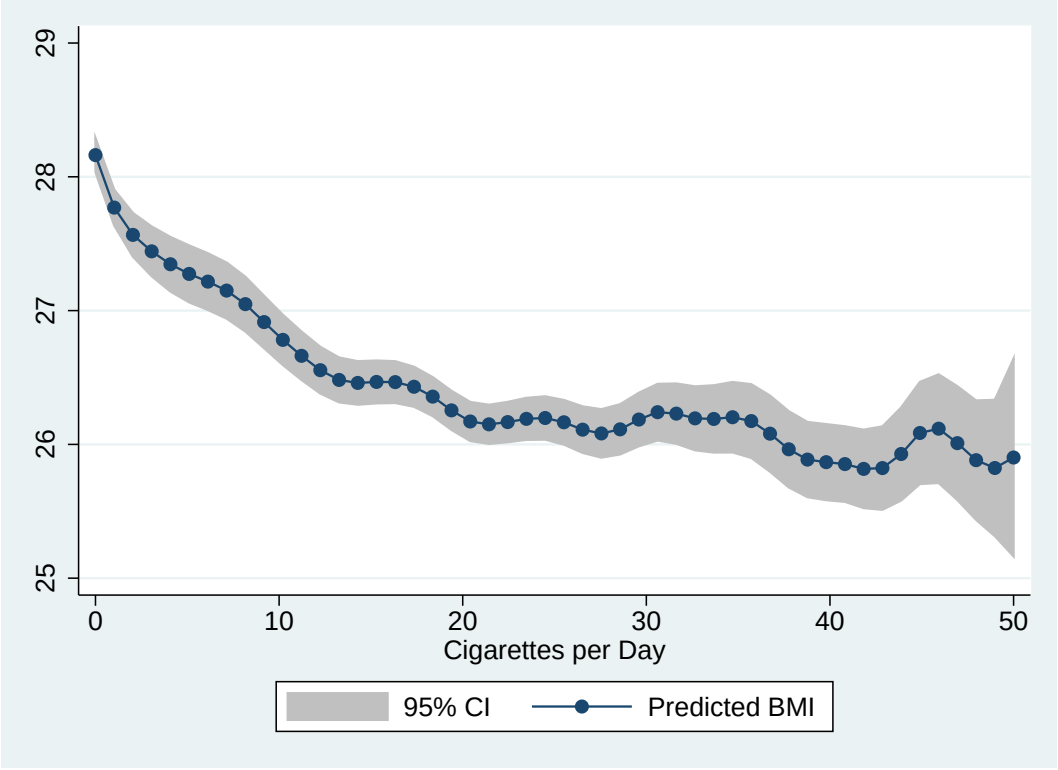


Figures 4 and 5 present the short-run semi-parametric IV results for $cigday$ and CO, respectively. The graphs display both the point estimates for each integer level of smoking and the 95% confidence intervals.Figure 2 shows that the short-run relationship between cigarettes smoked per day and BMI is highly nonlinear. Specifically, smoking has a diminishing marginal effect on BMI throughout most of the distribution, with the shape of the curve being approximately quadratic. Quitting smoking from levels of 10, 20, 30, and 40 cigarettes per day is predicted to lead to weight gains of 1.22, 1.58, 1.66, and 1.94 BMI units, respectively. Most of the effect of smoking on weight therefore appears to occur at levels below 20 cigarettes per day. Taken literally, this would suggest that heavy smokers could cut back to a pack a day without fear of substantial weight gain. Figure 3 shows that the short-run effect of CO on BMI is less obviously non-linear than the effect of cigarettes per day. The curve is somewhat flat at very low levels of CO – specifically two to five ppm – but recall that even non-smokers often have non-zero CO so changes at such low levels probably do not reflect changes in smoking behavior. Starting at five ppm, the graph begins to take

a quadratic shape, but unlike the graph for cigarettes per day we do not observe a complete leveling off until the far right tail of the distribution.

Figures 6 and 7 turn to the long-run results using simple averages of the smoking measures; the graphs using weighted averages are very similar and are available upon request. Figure 6 shows that the shape of the long-run relationship between cigarettes per day and BMI is roughly similar to the shape of the short-run relationship, as it is approximately quadratic and levels off at around 20 cigarettes per day. Figure 5 displays a similar pattern of results for CO level. CO has a diminishing marginal effect on BMI, and most of the weight gain from reduced CO comes at levels below about 20 ppm. The long-run relationship between CO and BMI therefore flattens out more quickly than the short-run relationship.[46]

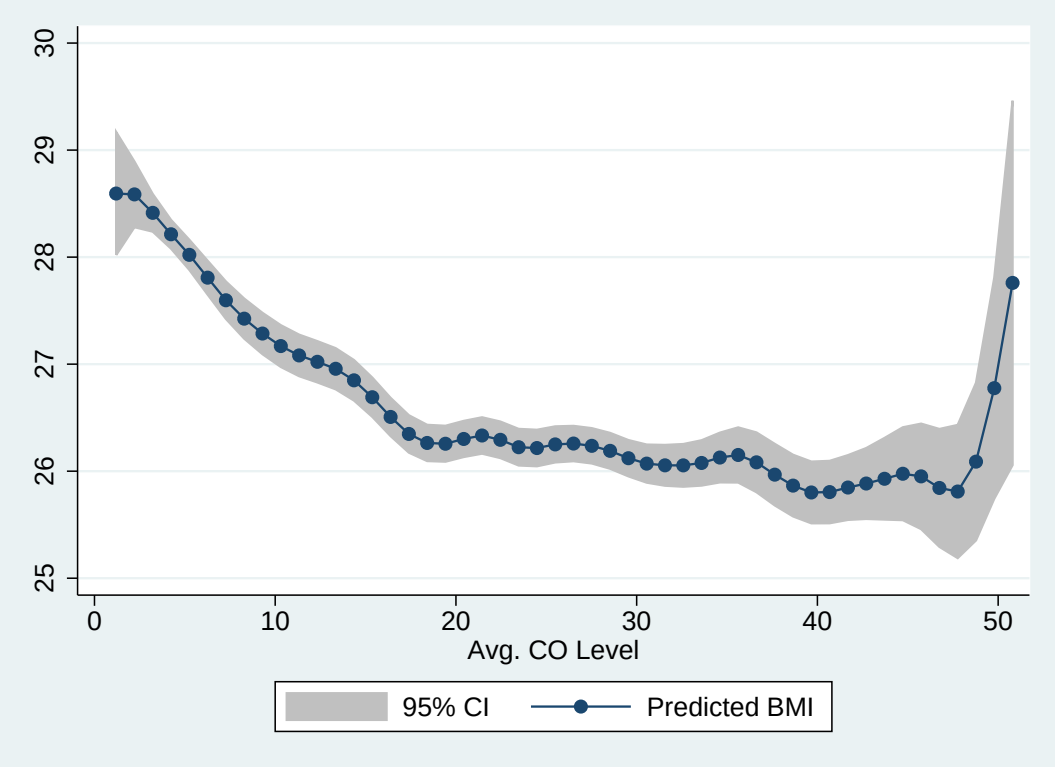
Figure 6: Estimated Long-Run Effect of Simple Average Cigarettes Per Day on BMI from Semi-Parametric Model



[46]Note that there is some evidence that additional CO actually leads to higher BMI at the far right tail of the distribution: CO levels of around 47-50 ppm. However, this should be interpreted with caution as it is based on a very small number of individuals. Accordingly, the confidence intervals in this portion of the distribution are quite large.

The average effects of quitting smoking on BMI implied by these semi-parametric graphs are generally similar to those from the parametric specifications. Using cigarettes per day, the average effect of quitting is 1.67 BMI units in the short run and 1.93 in the long run, compared to 1.52 and 1.91 from the corresponding parametric regressions. For CO, the average effect is 1.80 in the short run and 1.99 in the long run, compared to the parametric regressions' estimates of 1.33 and 1.81. The results presented in this section suggest that the marginal effect of smoking on weight is likely to be modest for levels of smoking above 20 cigarettes a day, which would be impossible to detect using linear specifications.

Figure 7: Estimated Long-Run Effect of Simple Average CO on BMI from Semi-Parametric Model



Subsample Analyses

We next conduct subsample analyses to evaluate whether the effect of smoking on BMI differs by age, gender, education, or baseline BMI. Heterogeneous effects could occur because of differences in either the biological effects of nicotine on appetite or metabolism or the behavioral responses to these biological effects. Given the complicated nature of these relationships, we make no *ex ante* predictions about the patterns of heterogeneity. For age, we split the sample into three groups: those under 45, 45-54, and 55 and over at baseline. We use these splits because there are no individuals under 35 or over 64 in the LHS. For education, we consider subsamples of those with no college education, some college, and a four-year college degree or greater. There are not enough individuals with less than a high school degree or greater than a college degree to enable further stratification. For baseline BMI, we are interested in whether quitting smoking leads to larger weight gains among those who were already at risk of weight-related ailments prior to the intervention. We therefore split the sample into those with healthy weights at baseline (BMI<25) and those who were already overweight or obese (BMI$\geq$25).

Table 14 displays the results for cigarettes per day (Panel A) and CO (Panel B). For brevity, Table 14 contains only the results from long-run parametric IV regressions using simple averages of the smoking measures. Semi-parametric graphs for each subsample are available in Appendix Tables A1-A4. Short-run estimates and those using weighted rather than simple averages lead to broadly similar conclusions and are available upon request.

The results suggest that the effect of smoking on weight is strongest for younger individuals, women, those without a college degree, and those with healthy baseline BMIs. Quitting smoking leads to an average weight gain of 2.19-2.21 BMI units for those under 45, 1.88-1.98 for 45-54 year olds, and 1.33-1.45 for those 55 and older. One possible explanation is that the health consequences from obesity become more salient with age, so older individuals may have a stronger incentive than others to mitigate weight gain after smoking cessation. Again averaging over the four specifications, the average effect of quitting smoking on BMI is 2.04-2.45 units for women compared to 1.56-1.59 for men. Stratifying by education, the average effects of smoking cessation are around

1.83-1.98 for those with no college education, 1.82-1.93 for those with some college but no degree, and 1.64-1.73 for those with a college degree. There is therefore some evidence of a small reduction in the effect of smoking on weight as education rises, with the largest gap being between those without a college degree and those with a degree. Perhaps education enables individuals to limit weight gain through an improved understanding of nutrition and exercise. Alternatively, education is correlated with income, and additional income may enable the purchase of healthier foods, gym memberships, or over-the-counter products that can help counteract weight gain. Finally, the average weight gain from quitting smoking is 1.98-2.05 BMI units for people with healthy baseline BMI levels compared to 1.7-1.82 for those who started the study overweight or obese. This result suggests that individuals who are at higher risk of health consequences from weight gain take more steps than others to limit the amount of weight gained after smoking cessation.

In all, though, perhaps the most striking results from Tables 14-15 is that, while some heterogeneity appears to exist, the overall amount of heterogeneity is relatively small. Negative and highly significant effects of smoking on weight are evident for all subsamples. The smallest average effect of quitting smoking on BMI from any specification (55 and over, parametric, CO) is a still sizable 1.33. The lack of substantial heterogeneity in the effect within the sample provides perhaps some assurances that the results are generalizable outside the sample. The next section evaluates the generalizability issue in more detail.

External Validity

We next perform some checks related to external validity. One obvious concern about the generalizability of the results is that the LHS was conducted in the early 1990s, raising the question of the relevance for current policy debates. (With that said, many of the frequently cited associational estimates are from studies using data that are as old or older.) Another concern related to generalizability is that the LHS' participants are not a random sample of smokers: participants had to desire to quit smoking, have mild (but not major) lung function impairment, and live within reasonable proximity of the locations for follow-up visits. As discussed in the Data section, the end result was

Table 14: Subsample Results for Age and Gender

(IV with Year 5 BMI Only)

		Age			Gender	
		<45	45 to 54	≥55	Women	Men
Average Cigarettes		-0.074***	-0.066***	-0.052***	-0.091***	-0.051***
		(0.014)	(0.010)	(0.015)	(0.015)	(0.008)
		[2.21]	[1.98]	[1.45]	[2.45]	[1.56]
	N	1514	2275	1177	1894	3072
Average CO		-0.093***	-0.083***	-0.069***	-0.101***	-0.069***
		(0.021)	(0.015)	(0.021)	(0.020)	(0.012)
		[2.19]	[1.88]	[1.33]	[2.04]	[1.59]
	N	1374	2065	1078	1714	2803

See notes from Table 12.

a sample that was almost exclusively white (97%) and exclusively middle-aged (starting age 35-59, ending age 40-64).

We attempt to at least somewhat alleviate these concerns by conducting additional analyses with the National Health Interview Survey (NHIS), a large nationally representative survey conducted annually by the Centers for Disease Control and Prevention. The NHIS contains self-reported data on smoking, weight, and height, along with the same control variables used in our LHS analyses (except for baseline BMI, since the NHIS is not a panel). We use the NHIS to see if the association between cigarettes smoked per day and BMI varies along the dimensions of the generalizability issues: time period, race, and age. Obviously a causal analysis is not possible with the NHIS, but verifying that the association between smoking and weight is not particularly unique among the LHS population should provide at least some assurance that the causal effect is not likely to be unique either. We first estimate the association among the NHIS' best available analog to the LHS sample: white 35-64 year olds in 1990-1994 (the years of the five LHS follow-up waves). We then evaluate whether this association has changed over time by estimating the same model among 35-64 year old whites in the five most recent NHIS waves currently available: 2009-2013. Next, we examine the issue of lack of representativeness by race by returning to the 1990-1994 NHIS waves and restricting the sample to 35-64 year old non-whites. Finally, we estimate the model for

Table 15: Subsample Results for Education and Baseline BMI

(IV with Year 5 BMI Only)

	Education			Baseline BMI	
	No College	Some College	College Graduate	< 25	≥ 25
Average Cigarettes	-0.068***	-0.065***	-0.059***	-0.072***	-0.061***
	(0.011)	(0.014)	(0.013)	(0.010)	(0.011)
	[1.98]	[1.93]	[1.73]	[2.05]	[1.82]
N	2082	1722	1162	2369	2597
Average CO	-0.088***	-0.080***	-0.071***	-0.090***	-0.077***
	(0.016)	(0.021)	[0.019]	(0.014)	(0.016)
	[1.83]	[1.82]	[1.64]	[1.98]	[1.70]
N	1895	1555	1067	2161	2356

See notes from Table 12.

whites of an age outside of the 35-64 range (i.e. 18-34 year olds combined with those 65+) in order to evaluate the implications of the lack of representativeness by age.[47]

Table 16 reports the results. The first column shows that, in the sample most comparable to the LHS, each additional cigarette smoked per day is associated with a reduction in BMI of 0.038 units. This implies an average weight gain from quitting smoking of 0.8 BMI units. The second column shows that the association between cigarettes smoked per day and BMI is stronger in the 2009-2013 sample than the 1990-1994 sample (-0.061 compared to -0.038), but the average effects of quitting smoking are nonetheless fairly similar (0.91 BMI units compared to 0.8) which is similar to the short-run OLS estimate from the LHS. This is because the average number of cigarettes smoked among smokers has dropped over the past two decades. In other words, β_2 may have grown over time but $cigday$ has shrunk for the average smoker, leaving $\left(\sum_{i=1}^{N}\beta_2 cigday_{i0}\right)/N$ roughly constant. Next, the third column provides evidence that the association between smoking and BMI for non-whites is stronger than for whites, but the implied average effects of quitting smoking are similar. Again, this is because on average non-white smokers consume fewer cigarettes than white smokers.

[47] The associations of the control variables with BMI are very different for the 18-34 year old age group and the 65+ age group. Therefore, in the regression combining 18-34 year olds with those 65 and older, we include as additional covariates the interactions of each control with an indicator for whether the individual is in the 18-34 portion of the sample or the 65+ portion.

Table 16: Comparison of Associations Between Smoking and BMI in Different NHIS Samples

(OLS Results Only)

	1990-1994 White Age 35-64	2009-2013 White Age 35-64	1990-1994 Non-White Age 35-64	1990-1994 White Age Not 35-64
Cigarettes	-0.038***	-0.061***	-0.059***	-0.040***
	(0.002)	(0.0 05)	(0.007)	(0.002)
	[0.80]	[0.91]	[0.79]	[0.68]
N	51253	53136	10114	60424

Notes: The controls for education, gender, marital status, and age are included in all regressions. NHIS sampling weights are used. See other notes from Table 12.

The final column shows that the association between smoking and BMI among those who are not between the ages of 35 and 64 is virtually identical to the association among those who are in this age range. The average effect of quitting smoking is, however, slightly smaller among the non-35-to-64 sample due to a lower number of cigarettes smoked among smokers.

In sum, though there is likely some heterogeneity across age, race, and time, these results provide at least some assurance that the lack of representativeness of the LHS is not driving our conclusions. Smoking is inversely associated with weight in all NHIS subsamples. The associations between cigarettes smoked per day and BMI all fall within a reasonably tight range of -0.038 to -0.061. The implied average effects of quitting smoking are all between 0.68 and 0.91 BMI units, which equate to 4.3 to 5.7 pounds. These magnitudes are within the range found in the associational literature and are well below the estimates from our LHS IV specifications. This underscores the importance of accounting for endogeneity when evaluating the relationship between smoking and weight.

Reconciling Our Results with Prior Literature

We close our empirical analysis by reconciling our results with those of EQ, who used previously published LHS summary statistics from O'Hara et al. (1998) to estimate a very large 21.4 lb av-

erage weight gain from smoking cessation. We first replicate EQ's results and then show that our finding of a considerably smaller effect can be attributed to our use of more comprehensive smoking measures.

We replicate EQ by computing a Wald IV estimate of the form

$$\hat{\beta}_{WALD} = \frac{\overline{bmi}_1 - \overline{bmi}_0}{\overline{quit}_1 - \overline{quit}_0} \tag{10}$$

where subscript one indicates the treatment group (combination of the SI-A and SI-P groups) and zero the control group (UC). $\overline{bmi}_1$ and $\overline{bmi}_0$ are average BMIs among the treatment and control groups, respectively, at the end of the study period (year 5). $quit$ represents EQ's measure of quitting smoking, called "sustained quitting," which is a dummy variable equal to one if and only if the individual was a medically verified non-smoker in *all five* follow-up waves. This is a very stringent measure, as anyone who smokes any amount in any of the five follow-up years is classified as a non-quitter.

The validity of the Wald estimator hinges on the assumption that the intervention only affected the weight of individuals for whom $quit = 1$. To the extent that the intervention also affected the weight of any other individuals (i.e. those with $quit = 0$), the denominator will effectively be too small. The observed difference in average weight between the treatment and control groups will therefore be scaled by too small a number, and the estimated effect of quitting smoking on weight will consequently be overstated. We suspect that the Wald estimator's identifying assumption is violated since there are two types of individuals categorized by EQ as having $quit = 0$ whose smoking behavior (and therefore weight) likely responded to the intervention to at least some extent.

The first type consists of those who quit smoking for part but not all of the 5-year follow-up period. If, for instance, someone quit smoking for the first two years, relapsed in year three, and then quit again for years four and five, this person is not classified as a quitter by EQ, but it seems likely that they would have gained almost as much weight as someone who quit for all five years. There are 1114 people in the treatment group who quit smoking in at least one follow up

wave but were not sustained quitters. Therefore, not accounting for this group has the potential to substantially impact the results.

The second type consists of those who reduced smoking but did not quit entirely. Given the highly addictive nature of cigarettes, it seems likely that there are at least some people who were able to cut back on their cigarette intake as a result of the intervention but were unable to quit completely. Indeed, among those in the treatment group who never quit in any of the five follow-up waves, average cigarettes smoked per day still fell from 31 to 22. There is no reason to suspect that the biological pathways through which smoking affects weight occur only along the extensive margin of smoking, so people who cut back on smoking would likely experience at least some amount of weight gain. Additionally, some people may also be a blend of the two types; e.g. someone who responds to the intervention by gradually cutting back on smoking until successfully quitting at the end of the third year.

After replicating EQ's results using the "sustained quitting" variable, we then re-compute the Wald estimate using our more nuanced long-run smoking measures discussed earlier in Section 3.1.2. Our "average quitter" measure addresses the issue of people who quit in some but not all follow-up years. The average cigarettes per day and average CO variables also address the issue of cutting back but not quitting entirely.

Table 17 reports the results. The first column shows that, replicating EQ's Wald estimator, we obtain an average estimated weight gain from quitting smoking of 3.196 BMI units, or 20.13 lbs at the average height. This is very similar to the result obtained by EQ, differing slightly because EQ used weight as the dependent variable rather than BMI. (We are unable to directly use weight because the LHS microdata suppress height and weight and only provide BMI.) The second column shows that using simple average quitter rather than sustained quitter reduces the average estimated weight gain from quitting smoking by about 17% to 2.655 BMI units. In the last two columns, we see that using the simple averages of cigarettes per day and CO attenuates this magnitude even further, to 1.84 and 1.58 BMI units, respectively. Ultimately, then, accounting for both temporary/delayed quitting and smoking intensity reduces the estimated average weight gain

Table 17: Reconciling Our Results with those of EQ

(IV with Year 5 BMI Only)

	Sustained Quit (EQ)	Average Quit	Average Cigarettes	Average CO
Smoking Variable	3.196***	2.655***	-0.063***	-0.072***
	(0.736)	(0.605)	(0.015)	(0.021)
			[1.84]	[1.58]
N	5446	5446	4966	4517

Notes: No control variables are included. See other notes for Table 12.

from smoking cessation by 42%-51% relative to using the naive sustained quitter measure. Since the Wald estimates using our preferred smoking measures from Table 17 are quite similar to those from our preferred long-run specifications in Table 12, we conclude that the difference between our results and those of EQ is due to the different smoking measures rather than our use of a covariate-adjusted regression model in Table 12. This is not surprising given the randomized design. Note, however, that the standard errors are lower in Table 12, so including covariates is still beneficial in that it improves the precision of the estimates.

Conclusion

This paper aimed to provide the most credible answers to date to several questions related to the relationship between smoking and weight. First, what is the average short-run causal effect of quitting smoking on body weight? Our preferred estimates suggest that this effect is around 1.5-1.7 BMI units, or 10-11 pounds at the average height. Second, does the weight gain from quitting smoking disappear over time? The answer appears to be no, as the weight gain actually becomes slightly larger in the long run. The long-run effect is around 11-12 pounds, which implies that the fall in smoking explains about 14% of the rise in obesity in recent decades. Third, how does the impact of smoking on weight vary across the smoking distribution? We find evidence of a diminishing marginal effect, with additional smoking having little long-run impact beyond about a pack of cigarettes per day or a CO level of 20 ppm. Finally, how does the effect of smoking on weight vary by age, gender, education, and baseline BMI? Our results suggest that, while quitting smoking

leads to sizeable weight gain for all subsamples, the impacts are largest for younger individuals, females, those with no college degree, and those with healthy baseline BMI levels.

Our estimated average effects of quitting smoking on weight fall within the range of estimates from the associational public health literature, albiet toward the high end of the range. It is clear, though, that our estimates are markedly smaller than those of EQ despite the fact that they utilized the same randomized intervention. The fact that our results are closer to the associational estimates than to those of EQ illustrates a broader methodological point about the dangers of using IV estimation uncritically even when the instrument is randomized or as-good-as-randomized. The randomization merely ensures the validity of the estimated reduced-form relationship between the instrument and outcome. Obtaining a reliable second-stage estimate requires the assumption that the endogenous variable is the only pathway through which the randomized instrument affects the outcome. This can be a difficult assumption to satisfy. As our paper shows, even if conceptually there is only one pathway through which the intervention can plausibly impact the outcome, careful measurement of that pathway is critical.

Our results also have interesting implications for the economics literature on tobacco control policies. As discussed in the introduction, the literature on the effect of cigarette costs (prices or taxes) on BMI reaches conflicting conclusions, with several studies suggesting the effect is either very small or negative – implying that quitting or reducing smoking actually leads to weight *loss*. Is it possible that the causal effect of an aggressive smoking cessation program is to increase BMI while the casual effect of higher cigarette costs is either zero or negative? This seems conceivable for two reasons.

First, the LATE from a price-induced reduction in smoking may differ from the LATEs from smoking ban-induced or aggressive smoking cessation program-induced reductions in smoking. Different smokers could be affected by these different types of interventions; for instance, it seems reasonable to think marginal smokers would be the ones to respond to cigarette price increases whereas those with strong addictions would be the ones to volunteer for a comprehensive program. Perhaps those with strong addictions are relatively more likely to "quit at all costs", even if it means

gaining a substantial amount of weight.

Second, perhaps cigarette prices/taxes affect BMI through pathways besides smoking behavior. In other words, people who quit smoking in response to higher cigarette prices may gain weight, but this could be counteracted by weight losses among those whose cigarette consumption is unchanged – a large share of the population given the price inelasticity of cigarettes (Chalopuka and Warner, 2000). Smokers who do not reduce their consumption when prices rise experience potentially sizeable negative income effects, which could lead to weight loss by reducing overall food consumption or frequency of eating out at restaurants. Moreover, cigarette taxes generate revenue for the state, which can be used to provide funding for nutrition education or health-related programs such as Medicaid. These, in turn, could reduce the BMIs of even non-smokers.

To close, we should emphasize that our results should not be interpreted as suggesting that individuals should be reluctant to quit smoking out of fear of gaining weight. The large body of epidemiologic evidence that smoking is bad for health implies that any increase in obesity-related ailments after quitting smoking is far outweighed by the health improvements along other dimensions. Instead, our findings should be interpreted as a call for further investigation into medical and policy interventions that can limit the weight gain from smoking cessation. If this side effect of an otherwise healthy decision could be eliminated, it stands to reason that the health gains from tobacco control efforts would become even greater.

To that end, Farley et al. (2012) provide a review of the literature on the effectiveness of various interventions in limiting the weight gain after quitting smoking. Some evidence suggests that the drugs dexfenfluramine, phenylpropanolamine, naltrexone, bupropion, and fluoxetine can reduce weight gain in the short run, though there is insufficient evidence to draw clear conclusions about whether the effects persist after the drugs are discontinued. Weight management education alone does not seem to reduce weight gain and might actually hinder efforts to quit smoking, whereas weight management education combined with personalized support appears more successful. For all types of interventions reviewed by Farley et al. (2012), their ability to draw clear conclusions was hindered by a lack of available research and small sample sizes in the studies that do exist.

Much more research is needed regarding which interventions can help limit weight gain following smoking cessation as well as how best to incentivize (e.g. more generous insurance coverage) interventions that prove effective.

III Misreporting Smoking Status and Consequences for Self-Reported Survey Data

Introduction

Smoking has been the number one cause of preventable death in the last 60 years. Today, the estimated cost of smoking totals 300 billion dollars per year in the US alone (2014 Surgeon General Report). As a result, federal and state governments invest large amounts of resources to decrease smoking participation. Over the past 50 years, smoking participation among U.S. adults has decreased by more than half, from around 40% in the mid-1960s to 20% by 2015 (CDC 2014).

Scientists and policymakers are interested in understanding how public policies contributed to the decline in smoking. A large body of literature suggests that policies, such as cigarette tax increase and the introduction of smoke free zones, contributed substantially to the decline in smoking (Carpenter et al. 2008, DeCicca et al. 2008, Levy et al. 2003, Evans 1999,). However, most of the studies utilize self-reported smoking information. Self-reported smoking information can be misreported and therefore may lead to biased estimates. Overall, not much is known about the level of misreporting and who misreports smoking status.

This paper analyzes misreporting in smoking status among adults. We utilize data from the Lung Health Study (LHS), a randomized smoking cessation study that includes self-reported and objectively verified smoking information. The data allows us to generate a binary misreporting variable capturing discrepancies between the self-reported and objectively verified smoking variables.[48] In our initial methodological approach we take the clinical smoking measure as the

[48]Objectively verified smoking status is estimated based on Cotinine saliva and Carbon Monoxide (CO) tests during the hospital visit.

gold-standard. First, we address whether misreporting is a pure mechanical result of strict Carbon Monoxide (CO) and Cotinine cutoff levels, where clinicians classify participants either as smokers or non-smokers.[49] Second, we evaluate whether misreporting differs by treatment group. In other words, do participants who receive treatment misreport at different levels than those in the control group. Third, we regress our misreporting variable on demographic characteristics including household smoking attributes. We also perform sub-sample analyses by treatment group.

In our second methodological approach we take an agnostic view on whether the objective smoking measure should be preferred to the self-reported measure. We use objectively verified BMI, CO and Cotinine data to inform whether a person is in reality a "true" smoker. We estimate a Bayesian mixture model which relaxes the assumption that observations are drawn from one distribution, but rather two underlying distribution. We believe that the LHS is a well suited dataset for this exercise for several reasons. First, the LHS consists of a sample of heavy smokers at the beginning of the trial. Second, upon receiving the very effective smoking cessation treatment, many participants quit smoking, leading to significant weight gain and reduction in CO and Cotinine levels in a short amount of time (Courtemanche et al. 2016). Thus, our post-treatment data comprises a mixture of smokers and non-smokers, where on average, quitting smoking leads to a significant increase in BMI, and reduction in CO and Cotinine levels relative to continuous smokers. As a result, we infer whether participants who misreport smoking status belong to the higher BMI distribution, and lower CO and lower Cotinine distributions. We also estimate whether misreporters are more likely to be part of the non-smoking distribution of changes in BMI, CO, and Cotinine levels relative to pre-treatment levels.

We believe that this paper contributes significantly to the economics literature. Currently there is limited and mixed evidence on whether smokers misreport. As a result, papers analyzing smoking behavior rarely discuss the problem of applying misreported smoking data. Most research uses self-reported information to identify causal effects in topics such as the relationship between cigarette taxes and smoking participation (Tauras 2006, Chaloupka et al. 2000). This could potentially be a

[49]CO is measured in parts-per-million (ppm) of exhaled air using either the MiniCO (Catalyst Research) or the EC50 (Vitalograph). Cotinine is measured from a saliva sample taken during the meeting

problem since self-reported smoking status can be misreported between 3% to 70% (Brachet 2008, Webb et al. 2003, Boyd et al. 1998). Similarly, Bound (1991) discusses that many self-reported survey responses may be misreported in the same survey. Therefore, any analysis comparing several self-reported variables applies mismeasured independent and dependent variables. As a result, conclusions from such data can be biased.[50]

Misreporting levels in smoking participation are less understood. In contrast to self-reported public program participation which can be matched with administrative data, smoking information can normally not be matched with administrative smoking information. There are several reasons for missing administrative smoking information. First, governments have public program participation information, as they need to know who they are paying benefits to, but collecting information on "correct" smoking participation requires tax resources without any tangible benefits to the government. Second, surveys collecting self-reported information are much more cost effective in accumulating information for a large sample of the population. As a result, self-reported and clinically verified smoking information are most commonly available in Randomized Controlled Trials (RCT).

Our descriptive analysis reveals that the number of treatment group participants misreporting smoking status is twice as large as in the control group. We speculate that a treatment group participant's knowledge of receiving treatment and the intended treatment goal may explain at least a portion of the higher misreporting level in the treatment group. Second, our results show that participants misreport across the CO and Cotinine level distributions, implying that clinicians take into account additional information beyond CO and Cotinine levels to make decisions regarding smoking status.

Our regression results show that being male, married and of older age significantly increases the

[50]There is also a large literature analyzing the level of misreporting in social welfare programs such the Supplemental Nutrition Assistance Program (SNAP) and Medicaid by matching self-reported survey information, such as the Current Population Survey (CPS), with administrative data. Several papers show that misreporting, in this case under-reporting, in SNAP participation can reach over 30% (Meyer et al. 2009). Additionally, The recognition of large levels of misreporting lead to the development of new estimation strategies to recover unbiased coefficients under the presence of misreporting (Kreider et al. (2012), Almada et al. 2015, Lewbel 2007, Nguimkeu et al. 2016).

probability of misreporting, while having a higher BMI[51] and consuming nicotine gum decreases the probability of misreporting. For example, being male or married increases the probability of misreporting by 10 percentage points. Our regression sub-sample analyses by smoking cessation group present heterogeneous effects. We only find a significant relationship between misreporting, gender, marital status, nicotine gum use and BMI for those participants randomized into the smoking cessation treatment group.

Our second methodological approach reveals two distinct distributions separating smokers from non-smokers across outcome measure. For example, utilizing the change in BMI from baseline to year 5 as the dependent variable, we find that the first distribution has an average of 0.21 BMI units and the second distribution has an average of 1.56 BMI units, where smokers are much more likely to be part of the first, small change in BMI, distribution (84%) and non-smokers are much more likely to be part of the second higher BMI distribution (97%). Our analysis comparing the distribution probabilities for those participants who were objectively classified as smokers, but self-reported to be non-smokers, shows a 77% average probability of belonging to the high, large change in BMI, distribution . As a result, even though they are objectively classified as smokers, the participants are much more likely to be part of the higher BMI distribution. We rely on descriptive statistics and the mixture model results to calculate that at least 29% of misreporters may be in fact falsely classified as smokers.

The rest of the paper is organized as follows. Section 2 discusses the literature on misreporting with an emphasis on under-reporting of program participation. Section 3 discusses the data. Sections 4-5 present the parametric and non-parametric estimation strategies and present the results. Section 6 concludes.

[51] Body Mass Index (BMI) is equal to weight in kg divided by height squared in centimeters. Revealing a measure if someone's weight is considered underweight, normal, overweight, or obese.

Literature Review

There is a large empirical literature estimating misreporting levels in public program participation and a smaller literature on smoking misreporting levels. Similarly, a growing literature proposes methodology to recover unbiased estimates for misreported data. We discuss all of these aspects with an emphasis on under-reporting also referred to as false negatives.

Many studies report that people misreport public program participation. The seminal study by Meyer, Mok and Sullivan (2009) compares ten transfer programs in five different surveys.[52] They compare weighted self-reported survey data to records of administrative aggregates and find that overall under-reporting is significant and has been increasing over time. It is particularly severe for the workers compensation and the AFDC/TANF program where under-reporting reaches 50% in surveys such as the PSID and CPS. Similarly, under-reporting for the food stamp program is close to 40% in the PSID and SIPP. Other studies, such as Meyer & Goerge (2011) and Marquis & Moore (1990 & 2010), find similar levels of under-reporting. Additionally, Bollinger and David (1997) find that participation in food stamp programs is a particularly prone to being misreported because of the social "stigma" associated with its program participation.

Under-reporting is not only confined to public program participation. Bound (1991) discusses that self-reported survey responses may be misreported across the board. One explanation for that can be that people misreport to mimic socially desirable behavior (Ansolabehere and Hersh 2012). Similarly, Card's (1996) study finds significant under-reporting in union coverage as well. Ezzati et al. (2006) show that women under-report their weight and middle aged men over-report their height. Even in situations where socially desirable behavior might not be identifiable, under-reporting is still large. Wolf (2004) takes advantage of GPS data and shows that people under-report

[52]The programs they look at are Unemployment Insurance (UI), Workers' Compensation (WC), Social Security Retirement and Survivors Insurance (OASI) and Social Security Disability Insurance (SSDI), Supplemental Security Income (SSI), the Food Stamp Program (FSP), the Earned Income Tax Credit (EITC), Aid to Families with Dependent Children/Temporary Assistance for Needy Families (AFDC/TANF), the Special Supplemental Nutrition Program for Women, Infants and Children (WIC) program and the National School Lunch Program (NSLP). The surveys are the Current Population Survey – Annual Demographic File/Annual Social and Economic Supplement (CPS), the Survey of Income and Program Participation (SIPP), the Panel Study of Income Dynamics (PSID), the American Community Survey (ACS), and the Consumer Expenditure Interview Survey (CE Survey).

the number of trips taken by car by 11% to 81%.

Less is known regarding the level of under-reporting of smoking behavior in the population. Studying the level of misreporting relies on a comparison between self-reported and clinical smoking information, such as carbon monoxide or cotinine levels, nicotine's main metabolite. However, even if an objective measure is available, a clinician has to make a judgment call by defining cotinine and carbon monoxide thresholds after which a person is classified as a smoker. As a result, the evidence on the level of misreporting is mixed. Evidence from samples of pregnant women show that under-reporting reaches 35% while over-reporting reaches 10% (Brittonet al. 2004). Similarly, Boyd et al.(1998) & Bardy et al.(1993) report smoking under- and over-reporting in a sample of pregnant women and neonatal mothers between 26-38% and 3-14%, respectively. One possible explanation for under- and over-reporting may lie in the unequal thresholds that define a smoker relative to a non-smoker. Nevertheless, it is surprising that all samples observe over-reporting in smoking status, because we find no evidence in our sample. It could be explained by the demography of the sample. Another explanation for over-reporting may be that pregnant women only quit smoking during pregnancy, and therefore still classify themselves as smokers.

A separate growing literature produces methodological work to recover unbiased estimators in the presence of misreporting (Aigner 1973, Bollinger & David 1997, Hausman et al. 1998, Brachet 2008). Mahajan (2006) and Hu & Schennach (2008) formalize a model for a mismeasured binary independent variable and show that under certain assumptions correct marginal effects can be obtained. Hug (2010) proposes a solution to misreporting with linear and non-linear estimation techniques. McCarthy and Tchernis (2011) apply Bayesian methods to estimate consistent and unbiased coefficients under misreported program participation. Lewbel (2007) as well as Mahajan (2006) rely on an instrumental variable approach without the typical conditional independence assumptions. Brachet (2008) proposes a two-stage estimator, similar to a two-part model, where treatment effects are estimated in the second stage. Berg and Lien (2006) introduce a model that can handle misreporting and non-response simultaneously. Lastly, Kreider et al. (2012) follow Manski (1995) & Pepper (2000) to identify an average treatment effect bound rather than the parameter.

Our paper contributes to both literatures.

Data

We use data from the the LHS, a randomized smoking cessation trial, to observe changes in the severity of chronic obstructive pulmonary disease (COPD). The study recruited 5887 smokers with initial ages between 35 and 59 at 10 clinical centers in the US and Canada. Recruitment lasted from 1986 to 1989 and the clinical trial ended in 1994. Participants were interviewed annually at a clinic near their residence. The data includes the baseline year (1989) and the five annual follow-up years (1990-94).[53] Eligible smokers had to show signs of mild lung function impairment, have no history of certain medications, consume less than 25 drinks per week, have a BMI under 40, and have no severe illnesses or chronic medical conditions.

Participants were randomly assigned into treatment and control groups, with an additional randomization arm within the treatment group to receive either a bronchodilator with medical benefits (SI-A) or a placebo inhaler (SI-P) with no medical benefits. Both treatment groups received a special intervention (SI) focusing on a behavior intervention consisting of a physician message, interventionist meetings, an intensive quit week, and frequent contact with support personnel with invitations to bring a spouse or a relative to the meetings. Additionally, both treatment groups received free nicotine gum. Most of the intensive intervention treatments were completed within the first 4 months of the study. The control group, or usual care (UC) group, received no intervention and group members were referred to their own private sources for medical care.

We take advantage of information on BMI, Carbon Monoxide levels, Cotinine levels, smoking behavior, family smoking habits, and demographic characteristics. BMI was calculated from measured weight and height by medical staff at the participants' clinic visits. Information on smoking behavior includes self-reported smoking status as well as self-reported average cigarette consumption per day. The data also includes a binary indicator measuring the objective smoking status measured by a clinician who has access to information on CO and Cotinine levels taken from par-

[53] Attrition was low, only 5% of participants did not interview in the final follow-up

ticipants at each of the annual follow-up visits.[54] We also utilize variables measuring self-reported smoking participation in the household. Specifically, we have binary variables identifying if the spouse smokes, if the child smokes, if anyone else smokes, and a continuous variable measuring the number of smokers living with the participant. Lastly, we use information on education level, gender, age, and marital status.

We utilize self-reported and objectively verified smoking variables to generate a measure of a misreporting, a dummy variable quantifying whether the self-reported and objective smoking information present conflicting evidence after the first annual follow-up visit. For example, the misreporting variable equals one if participants report being a non-smoker while the objectively verified measure suggests that the participant is a smoker and vice versa. The data show a clear sign of under-reporting in smoking status (false negative) with no-one over-reporting smoking status. In other words, some smokers report being a non-smoker while the objective measure suggests otherwise.

Table 18 presents descriptive statistics for the treatment and control groups at the time of randomization. Across all groups the variables are balanced, suggesting that randomization was successful. The sample represents heavy smokers, consuming about 30 cigarettes per day and about 40% of participants had someone smoking in the household. This is expected because approximately 65% of participants were married and heavy smokers are more likely to have a spouse who is also a smoker. The sample is on average more highly educated, 13.57 of completed years of education, than the general population of the early 90s. Additionally, the average age is 48, indicating that we are most likely looking at heavy smokers that have been smoking for many years.

Table 19 presents descriptive statistics for the self-reported, the objective verified smoking status, and the dependent variable measuring if those two aforementioned variables present conflicting results. In total 316 individuals are classified as misreporters, meaning that they self-report that they do not smoke while the objective measure suggests that they are smoking. About 7% in the treatment groups misreport smoking status and only 3% in the control group misreport smoking

[54] Again, CO is measured in parts per million (ppm) and Cotinine is measured in nanograms per milliliter (ng/mL)

Table 18: Summary Statistics by Treatment Group Assignment at Randomization Meeting

Variable	SI-A	SI-P	UC
Cigarettes per Day	29.59	29.49	29.53
	(14.08)	(13.60)	(14.11)
Cotinine Level	367.07	369.42	363.11
	(203.76)	(200.04)	(204.82)
Carbon Monoxide	25.97	26.69	28.98
	(13.48)	(12.67)	(12.67)
BMI	25.42	25.67	25.55
	(3.91)	(3.92)	(3.92)
Smoker in HH	0.39	0.40	0.39
	(0.49)	(0.49)	(0.49)
HH Spouse Smokes	0.27	0.29	0.27
	(0.45)	(0.45)	(0.44)
HH Child Smokes	0.13	0.11	0.12
	(0.34)	(0.31)	(0.32)
HH Other Smokes	0.06	0.05	0.06
	(0.24)	(0.23)	(0.23)
Education Level	13.57	13.56	13.68
	(2.84)	(2.84)	(2.79)
Male	0.60	0.64	0.64
	(0.49)	(0.48)	(0.48)
Married	0.65	0.67	0.65
	(0.47)	(0.47)	(0.47)
Age	48.41	48.55	48.43
	(6.84)	(6.83)	(6.80)
Observations	1961	1962	1964

Notes: Standard Deviations in Parentheses. SI-A and SI-P are the treatment arms and UC is the control.

status. The analysis by group assignment suggests that misreporting is related to treatment status. The difference in misreporting levels by treatment group may also reflect differences in observable characteristics among misreporters. For example, heavy smokers in the treatment group may be more likely to misreport smoking status than those in the control group.

Table 19: Summary Statistics at First Follow-Up Visit

Variable	SI-A	SI-P	UC
Self-reported Smoking Status	0.56	0.57	0.88
	(0.50)	(0.50)	(0.33)
Objectively verified Smoking Status	0.63	0.65	0.91
	(0.48)	(0.48)	(0.29)
Mismeasured Smoking Status	0.07	0.07	0.03
	(0.26)	(0.25)	(0.16)
Observations	1867	1866	1865

Notes: Standard Deviations in Parentheses. SI-A and SI-P are the treatment arms and UC is the control.

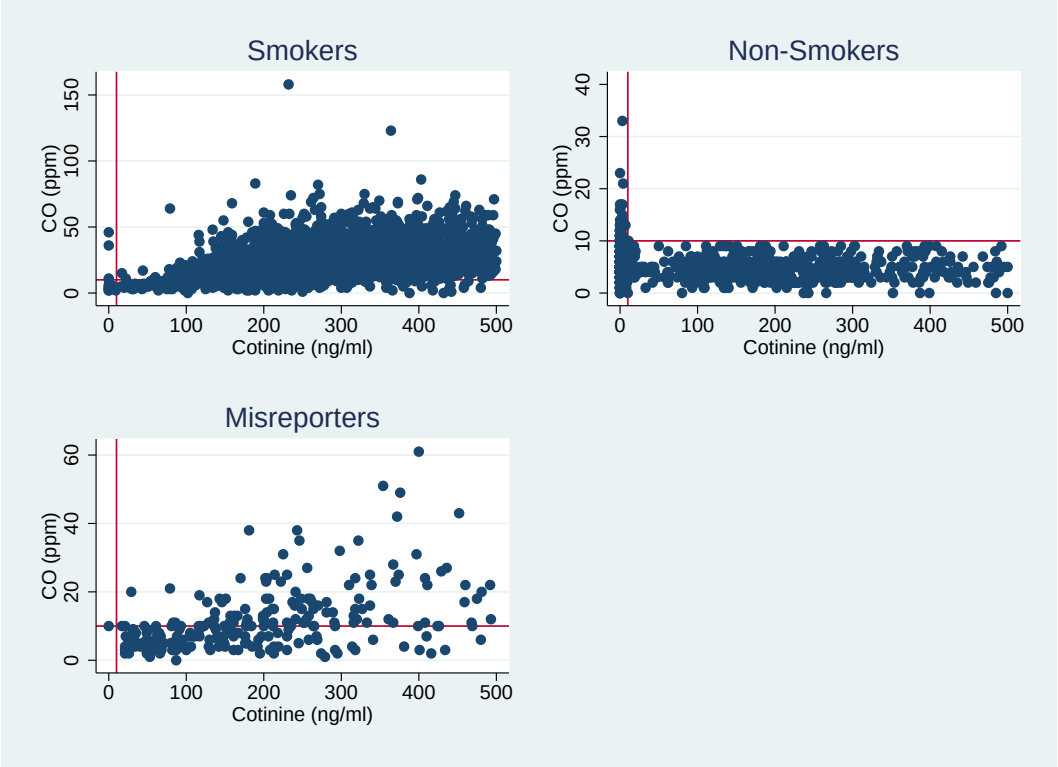
There are concerns that the CO and Cotinine measures only represent short-term measures of smoking status or possible exposure to second hand smoke. CO has a half-life of about 5 hours while Cotinine has a half-life of about 7 days (Jo & Oh 2003). Therefore Cotinine levels should reveal more accurate long-term smoking patterns. To this end, a typical approach in the literature is to classify smokers as having a CO and Cotinine levels above 10ppm and 10 ng/mL, respectively (Carey & Abrams 1988). The Lung Health Study utilized thresholds of 10 ppm for CO and 20ng/mL for Cotinine (Murray et al. 1993). However, it is possible to have smokers with lower concentrations, and e non-smokers with above threshold concentrations.[55] A typical non-smoker has a CO level below 4ppm (Deveci et al. 2004), but can have Cotinine levels of up to 30mg/ml (Laugesen et al. 2009). Moreover, having non-smokers consume nicotine gum can also lead to high levels of saliva Cotinine (Gross et al. 1995). Therefore, it is not easy to clearly separate non-smokers from smokers considering that many factors affect CO and Cotinine levels.

We begin by comparing CO and Cotinine levels among objectively classified smokers, non-smokers, and misreporters at the first annual follow-up visit, because this is the first time we can

[55]For more information see http://www.healthnz.co.nz/CObreath.htm

generate our misreporting measure. Figure 8 includes three graphs that includes reference lines for the 10 ppm and 10ng/ml threshold limits for CO and Cotinine levels, respectively. The top left graph shows that many smokers tend to have high CO and Cotinine levels, but there are also a few smokers with CO and Cotinine levels below the thresholds of 10. Moreover, among non-smokers (top right graph) we find very few participants with CO levels above 10 ppm, but a relatively uniform distribution of Cotinine levels. The distribution of CO and Cotinine levels among misreporters mimics more closely the distribution of smokers, however, there are many misreporters with CO levels below the typical 10 ppm threshold. Overall, Figure 8 suggests that CO seems to be an important clinician measure whether somebody is classified as a smokers or not. Moreover, the graphical evidence suggests that there are many misperoters with CO levels below the threshold that classifies them as smokers. Next we discuss our empirical strategy to identify whether misreporters differ from non-smokers at the first annual follow-up visit.

Figure 8: CO and Cotinine Levels by Smoking and Misreporting Status at the First Annual Follow-Up Visit



Econometric Analysis

Regression Analysis

In our first approach we take the objective smoking status variables as accurately representing each participant's smoking status. We estimate regression models to identify how demographic variables, CO, and Cotinine levels correlate with misreporting status. Our motivation for this approach is the following - if household smoking variables are positively correlated with misreporting, then this could indicate that clinicians may misinterpret CO and Cotinine information. It is possible that a non-smoker exposed to second-hand smoking reports a high CO level. Additionally, a non-smoker consuming nicotine gum increases Cotinine levels significantly. Thus, if clinicians do not account for these influences participants may be wrongly classified as misreporters. To this end, we include a dummy variable for whether the participants consumed nicotine gum in the last 23 hours before the annual follow-up visit. We also include different types of dummy variables measuring household smoking. First, we include whether the person reports that they have a smoker living with them. Second, we include a variable measuring the number of smokers in the household. Third, we include dummy variables for whether the spouse, child, or someone else smokes in the household. Given the nature of our binary dependent variable, we estimate Probit and Linear Probability Models (LPM). The basic LPM specification is given here:

$$S_{i1} = \gamma_0 + \gamma_1 HH_i + \gamma_2 X_i + \mu_{i1} \tag{11}$$

where S_{i1} is a binary indicator measuring misreporting smoking status for individual i at the first follow-up visit. In all of the regressions our references group are verified non-smokers, i.e non-misreporters. HH_i reflects a vector of binary household smoking information as outlined in the data section and summarized in Table 18. We include different combinations of household smoking variables in our regressions to test the robustness of our results. X_i includes education level, marital status, age, nicotine gum use, and BMI of the individual. Similarly, we estimate equation (11) with Probit and Logit models to estimates if the LPM results are robust to model specification.

Parametric Results

Table 20 presents six regression estimates for equation (11), where each regression includes a different set of household smoking characteristics. Column 1 includes the baseline model with no household smoking variables, and columns 2-6 include different variables of household smoking participation. In unreported Probit regressions we find that the marginal effects are not statistically different from the LPM results.

Across specification we find that being older, married, and male increases the probability of misreporting, while using nicotine gum and having a higher BMI decrease the probability of misreporting. Specifically, we find that being a married male increases the probability of misreporting by about 10%. In contrast, consuming nicotine gum decreases the probability of misreporting by 7%. To put all results in perspective, about 14.9% of non-smokers are misreporting, suggesting that the marginal effect presents large changes relative to the baseline misreporting level. To our surprise all observable characteristics only explain a fraction of the variation in misreporting as documented by the R-squared of only 0.03.

Next, we estimate sub-sample regression of equation (11) by treatment status. Table 20 presents evidence that a larger proportion of treatment group participants misreport relative to the control group. Thus, we also report results restricting the sample by treatment group in Appendix Tables A6-8. Ex ante, we believe that the relationship between demographic characteristics and misreporting status can be different for these sub-samples for reasons discussed previously in the data section.

Among the sub-sample results by treatment assignment, we find evidence in both treatment group regressions that being male increase the probability of misreporting, while the use of nicotine gum and having a higher BMI decreases the probability of misreporting. Interestingly, we find that the positive relationship between age and misreporting is only significant in the SIA treatment group sub-sample in Table A6 and that the positive effect of being married on misreporting is only significant in the SIP treatment group in Table A7. In the control group, we find that being of older age and having a smoking spouse decreases misreporting, with some regression specifications pre-

Table 20: Regression Estimates on the Effect of Demographic Variables on Misreporting

Dependent Variable: Misreporting Smoking Status						
Male	0.0651***	0.0646***	0.0650***	0.0650***	0.0667***	0.0669***
	(0.0177)	(0.0177)	(0.0177)	(0.0177)	(0.0177)	(0.0178)
educ	-0.000531	-0.000714	-0.000584	-0.000608	-0.000221	-0.000279
	(0.00310)	(0.00310)	(0.00310)	(0.00309)	(0.00307)	(0.00307)
Married	0.0297*	0.0305*	0.0300*	0.0325*	0.0323*	0.0292
	(0.0179)	(0.0179)	(0.0179)	(0.0182)	(0.0182)	(0.0185)
Age	0.00344***	0.00343***	0.00344***	0.00345***	0.00342***	0.00339***
	(0.00126)	(0.00126)	(0.00126)	(0.00126)	(0.00126)	(0.00127)
BMI	-0.00960***	-0.00957***	-0.00960***	-0.00960***	-0.00959***	-0.00960***
	(0.00214)	(0.00214)	(0.00214)	(0.00214)	(0.00214)	(0.00214)
Nicotine Gum	-0.0715***	-0.0723***	-0.0717***	-0.0730***	-0.0729***	-0.0724***
	(0.0171)	(0.0172)	(0.0171)	(0.0172)	(0.0173)	(0.0172)
HH Smoker		-0.0146				
		(0.0191)				
# Smokers in HH			-0.00355			
			(0.0147)			
HH Spouse Smokes				-0.0227	-0.0269	-0.0265
				(0.0232)	(0.0231)	(0.0231)
HH Child Smokes					0.0435	0.0450
					(0.0301)	(0.0302)
HH Other Smokes						-0.0510
						(0.0422)
Constant	0.207**	0.213**	0.209**	0.210**	0.201**	0.207**
	(0.103)	(0.103)	(0.103)	(0.103)	(0.102)	(0.103)
Observations	1,803	1,803	1,803	1,803	1,803	1,803
R-squared	0.030	0.030	0.030	0.030	0.031	0.032

Notes: Regressions include Heteroscedasticity robust standard errors in parentheses. Each column presents esti mates for a different regression. Sample includes non-smokers and misreporters.

senting a positive effect of being married on misreporting. We conclude that this is a profound result because it implies that endogenous misreporting occurs by treatment group assignment. We tested the hypothesis whether the marginal effects in the treatment group sub-samples are different from the control group, and find significant differences in the nicotine gum and HH spousal smoking status variables.

In conclusion, we find that under-reporting in smoking participation is prevalent in the LHS where participants know that they will be subjected to clinical tests. Our regression estimates indicate that being male, married, and of older age increases the probability of misreporting, while having a higher BMI and consuming nicotine gum decreases the probability of misreporting. Our sup-sample analysis presents evidence for endogenous misreporting, emphasizing the need to apply models that correct for endogenous misreporting. We believe that LHS treatment group participants may feel social pressure to show that the smoking cessation program worked and therefore are more likely to misreport than control group participants. However, overall the observable variables only account for a small fraction of the variation that explains misreporting. Therefore, future research needs to expand the set of control variables to explore other reasons for misreporting. In our next step we analyze whether the objective smoking measure does a good job at identifying misreporters.

Are Misreporters Different from Non-Smokers and Smokers?

We begin the second part of the paper to analyze whether misreporters are different from smokers and non-smokers. Specifically, we compare the means of several smoking variables (Cigarettes per Day, CO and Cotinine levels at baseline) for misreporters to smokers and non-smokers. If misreporters are more similar to smokers than non-smoker, then this provides some evidence that clinicians correctly identified misreporters.

Table 21 summarizes observable smoking characteristics at the baseline meeting for all misreporters by treatment group and compares them to baseline characteristics of participants classified as smokers and non-smokers by the first annual follow-up visit. We find that the mean cigarette consumption of misreporters is statistically different from the mean cigarette consumption of smok-

ers. Specifically, misreporters consume about 2 cigarettes less cigarettes at the baseline interview compared to those who smoke. There is also evidence, that misreporters consume less cigarettes per day than non-smokers. Since misreporters consume less cigarettes at baseline than smokers and non-smokers it is conceivable that a good proportion of misreporters was able to quit smoking or reduce cigarette consumption significantly.

Table 21: Summary Statistic at Randomization for Participants who Misreport by Treatment Group

	SI-A	SI-P	UC	Non-Smokers	Smokers
Cigarettes per Day	26.76*	25.66*	23.83*	27.95	30.45
	(12.87)	(14.11)	(13.18)	(14.39)	(13.69)
CO Baseline Visit	33.46	32.06	N/A	30.88	34.15
	(16.11)	(17.85)	N/A	(17.47)	(16.48)
CO Screening Visit*	26.3	27.5	30.5	25.28	26.52
	(14.09)	(13.79)	(14.46)	(13.57)	(12.58)
Cotinine level	368.54	368.34	433.02*	341.33	375.52
	(180.73)	(203.70)	(341.86)	(207.80)	(198.63)
Observations	138	130	48	1539	4019

Notes: Standard deviation in parentheses. SI-A and SI-P are the treatment arms and UC is the control arm. * implies that only a fraction of the misreporters observations are available 25,18,10, respectively, and also that the mean is significantly different at the 5% level from the mean of smokers. Smokers and non-smokers are defined by their smoking status at the first annual follow-up visit.

We continue our descriptive analysis by comparing the changes from the baseline year to the first annual follow-up visit in average BMI, Cotinine, and CO levels for misreporters to smokers and non-smokers. Table 22 summarizes the changes for all variables by group. T-tests suggest that almost all means among misreporters are significantly different from non-smoker means. Specifically, we find that the changes in cigarette consumption, CO, and Cotinine are all smaller than for those classified as non-smokers. However, it is noteworthy that among misreporters we find significant reductions in CO and Cotinine levels which are relatively closer to non-smokers than smokers. Moreover, misreporteres saw a threefold increase in BMI relative to smokers with an average gain of about 0.9 BMI units, but an almost 50% smaller increase in BMI than non-smokers. We also find evidence that misreporters consume nicotine gum at significantly lower rates than

non-smokers, but also at significantly higher rates than smokers. We have also compared changes across the same variables after five years and find an even larger decreases in Cotinine level among misrepoters. Thus, misreporters seem to be descriptively closer to non-smokers than smokers, however, the changes in the smoking variables are all lower than among non-smokers.

Table 22: Changes in Smoking Variables from Baseline to the First Annual Follow-up Visit for Misreporters Compared to Smokers and Non-Smokers

Variable	SI-A	SI-P	UC	Non-Smokers	Smokers
Change in BMI	0.89*	0.95*	0.82*	1.61	0.29
	(1.60)	(1.45)	(1.15)	(1.66)	(1.29)
Change in Cotinine	-137.72*	-154.38*	-270.06	-252.17	-61.53
	(177.91)	(250.82)	(365.77)	(240.64)	(211.95)
Change in CO (s2)	-11.17*	-17.58	-6.68*	-20.22	-1.96
	(10.86)	(13.66)	(15.63)	(13.72)	(13.03)
Change in CO Baseline	-19.95*	-20.95*	N/A	-25.99	-8.74
	(15.99)	(16.00)	N/A	(17.58)	(17.41)
Change in Cigarettes	-26.76	25.66*	-23.83*	-27.95	-6.79
	(12.87)	(14.11)	(16.08)	(14.39)	(11.95)
Nicotine Gum	0.19*	0.23*	0.15	0.34	0.09
	(0.40)	(0.42)	(0.36)	(0.47)	(0.29)
Observations	138	130	48	1533	3542

The changes in CO are based on different reference dates. The "s2" refers to changes from the second screening interview where a few participants were voluntarily tested. However there are only few observations per group available, 20,17,8,249, and 517, respectively. Baseline refers to the change from the randomization visit where only treatment participants were interviewed. However the sample is significantly larger with 110,105, N/A, 1286, and 1816 observations in each column, respectively. * indicates that means are significantly different from the means of non-smokers at the 10% level.

The descriptive analysis presents evidence that misreporters reduce cigarette consumption, CO, and Cotinine levels significantly. However, it is unclear whether those reductions lead to cessation or just to lower levels of consumption. We calculate a back of the envelope estimate on the level of miss-classification of misreporters by counting all misreporters whose change in either BMI, Cotinine or CO level are larger or equal to the mean change for non-smokers. This rough calculation reveals that 206 misreporters may be in fact non-smokers. However, since many non-smokers can show high levels of CO and Cotinine due to external influences, such as household smoke

and nicotine gum use, CO and Cotinine levels may have been elevated at the first examination. Therefore, we revisit the objective smoking measure in the fifth annual-follow-up visit and find that 108 of the misreporters are now classified as non-smokers. Since, the intervention and most changes in the smoking measures persist after five years it is unlikely that many of the misreporters quit smoking after the first annual follow-up visit. To get a better estimate than the aforementioned rough missclassifcation calculation our next section models a latent variable framework that tries to distinguish smokers from non-smokers.

Bayesian Mixture Model

In our second methodological approach we take an agnostic view on whether the objectively verified smoking measure should be preferred to the self-reported smoking measure. Specifically, we have no prior on whether the objective smoking variable presents the underlying "true" smoking status. Rather our goal is to identify if the objectively verified smoking variable has information that will more accurately predict the change in BMI, CO and cotinine levels. Our data is well suited for such exercise, because the data includes a sample of smokers with a large proportion quitting smoking after receiving treatment within the first year. Courtemanche et al. (2016) show that quitting smoking leads to significant weight gain, therefore people misreporting smoking status should not have experienced a large increase in weight, because they are in reality smokers. On the other hand, if they quit smoking then they should have experienced significant decreases in CO and cotinine levels.

We believe that a mixture model is a well suited approach. Broadly, the idea behind the mixture model is to separate a population into subpopulations, where we only use information from the overall population to make inference about the subspopulation. The approach allows us to derive properties such as distribution means and probability assignment for each participant to each distribution. In our case, we have a strong smoking cessation treatment that generates an immediate change in BMI, CO and Cotinine levels, that separates smokers from non-smokers.

Specifically, we apply a Bayesian mixture model (Koop et al. 2007) that implements a latent

variable framework where it estimates the probability for each individual belonging to each mixture component. We, generally utilize the objective smoking variable as an "instrument" that assigns each person to one of the two distributions. Ideally the objective smoking variable helps assign each smoker to their correct distribution made up of smokers, rather than the distribution of non-smokers as measured by their BMI, CO and Cotinine levels.

We also estimate the model in levels rather than changes. Since a participant's demographic characteristics define a participant's BMI, and BMI is a stock that evolves over time, we include an extensive set of control variables that influence the mixture component assignment. For example, it is important to include gender and age information, because males and older participants weigh more than their female and younger counterparts. To this end we also include the baseline BMI, CO, and Cotinine levels to account for the pre-treatment outcome levels where appropriate. The other explanatory variables are marital status, age, education level, presence of a household smoker, and nicotine gum use. The general density setup of the mixture model is the following:

$$p(Y_i|X_i,\theta) = [1-\Phi(X'\beta)]\phi(Y_i;\mu_1;\sigma_1^2) + \Phi(X'\beta)\phi(Y_i;\mu_1+\alpha;\sigma_2^2) \tag{12}$$

where the key components of the mixture model are the Y means μ_1 and $\mu_{1+\alpha}$ and variances σ_1^2 and σ_2^2 of each mixture component ϕ, respectively. Y_i represents the dependent variable for each individual i at the first follow-up visit. The probability of being in each component of the mixture, Φ, adjusts given the variables summarized in X_i. The mixture assignment probability can be rewritten in a latent variable framework:

$$p(Y_i|X_i,\theta,z_i^*) = \phi(Y_i;\mu_1;\sigma_1^2)^{I(z_i^*<0)} + \phi(Y_i;\mu_1+\alpha;\sigma_2^2)^{I(z_i^*>0)} \tag{13}$$

where the latent probability, z_i^*, is determined by the independent variables:

$$z_i^* = X'\beta + \epsilon_i, \epsilon_i \sim N(0,1) \tag{14}$$

In this case the latent variable z_i^* has to be greater than zero to include an individual in the sec-

ond mixture component. For example, if X_i includes strong positive predictors suggesting that a person has a higher BMI, then the latent variable z_i^* will be more likely to have a value greater than zero. This may result in a separation in the mixture component assignment for differential X_i characteristics. To complete the model, the following flat prior specifications are included:

$$\mu_1 \sim N(\underline{\mu}, V_\mu)$$
$$\alpha \sim N(\mu_\alpha, V_\alpha)$$
$$\sigma_j^2 \sim IG(a_i, b_i),\ j = 1, 2$$
$$[\beta] \sim N(\mu_\beta, V_\beta)$$

We assume that α is drawn from a normal distribution. To identify the mixture model there has to be a restriction on either the priors, the mixture component, or the hyper parameters (parameters in the model besides those defined in the priors or mixture model so far). In this model we restrict the mean of the second mixture component to be larger than the first component. The joint posterior distribution for this model is proportional to:

$$\left(\prod_{i=1}^{N}\left[\phi(Y_i;\mu_1;\sigma_1^2)^{I(z_i^*<0)}\phi(Y_i;\mu_1+\alpha;\sigma_2^2)^{I(z_i^*>0)}\right]\phi(z_i^*;X'\beta,1)\right)\times p(\beta)p(\mu_1)p(\alpha)p(\sigma_1^2)p(\sigma_1^2)$$

We estimate the posterior conditionals based on the following specification:

$$\mu_1|\theta_{-\mu_1}, Bmi \sim N(D_{\mu_1}d_{\mu_1}, D_{\mu_1})$$

$$D_{\mu 1} = (\iota'\Sigma^{-1}\iota + V_\mu^{-1})^{-1}$$

$$d_{\mu 1} = \iota'\Sigma^{-1}(Bmi - z\alpha) + V_\mu^{-1}\underline{\mu})$$

Note ι is a Nx1 vector of ones, and z is equal to:

$$z_i = I(z_i^* > 0) \ \ and \ \ \Sigma \equiv Diag\{(1 - z_i)\sigma_1^2 + z_i\sigma_2^2\}$$

The mean of the second component of ϕ includes α, and α is drawn from:

$$\alpha|\theta_{-\alpha}, Y \sim TN_{(0,\infty)}(D_\alpha d_\alpha, D_\alpha)$$

$$D_a = (n_2/\sigma_2^2 + V_\alpha^{-1})^{-1}$$

$$d_\alpha = \sum_{i=1}^{N} z(Y_i - \mu_1)/\sigma_2^2 + V_a^{-1}\mu_\alpha$$

Note that $n_2 = \sum_i z_i$ and $n_1 = N - n_2 = \sum_i (1 - z_i)$

The Betas are drawn from:

$$\beta|\theta_{-\beta}, Y \sim N(D_\beta d_\beta, D_\beta)$$

with $D_\beta = (X'X + V_\beta^{-1})^{-1}$ and $d_\beta = X'z^* + V_\beta^{-1}\mu_\beta$

The variances σ_1^2 and σ_2^2 are drawn from:

$$\sigma_1^2|\theta_{-\sigma_1^2}, Y \sim IG\left(\frac{n_1}{2} + \alpha_1, \left[b_1^{-1} + .5\sum_{i=1}^{N}(1 - z_i)(Y_i - \mu_1)^2\right]^{-1}\right)$$

$$\sigma_2^2|\theta_{-\sigma_2^2}, Y \sim IG\left(\frac{n_2}{2} + \alpha_2, \left[b_2^{-1} + .5\sum_{i=1}^{N} z_i(Y_i - \mu_1 - \alpha)^2\right]^{-1}\right)$$

\

Lastly, we draw z_i^* with initially drawing U from a two point distribution $U\epsilon\{0, 1\}$ from:

$$Pr(U = 1) = \frac{[1 - \Phi(X'\beta)]\phi(Y_i; \mu_1; \sigma_1^2)}{[1 - \Phi(X'\beta)]\phi(Y_i; \mu_1; \sigma_1^2) + \Phi(X'\beta)\phi(Y_i; \mu_1 + \alpha, \sigma_2^2)}$$

This gives the relative probability of the distribution in the first vs the second component. We subtract $Pr(U = 1)$ from a random variable drawn from the uniform distribution of the interval (0,1). If the probability of U becomes negative, we draw z_i^* truncated $\{0, \infty\}$ and vise versa:

$$z_i^*|U, \theta_{-z^*}, Y \begin{cases} TN_{(-\infty,0)}(X'\beta, 1) & if\ U = 1, \\ TN_{(0,\infty)}(X'\beta, 1) & if\ U = 0 \end{cases}$$

Lastly, we use hyperparameter values for the priors on μ, α, σ_j^2 and β. Specifically, $\underline{\mu} = 5.5, V_\mu = .4$, $\mu_\alpha = .4, V_\alpha = .4$, $a_1 = a_2 = 2, b_1 = b_2 = 100$,$\mu_\beta = 0_{kx1}, V_\beta = 4I_2$. We estimate the model for 10000 iterations and drop the first 2500 simulations.

Mixture Model Results

We present results for six estimations in Table 23 – three mixture models where the dependent variables are BMI, CO, and Cotinine level, and three models where the dependent variables are the changes in BMI, CO, and Cotinine level. All models include an extensive set of explanatory variables as described in the previous section. Table 23 presents posterior means and standard deviations in parentheses for the first and second mixture components. The average BMI for the first distribution is 23.21 BMI units, while the second component has an average BMI of about 29 BMI units ($\mu_1 + \alpha$). The means are significantly different in each distribution, as measured by the distance from their standard deviations. Similarly, we find that the model with the change in BMI as the dependent variable presents two positive means, suggesting that overall participants gain weight by the first year. Among our estimates including the CO or Cotinine measures, we find that the first component mean is significantly lower than the second component mean. Moreover, we find that the estimates utilizing the changes in CO or changes in Cotinine level present first component means that are negative, indicating a population that decreased CO and Cotinine levels from the baseline year.

Table 23 does not reveal whether the mixture model assigns smokers correctly to the first mix-

ture component while assigning non-smokers to the second, higher mixture component. Table 24 reports the posterior probabilities of non-smokers and smokers belonging to each distribution.[56] Ex ante, we expect that smokers are more likely to have a lower BMI, have higher levels of CO and Cotinine, and also have less changes in BMI units, CO, and Cotinine levels. In other words, we expect to have more smokers in the first component of the mixture models with the dependent variable measuring BMI and changes in BMI. We expect that because smokers do not change behavior and therefore do not expect large changes in our dependent variables. On the other hand, we expect to have more smokers in the second component of in the model with the dependent variables of CO level, Cotinine level, change in CO level, and change in Cotinine level.

The first panel of Table 24 reveals that the average probability that a non-smoker belongs to the high BMI distribution is 67%. In contrast, the probability for a smoker to belong to the high BMI distribution is 47%. Thus, there is is still significant mixture of smokers and non-smokers in the high BMI distribution. One reason may be that we are missing important explanatory variables that assigns smokers from non-smokers into each distribution. The second panel presents results for the distribution probabilities for smokers with the dependent variable measuring the change in BMI. Here we find a clear separation, smokers and non-smokers belong to the second BMI distribution with a large gain in BMI, as reported in Table 24, with an average probability of 16% and 97%. Thus, as previously expected most participants with a large gain in BMI are non-smokers. Across the other four estimated models we find a similar expected separation of smokers and non-smokers by distributional means. Those with low levels of CO and Cotinine as well as large decreases in CO and Cotinine levels are more likely to be non-smokers, while smokers are more likely to belong to the higher mean CO, and Cotinine distributions and the distributions with smaller changes in CO and Cotinine levels.

Next, we focus on the distribution assignment of participants that misreport smoking status. As discussed, those individuals enter the model as smokers and it is of pertinent interest to understand

[56]The Posterior probability of a smoker belonging to the higher distribution (delayed component) is equal to $1 - \Phi(X\beta)$. Where beta includes the constant and the coefficient on smoking. Similarly, the probability can also be calculated by dividing the sum of zi by the number of iterations among smokers.

Table 23: Posterior Means and Variance of all Mixture Models

Dependent Variable BMI (N=5337)

Mean μ_1	Mean μ_2	Mean σ_1	Mean σ_2
23.21	28.84	2.49	3.73
(1.68)	(0.75)	(0.53)	(0.73)

Dependent Variable ΔBMI (N=5337)

Mean μ_1	Mean μ_2	Mean σ_1	Mean σ_2
0.21	1.56	1.00	1.86
(0.12)	(0.15)	(0.08)	(0.04)

Dependent Variable CO (N=3312)

Mean μ_1	Mean μ_2	Mean σ_1	Mean σ_2
4.80	24.12	2.28	13.91
(0.45)	(0.50)	(0.19)	(0.26)

Dependent Variable ΔCO (N=3312)

Mean μ_1	Mean μ_2	Mean σ_1	Mean σ_2
-25.63	-8.52	19.12	15.46
(1.01)	(0.94)	(0.39)	(0.81)

Dependent Variable Cotinine (N=5040)

Mean μ_1	Mean μ_2	Mean σ_1	Mean σ_2
5.53	275.52	6.52	163.30
(27.89)	(22.08)	(32.64)	(4.36)

Dependent Variable ΔCotinine (N=5040)

Mean μ_1	Mean μ_2	Mean σ_1	Mean σ_2
-136.61	-24.61	293.67	119.28
(7.83)	(5.63)	(11.75)	(10.14)

Notes: $\alpha + \mu_1$ presents the mean of the second mixture component. Standard deviation in parentheses.

Table 24: Posterior Probability Assignment to Lower and Higher Distributions for Smokers and Non-Smokers at the First Annual Follow-up Visit

Dependent Variable BMI

Prob. Low BMI		Prob High BMI	
Smoker=1	Smoker=0	Smoker=1	Smoker=0
0.53	0.33	0.47	0.67

Dependent Variable ΔBMI

Prob. Low ΔBMI		Prob High ΔBMI	
Smoker=1	Smoker=0	Smoker=1	Smoker=0
0.84	0.03	0.16	0.97

Dependent Variable CO

Prob. Low CO		Prob High CO	
Smoker=1	Smoker=0	Smoker=1	Smoker=0
0.01	0.99	0.99	0.01

Dependent Variable ΔCO

Prob. Low ΔCO		Prob High ΔCO	
Smoker=1	Smoker=0	Smoker=1	Smoker=0
0.08	0.98	0.92	0.02

Dependent Variable Cotinine

Prob. Low Cotinine		Prob High Cotinine	
Smoker=1	Smoker=0	Smoker=1	Smoker=0
0.01	0.52	0.99	0.48

Dependent Variable ΔCotinine

Prob. Low ΔCotinine		Prob High ΔCotinine	
Smoker=1	Smoker=0	Smoker=1	Smoker=0
0.33	0.99	0.67	0.01

which mixture component they are most likely part of. For example, if they are in fact smokers, we suspect that they are assigned with a higher probability to the low BMI and low change in BMI distribution. However, if those smokers are indeed non-smokers then we expect to find possibly mixed results, i.e. that a large portion is assigned to the high BMI distribution. Ideally, if the objective measure corrects for misreporting, then this additional smoking information should help identify the mixture component assignment. On the contrary, including self-reported smoking information should add more noise into the model because it assumes they are non-smokers while in fact they are not. In summary, we focus on the distribution assignments of the misreporters to understand whether they are more likely to be non-smokers than smokers.

Specifically, we analyze how 270 individuals, who self-reported to be non-smokers but the objective measures indicates that they are smokers, are assigned to the distributions across all six models.[57] Since we only have a sub-sample reporting CO levels, the results reflect estimates among 215 misreporters. Table 25 presents the distribution probabilities for misreporting participants in all mixture models.

The first panel presents the probability assignment to the BMI distributions. We find an average probability of belonging to the high BMI distribution of 48%. The second model reveals that misreporters belong to the high change BMI distribution with a probability of 77%. As a result, even when we utilize the objective smoking variable in the mixture models, there are many misreporters that belong to the high BMI distribution that is much more likely to consist of non-smokers. However, the probability assignment for the CO and Cotinine level mixture models estimates show that misreporters are much more likely to be part of smoking distributions with 97% and 99%, respectively. The relatively high levels of CO and Cotinine levels among many smokers and non-smokers may lead to such an probabilistic assignment. Thus, we concentrate on the changes in CO and Cotinine models that show a similar probability assignment for misreporters to the distributions that see large changes in the levels. Misreporters have an average probability of being part of the large decrease in CO and Cotinine distributions with 14% and 42%, respectively. We derive from this

[57] 46 individuals who are also misreporters have no BMI information in the sample

result that the objective smoking measure may not necessarily be more accurately measuring true smoking status. We reiterate that the descriptive statistics and the mixture model estimates suggests that there is a good amount of misreporters that ”mimic” features of non-smokers.

We proceed by estimating a more sophisticated level of misreporting. Our previous descriptive evidence suggested that 206 misreporters may in fact be not misreporting. We now add an additional layer of identification. Specifically, we condition on whether any of the 206 participants have a probability assignment of greater than 90% of belonging to the distributions mostly made up of non-smokers as defined by mixture models. For simplicity, we count how many misreporters belong to the high change in BMI distribution, low change in CO distribution and low change in Cotinine distributions. As an example, Figure 9 presents the probability assignment for misreporters to the high change in BMI distribution. Our final new calculation finds that between 68 and 109 misreporters may in fact be not misreporting smoking status.[58] Thus this indicates that depending on the sample used, between 26% and 51% of misreporters may not be misreporting.

Nevertheless, it is possible that most of the misreporters only reduced cigarette consumption, but never quit completely. We argue that this is only of limited concern, because there is ample evidence that many of the misreporters are in fact non-smokers. By the second annual follow-up 95 of the misreporters are objectively classified as non-smokers. It is unlikely that all of them quit in year two, because there was no smoking cessation treatment or hospital support after the first four months. Instead, we believe that it is more likely that most of them are misclassified, because there were no objective clinical measures that clearly separated non-smokers from smokers – as reported in Figure 9.

[58] Each separate calculation finds that 109 misreporters in the low CO change distribution, 78 in the high BMI change distribution and 68 in the low change Cotinine distribution satisfy the criteria.

Table 25: Posterior Probability Assignment to the Lower and Higher BMI Component for Misreporters in each Model

Dependent Variable BMI (N=270)

Prob. Low BMI	Prob High BMI
Misreporting=1	Misreporting=1
0.52	0.48

Dependent Variable ΔBMI (N=270)

Prob. Low ΔBMI	Prob High ΔBMI
Misreporting=1	Misreporting=1
0.23	0.77

Dependent Variable CO (N=215)

Prob. Low CO	Prob High CO
Misreporting=1	Misreporting=1
0.03	0.97

Dependent Variable ΔCO (N=215)

Prob. Low ΔCO	Prob High ΔCO
Misreporting=1	Misreporting=1
0.14	0.86

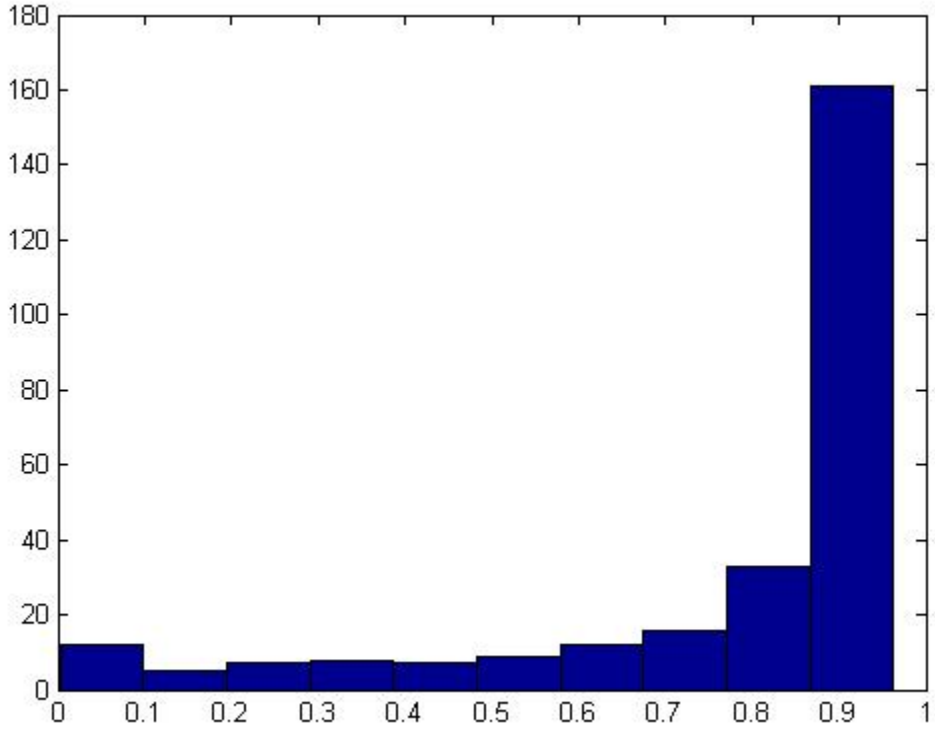
Dependent Variable Cotinine (N=263)

Prob. Low Cotinine	Prob High Cotinine
Misreporting=1	Misreporting=1
0.01	0.99

Dependent Variable ΔCotinine (N=263)

Prob. Low ΔCotinine	Prob High ΔCotinine
Misreporting=1	Misreporting=1
0.42	0.58

Figure 9: Posterior Probability of All Misreporting Participants Belonging to the Higher Mean Change BMI Distribution



In summary, our Bayesian mixture model approach utilizing information on BMI, CO, and Cotinine level presents evidence that smokers belong to a different distribution than non-smokers in all measures. Moreover, we find that a significant portion of the misreporters have a high probability of belonging to the distribution made up of mostly non-smokers. In the end we estimate that about 30% to 50% of misreporters may in fact be not misreporting smoking status. We believe that this finding shows the importance of accounting for misreporting and we believe that descriptive evidence and the mixture model approach allows researchers to identify "true" misreporters.

Conclusion

This paper provides new evidence to several questions regarding misreporting in self-reported smoking information. Descriptive statistics show that smoking cessation treatment group partici-

pants misreport twice as much as participants in the control group. We suspect that being in the treatment group contributes to the higher rate of misreporting by treatment group assignment. It is also conceivable that the participant's perception of being treated to achieve a personal and socially desirable outcome magnifies the misreporting level.

Our results indicate that being male, married, and of older age increases the probability of misreporting significantly. Our preferred estimates suggest that being male and married increases the probability of misreporting by about 10 percentage points. Sub-sample analyses by treatment group show that the relationship between demographic characteristics and misreporting is larger in the treatment groups than in the control group.

We also provide evidence that the objectively verified smoking measure may not be necessarily "better" compared to the self-reported smoking measure. In our second methodological approach we take an agnostic view on whether participants classified as misreporters are in fact misreporting smoking status. We utilize BMI, CO and Cotinine level information to inform if someone is a smoker and focus on those classified as misreporters. We apply a two distribution mixture model and our estimates suggest that there are two separate distributions. For example, our results indicate that when we model the two distributions of changes in BM, from the pre-treatment BMI level to the first annual follow-up visit, that the high change in mean BMI distribution is mostly made up of non-smokers. Based on descriptive evidence and the mixture model results, we believe that at least 30% of misreporters may in fact not be misreporting smoking status, reaching as high as 50%. Moreover, our results imply that common explanatory variables can help reveal wrongly misreported smoking participation in a mixture model framework.

Our results also have implications for the smoking literature. We show that participants who receive a smoking cessation treatment misreport at higher levels than in the control group. It is possible that the knowledge or perception of treatment may also occur in government programs. One such treatment could be a state's cigarette tax increase. If smokers are aware of a tax increase they may be more likely to under-report smoking participation in survey responses. In this case, the negative correlation between state cigarette taxes, and smoking participation may be overstated.

As a result, claims that higher taxes reduce smoking participation should be revisited.

We also provide evidence on the level of misreporting in smoking. We find relatively low levels of under-reporting in our adult sample. Clinicians classify 8% of self-reported non-smokers as smokers. Evidence from RCTs presents under reporting smoking levels up to 35% among pregnant women. Additionally, we find no evidence of over-reporting smoking participation. We conclude that under-reporting of smoking participation is significantly lower than in other studies. Additionally, we find evidence suggesting that a significant portion of misreporters may in fact not be misreporting their status at all. Thus, misreporting levels may in fact be even lower than the reported 8%.

Lastly, we provide evidence that can be relevant for the methodological literature addressing misreporting. We find that misreporting is endogenous along different demographic dimensions. Our results provide a benchmark to identify which characteristics need to be controlled for to achieve exogenous misreporting, a condition required by many papers to recover unbiased estimators. We also question the validity of estimators relying on exogenous misreporting if they cannot account for the demographic characteristics that affect misreporting, such as the presence of additional household smokers.

Appendix To Chapter I

Table A1: Cigarette Consumption per Day and Smoking Status by Month Prior to the First Annual Follow-up Visit

By Month	Mean Smokers	Mean Cigarettes per Day
12 Months Before Visit	0.88	26.77
	(0.32)	(16.99)
11 Months Before Visit	0.79	22.54
	(0.40)	(17.56)
10 Months Before Visit	0.70	18.53
	(0.46)	(17.20)
9 Months Before Visit	0.63	16.00
	(0.48)	(16.66)
8 Months Before Visit	0.62	15.62
	(0.48)	(16.50)
7 Months Before Visit	0.63	15.76
	(0.48)	(16.37)
6 Months Before Visit	0.65	16.02
	(0.47)	(16.19)
5 Months Before Visit	0.66	16.11
	(0.47)	(16.05)
4 Months Before Visit	0.66	16.21
	(0.47)	(15.97)
3 Months Before Visit	0.66	16.27
	(0.47)	(15.95)
2 Months Before Visit	0.67	16.27
	(0.47)	(15.93)
1 Month Before Visit	0.68	16.23
	(0.46)	(15.76)

Standard errors are in parenthesis. Columns 2 presents an average smoking status for a binary variable in each months. Column 3 presents self-reported average cigarette consumption in each month.

Table A2: The Effect of the Length of Abstaining from Smoking on Alcoholic Beverage Consumption at the First Annual Follow-up Visit

	Dep. Var. Drinks per Week	
Length of Non-Smoking	OLS	2SLS
12 Months Sustained Non-Smoker	-1.251***	-1.183***
	(0.403)	(0.428)
11 Months Sustained Non-Smoker	-0.614	-0.695
	(0.451)	(0.454)
10 Months Sustained Non-Smoker	-0.578	-0.571
	(0.356)	(0.363)
9 Months Sustained Non-Smoker	-0.587	-0.630*
	(0.360)	(0.370)
8 Months Sustained Non-Smoker	-0.284	-0.112
	(0.549)	(0.588)
7 Months Sustained Non-Smoker	-0.770	2.096***
	(0.958)	(0.676)
6 Months Sustained Non-Smoker	-0.232	-0.958
	(1.042)	(1.306)
5 Months Sustained Non-Smoker	0.522	1.249
	(1.438)	(2.160)
4 Months Sustained Non-Smoker	0.346	89.03
	(1.105)	(73.67)
3 Months Sustained Non-Smoker	-2.003**	-0.626
	(0.785)	(1.188)
2 Months Sustained Non-Smoker	-1.963***	-2.812
	(0.718)	(0.616)
1 Month Sustained Non-Smoker	0.298	0.049
	(0.926)	(1.263)

Robust standard errors are in parenthesis. ***,**, and * indicate significance at the 1, 5, and 10 percent level. Columns 2-3 present OLS and 2SLS estimates for a binary variable equal to one if a person becomes a sustained quitter for the rest of the year and zero for a continuous smokers. Each row entry presents a separate marginal effect.

Table A3: The Effect of the Length of Abstaining from Smoking on Alcoholic Beverage Consumption at the First Annual Follow-up Visit

	Dep. Var.Drinks per Week	
Length of Non-Smoking	NB	CF-NB
12 Months Sustained Non-Smoker	-1.49***	-1.39**
	(0.479)	(0.548)
11 Months Sustained Non-Smoker	-0.769	-0.862*
	(0.484)	(0.497)
10 Months Sustained Non-Smoker	-0.557	-0.572
	(0.375)	(0.383)
9 Months Sustained Non-Smoker	-0.558	-0.608
	(0.383)	(0.398)
8 Months Sustained Non-Smoker	-0.225	-0.076
	(0.553)	(0.575)
7 Months Sustained Non-Smoker	-0.786	-2.62***
	(1.02)	(1.00)
6 Months Sustained Non-Smoker	-0.288	-0.892
	(1.034)	(1.46)
5 Months Sustained Non-Smoker	0.122	0.617
	(1.26)	(1.69)
4 Months Sustained Non-Smoker	0.114	-0.048
	(1.105)	(1.47)
3 Months Sustained Non-Smoker	-2.06*	-0.404
	(1.21)	(1.49)
2 Months Sustained Non-Smoker	-2.749**	-5.01***
	(1.18)	(1.63)
1 Month Sustained Non-Smoker	0.017	-0.30
	(0.860)	(1.35)

Robust standard errors are in parenthesis. ***,**, and * indicate significance at the 1, 5, and 10 percent level. Columns 2-3 represent Negative Binomial and Control function approach Negative Binomial marginal effects for a binary variable equal to one if a person becomes a non-smoker for the rest of the year and zero for a continuous smokers. Each row entry presents a separate average marginal effect.

Table A4: The Effect of the Length of Abstaining from Smoking on Current Alcohol Consumption at the First Annual Follow-up Visit

	Female		Male	
Length of Non-Smoking	NB	CF-NB	NB	CF-NB
12 Months Sustained Non-Smoker	-1.983***	-2.127***	-1.142*	-0.882
	(0.538)	(0.614)	(.683)	(0.790)
11 Months Sustained Non-Smoker	-1.170*	-1.286**	-0.567	-0.658
	(0.632)	(0.661)	(0.641)	(0.652)
10 Months Sustained Non-Smoker	-0.399	-0.528	-0.751	-0.629
	(0.449)	(0.467)	(0.539)	(0.544)
9 Months Sustained Non-Smoker	-0.198	-0.298	-0.931*	-0.936*
	(0.475)	(0.513)	(0.547)	(0.557)
8 Months Sustained Non-Smoker	-0.0263	-0.245	-0.349	0.0276
	(0.675)	(0.738)	(0.765)	(0.775)
7 Months Sustained Non-Smoker	0.112	-3.729**	-1.437	-2.395**
	(1.478)	(1.813)	(1.250)	(1.215)
6 Months Sustained Non-Smoker	-1.460	-0.642	0.358	-0.961
	(1.353)	(1.780)	(1.382)	(2.006)
5 Months Sustained Non-Smoker	-1.494	-65.52***	0.911	1.584
	(1.681)	(4.437)	(1.657)	(1.907)
4 Months Sustained Non-Smoker	0.226	0.472	-0.211	-0.571
	(1.167)	(1.944)	(1.687)	(1.803)
3 Months Sustained Non-Smoker	0.0982	0.093	-4.392***	-1.264
	(1.117)	(1.923)	(1.928)	(1.549)
2 Months Sustained Non-Smoker	-3.086***	-4.274***	-2.023	-5.139*
	(0.802)	(1.249)	(1.862)	(2.817)
1 Month Sustained Non-Smoker	-1.730	-1.520	0.661	0.580
	(1.577)	(2.294)	(1.047)	(1.536)

This table presents presents Negative Binomial and Negative Binomial with control function approach average marginal effects for a binary indicator equal to one for a person becoming a sustained non-smoker in a month and zero if he is continuous smoker. This is a supplement for Table 8 presenting OLS and 2SLS results.

Appendix To Chapter II

Table A5: Sensitivity of Estimates to Different BMI Depreciation Rates

(IV With Year 5 BMI Only)

	Long Run (BMI Year 5)		
	Average Quit	Average Cigarettes	Average CO
$\delta = 0$	2.646***	-0.065***	-0.082***
	(0.285)	(0.007)	(0.011)
		[1.91]	[1.81]
$\delta = 0.05$	2.675***	-0.066***	-0.083***
	(0.288)	(0.007)	(0.011)
		[1.94]	[1.83]
$\delta = 0.1$	2.706***	-0.067***	-0.084***
	(0.291)	(0.007)	(0.011)
		[1.97]	[1.86]
$\delta = 0.15$	2.738***	-0.068***	-0.086***
	(0.294)	(0.008)	(0.011)
		[2.00]	[1.89]
$\delta = 0.2$	2.772***	-0.069***	-0.088***
	(0.297)	(0.008)	(0.011)
		[2.03]	[1.93]
$\delta = 0.25$	2.808***	-0.070***	-0.089***
	(0.300)	(0.008)	(0.011)
		[2.07]	[1.96]
$\delta = 0.30$	2.844***	-0.072***	-0.091***
	(0.304)	(0.008)	(0.012)
		[2.10]	[2.00]
N	5446	4966	4517

See notes for Table 2.

Figure A1: Semi-Parametric Graphs of Long-Run Effects of Simple Average Cigarettes per Day on BMI for Education and Age Subsamples

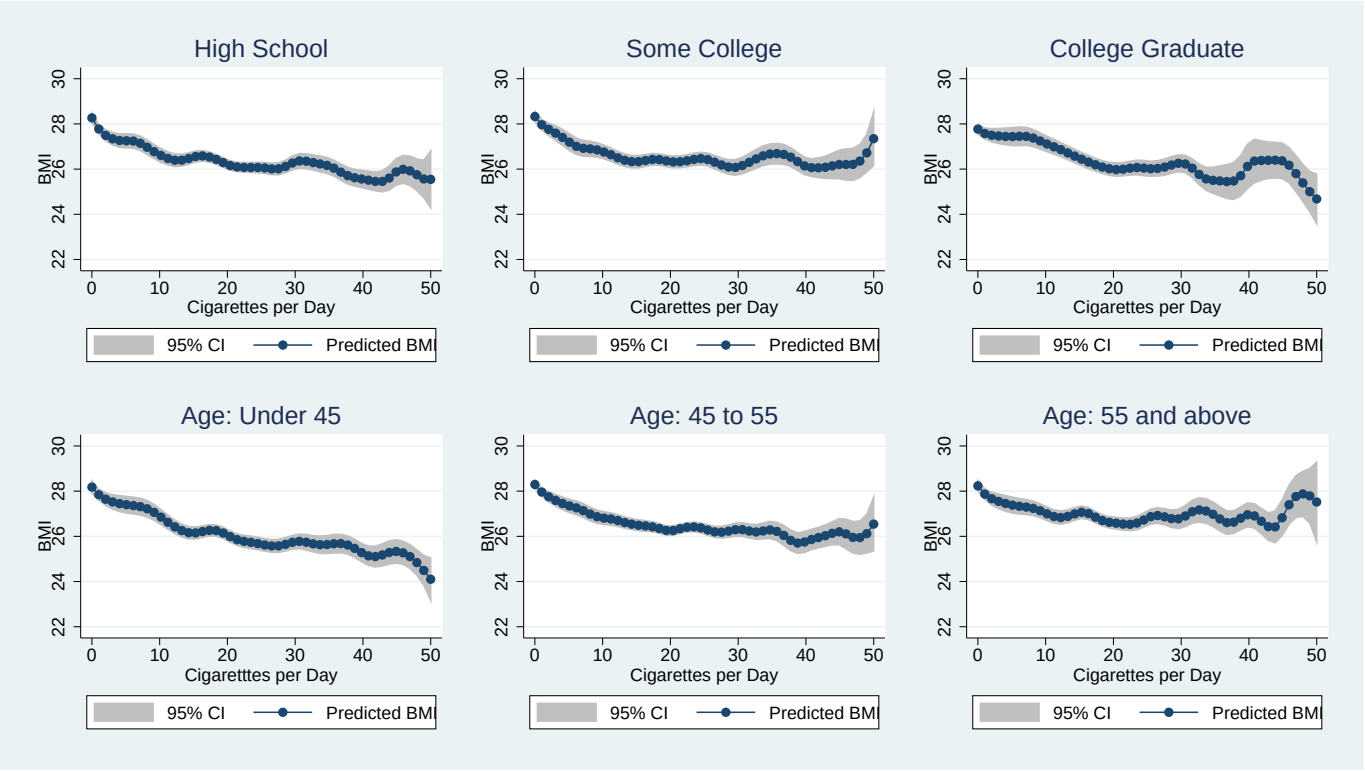


Figure A2: Semi-Parametric Graphs of Long-Run Effects of Simple Average Cigarettes per Day on BMI for Gender and Baseline BMI Subsamples

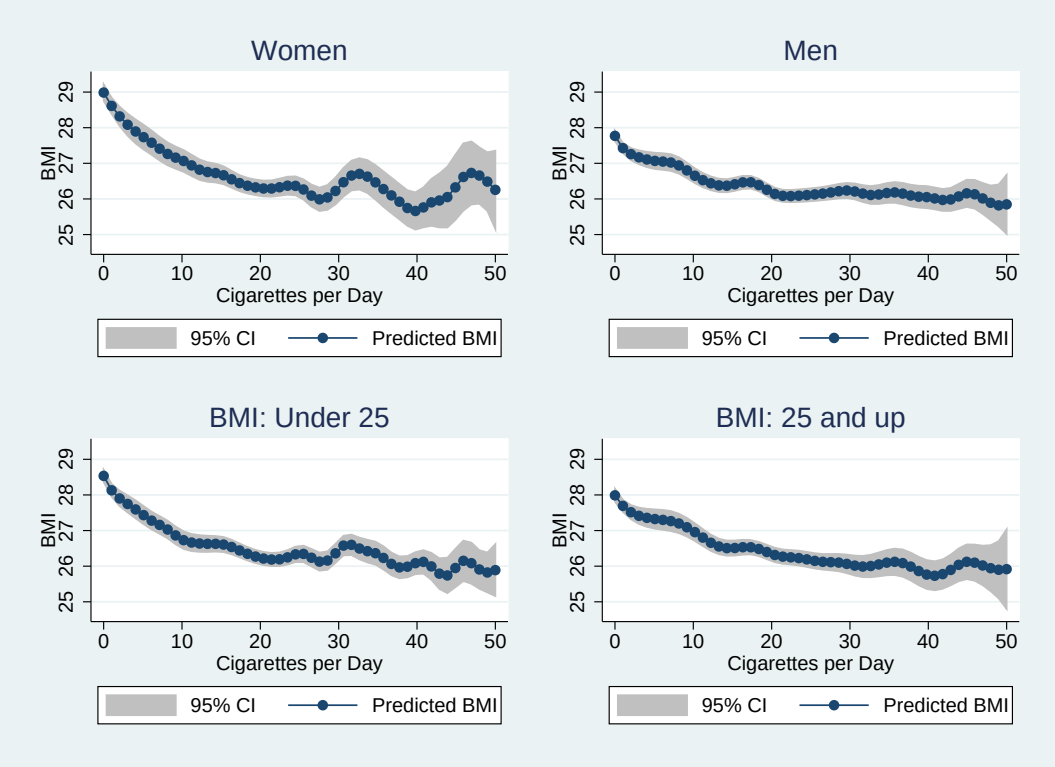


Figure A3: Semi-Parametric Graphs of Long-Run Effects of Simple Average CO on BMI for Education and Age Subsamples

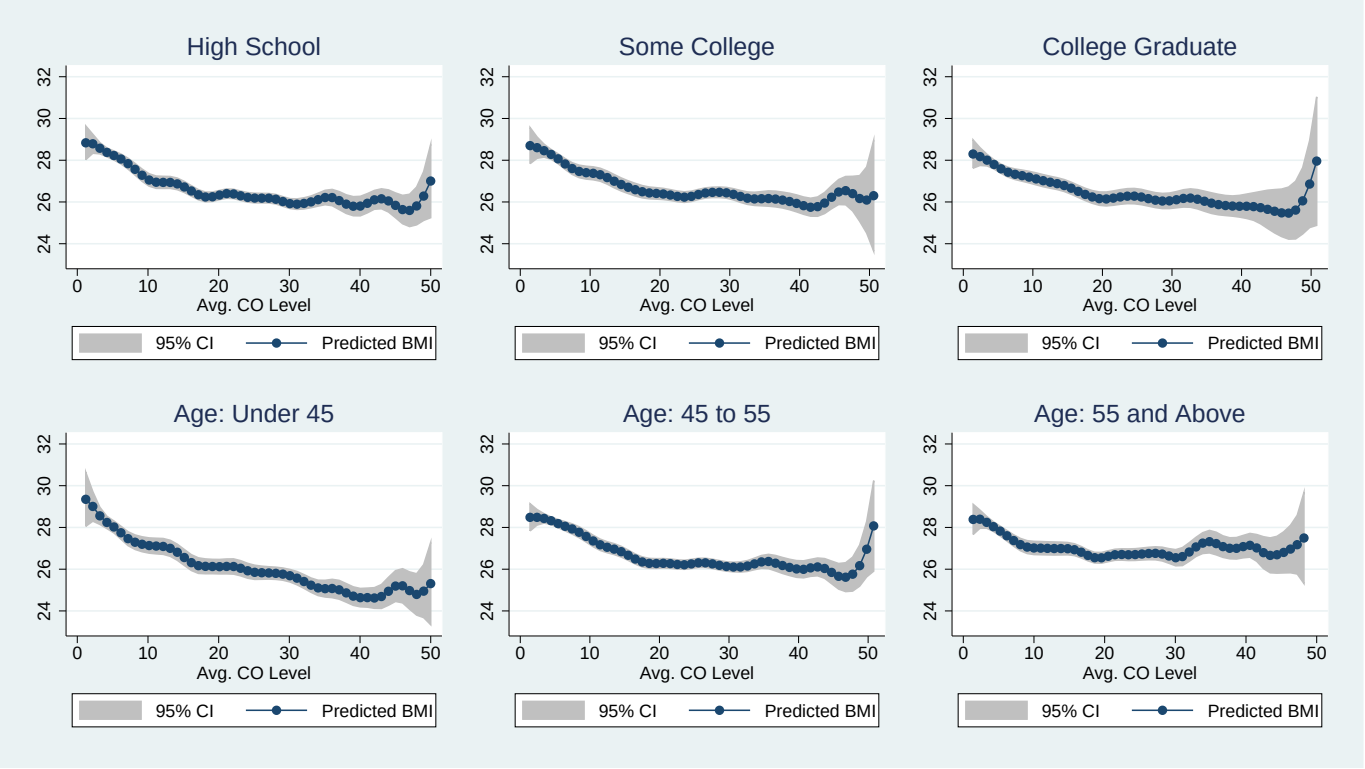
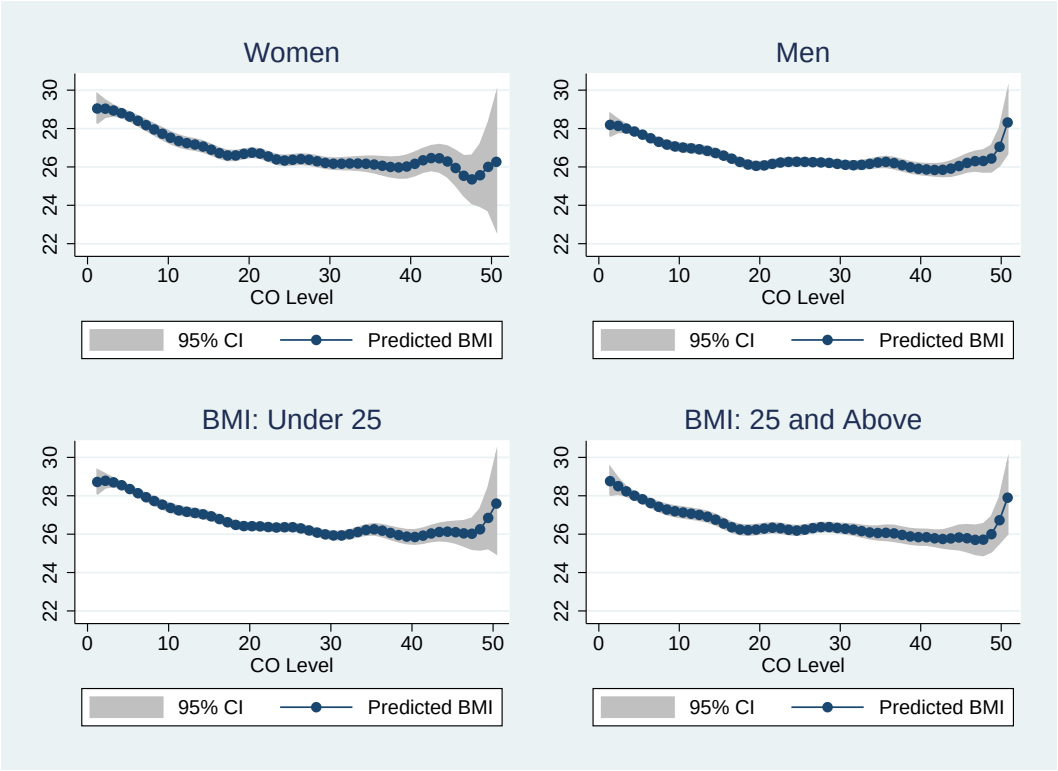


Figure A4: Semi-Parametric Graphs of Long-Run Effects of Simple Average CO on BMI for Gender and Baseline BMI Subsamples



Appendix To Chapter III

Table A6: Regression Estimates for Treatment Group 1 (SIA) Subsample

Dependent Variable: Misreporting Smoking Status

Male	0.0764***	0.0757***	0.0761***	0.0772***	0.0786***	0.0792***
	(0.0261)	(0.0261)	(0.0262)	(0.0261)	(0.0261)	(0.0261)
Education	-0.00541	-0.00610	-0.00577	-0.00585	-0.00532	-0.00541
	(0.00441)	(0.00446)	(0.00444)	(0.00442)	(0.00438)	(0.00437)
Married	0.00103	0.00311	0.00278	0.00819	0.00904	0.00552
	(0.0281)	(0.0280)	(0.0281)	(0.0283)	(0.0283)	(0.0286)
AGe	0.00485**	0.00488**	0.00491**	0.00483**	0.00472**	0.00473**
	(0.00199)	(0.00199)	(0.00199)	(0.00198)	(0.00198)	(0.00198)
BMI	-0.0118***	-0.0116***	-0.0117***	-0.0117***	-0.0118***	-0.0120***
	(0.00316)	(0.00316)	(0.00316)	(0.00317)	(0.00316)	(0.00316)
Nicotine Gum	-0.116***	-0.118***	-0.117***	-0.119***	-0.119***	-0.118***
	(0.0244)	(0.0245)	(0.0245)	(0.0247)	(0.0247)	(0.0247)
HH Smoker		-0.0347				
		(0.0276)				
# HH Smokers			-0.0149			
			(0.0208)			
HH Spouse Smokes				-0.0508	-0.0548*	-0.0524
				(0.0327)	(0.0326)	(0.0325)
HH Child Smokes					0.0384	0.0430
					(0.0412)	(0.0413)
HH Other Smokes						-0.0995*
						(0.0595)
Constant	0.299*	0.312*	0.304*	0.309*	0.304*	0.313**
	(0.157)	(0.159)	(0.158)	(0.158)	(0.157)	(0.157)
Observations	804	804	804	804	804	804
R-squared	0.051	0.052	0.051	0.053	0.054	0.056

Notes: Regressions include Heteroscedasticity robust standard errors in parentheses.

Table A7: Regression Estimates for Treatment Group 2 (SIP) Subsample

Dependent Variable: Misreporting Smoking Status

Male	0.0628**	0.0631**	0.0631**	0.0631**	0.0644**	0.0646**
	(0.0265)	(0.0265)	(0.0265)	(0.0265)	(0.0266)	(0.0268)
Education	0.00208	0.00215	0.00224	0.00217	0.00236	0.00234
	(0.00481)	(0.00482)	(0.00483)	(0.00483)	(0.00483)	(0.00484)
Married	0.0473*	0.0473*	0.0474*	0.0454*	0.0447*	0.0419
	(0.0261)	(0.0262)	(0.0262)	(0.0266)	(0.0266)	(0.0271)
Age	0.00146	0.00146	0.00147	0.00144	0.00149	0.00146
	(0.00187)	(0.00188)	(0.00187)	(0.00187)	(0.00187)	(0.00188)
BMI	-0.00874***	-0.00876***	-0.00873***	-0.00873***	-0.00876***	-0.00871***
	(0.00321)	(0.00322)	(0.00321)	(0.00321)	(0.00321)	(0.00321)
Nicotine Gum	-0.0525**	-0.0519**	-0.0516**	-0.0504*	-0.0503*	-0.0498*
	(0.0261)	(0.0263)	(0.0262)	(0.0264)	(0.0264)	(0.0263)
HH Smoker		0.00782				
		(0.0304)				
# HH Smokers			0.0123			
			(0.0239)			
HH Spouse Smokes				0.0233	0.0209	0.0208
				(0.0383)	(0.0380)	(0.0382)
HH Child Smokes					0.0411	0.0419
					(0.0505)	(0.0507)
HH Other Smokes						-0.0325
						(0.0602)
Constant	0.228	0.225	0.221	0.225	0.217	0.220
	(0.151)	(0.151)	(0.151)	(0.151)	(0.150)	(0.152)
Observations	792	792	792	792	792	792
R-squared	0.024	0.024	0.024	0.024	0.025	0.026

Notes: Regressions include Heteroscedasticity robust standard errors in parentheses.

Table A8: Regression Estimates for Control Group (UC) Subsample

Dependent Variable: Misreporting Smoking Status

Male	0.0567	0.0524	0.0504	0.0468	0.0508	0.0514
	(0.0513)	(0.0523)	(0.0526)	(0.0515)	(0.0531)	(0.0530)
Education	0.00516	0.00498	0.00528	0.00723	0.00789	0.00815
	(0.00936)	(0.00937)	(0.00939)	(0.00934)	(0.00921)	(0.00927)
Married	0.0647	0.0707	0.0744	0.0917*	0.0880*	0.0910*
	(0.0476)	(0.0491)	(0.0496)	(0.0504)	(0.0507)	(0.0504)
Age	0.00588*	0.00580*	0.00586*	0.00630**	0.00638**	0.00656**
	(0.00317)	(0.00317)	(0.00316)	(0.00315)	(0.00316)	(0.00319)
BMI	-0.00657	-0.00694	-0.00719	-0.00700	-0.00630	-0.00625
	(0.00617)	(0.00609)	(0.00606)	(0.00606)	(0.00612)	(0.00614)
Nicotine Gum	0.131	0.129	0.133	0.143	0.143	0.146
	(0.101)	(0.101)	(0.101)	(0.0997)	(0.1000)	(0.100)
HH Smoker		-0.0389				
		(0.0589)				
# HH Smokers			-0.0421			
			(0.0503)			
HH Spouse Smokes				-0.144**	-0.157**	-0.157**
				(0.0669)	(0.0663)	(0.0665)
HH Child Smokes					0.0640	0.0659
					(0.0927)	(0.0930)
HH Other Smokes						0.0627
						(0.139)
Constant	-0.127	-0.102	-0.101	-0.158	-0.195	-0.213
	(0.272)	(0.268)	(0.267)	(0.270)	(0.270)	(0.271)
Observations	207	207	207	207	207	207
R-squared	0.046	0.048	0.050	0.063	0.065	0.066

Notes: Regressions include Heteroscedasticity robust standard errors in parentheses.

References

Angrist, J., and J.S. Pischke, (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press: Print.

Aigner, D., (1973). "Regression with a Binary Independent Variable Subject to Errors of Observation." *Journal of Econometrics*, 1.1 : 49-59.

Almada, L., McCarthy, I., and R. Tchernis, (2015). "What Can We Learn about the Effects of Food Stamps on Obesity in the Presence of Misreporting?" *NBER Working paper No. w21596*, National Bureau of Economic Research. Cambridge MA.

Ansolabehere, S. and E. Hersh, (2012). "Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate." *Political Analysis*.

Ashley M., Olin, J., le Riche, W., Kornaczewski, A., Schmidt, W., and J. Rankin, (1977). "Morbidity in Alcoholics. Evidence for Accelerated Development of Physical Disease in Women." *Arch Intern Med.*, 1977 Jul;137(7):883-7.

Aubin, H., Farley, A., Lycett, D., Lahmek, P., and P. Aveyard, (2012). "Weight Gain in Smokers after Quitting Cigarettes: Meta-Analysis." *BMJ*, 345: e4439.

Bardy, A., Seppälä,T., Lillsunde, P. , Kataja, J., Koskela, P., Pikkarainen, J. , and V. Hiilesmaa, (1993). "Objectively Measured Tobacco Exposure during Pregnancy: Meonatal Effects and Relation to Maternal Smoking." *BJOG: An International Journal of Obstetrics & Gynaecology*,100(8), 721-726.

Barrett, S.P., Tichauer, M., Leyton, M., and R.O. Pihl, (2006). "Nicotine Increases Alcohol Self-Administration in Non-Dependent Male Smokers." *Drug and Alcohol Dependence*, 81:197–204.

Bask, M., and M. Melkersson, (2003b). "Should One Use Smokeless Tobacco in Smoking Cessation Programs? A Rational Addiction Approach." *The European Journal of Health Economics, formerly: HEPAC*, 4.4 (2003): 263-270.

Bask, M., and M. Melkersson, (2004). "Rationally Addicted to Drinking and Smoking?" *Applied Economics*, 36 4: 373-381.

Baum, C., (2009). "The Effects of Cigarette Costs on BMI and Obesity." *Health Economics*, 18(1): 3-19.

Becker, G., and K. Murphy, (1988). "A Theory of Rational Addiction." *Journal of Political Economy*, Vol. 96, No. 4 pp. 675-700.

Ben-Porath,Y. (1967). "The Production of Human Capital and the Life Cycle of Earnings." *Journal of Political Economy*, Vol. 75, No. 4, Part 1 (Aug., 1967), pp. 352-365.

Berg, N., and D. Lien, (2006). ”Same-Sex Sexual Behavior: US Frequency Estimates from Survey Data with Simultaneous Misreporting and Non-Response.” *Applied Economics*.Taylor and Francis Journals, vol. 38(7), pages 757-769.

Bien, T. H. and R. Burge, (1990). ”Smoking and Drinking: A Review of the Literature.” *Substance Use & Misuse*, 25.12 (1990): 1429-1454.

Blundell, R., and J. Powell, (2004). ”Endogeneity in Semiparametric Binary Response Models.” *Review of Economic Studies*, 71: 655-679.

Bobo, J.,Gilchrist, L., Schilling II, R., Noach, B. and P. Schinke, (1987).”Cigarette Smoking Cessation Attempts By Recovering Alcoholics.” *Addictive Behaviors*, Vol. 12 pp. 209-215.

Bobo, J., and C. Husten, (2000). ”Sociocultural Influences on Smoking and Drinking.” *Alcohol Reserach & Health*, Vol 24, No 4, pp. 225-232.

Bouchery, E.E., Harwood, H.J., Sacks, J.J., Simon, C.J. and R.D. Brewer, (2013). ”Economic Costs of Excessive Alcohol Consumption in the U.S., 2006.” *Am J Prev Med.*, Nov;41(5):516–524.

Bollinger, C., and M. David, (1997). ”Modeling Discrete Choice with Response Error: Food Stamp Participation.” *Journal of the American Statistical Association*, 92:827-835.

Bound, J., and A. Krueger, (1991). ”The Extent of Measurement Error in Longitudinal Earnings Data: Do two wrongs make a right?” *Journal of Labor Economics*, 1-24.

Boyd, N., Windsor, R., Perkins,L. , and J. Lowe, (1998). ”Quality of Measurement of Smoking Status by Self-Report and Saliva Cotinine Among Pregnant Women.” *Maternal and child health journal* 2, no. 2 : 77-83.

Brachet, T., (2008). ”Maternal Smoking, Misclassification, and Infant Health.” *Working Paper*. University of Pennsylvania.

Britton, G., Brinthaupt, J., Stehle, J., and G. James, (2004). ”Comparison of Self-Reported Smoking and Urinary Cotinine Levels in a Rural Pregnant Population.” *Journal of Obstetric, Gynecologic, & Neonatal Nursing*, 33(3), 306311.

Cameron, A. C., and P. K. Trivedi, (2005). ”Microeconometrics: Methods and Applications.” Cambridge university press.

Cameron, L., and J. Williams, (2001). ”Cannabis, Alcohol and Cigarettes: Substitutes or Complements?” *Economic Record*, 77 236: 19-34.

Card, D., (1996). ”The Effect of Unions on the Structure of Wages.” *Econometrica*, 64:957-979.

Carey, K. B., and D. B. Abrams, (1988). ”Properties of Saliva Cotinine in Young Adult Light Smokers.” *American journal of public health*, 78(7), 842-843.

Carmelli, D., Swan, G.E., and D. Robinette, (1993). ”The Relationship Between Quitting Smoking and Changes in Drinking in World War II veteran Twins.” *Journal of Substance Abuse* 5:103–116.

Carpenter, C., and P. Cook, (2008). "Cigarette Taxes and Youth Smoking: New Evidence from National, State, and Local Youth Risk Behavior Surveys." *Journal of Health Economics*, 27.2 : 287-299.

Cawley, J., and C. Meyerhoefer, (2012). "The Medical Care Costs of Obesity: An Instrumental Variables Approach." *Journal of Health Economics*, 31: 219-230.

Center for Disease Control News Release, (2014). accessed June 1st, 2015. http://www.cdc.gov/media/releases/2014/p1210-tips-roi.html

Center for Disease Control, (2016). http://www.cdc.gov/chronicdisease/resources/publications/aag/pdf/2015/alcoh aag.pdf

Center for Disease Control, (2014). Trends in Current Cigarette Smoking Among High School Students and Adults, United States, 1965–2014. Accessed March 16, 2016: http://www.cdc.gov/tobacco/data_statistics/tables/trends/cig_smoking/index.htm

Chaloupka, F. J., (1991). "Rational Addictive Behavior and Cigarette Smoking." *Journal of Political Economy*, August 1991, 722–42.

Chaloupka, F. J., and K. Warner, (2000). The economics of smoking. Culyer, A. and J. Newhouse, (Eds.) Chapter 29 in *Handbook of Health Economics*, Amsterdam: North-Holland, 1539-1627.

Chen, Y., Horne, S., and J. Dosman, (1993). "The Influence of Smoking Cessation on Body Weight May Be Temporary." *American Journal of Public Health*, 83(9): 1330-1332.

Chou, S , Grossman, M., and H. Saffer, (2004)." An Economic Analysis of Adult Obesity: Results from the Behavioral Risk Factor Surveillance System." *Journal of Health Economics*, 23(3): 565–87.

Connett, J. E., Kusek, J. W., Bailey, W. C., O'Hara, P., and M. Wu, (1993). "Design of the Lung Health Study: A Randomized Clinical Trial of Early Intervention for Chronic Obstructive Pulmonary Disease." *Control Clin Trials*, 14(2 Suppl): 3S-19S.

Corrigall, W.A., Coen, K.M., and K.L. Adamson, (1994). "Self-Administered Nicotine Activates the Mesolimbic Dopamine System through the Ventral Tegmental Area." *Brain Research*, 653:278–284.

Courtemanche, C., (2009). "Rising Cigarette Prices and Rising Obesity: Coincidence or Unintended Consequence?" *Journal of Health Economics*, 28(4): 781–98.

Courtemanche, C., Pinkston, J., Ruhm, C., and G. Wehby, *forthcoming*. "Can Changing Economic Factors Explain the Rise in Obesity?" *Southern Economic Journal*.

Courtemanche, C., Pinkston, J., and J. Stewart, (2015). "Adjusting Body Mass for Measurement Error with Invalid Validation Data." *Economics and Human Biology*, 19: 275-293

Courtemanche, C., R. Tchernis, and B. Ukert, (2016). "The Effect of Smoking on Obesity: Evidence from a Randomized Trial." *NBER Working paper No. w21937*. National Bureau of Economic Research. Cambridge MA.

DeCicca, P., and L. McLeod, (2008). "Cigarette Taxes and Older Adult Smoking: Evidence from Recent Large Tax Increases." *Journal of Health Economics*, 27.4: 918-929.

Decker, S. L., and A. E. Schwartz, (2000). "Cigarettes and Alcohol: Substitutes or Complements?" *NBER Working Paper no. 7535.* National Bureau of Economic Research. Cambridge MA.

Deveci, S. E., F. Deveci, Y. Açik, and A. Ozan, (2004). "The Measurement of Exhaled Carbon Monoxide in Healthy Smokers and Non-Smokers." *Respiratory medicine*, 98(6), 551-556.

Dierker, L., Lloyd-Richardson, E., Stolar, M., Flay, B., Tiffany, S., Collins, L., and Tobacco Etiology Research Network, (2006). "The Proximal Association Between Smoking and Alcohol Use Among First Year College Students." *Drug Alcohol Depend* 81(1): 1-9.

Eisenberg, D., and B. C. Quinn, (2006). "Estimating the Effect of Smoking Cessation on Weight Gain: An Instrumental Variable Approach." *Health Services Research*, 41(6): 2255–2266.

Enoch, M.A., and D. Goldman, (2001). "The Genetics of Alcoholism and Alcohol Abuse." *Current Psychiatry Reports*, 3:114–151,

Evans, W., (1999). "Do Workplace Smoking Bans Reduce Smoking?" *American Economic Review*, 89(4), 176-194.

Ezzati, M., Martin, H., Skjold, S., Vander Horn, S., and C. Murray, (2006). "Trends in National and State-Level Obesity in the USA after Correction for Self-Report Bias: Analysis of Health Surveys." *J R Soc Med.*, May; 99(5): 250-257.

Farley, A.C., Hajek, P., Lycett, D., and P. Aveyard, (2012). "Interventions for Preventing Weight Gain after Smoking Cessation." *Cochrane Database of Systematic Reviews*, Issue 1, Art. No.: CD006219.

Finkelstein, E. A., Trogdon, J. G., Cohen, J. W., and W. Dietz, (2009). "Annual Medical Spending Attributable to Obesity: Payer-and Service-Specific Estimates." *Health Affairs*, 28(5): 822-831.

Flegal, K., Troiano, R., Pamuk, E., Kuczmarski, R., and S. Campbell, (1995). "The Influence of Smoking Cessation on the Prevalence of Overweight in the United States." *New England Journal of Medicine*, 333(18): 1165-1170.

Flegal, K., Carroll, M., Kuczmarski, R. and C. Johnson, (1998). "Overweight and Obesity in the United States: Prevalence and Trends, 1960-1994." *International Journal of Obesity*, 22: 39-47

Flegal, K., Graubard, B., Williamson, D. and M. Gail, (2005). "Excess Deaths Associated with Underweight, Overweight, and Obesity." *Journal of the American Medical Association*, 293: 1861-1867.

Fletcher, J., (2014). "The Effects of Smoking Cessation on Weight Gain: New Evidence Using Workplace Smoking Bans." *Forum for Health Economics and Policy*, 17(2): 105-129.

Funk, D., Marinelli, P.W., and A.D. Lee, (2007). "Biological Processes Underlying Co-Use of Alcohol and Nicotine: Neuronal Mechanisms, Cross-Tolerance, and Genetic Factors." *Alcohol Research & Health*, 29(3):186–190.

Gardner, M. N., and A. M. Brandt. (2006). ""The Doctors' Choice Is America's Choice" The Physician in US Cigarette Advertisements, 1930-1953." *American journal of public health* 96, no. 2 (2006): 222-232.

Goel, R. K., and M. J. Morey, (1995). "The Interdependence of Cigarette and Liquor Demand." *Southern Economic Journal*, 62 2: 451-459.

Gregory, C. A., and C. J. Ruhm, (2011). "Where does the Wage Penalty Bite?" Grossman, M and Mocan, N., (Eds.) in *Economic Aspects of Obesity*, University of Chicago Press: 315-347.

Gruber, J., Sen, A. and M. Stabile, (2003). "Estimating Price Elasticities When There Is Smuggling: The Sensitivity of Smoking to Price in Canada." *Journal of Health Economics*, 22 5: 821-842.

Gruber, J., and M. Frakes, (2006). "Does Falling Smoking Lead to Rising Obesity?" *Journal of Health Economics*, 25(2): 183-197.

Gross, J., Johnson, J., Sigler, L., and M. L. Stitzer, (1995). "Dose Effects of Nicotine Gum." *Addictive behaviors*, 20(3), 371-381.

Hausman, J., Abrevaya, J., and F. Scott-Morton, (1998). "Misclassification of the Dependent Variable in a Discrete-Response Setting." *Journal of Econometrics*, 87:239-69.

Heckman, J., (1979). "Sample selection Bias as a specification Error." *Econometrica*, 47(1): 153-161.

Heckman, J., and R. Robb, (1986). "Alternative Methods for Solving the Problem of Selection Bias in evaluating the Impact of Treatments on Outcomes." *Drawing Inferences from Self-selected Samples*, 63-107.

Hedenstrom, P., Malmberg, H., and H. V. Fridriksson, (1986). "Reference Values for Lung Function Tests in Men: Regression Equations with smoking Variables." *Upsala Journal of Medical Science*, 91: 299-310.

Hu, Y., and S. Schennach, (2008). "Instrumental Variable Treatment of Nonclassical Measurement Error Models." *Econometrica*, 76.1 : 195-216.

Hug, S., (2010). " The Effect of Misclassifications in Probit Models: Monte Carlo Simulations and Applications." *Political Analysis*, 18:78-102.

Jones, A., (1989). "A System Approach to the Demand for Alcohol and Tobacco." *Bulletin of economics Research*, 41:2 85-105.

Jo, W. K., and J. W. Oh, (2003). "Evaluation of CO Exposure in Active Smokers While Smoking Using Breath Analysis Technique." *Chemosphere*, 53(3), 207-216.

Kenkel, D., Mathios, A., and R. Pacula, (2001). "Economics of Youth Drug Use, Addiction and Gateway Effects."*Addiction*, 96:151-164.

Koop, G., Poirier, D. J., and J. L. Tobias, (2007). *Bayesian econometric methods*, Cambridge University Press.

Kreider, B., Pepper, J., Gundersen, C., and D. Jolliffe, (2012). "Identifying the Effects of SNAP (food stamps) on Child Health Outcomes when Participation is Endogenous and Misreported." *Journal of the American Statistical Association*, 107, no. 499: 958-975.

Laugesen, M., Epton, M., Frampton, C. M., Glover, M., and R. A. Lea, (2009). "Hand-Rolled Cigarette Smoking Patterns Compared with Factory-Made Cigarette Smoking in New Zealand Men." *BMC Public Health*, 9(1), 1.

Larsson, A., and J.A. Engel, (2004) "Neurochemical and Behavioral Studies on Ethanol and Nicotine Interactions." *Neuroscience and Biobehavioral Reviews*, 27:713–720.

Lee. S., (2007). "Endogeneity in Quantile Regression Models: A Control Function Approach." *Journal of Econometrics* 141: 1131-1158.

Levy, D., and K. Friend, (2003). "The Effects of Clean Indoor Air Laws: What Do We Know and What Do We Need to Know." *Health Education Research*, 18(5): 592-609.

Lewbel, A., (2007). "Estimation of Average Treatment Effects with Misclassification." *Econometrica*, 75.2: 537-551.

Mahajan, A., (2006). "Identification and Estimation of Regression Models with Misclassification." **Econometrica**, 74(3), 631-665

Manson, J. E., Colditz, G. A., Stampfer, M. J., Willett, W. C., Rosner, B., Monson, R. R., Speizer, F. E., and C. H. Hennekens, (1990). "A Prospective Study of Obesity and Risk of Coronary Heart Disease in Women." *New England Journal of Medicine*, 322(13): 882-889.

Marquis, K. H., and J. Moore, (2010). "Measurement Errors in SIPP Program Reports." *Survey Methodology*, 1.

Marquis, K.,and J. Moore, (1990). "Measurement Errors in SIPP Program Reports." *Proceedings of the Bureaue of the Census Annual Research Conference*, Washington, D.C.: Bureau of the Census.

McCarthy, I., and R. Tchernis, (2011). "On the Estimation of Selection Models when Participation is Endogenous and Misclassified." *Advances in Econometrics, Missing-Data Methods: Cross-sectional Methods and Applications*, 27: 179-207.

Meyer, B., Mok, D. W., and J. X. Sullivan, (2009), "The Under-Reporting of Transfers in Household Surveys: Its Nature and Consequences." *NBER Working Paper No. 15181*, National Bureau of Economic Research, Cambridge, MA.

Meyer, B. and R. Goerge, (2011). "Errors in Survey Reporting and Imputation and their Effects on Estimates of Food Stamp Program Participation." *US Census Bureau Center for Economic Studies Paper No. CES-WP-11-14*.

Mizoue, T., Ueda, R., Tokui, N., Hino, Y., and T. Yoshimura, (1998). "Body Mass Decrease after Initial Gain Following Smoking Cessation." *International Journal of Epidemiology*, 27(6): 984-989.

Mokdad, A. H., Bowman, B. A., Ford, E. S., Vinicor, F., Marks, J. S., and J. P. Koplan, (2001). "The Continuing Epidemics of Obesity and Diabetes in the United States." *Journal of the American Medical Association*, 286(10): 1195-1200.

Murray, R. P., Connett, J. E., Lauger, G. G., and H. T. Voelker, (1993). "Error in Smoking Measures: Effects of Intervention on Relations of Cotinine and Carbon Monoxide to Self-Reported Smoking." The Lung Health Study Research Group. *American journal of public health*, 83(9), 1251-1257.

Nguimkeu,P., Denteh, A., and R. Tchernis, (2016). " On the Estimation of Treatment effects with Endogenous Misreporting." Working paper.

Nonnemaker, J., Finkelstein, E., Engelen, E., Hoerger, T., and M. Farrelly, (2009). "Have Efforts to Reduce Smoking Really Contributed to the Obesity Epidemic?" *Economic Inquiry*, 47(2): 366-376.

Ogden, C.L., Carroll, M., Kit, B.K., and K. M. Flegal, (2014). "Prevalence of Childhood and Adult Obesity in the United States, 2011-2012." *Journal of the American Medical Association*, 311(8): 806-814.

Ogden, C. and M. Carroll, (2010). "Prevalence of Overweight, Obesity, and Extreme Obesity among Adults: United states, Trends 1960–1962 through 2007–2008." *Division of Health and Nutrition Examination Surveys*.

Ogden, C., Carroll, M., Kit, B., and K., Flegal, (2012). "Prevalence of Obesity in the United States, 2009–2010." *NCHS data brief*, no 82.

O'Hara, P., Grill, J., Rigdon, M. A., Connett, J. E., Lauger, G. A., and J.J. Johnston, (1993). "Design and Results of the Initial Intervention Program for the Lung Health Study. The Lung Health Study Research Group." **Preventive medicine**, 22(3): 304-315.

O'Hara, P., Connett, J. E., Lee, W. W., Nides, M., Murray, R., and R. Wise, (1998). "Early and Late Weight Gain Following Smoking Cessation in the Lung Health Study." **American journal of epidemiology**, 148(9): 821-830.

Pacula, R. L. (1998). "Does Increasing the Beer Tax Reduce Marijuana Consumption?" *Journal of Health Economics*, 17 5: 557-585.

Philipson, T., and R. Posner, (1999). "The Long-Run Growth in Obesity as a Function of Technological Change." *NBER Working Paper No. 7423*, National Bureau of Economic Research, Cambridge, MA.

Picone, G. A., Sloan F, JG.Trogdon, (2004). "The Effect of the Tobacco Settlement and Smoking Bans on Alcohol Consumption." *Health Economics*, 13 10: 1063-1080.

Pierani, P., and S. Tiezzi, (2009). "Addiction and Interaction between Alcohol and Tobacco Consumption." *Empirical Economics*, 37 1: 1-23.

Pinkowish, M., (1999). ”Hand in Glove: Smoking Cessation and Weight Gain.” *Patient Care*, 33(2): 134.

Potthoff, A.D., Ellison, G., and L. Nelson, (1983). ”Ethanol Intake Increases during Continuous Administration of Amphetamine and Nicotine, but Not Several Other Drugs.” *Pharmacology Biochemistry and Behavior*, 18:489–493.

Rashad, I., Grossman, M., and S. Chou, (2006). ”The Super Size of America: An Economic Estimation of Body Mass Index and Obesity in Adults.” *Eastern Economic Journal*, 32(1): 133-148.

Rexrode, K. M., Hennekens, C. H., Willett, W. C., Colditz, G. A., Stampfer, M. J., Rich-Edwards, J. W., Speizer, F. E., and J. E. Manson, (1997).” A Prospective Study of Body Mass Index, Weight Change, and Risk of Stroke in Women.” *Journal of the American Medical Association*, 277(19): 1539-1545.

Robinson, P., (1988). ”Root-N Consistent Semiparametric Regression.” *Econometrica*, 56(4): 931-954.

Rose, J.E., Brauer, L.H., Behm, F.M., Cramblett M, Calkins K, D. Lawhon, (2004). ”Psychopharmacological Interactions between Nicotine and Ethanol.” *Nicotine and Tobacco Research*, 6:133–144.

Russell, M., P. Cole, and E. Brown, (1973).”Absorption by Non-smokers of Carbon Monoxide from Room Air Polluted by Tobacco Smoke.” *The Lancet*, 301.7803: 576-579.

Sacks, J. J., Gonzales, K. R., Bouchery, E. E., Tomedi, L. E., and R. D. Brewer, (2015). ”2010 National and State Costs of Excessive Alcohol Consumption.” *Am J Prev Med*, 49(5): 73-79.

Schoenborn C.A. (2004). ”Marital Status and Health: United States, 1999–2002. Advance Data from Vital and Health Statistics.” no 351. Hyattsville, Maryland: National Center for Health Statistics.

Soderpalm, B., Ericson, M., Olausson, P., Blomqvist, O., and J.A. Engel, (2000). ”Nicotinic Mechanisms Involved in the Dopamine Activating and Reinforcing Properties of Ethanol.” *Behavior and Brain Research* 113:85–96.

Smith, B.R., Horan, J.T., Gaskin, S., and Z. Amit, (1999). ”Exposure to Nicotine Enhances Acquisition of Ethanol Drinking by Laboratory Rats in a Limited Access Paradigm.” *Psychopharmacology*, 142:408–412.

Stock, J. H., Wright, J.H., and M. Yogo, (2002). ”A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments.” *Journal of Business & Economic Statistics*, 20.4.

Swan, G.E., Carmelli, D., and L.R. Cardon, (1997). ”Heavy Consumption of Cigarettes, Alcohol and Coffee in Male Twins.” *Journal of Studies of Alcohol*, 58:182–190.

Tauras, J., (2006). ”Smoke-free Air Laws, Cigarette Prices and Adult Cigarette Demand.” *Economic Inquiry*, 44 (2), pp. 333–342.

Terza, J., Basu, A., and P. Rathouz, (2008). "Two-Stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling." *Journal of Health Economics*, 27(3): 531-543.

Thun, M.J., Carter, B.D., Feskanich, D., Freedman, N.D., Prentice, R., Lopez, A.D., Hartge, P. and S.M. Gapstur, (2013). "50-year trends in smoking-related mortality in the United States." *New England Journal of Medicine*, 368(4), pp.351-364.

Travier, N., Agudo, A., May, A., et al., (2012). "Longitudinal Changes in Weight in Relation to Smoking Cessation in Participants of the EPIC-PANACEA study." *Preventive Medicine*, 54: 183-192.

U.S. Department of Health and Human Services, (1964). Smoking and Health: Report of the Advisory Committee of the Surgeon General of the Public Health Service. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Coordinating Center for Health Promotion, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

U.S. Department of Health and Human Services, (1990). The Health Benefits of Smoking Cessation: A Report of the Surgeon General. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Coordinating Center for Health Promotion, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

U.S. Department of Health and Human Services. (2001). Women and Smoking: A Report of the Surgeon General. Atlanta: 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,

U.S. Department of Health and Human Services. (2012). Preventing Tobacco Use Among Youth and Young Adults: 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.

U.S. Department of Health and Human Services, (2012). Health, United States, 2011: With Special Feature on Socioeconomic Status and Health. U.S. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. National Center for Health Statistics.

U.S. Department of Health and Human Services. (2014). The Health Consequences of Smoking — 50 Years of Progress: A Report of the Surgeon General. Atlanta: 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.

Verardi, V., and N. Debarsy, (2012)." Robinson's Square Root of N Consistent Semiparametric Regression Estimator in Stata". *The Stata Journal*, 12(4): 726–735.

Wang,Y. C., McPherson, K., Marsh, T., Gortmaker, S. L., and M. Brown, (2011). "Health and Economic Burden of the Projected Obesity Trends in the USA and the UK." *Lancet*, 378(9793): 815-825.

Webb, D., N. Boyd, D. Messina, and R. Windsor, (2003). ”The Discrepancy Between Self-Reported Smoking Status and Urine Cotinine Levels Among Women Enrolled in Prenatal Care at Four Publicly Funded Clinical Sites.” *Journal of Public Health Management and Practice*, 9, no. 4: 322-325.

Wehby, G., and C. Courtemanche, (2012). ”The Heterogeneity of the Cigarette Price Effect on Body Mass Index.” *Journal of Health Economics*, 31(5): 719-729.

Williamson, D. F., Madans, J., Anda, R. F., Kleinman, J. C., Giovino, G. A., and T. Byers, (1991). ”Smoking Cessation and Severity of Weight Gain in a National Cohort.” *New England Journal of Medicine*, 324(11): 739-745.

Wolf, J., (2004). ”Applications of New Technologies in Travel Surveys.” *7th International Conference on Travel Survey Methods*.

Wooldridge, J. M., (2010). ”Econometric Analysis of Cross Section and Panel Data.” *MIT press*.

Zhao, X., and M. N. Harris, (2004). ”Demand for Marijuana, Alcohol and Tobacco: Participation, Levels of Consumption and Cross-Equation Correlations.” *Economic Record*, 80 251: 394-410.

Vita

Benjamin David Ukert was born on May 25, 1985 in Hamburg, Germany. After graduating from High School in Hamburg Germany he attended the University of North Florida in Jacksonville, FL, graduating in 2009 with a B.B.A. in Economics and a B.B.A in Finance.

Prior to his graduate studies, Benjamin attended the University of Duisburg-Essen in Essen, Germany, pursuing a M.Sc. in Finance and Energy.

Benjamin began his doctoral studies at the Andrew Young School of Policy Studies at Georgia State University in the fall of 2011. He initially worked as a research assistant for Dr. Erdal Tekin, and beginning in fall 2014 for Dr. Rusty Tchernis. His work was supported by the CEAR's Fellowship, and the George J. Malanos Scholarship.

During his graduate studies, Benjamin presented papers on the topic of risky behaviors at the Western Economics Association and the Southern Economic Association's annual conference. He also presented research several times to internal seminar audiences at the Andrew Young School.

In addition to his research activities, he served as an instructor to several undergraduate and graduate economics course, including the undergraduate Introductory Economics class. He also served as vice-president of the Graduate Student Association. He earned his M.A. in Economics from Georgia State University in 2014.

Benjamin was awarded a Ph.D. in Economics by Georgia State University in 2016. Upon graduation he will begin working as a post-doctoral researcher at the Leonard Davis Institute of Health Economics in the Wharton School, University of Pennsylvania.