LABOR MARKET EFFECTS OF OBESITY, SMOKING, AND ALCOHOL USE

By

ILKER DASTAN

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Economics

written under the direction of

Louise B. Russell

and approved by

____

____

____

____

New Brunswick, New Jersey

October, 2010

Abstract Of The Dissertation

Labor Market Effects of Obesity, Smoking, and Alcohol Use

By Ilker Dastan

Dissertation Director:

Louise B. Russell

This dissertation analyzes the joint effects of obesity, smoking, and binge drinking on wages and on unemployment by using the National Longitudinal Survey of Youth data set. The main objective of this study is to show that the effects of these behaviors on wages and unemployment may not be measured accurately in analyses that consider only one or two since these behaviors are correlated or tend to cluster. My results illustrate that failing to include one or more of the health behaviors in wage or unemployment regression would lead to an underestimation of the impact of being obese and an overestimation of the effect of binge drinking for both genders. However, when endogeneity is addressed by employing the Hausman-Taylor instrumental variable (HTIV) method in wage analysis and the multivariate probit method in unemployment analysis, I find that the estimated parameters of obesity or binge drinking are not statistically significantly different whether these behaviors are considered individually or simultaneously.

This study also conducts several sensitivity analyses. Firstly, the results reveal that the effects of these behaviors are not interactive. Secondly, the paper illustrates that the wage penalties for daily smoking are fairly constant over the wage distribution for both genders, but obesity affects the wages of males and females relatively more at lower quantiles of wages, and there is no wage penalty for being a binge drinker for either gender. Further, it is found that smokers are a heterogeneous group of people. In particular, the wage and unemployment effects of persistent smokers are different than beginning smokers and quitters. Moreover, obesity affects the wages and the likelihood of being unemployed of males only at the extremes of obesity. Lastly, I find evidence of wage penalties for being obese or a smoker in private sector jobs, but in the public sector only male smokers face lower wages.

Acknowledgements

I gratefully acknowledge my dissertation advisor, Prof. Louise B. Russell, for her guidance and support during my whole graduate study life. She first introduced me to health economics, guided and mentored me in many ways, and spent countless hours sharing her valuable and professional knowledge about economics and health economics with me. This dissertation would have been impossible without her invaluable and ongoing support.

I am also grateful to my committee members Jeffrey Rubin, Mark Killingsworth, and Irina Grafova for their insightful comments and help. Prof. Killingsworth's Applied Microeconometrics course and our discussions outside class taught me the fundamentals that are essential to my research. I greatly benefited from discussing my research and healthcare economics with Prof. Rubin. He was also an excellent Undergraduate Director and working with him helped me a lot to be a better instructor. I also tremendously learned from Prof. Grafova about basics of risky behaviors. She provided enormous insight about issues in my research and my future plans.

I would like to thank my friends Cem Iyigun, Costanza Biavaschi, Sami Demiroluk, and Volkan Cetinkaya who spent long hours to help my research. I have been fortunate to have their friendship and to have all the great comments and suggestions from them. I would also like to thank my other friends Abdullah Karaman, Evin Acan, Ibrahim Bakir, Ibrahim Hokelek, Kaveh Akram, Kemal Karakayali, Mustafa Karakayali, Sezgin Kiren for making me feel that I was not alone during this journey. I would also like to thank Dorothy Rinaldi for her support from the first day of my Ph.D. process.

Most importantly, none of this would have been possible without support, assistance, and patience of my parents, my sisters, and my brother. I will never find words of gratitude to my family.

Dedication

To My Parents, Recayi and Meliha Dastan

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

Current studies show that about 35% of the US population is obese, approximately 20% of Americans are smokers, and around 15% are excessive alcohol drinkers (CDC, 2007)[1]. Although there has been a decline in smoking rates in recent decades, the prevalence of obesity rates has risen rapidly while drinking rates have remained constant. Obesity is an important risk factor for numerous health problems including coronary heart disease, diabetes, osteoarthritis, hypertension, and stroke. Excessive alcohol consumption can lead to liver problems, infertility, cancer, and unintentional injuries due to accidents. Smoking has deadly consequences, including lung, larynx, esophageal, and oral cancers (WHO, 2002). Additionally, these health behaviors are widely documented to cause major problems with potentially important social and economic consequences, including increases in medical expenditures, lost productivity, work absence, unemployment, social penalties and discrimination (Harwood, 2000; Sturm, 2002).

Economists have given considerable attention to the effects of smoking, obesity, and excessive alcohol use on labor market outcomes. However, their results on the effects of these health behaviors on workers' wages differ widely. Despite the potentially large

---

[1] Obesity is defined as BMI (kg/m^2) of 30 or more, and identification for those under age 21 is through age- and gender-specific. Smoking is defined as smoking cigarettes daily or smoking an average of 1 or more cigarettes daily. Binge drinking is defined as drinking 6 or more alcoholic drinks in 4 or more occasions past month.

negative effects of health behaviors on unemployment, surprisingly little research have been done on this issue. Further, none of the earlier studies considered the effects of all three health factors on labor market outcomes in the same analysis. Cigarette smoking, excessive alcohol use, and poor eating habits tend to reinforce each other (Betts, 2000), and it is known that smokers are less likely to be obese. Hence, when these health factors are considered individually, estimates of their effects on wages or unemployment may be biased since the behaviors may be correlated. Therefore, the main focus of this dissertation is to estimate the joint effects of obesity, smoking and excessive alcohol use on wages and on unemployment using the National Longitudinal Survey of Youth (NLSY) data.

The NLSY consists of a nationally representative sample of 12,686 young men and women who were 14-21 years old when they were first surveyed in 1979. Data were collected yearly from 1979 to 1994, and biennially from 1996 to the present. The NLSY gathers information about this large group of young adults as they make the transition from school to work, and it provides a comprehensive overview of respondents' labor force experiences, career changes, labor market attachment, marriage, education investments, etc. (NLSY79 User's Guide, 2002). The use of a longitudinal data set enables the estimation of panel data methods, e.g. fixed effects model, which help avoid the potential bias caused by unobserved individual factors not captured in cross-sectional models. The use of panel data methods is an important extension to the literature because the data sets used in previous studies are mostly cross-sectional.

This study offers myriad additional contributions to literature: Firstly, the existing literature suggests that these health behaviors may be endogenous or, more specifically, missing or unobservable determinants of both behaviors and labor market outcomes may be correlated. Because the present data set contains extensive information on demographics, health behaviors, and related lifestyle behaviors, the analysis is able to examine health behaviors/labor market outcomes relationships while also addressing the potential endogeneity of the behaviors via Hausman-Taylor Instrumental Variable model in wage analyses and multivariate probit model in unemployment analyses, which appear to be the most appropriate methods for the purpose and the sample.

Another goal of this study is to examine whether the effects of these health behaviors on wages and unemployment are interactive or additive. For example, the effects of smoking and excessive drinking are interactive if the effects of excessive drinker smokers differ from the two effects added together, e.g. if they are due to the same unobserved factors. However, the effects would be additive if the health of the individual worsens more when the individual is attached to another behavior in addition to the current behavior. For this purpose, interaction terms are included in the analyses and are tested for statistical significance.

Thirdly, the present study separates males and females, and blacks, Hispanics and whites into subsamples to conduct a richer examination of gender and racial/ethnic differences. Further, one may argue that obese persons, smokers or binge drinkers may tend to be in public sector jobs since workers are more exposed to wage discrimination in the private sector. The effects of obesity, smoking and binge drinking on wages are examined for private and public sector workers separately.

Additionally, one may argue that obese, smoker, or binge drinkers may tend to be in less restricted jobs, i.e. part-time jobs or self-employed jobs, due to employer, customer or co-worker based discriminations. Moreover, since obesity, smoking, or binge drinking has adverse effects on health, people with these behaviors may observe health limitations to work full-time and may choose to work part-time or to be self-employed. Hence, the effects of three health factors on full-time/part-time employment probabilities and on self-employment/employment-not-self probabilities are also examined.

Lastly, higher wage workers may be less affected by these health behaviors, possibly because of the nature of their jobs, than lower wage workers or workers with less education, or vice versa. Hence, I test whether the effects of the three factors on wages vary across the wage structure.

The results reveal that the effect of obesity is underestimated and the effect of binge drinking is overestimated for both genders if one fails to control for all health behaviors in the same analysis. However, the results also show that once endogeneity is controlled for, the estimated parameters of health behaviors are not statistically significantly different whether these behaviors are considered individually or all together in both wage and unemployment analyses. Moreover, obesity has negative effects only on the wages of females (a penalty of 4.3%), and smoking wage penalties are 4.8% and 2.5% for males and females, respectively. Being obese increases the probability of being unemployed by 1.8% for females only, smoking increases the likelihood of being unemployed by 2.5% and 1.7% for males and females, respectively. Binge drinking has no effect on wages or on unemployment for either gender.

Sensitivity analyses illustrate that none of the interaction terms of health behaviors are statistically significant, suggesting that the effects of health behaviors are not interactive. Furthermore, the results from ethnic/racial subsamples illustrate that although smoking has wage penalties for all subsamples, obesity affects the wages of only white females, and binge drinking affects wages of only white males. Also, I find evidence of wage penalties for obesity or smoking in private sector jobs. However, in the public sector, only males face lower wages due to smoking.

Moreover, the results reveal that the wage penalties for daily smoking are fairly constant over the wage distribution for both genders. However, obesity affects the wages of males and females relatively more at lower quantiles, and there is no wage penalty for being a binge drinker for males and females at any quantile.

Multinomial logit models provide contradicting results: binge drinker males are less likely to be out of labor force than employed and they are less likely to be part-time employed than full-time employed. Daily smoker males and females are less likely to be self-employed than employed by others and obese males are less likely than their counterparts to be self-employed than employed by others. One possible explanation could be that people having these health behaviors who face wage penalties may move to the sectors or full-time employment where the wage structure is fixed, even if they receive lower wages. Alternatively, these individuals may tend to misreport their true weight, height, smoking and drinking behaviors.

I also estimated wage and unemployment models using more detailed measures of health behaviors. The results, when endogeneity is addressed, demonstrate that smokers are a heterogeneous group of people: the wage and unemployment effects of persistent smokers are different than starters, quitters or young experimenters. Furthermore, obesity appears to affect the wages and the likelihood of being unemployed of males only at the extremes[2]. Moreover, contrary to the previous literature, when endogeneity is accounted for, drinking is found to have no positive effect on wages or on unemployment for either gender.

The structure of the dissertation is as follows: conceptual rationales for why the behaviors may affect labor market outcomes and the literature review are presented in chapter 2. The empirical methodology is discussed in chapter 3. The data is described in chapter 4 together with the descriptive statistics. The wage and the unemployment effects are presented in chapters 5 and 6, respectively, and the final chapter of the dissertation presents conclusions and a discussion of limitations and opportunities for further research.

[2] Morbid obesity is defined as having a BMI or 35 or more.

CHAPTER 2

Conceptual Rationales and Literature Review

2.1.Conceptual Rationales:

During the past 20 years in the United States, the percentage of the population that reports drinking and binge drinking has stayed the same and the percent of smokers has been decreasing, but there has been a remarkable rise in obesity (CDC, 2003, 2007). The 2003-2004 National Health and Nutrition Examination Survey (NHANES) shows that the percentage of US population that is obese increased from 15% in 1976-1980 to 23% in 1988, 31% in 1999-2000 and 33% in 2003-2004. According to a recent study released by the Centers for Disease Control and Prevention (CDC), obesity prevalence has not noticeably increased in the past few years but the level is still high at 34%. Janet Collins, the director of CDC's National Center for Chronic Disease Prevention and Health Promotion says "CDC made the prevention of obesity one of its top public health priorities in view of these alarmingly high rates of obesity in all population groups" (CDC 2007).

According to recent national surveys, more than half of the adults in the US drank alcohol in the past 30 days (CDC 2007). CDC's Behavioral Risk Factor Surveillance System data show that the prevalence of binge drinking (consuming 5 or more drinks on an occasion during the past 30 days) among adults in the US has remained around 15%

and heavy drinking (consuming an average of two or more drinks per day) has been stayed around 5% between 1993 and 2007.

The CDC report also demonstrates that the percentage of US adults who were smokers has decreased from 42% in 1965 to 25% in 1995 and to about 20% in 2007. Along with a decrease in the prevalence of smoking, cigarettes per capita consumption decreased from 3,849 in 1980 to 2,474 in 1995 and to 1,814 in 2004 (Department of Agriculture's Tobacco Outlook, 2006). Also, between 1990 and 2007, cigarette production decreased by about 34%, exports decreased by 34% as well and consumption fell by 31% (CDC, 2007).

According to reports from the Center for Disease Control and Prevention (CDC) smoking, obesity and excessive alcohol consumption are the three leading lifestyle-related causes of death in the US each year (CDC, 2007). All three health behaviors have impacts that can increase the risk of many detrimental health conditions. For instance, excessive alcohol consumption can lead to liver problems, infertility, cancer and unintentional injuries due to accidents. Smoking has deadly consequences, including lung, larynx, esophageal, and oral cancers. Obesity is an important risk factor for a number of major diseases including coronary heart disease, diabetes, osteoarthritis, hypertension and stroke (WHO, 2002).

Annually about 80,000 deaths are attributed to excessive alcohol use, 8.6 million people have a serious disease caused by smoking, approximately 443,000 people die prematurely from smoking or exposure to secondhand smoke, and around 300,000 deaths each year are associated with obesity in the US (CDC, 2007). The medical costs of these health behaviors are large. Existing studies mention that every year about 6-11% of medical care costs are estimated to be attributable to smoking in the US during past 20 years and annual medical expenditures for alcohol related problems was $26.3 billion in 2000 (Rice, 1999; Harwood, 2000). The studies also show that 6-8% of total health spending in the US is attributable to obesity each year and the medical costs of obesity were $50 billion in 1995 when the obesity rates were lower than current levels (Wolf et al., 1998; Finkenstein et al., 2003).

Obesity, smoking, and excessive alcohol use have also substantial negative social and economic effects. The losses due to health behaviors come in the form of adverse labor supply effects and reduced on-the-job performance, i.e. lost productivity, possible absence from work, increased probability of unemployment, reduced earnings as a result of fewer and lower-quality job opportunities, social stigmatization and discrimination. In 1992 alcohol abuse and dependence imposed an estimated $246 billion in costs on the US economy, while the estimated lost productivity of workers suffering from alcohol problems accounted for $176 billion (Harwood, 2000). An economic burden of $97 billion per year is results from lost productivity attributable to smoking and in 2000 the aggregate cost of obesity was estimated to be $117 billion (Wolf, 2001).

Existing studies explain the correlation between health behaviors and labor market outcomes in four ways (Morris, 2007):

(1) Health behaviors may cause lower wages or unemployment. This might occur for the following four reasons:

a) First, obesity, smoking or binge drinking can be weakening health conditions. Thus, the obese, smokers or binge drinkers are likely to be less productive than their counterparts and consequently less likely to be employed or more likely to receive lower wages.

Obesity is a major health problem related to a number of serious diseases. Burkhauser and Cawley (2004) assert that increases in weight raise the probability of health-related work limitations. Furthermore, obesity can lead to some psychological problems occurring from discrimination, humiliation, social rejection or unhappiness with oneself. Existing research also shows that excessive alcohol use and smoking are associated with a long list of physical, psychological, and cognitive impairments. These problems includs liver and heart damage, cancer, significant time spent consuming and seeking alcohol and cigarettes, hangovers from drinking, legal and social consequences of drinking, reduced hand-eye coordination, preoccupation with cigarette and alcohol, unusual or unstable behavior, increased accidents and injuries (NIAAA, 1993). These problems at the personal level may also affect the individual's relationship to the labor market. Hence, obesity, smoking, and binge drinking may reduce work ability and increase absence from work, thereby reducing the individual's productivity, which in turn leads to lower wages and a lower likelihood of employment. (Mullahy and Sindelar 1993, 1996; Kenkel and Ribar 1994; Burkhauser and Cawley, 2004).

Furthermore, research shows there is a J-shaped association between alcohol consumption and the risk of cardiovascular disease, which suggests that alcohol consumption at moderate levels may be beneficial for health by relieving stress and reducing the incidence of heart disease (Sesso, 2001; Baum and Ford, 2006). Therefore, light or moderate drinking may be favorable to an individual's labor market productivity and might have positive effects on wages and on employment.

One may think that adverse health effects of these three behaviors generally appear late in life, and as a result these behaviors would not have any effect on labor market outcomes for people in our sample since most of the respondents in NLSY data are young adults. However, earlier literature also shows that smoking, drinking or obesity has some adverse health effects on young people (Conway and Cronan, 1992, Hoad and Clay, 1992).

b) Second, there may be discrimination against the obese, smokers or binge drinkers, which lessens their likelihood of being employed or negatively impacts their wages.

There may be prejudice by employers, employees, or customers reflecting their distaste or negative preferences for workers with certain unhealthy behaviors, particularly against obese women and smokers (Becker, 1971; Moon and McLean, 1980). It has been also documented in numerous experimental studies that there is weight-based discrimination at every stage of employment, from the hiring decision through wage-setting and promotion (Puhl and Brownell, 2001). For example, the Obesity Action Coalition (OAC) asserts that Wal-Mart treats obese employees unfairly, because they claim that Wal-Mart restricts the hiring of obese and out of shape workers by structuring jobs in a way that it is difficult for obese workers to do the manual task (OAC, 2005). Also, workers with these health behaviors in customer oriented services or sales occupations, which need direct public contact and interaction, could experience a wage penalty through the interaction with customers. Customers might be reluctant to interact with obese workers, and this would lead to a decrease in wages for these workers (Baum and Ford, 2004).

Furthermore, there might be discrimination in training opportunities: people with health behaviors would not have equal opportunity to train due to prejudice and this may lead to lower wages. Although smoking is prohibited in many work places, in the places where it is permitted, the adverse effects of second-hand smoke could make it possible for co-workers or employers to object to working with smokers and subsequently cause discrimination against these smoker workers (Levine et al., 1997).

Alternatively, there may be stereotyping by employers, arising from beliefs that the people with health behaviors are lazier and less intellectually or socially skilled, thus less productive (Everett, 1990). But this discrimination may have statistical validation if the obese, smokers or binge drinkers have more sick days, higher quit rates, and lower productivity than their non-obese, non-smoker or non-binge drinker peers.

In addition, discrimination against the individuals with health behaviors may arise through uncertainty, or a lack of knowledge about the productivity of these individuals (Pagan and Davila, 1997). Not all obese, smokers or binge drinkers are unhealthy even though obesity, smoking and binge drinking are related to poor health. But when an employer has imperfect information about the true productivity and health of a person with health behavior, the employer may choose to hire a non-obese, non-smoker or non-binge drinker person to maximize profits, given the evident that there is statistical relationship between increasing weight, cigarette smoking and excessive alcohol use and health problems. Finally, obesity is negatively valued in wealthy societies.

c) Third, these workers could be more costly for employers who provide health care. It is apparent that obese persons, smokers or drinkers will need more medical care than their counterparts. Economic theory implies that wages are decreasing in fringe benefits. Therefore, higher health care costs would lead to lower wages if employers provide health insurance benefits. The health problems associated with smoking, drinking or obesity may lead smokers, drinkers or obese workers to prefer jobs that include employer provided health insurance at the cost of a lower wage. This would result in a negative wage outcome for smoking, drinking or obesity (Levine et al., 1997; Baum and Ford, 2004).

d) Fourth, obesity, smoking or binge drinking may indirectly affect labor market outcomes. Studies also show that individuals who have these behaviors delay marriage, have reduced labor market experience, lower educational attainment, and greater probability of divorce (Yamaguchi and Kandel, 1985; Kenkel and Ribar, 1994), which indirectly affect their labor market success and reduce their productivity.

(2) Unemployment or lower wages may cause obesity, smoking or binge drinking:

Unemployment, low income level, social isolation or economic constraints may restrict access to healthy food and safe exercise, and may increase the likelihood of weight changes (Cawley, 2004). Unemployment may also affect weight if there is a relationship between long-term unemployment and mental health, and mental health is correlated with obesity (Bove and Olson, 2005). Additionally, sociology and alcohol research literatures state that unemployment causes emotional and financial stress, and stress is known to be related to weight gain, increased alcohol consumption, and smoking (Forcier, 1985; Harris et al., 1998).

On the other hand, labor market outcomes could influence smoking, drinking, or obesity through the income effect: people may spend more money on drinking or smoking when they earn more. Petry (1995) asserts that alcohol is a normal good and as income increases, expenditures on drinking increases. However, we cannot be sure that people whose income increases drink more as they earn more. Expenditures on drinking might be increasing because people spend more money on drinking when they earn more but they may consume the same amount since they prefer higher quality alcohol now. Heien (1996) finds that income is not a significant determinant of amount of alcohol consumed. Furthermore, it is unlikely that changes in wages affect the drinking decision since the drinking decision is for the most part not based on monetary concerns; alcohol is often available for free in work gatherings or at parties. Additionally, Hamilton and Hamilton (1997) find that price is not a determinant of whether or not an individual drinks.

Becker et al. (1994) find that the demand for cigarettes is related to both individual and family income. Becker et al., and Lewit and Coate (1982) find positive income elasticity in studies that estimate the demand for cigarettes, but Keeler et al. (1988) and Wasserman et al. (1991) find income elasticity that is not significantly different from zero. If one assumes that cigarettes are normal good, then higher wages would lead to more smoking and this result may underestimate the true negative effect of smoking on wages. However, this assumption may not be valid since earlier literature has contrasting results.

To date, very few studies have been able to estimate the causal impact of income on weight. Bhattacharya and Sood (2005) claim that weight loss does not always increase or decrease with income, but McArthur et al. (2001) and Jahns et al. (2003) find that there is a significant inverse relationship between socioeconomic status and income and obesity.

(3) There may be unobserved variables that are correlated with both employment/wages and health behaviors:

A third factor may explain the relationship between health behaviors and employment/unemployment, e.g. time preference, self-control, sociability, reduced aspirations and lower commitment to values underlying the desire for success, etc. For instance, this unobserved factor could reflect preferences for current time as opposed to preferences for the future. Hence, people with poor health behaviors are less likely to be concerned about the future and possible negative effects of these behaviors, and may invest less in human capital now and may be less likely to engage in job training. As a result, they would have a low likelihood of being employed or a lower earning profile (Fuchs 1974; Becker et al. 1994).

Alternatively, another unobserved factor, such as sociability, may be positively correlated both with labor market success and moderate drinking. For instance, alcohol may help an individual network during the time spent with colleagues and could help the person get additional information about promotion opportunities or may serve as a signal of commitment to the firm at company meetings (Montgomery, 1991; MacDonald and Shields, 2001).

These factors are generally difficult to observe and quantify, leading to an omitted variable or unobserved heterogeneity problem in the statistical estimations of the health behaviors and labor market outcomes regardless of the perspective taken on causation.

(4) Health behaviors may be measured systematically with error due to unobserved factors correlated with employment or labor market success. For example, this may occur if individuals in lower socioeconomic groups are more likely to underreport their weights, cigarette use or alcohol use, or over report their heights, etc. (Greeve, 2007).

The goal of the dissertation is to identify the first effect and to produce unbiased estimates that are not contaminated by the other effects.

2.2. Literature Review:

In the economics literature, there are many studies that investigate the effects of obesity, smoking or drinking on labor market outcomes and a considerable number of studies employ NLSY79 data. I employ the same data in my analysis. The NLSY contains a larger sample, a longer time period, and more health behavior and work information of individuals than any other data source. It also gives one the opportunity to use a large number of exogenous background characteristics in the analysis. Therefore, I focus more on the studies that used the NLSY data.

In the literature, each study uses a different sample from the NLSY. Most of the studies use wage levels methods, probit/logit models, while some employ wage changes methods, and a few use panel data techniques or instrumental variable methods. The wage levels method uses cross-sectional data whereas the wage changes method focuses on the characteristics of panel data to correct for unobserved heterogeneity and it assumes that change in the wages is a function of the change in the explanatory variables. Similarly, the other panel data techniques such as fixed-effects, between-effects or random-effects models help to avoid the possible bias caused by unobserved factors not captured in cross-sectional analyses. Potential endogeneity is also addressed by instrumental variables techniques. The estimates are generally greater in magnitude than OLS estimates but more exposed to biases due to weak identification. Therefore, the studies that use different methods are not measuring the same thing and it is reasonable to see different results.

2.2.1. Effects of Obesity:

The effects of obesity on labor market success, i.e. earnings, labor supply and occupation selection, have been analyzed in numerous US studies. The studies that do not use NLSY data find obesity wage penalties only for females (Sargent and Blanchflower, 1994; Behrman and Rosenzweig, 2001). The results of the studies that employ NLSY data and cross-sectional analyses report different results: some studies find wage penalties for both genders (Averett and Korenman, 1996; Maranto and Stenoien, 2000), some argue there is no penalty for either genders (Loh, 1993), and the remaining find that obesity reduces female wages, but not those of males (Register and Williams, 1990; Pagan and Davila, 1997). Cawley (2004) uses NLSY data and accounts for endogeneity by employing fixed effects and instrumental variable (IV) models and finds obesity wage penalty only for white females. Employing NLSY and individual and sibling fixed effects models, Baum and Ford (2004) find an obesity wage penalty for both genders but more for females.

Earlier studies that do not use NLSY79 data find obesity wage penalties for both genders, with greater penalties for female workers. Employing a data set from England, Sargent and Blanchflower (1994) find that both current and lagged (BMI of 7 years ago) obesity measures for females reduce their current wages, but the obesity wage penalty is much smaller and insignificant for male workers. Behrman and Rosenzweig (2001) inspect the obesity wage penalty for twin females. Using a fairly small sample of 808 twin pairs from Minnesota, they find no relationship between body mass index and wages or between weight and wages.

Other studies that use NLSY79 data also find obesity wage penalties but their results differ. Gortmaker et al. (1993) find that by regressing the wages on body weight from seven years earlier, after seven years of follow-up women who had been overweight had lower earnings than those who had not been overweight, but the results are statistically insignificant for males. Register and Williams (1990) investigate the effect of obesity on wages using cross-sectional data from the 1982 NLSY. Their results show that, among the 18-25 year old respondents, obesity reduces females' wages by 12% but has no significant effects for males. Pagan and Davila (1997) agree with Register and Williams. They employ similar methods to define the sample. After deleting all observations with missing values and keeping only respondents who reported working during 1988, they estimate cross-section wage levels models using 1989 NLSY data and

7,292 respondents and find that obesity reduces female wages, but not male wages. Using BMI, BMI^2, mild obesity (defined as 20% over standard weight) and morbid obesity (defined as 100% over standard weight) as risk factors, Maranto and Stenoien (2000) find obesity wage penalties for both genders with larger penalties for females.

Averett and Korenman (1996) use a cross-sectional sample from 1988 NLSY data including employed and non-employed persons in the sample. They find that men (about 8%) and women (10-24%) both suffer obesity wage penalties. Loh (1993) examines the effects of obesity on wages with an analysis of wage levels of full-time workers in the 1982 NLSY and of wage changes between 1982 and 1985. He finds no effect of obesity on male or female wage levels in 1982, while finding that wages grew about 5% less between 1982 and 1985 for obese men and women. Although these studies use the same data, their results differ. Differences might be occurring because the authors use different samples and different waves of the NLSY data.

The two most recent studies, Cawley (2004) and Baum and Ford (2004), use NLSY79 data to examine the effects of weight and the effects of being obese on wages. Employing wage level models using NLSY data from 1981 through 1998 waves with 45,120 female person-year observations and 51,899 male person-year observations, Cawley finds that weight has a negative effect on wages. To control for unobserved heterogeneity, he uses fixed effects and instrumental variable models using BMI of a sibling as an instrument and finds that the obesity wage penalty disappears for white males, Hispanics and blacks, but still exists for white females. Baum and Ford find similar results employing the same years of NLSY data as Cawley used with 6,437 respondents and 51,500 person-year observations. Employing individual and sibling fixed effects models, they find that obese female workers get 6% less than their non-obese female control group, but obesity lowers male wages by only around 3% only. I follow Cawley's and Baum and Ford's articles and use their methods to define my sample and to choose the obesity variable.

All studies delete the observations with missing values in height, weight or hourly wage information, but only Loh focuses on full-time workers. Averett and Korenman include even non-employed respondents in the sample. Using cross-section wage levels method, fixed effects or changes in wages method might be another reason why they have contrasting results. Loh finds that obesity affects wages among full-time male and female workers, but when Loh employs wage changes method, he finds no significant effect of obesity on workers' wages. It is reasonable to see a different result here, because these two methods measure different things; he finds that obesity does not affect wage growth but affects wage levels. Furthermore, Register and Williams use cross-sectional data from 1982 NLSY. The respondents were between the ages of 18 and 25 in 1982; thus their results do not generalize to the broader age population since the respondents are young and they do not have enough labor force experience.

The effect of obesity on employment has received less attention and generally been examined mainly in European studies. Some of the European studies find insignificant effects of obesity on employment for both genders (Harper, 2000), some find positive effects on the long-term unemployment only for females (Sarlio-Lahteenkorva and Lahelma, 1999), and some find weak evidence that obese individuals are more likely to be unemployed (Garcia and Quintana-Domeque, 2007). Morris (2007) addresses the endogeneity by employing the bivariate probit (IV) model and finds that obesity has a negative effect on employment for both males and females. Greeve (2008) also uses IV models and finds negative effects of BMI on employment. Cawley (2000a) is the only US study that examines employment probabilities. He uses a conventional IV regression approach and finds that BMI has a positive effect on employment for mothers.

The effect of obesity on employment/unemployment has generally been examined in European studies. In one of the British studies, Sargent and Blanchflower (1994) used the 1981 round of the National Child Development Study (NCDS) to examine the impact of obesity (BMI at the 90^{th} percentile of the sample or greater, and at the 99th percentile or greater) at age 16 years on unemployment at age 23 years. Their estimates of a logit model show that obesity has an insignificant effect on unemployment. Using a later (1991) round of the NCDS, Harper (2000) estimates the impacts of obesity (a BMI in the 80–89th percentiles and the 90–100th percentiles), physical appearance, attractiveness and height at age 23 years on unemployment at age 33 years. Logistic model results similarly showed that obesity had an insignificant effect on unemployment for both males and females.

Sarlio-Lahteenkorva and Lahelma (1999) use a 1994 Finnish data set to examine the effect of current obesity (measured as a BMI of 30 or more) on current employment and long term unemployment (being unemployed for two years or more in the previous five year period) using a logistic model. Their results show that obesity only has a significant positive effect on long-term unemployment for females; other results are insignificant. Another Finnish study by Laitinen et al. (2002) investigates the relationship between obesity (BMI greater than or equal to 30) in adolescence and a long history of employment or other adverse social outcomes at age 31 years. Employing a longitudinal data set of 9,754 subjects born in 1966 in Finland and using logistic regressions, they fail to find any correlation between overweight and obesity at 14 years old and a long history of unemployment at 31 years old.

Using a survey based data coming from the European Community Household Panel (ECHP), Sousa (2005) applies the propensity score technique in order to assess the effect of BMI on labor market outcomes in Europe. Pooling all countries together, she finds the average treatment effect for those having a BMI above 25 decreases labor force participation for women, but it increases male labor force participation. Employing the 1994-2001 period of the same data, Garcia and Quintana-Domeque (2007) use three different measures of body size, BMI, weight in kilograms and being obese (measured as BMI of 30 or more), to examine the associations between body size measures, employment status and wages for several European countries. The multinomial logit estimates show that there is weak evidence that obese individuals are more likely to be unemployed or tend to be more segregated in self-employment jobs than non-obese individuals. The relationship between obesity, unemployment and wages is different for men and women, and there is heterogeneity across countries. Lundborg et al. (2006) use the Survey of Health, Ageing and Retirement in Europe (SHARE) to analyze the effect of obesity on employment, hours worked and hourly wages in 10 European countries for people aged 50 and above. Pooling all countries, they find that obesity is negatively associated with being employed for both men and women and with female hourly wages. They also find that the effects differ across Europe.

None of these studies investigated the endogeneity of obesity and unemployment. Using data from two rounds, 1997-1998, of the Health Survey for England, Morris (2007) examines the effect of obesity on employment. To address the endogeneity of obesity, he employs an IV regression, i.e. bivariate probit, model and the prevalence of obesity in the area in which the respondent lives as the instrument. Although he does not find any endogeneity of obesity for males, he finds that obesity has a statistically significant and negative effect on employment in both males and females, and failing to account for endogeneity in females leads to underestimation of the impact of obesity. Using the same data Morris (2006), using OLS, finds that BMI has a positive and significant effect on occupational attainment in males and a negative and significant effect in females. He is unable to identify any endogeneity of BMI.

Using a dataset from a Danish panel survey from 1995 and 2000, covering 8,000 individuals, Greeve (2008) analyzes the relationship between three body weight measures and employment status and wages. Body weight measures are BMI, a measure of total body fat (TBF) and fat-free mass (FFM), through use of predictions from a recent Swedish medical study. The study also addresses the endogeneity of weight by employing 2SLS and fixed effects models, using instruments of whether the interviewee's parents were alive in 2000 and whether the interviewee's father or mother had been prescribed medication for obesity. The study finds that the effect of BMI and TBF on employment takes on a significant inverted U-shaped relationship for men and a significant negative linear relationship for women. However, when a measure of FFM is included in the model, there is a positive but insignificant relationship between FFM and the probability of being employed for women.

To my knowledge, the only US study on employment/unemployment is done by Cawley (2000a). Addressing the endogeneity using a conventional IV regression approach, he analyses the impact of BMI on wages and employment using the NLSY data over the period 1981 to 1998. The instruments for BMI are the BMI of a biological child aged six to nine years and interactions of this with the child's age and gender. The analysis is restricted to females who have borne children. Although in the probit model the BMI coefficient is not statistically significant, BMI is found to have a positive effect on employment that is statistically significant at the 10% level in the IV probit model.

The only study that employs both body measures and smoking behavior in the same model is done by Jusot et al. (2008). They jointly examine the roles of health and health-related behaviors, such as obesity (measured as BMI of 30 or more), smoker (smoking 20 cigarettes or less per day) and heavy smoker (smoking more than 20 cigarettes per day), as precursors of employment using a longitudinal data set from France. Logistic regressions show that after adjusting for self-rated health, obesity is found to be significantly related to unemployment for women and heavy smoking is related to unemployment for men only.

2.2.2. Effects of Alcohol Use:

Previous literature on the labor market effects of alcohol consumption concentrates on US data and differs in the alcohol consumption measures used. The choice of the alcohol measures is usually determined by data availability. The commonly used measures are binary indicators of alcohol consumption at different levels, frequency of consumption over some periods and clinical measures of alcohol consumption. Most of the studies use only cross-sectional samples and a few of them control for endogeneity or unobserved heterogeneity.

The results of the wage effects of alcohol use are contradictory. Some of the results indicate that light to moderate drinkers have higher wages than abstainers and heavy or binge drinkers (Berger, 1988; Hamilton and Hamilton, 1997; McDonald and Shields, 2001). Earlier, Shaper (1988) found a consistent result with the medical literature that moderate alcohol drinking may improve health, but he asserted that the results decline or disappear when correcting for endogeneity and controlling for earlier health status of the respondents. The other studies that do not employ NLSY79 found similar results that excessive alcohol consumption is associated with lower wages or income. Nonetheless, there is a disagreement on the wage penalties for moderate to heavy drinking. While some of the studies do not find any penalties for heavy or binge drinking (Zarkin et al., 1998a), others find no evidence for benefits in terms of higher wages from moderate drinking over abstinence (Bryant et al., 1992). Employing NLSY data and addressing endogeneity, Kenkel and Ribar (1994) find that binge and heavy drinkers earn less than abstainers and much less than moderate drinkers, but Peters (2004) finds positive significant effects of drinking on wages, and Renna (2008) uses first difference regressions and fails to find any association.

Studies that use NLSY79 data indicate that binge and heavy drinkers earn less than abstainers and much less than moderate drinkers, but moderate drinkers earn more than abstainers and heavy drinkers (Kenkel and Ribar, 1994, Peters, 2004). Kenkel and Ribar use three waves of NLSY data (1985, 1988 and 1989) and in addition to wage level models, they include sibling and individual fixed effects, and simultaneous equations, keeping men and women separate. They conclude that binge and heavy drinkers earn less than abstainers and much less then moderate drinkers.

Peters follows up Kenkel and Ribar with analysis of wage levels and wage changes of workers, and I follow Peters' paper to define my drinking variables. He uses data from 1982 to 1994 NLSY waves. The analysis focuses on full-time workers and he asserts that this might be problematic since there is no control for selection into full time working. However, he assumes that the decision to drink will generally not affect labor supply and earnings for the respondents who face a current market wage that is less than their reservation wage for working full time. He excludes the respondents who were in the armed forces or self-employed and the sample consists of 4,796 male and 3,976 female full-time workers. He first considers how current drinking affects wages. Then, binge drinking (consuming six or more drinks at least four times in a month), permanent drinking (being a current drinker in all eight survey years) and permanent abstaining (being not a current drinker in each of the eight survey years) are controlled for. He then allows the drinks per week variable to enter into the wage equation. Finally, to control for unobserved heterogeneity bias he employs individual-specific fixed effects model. Peters finds a positive significant coefficient on drinking even when a rich set of covariates is controlled for when the wage levels method is used. He finds statistically insignificant estimates when he uses the wage changes method.

Contrary to the literature on alcohol use and earnings, there has been little research on the association between alcohol use and employment/unemployment and the results are less clear. Employing US data and accounting for endogeneity by IV models, Mullahy and Sindelar (1993, 1996) and Bryant et al. (1996) find insignificant effects of problem drinking on the probability of employment, but Terza (2002) finds negative significant effects for males. Kenkel and Ribar (1994) analyze NLSY data and account for endogeneity, and find a positive and statistically significant association for females. Employing non-US data and clinical measures of alcohol consumptions, some studies find lower employment probabilities for alcohol dependent men (MacDonald and Shields, 2004), but others find no significant effect for either gender when endogeneity is addressed via IV and fixed-effects models (Tekin, 2004; Feng et al., 2001).

In an early study, Benham and Benham (1982) found no significant relationship between problem drinking and employment but their results are generally constrained by weak data. Significant contributions to the literature are reported in studies by Mullahy and Sindelar (1991, 1993, 1996). In their first study, they use a household data set from the Epidemiologic Catchment Area (ECA) study and find that a lifetime diagnosis of an alcohol disorder is generally associated with greater unemployment and lower wages for men and women. In their other two studies, employing data from the 1988 Alcohol

Supplement of the National Health Interview Survey they argue that problem drinking is potentially endogenous using conventional IV estimation, but they do not deal with the non-linearity of the regression structure. They find that problem drinking decreases the likelihood of employment and increases the probability of unemployment, but the effects are not statistically significant. Using the same data, Terza (2002) applies a non-linear multinomial logit model accounting for the potential endogeneity of alcohol use. Consistent with the Mullahy and Sindelar, he finds a large negative and significant effect of problem drinking on employment for the male sample. Employing the NLSY data, Bryant et al. (1996) analyzes the impacts of substance use on a person's propensity to be employed. While the cross-sectional logistic regression results show a negative association between alcohol use and employment, the longitudinal studies are not clear.

A few European studies find similar results to Mullahy and Sindelar, and Terza. A Finnish study by Lahelma et al. (1995) examines the association of drinking and drinking problems with employment status. Although the frequency of drinking and frequency of intoxication are unassociated with employment status, frequency of health problems due to drinking is found to be significantly correlated with unemployment among only men. Using a data set from the Health Survey of England and accounting for the endogeneity of problem drinking using parental smoking, partner smoking and non-chronic health condition as instruments, MacDonald and Shields (2004) find that medically oriented measure of problem drinking results in a reduction in the probability of working. Johansson et al. (2007) use a Finnish cross-sectional population survey and investigate the success of alcohol-dependent individuals and abstainers on labor market. Employing clinical measures of alcohol use, they find that alcohol dependent men have lower employment probabilities when endogeneity is accounted for and male abstainers have lower employment probabilities than non-abstainers.

The results of a few studies are contrary to the results of the existing literature. Kenkel and Ribar (1994) analyzed data from the 1989 panel of the NLSY and estimated a model of hours of labor supplied accounting for unobserved heterogeneity and endogeneity via IV and fixed-effects models. They find that the effect of alcohol abuse on the labor supply of males is statistically insignificant, but there is a positive and statistically significant association for females. The contrasting results could possibly be due to including only individuals with positive hours of labor in the analyses.

Using a data set from the Russian Longitudinal Monitoring Survey (RLMS), Tekin (2004) estimates cross-sectional and fixed effects models of the effects of three different clinical measures of alcohol consumption on employment using alcohol prices at the community level as instruments. Whereas the cross-sectional results show that alcohol use has an inverse U-shaped effect on employment propensity, fixed effects estimates state that there is no significant effect of alcohol use on employment for either gender. Feng et al. (2001) examined the relationship between a medically oriented problem drinking measure and employment using a data set of prime age men and women from six southern states in the US. Drinking is found to be endogenous only for males. Univariate probit model estimates for females and bivariate probit model results for males show that problem drinking has a negative and insignificant impact on employment. However, the authors also note that their results should be read with some caution since the sample in the study is potentially non-representative and the problem drinking data were collected by phone interviews.

2.2.3. Effects of Smoking:

The labor market effects of smoking have gained less attention than obesity or drinking but the results are quite similar. Employing international data, Lee (1999) finds that there is a wage penalty for smokers in all subgroups. Using the US data, Leigh and Berger (1989) and Viscusi and Hersh (2001) find similar results. Employing 1987 National Medical Expenditure Survey, they find that smokers receive lower wages although they tend to work in riskier jobs that pay higher wages. Levine et al. (1997) use NLSY data and account for endogeneity by using sibling fixed-effects. Their results support the earlier findings that smoking negatively affects wages of both genders. Using NLSY data, Grafova and Stafford (2009) find that persistent smokers have lower wages than non-smokers, but smokers who quit smoking in later years do not have a wage penalty even when they smoke.

To the best of my knowledge, the first study to examine the effect of smoking on wages using NLSY was conducted by Levine et al. (1997), and I use this study to define my smoking variables. They investigate the effects of smoking on wages by using data from 1984 to 1992 NLSY waves. The analysis focuses on full-time, full-year workers who are defined as those who have worked at least 50 weeks or 1,750 hours in a year and who were not in the armed forces. They exclude the respondents who did not have an AFQT score and the respondents who have missing data on family background characteristics, like parents' education and parents' employment status. These restrictions lower the sample size to 3,473.

The authors first estimate a standard, cross sectional earnings function for all persons including dummy variables indicating whether the respondent currently smokes daily, the respondent is a continuous smoker (being a daily smoker in both survey years) or continuous non-smoker (being a non-smoker in both survey years), and the respondent is a quitter or starter along with a set of personal and family background measures. They then use the wage differences method only for siblings. Finally, they estimate changes in wages as a function of changes in smoking behavior over time for all persons. The authors find that conditional on their other observed characteristics, workers who smoke earn 8% less than nonsmokers in the wage levels model and 4% less than nonsmokers in changes in the wages model. They state that these wage penalties could be due to the decreases in productivity, increases in health problems, discrimination and economic myopia.

Using NLSY data, Grafova and Stafford (2009) find that persistent smokers have lower wages than non-smokers. Smokers who quit smoking in later years of data set do not have a wage penalty even when they smoke. This explains why fixed effects would likely produce small and insignificant results. The identification in the fixed effects model largely comes from individuals who quit smoking. However, these individuals never experienced a wage gap to begin with.

Although there is no study that examines the individual effect of smoking on employment, the studies on substance use and employment/unemployment probabilities often included smoking in their analyses. While Zarkin et al. (1998) fails to find any impact of smoking on hours of labor supply of males, accounting for endogeneity via IV using the religiosity indicator as an instrument, French et al. (2001) find a negative effect of smoking on employment. Further, employing NLSY, Kandel and Davies (1990) find negative effects of smoking on job mobility, employment gaps, and duration of unemployment.

Kandel and Davies (1990) used the 1984-85 period of NLSY to examine the effects of four categories of substance use, i.e. cigarette, alcohol, marijuana, and cocaine use, on employment. Controlling for multicollinearity but not for endogeneity, their results indicate that illicit drug use, including cigarette smoking, have negative effects on job mobility, employment gaps, and duration of unemployment.

In order to investigate the relationships between hours worked of young men and marijuana use, cocaine use, cigarette use, alcohol use, and the use of other drugs, such as heroin, hallucinogens, and inhalants, Zarkin et al. (1998) use 1991 data of the National Household Survey on Drug Abuse, NHSDA. They find little evidence to support a negative impact of drug use in the past month on hours of labor supplied by young men.

French et al. (2001) use the same self-reported data to examine the relationship between various measures of substance use, i.e. cigarette, alcohol and illicit drug use, and labor force participation. Using a composite religiosity indicator as an instrument for drug use, they reject the endogeneity and find that drug use has a negative effect on employment. Hence, no clear conclusion can be drawn on the basis of two studies using NHSDA data.

2.2.4. Combined Effects of Alcohol Use and Smoking:

Studies that use international data (Lee, 1999; Auld, 2005) find that moderate drinking is associated with a higher income than drinking abstinence, while smoking is associated with lower income than non-smoking after correcting for endogeneity. Yet the effects of heavy drinking are still unclear. Moreover, it is found that failure to control for other health behaviors leads to a moderate underestimate of the effect of smoking or drinking.

2.2.5. Combined Effects of Obesity and Smoking:

Berger and Leigh (1989) find no wage penalty for either behavior, but Baum and Ford (2006) find an obesity wage penalty only for female workers when both health behaviors are included in the same analysis and when accounting for endogeneity. Controlling for both health factors, Jusot et al. (2008) find that obesity is significantly related to unemployment for women and heavy smoking is related to unemployment for men.

2.2.6. Shortcomings of the Literature:

Most of the studies in the literature have some shortcomings and these shortcomings may be the reasons why they provide contradictory results. Firstly, to my knowledge, none of the studies have used all three behaviors in the same regression analysis. Secondly, some articles use small and unrepresentative samples. For instance, several studies use the very early years of longitudinal data in which the respondents were very young, have lower wages and less labor market experience, and not representative of the general population. Additionally, although some of the studies account for endogeneity of health behaviors by employing instrumental variables (IV) methods, most of their identification strategies are somewhat poor, because they rely on very strong assumptions and the instruments they use may be weak. Furthermore, with the exception of a few studies, almost all articles rely on cross-sectional data and do not use panel data techniques to control for unobserved heterogeneity. Lastly, the results differ in the measures of employment and health behaviors used, since the choice of the measures is usually dictated by data availability.

CHAPTER 3

Empirical Methodology

The empirical work in this study is based on a static neoclassical choice framework developed by Mullahy and Sindelar (1993, 1996) and subsequently adopted by many economists. They offer the most commonly used and general model of labor supply that depends on wages, desired hours of work and the decision to participate. An individual maximizes his/her utility by allocating times for labor and leisure consumption activities, which yields the labor supply of the individual. But the individual's labor supply decision is turned into employment only if the individual is selected for employment, which is partly based on reservation wage, market wage and the employer's decision to offer employment. Hence, in short, the employment status of an individual is a function of human capital of the individual, economic conditions, demographic factors, and other factors that affect the individual's choice between labor and leisure consumption.

In the light of earlier studies, in the simplest form labor supply decisions (L), decisions about how much alcohol to consume (A), how many cigarettes to consume (C), and how much food (or fast food, food with high calories) to consume (F) are all choice variables that depend on a vector of all prices including the price of alcohol, the price of cigarette, and the price of food, wages (W), a vector of all observable factors that affect household production (X) including non-labor income, and unobservable household characteristics ($u, a, c, and f$, respectively), i.e.,

$$L_{it} = L(P_{it}, W_{it}, X_{it}, u_{it}) \tag{1}$$

$$A_{it} = A(P_{it}, W_{it}, X_{it}, a_{it}) \tag{2}$$

$$C_{it} = C(P_{it}, W_{it}, X_{it}, c_{it}) \tag{3}$$

$$F_{it} = F(P_{it}, W_{it}, X_{it}, f_{it}) \tag{4}$$

where i refers to the i^{th} individual and t represents the time. The individual would choose a corner solution with respect to labor force participation if his/her reservation wage exceeds market wage (W_{it}). Hence, labor force participation decision is determined by the same factors that affect labor supply decision and it can be expressed as:

$$E_{it} = E(P_{it}, W_{it}, X_{it}, u_{it}) \tag{5}$$

where E_{it} is a binary indicator of employment.

In these formulations, binge drinking (BD), cigarette smoking (S), and obesity (O) might be seen as adverse health behaviors or outcomes arising from health reduction functions in which E, A, C, F, and u, a, c, f are arguments,

$$BD_{it} = BD(A_{it}, E_{it}, X_{it}, a_{it}) \tag{6}$$

$$S_{it} = S(C_{it}, E_{it}, X_{it}, c_{it}) \tag{7}$$

$$O_{it} = O(F_{it}, E_{it}, X_{it}, f_{it}) \tag{8}$$

In order to relate obesity, smoking, and binge drinking to employment, equation 5 can be rewritten as a reduced-form relationship that can be expressed as:

$$E_{it} = E(O_{it}, S_{it}, BD_{it}, X_{it}, u_{it}) \tag{9}$$

These types of reduced-form equations eliminate the need to specify the underlying structure of the health behaviors and they are also valid in the presence of corner solutions. Assuming linearity, the econometric equivalent of equation 9 can be expressed as:

$$E_{it} = \beta_1 O_{it} + \beta_2 S_{it} + \beta_3 BD_{it} + \beta_4 X_{it} + u_{it} \tag{10}$$

where β's are the parameters to be estimated. The employment equation can be estimated by a standard probit or logit model:

$$E_{it}^{*} = \beta_1 O_{it} + \beta_2 S_{it} + \beta_3 BD_{it} + \beta_4 X_{it} + e_{it} \tag{11}$$

where E_i^* is unobservable, and $E_i = 1$ if $E_i^* > 0$. e is assumed to be a mean-zero, constant-variance random variable and it is uncorrelated with the explanatory variables. The coefficients can be obtained via maximum likelihood estimations.

Using the equations 1-8, wage equation can be formulated as a function of the binary indicators of the three health behaviors (O, S, BD), the set of all explanatory variables (X), and a set of unobservable characteristics (ε),

$$W_{it} = W(O_{it}, S_{it}, BD_{it}, X_{it}, \varepsilon_{it}) \tag{12}$$

Hence, assuming the linearity, the effects of the three health behaviors on wages can be estimated from the following 'wage level' model:

$$ln(W_{it}) = \beta_1 O_{it} + \beta_2 S_{it} + \beta_3 BD_{it} + \beta_4 X_{it} + \varepsilon_{it} \tag{13}$$

ε is assumed to be a mean-zero, constant-variance random variable and it is assumed to be uncorrelated with the explanatory variables in *X*. The coefficients can be estimated via ordinary least squares (OLS) regression.

Providing unbiased correlations between obesity, smoking, binge drinking and unemployment/wages is not straightforward because the health behaviors and unemployment/wages may simultaneously affect each other. In addition, an unobserved factor may be correlated with both health behaviors and labor market outcomes, such as time preference, self-control, sociability, etc. Technically these mean that the error term in unemployment or wage equation is correlated with obesity, smoking, or binge drinking binary variables. Therefore, a supplemental set of covariates is included into the model to control more widely for individual background characteristics that may affect labor market outcomes.

In a cross-sectional framework, previous studies have generally dealt with the problem of endogeneity with conventional two-stage IV models. In such an approach one needs to instrument obesity, smoking, and binge drinking with an instrument (or instruments) that is (are) correlated with health behaviors, but uncorrelated with labor market outcome.[3] However, the achievement of unbiased estimates via the IV method

[3] The same approach has also been employed in this study.

To model these endogenously determined outcomes, a traditional instrumental variable method is employed. Therefore, the system of the equations becomes:

$$\begin{aligned} ln(W_{it}) &= \beta_1 X_{it} + \gamma_1 O_{it} + \gamma_2 S_{it} + \gamma_3 BD_{it} + \varepsilon_{1it} \\ O_{it}^* &= \beta_2 X_{it} + \varphi_1 I_{it} + \varepsilon_{2it} \\ S_{it}^* &= \beta_3 X_{it} + \varphi_2 I_{it} + \varepsilon_{3it} \\ BD_{it}^* &= \beta_4 X_{it} + \varphi_3 I_{it} + \varepsilon_{4it} \end{aligned} \tag{14}$$

where X is the same vector of variables that affect wages, O^* is the propensity of being obese, S^* is the smoking propensity, BD^* is the binge drinking propensity, I is a vector of instruments that affect obesity, smoking, and binge drinking, but do not directly affect wages. The observed obesity, smoking, or binge depends essentially on the predictive power and validity of the instruments. If there is weak correlation between instruments and health behaviors, or if the instruments are correlated with labor market outcome, then the IV estimates could still be biased. Moreover, mean square errors with an IV method may be large which implies a trade-off between bias and variance (Bollen et al., 1995; Norton et al., 1998). Furthermore, IV estimation is less efficient, because it does not fully exploit the specification and non-linearity of the unemployment model (Feng, 2001).[4]

---

drinking statuses are indicators that $O_{it} = 1$ if $O^{*}>0$ for obesity, $S_{it} = 1$ if $S^{*}>0$ for smoking, and $BD_{it} = 1$ if $BD^{*}>0$ for binge drinking. In unemployment analyses, equation 10 is used instead of log linear wage equation.

Earlier studies have widely used state cigarette tax, state alcoholic drinks tax (Kenkel and Ribar, 1994), and the BMI of siblings, parents, children (Cawley, 2004), prevalence of obesity, smoking, and drinking across individuals living in the same health authority area (Morris, 2007) as instruments for the individual's smoking behavior, drinking behavior, and obesity. However, it has been argued that most of these instruments are weak since they may be correlated with wages/employment prospects as well, and adopting the health behavior of a sibling, child or parent would decrease the sample size largely. Moreover, the state taxes could not be used in this study, because information about the individuals' geographic location is not part of the public release version of NLSY79 data and these measures can only be obtained via Geocode data of NLSY79 data, which I could not acquire yet because of the strict policies of the BLS. A combination of other frequently used instruments are employed in the instrumental variables (IV) models: several religion and religious attendance dummies, Rotter test score, youngest age started drinking dummies, alcoholic father in childhood (only satisfied for males), alcoholic mother in childhood (only satisfied for females), and cumulative years lived with alcoholic relative variables (Berger and Leigh, 1988; Bryant et al., 1996; Heien, 1996; Pagan and Davila, 1997; Hamilton and Hamilton, 1997).

Two-stage instrumental variable (IV) technique is employed by using Stata command 'ivprobit'. In the first stage, each endogenous variable is regressed on all exogenous variables and a set of instruments, and the reduced form residuals are obtained by probit model. In the second stage, the reduced form residuals are included to the wage equation or unemployment equation. It is assumed that the residuals from the first stage regressions and from the wage or unemployment regression are independent and identically distributed. Wald tests are employed to test for exogeneity. For the validity of the instruments, first-stage equation F-tests and adjusted R^2s are used. Furthermore, IV for some panel data models (fixed effects and random effects models) in which obesity, smoking, and binge drinking are assumed to be endogenous are also employed.

[4] There is a large literature claiming that failing to control for selection into working could lead to sample selection bias in analyses of wage effects. To account for this bias, a two-step Heckman correction method is employed for males and females separately. In the first stage, a probit model is employed to estimate to specify an individual's decision to work or receive positive wage:

$$Work_{it}^{*} = \gamma \mathrm{Z_{it}} + \mathrm{z_{it}} \tag{15}$$

where $Work^{*}$ denotes receiving positive wage and $Work_{it}{}^{*} = 1$ if $W_{it} > 0$. It is assumed that the error terms in equation (13) and (15) are normally distributed and are correlated where $\rho_{\varepsilon z}$ indicates the

A longitudinal data set offers alternative solutions to the endogeneity problem: fixed-effects, random-effects, between-effects models.[5] The most commonly used model

correlation coefficient. Non-wage family income is served as an instrument for the propensity of the individual to work or receive positive wage. From the model, an inverse Mills Ratio correction term is calculated:

$$\mathrm{IMR}^{\mathrm{work}} = \frac{\phi(\pi)}{\Phi(\pi)} \tag{16}$$

where $\phi(\pi)$ terms are the density functions and $\Phi(\pi)$ are the cumulative density functions for working. In the second stage, the log wage equation (13) is estimated including the Mills Ratio correction terms. The Mills Ratio correction terms which are employed to control for selection into working are not statistically significant in the wage equation either for males or females. Furthermore, the correction does not change the sizes and significance levels of the three health behaviors. Moreover, instrumenting the health behaviors do not lead to any significant change in either the Mills Ratio term or in the estimated parameters of the health behaviors. Hence, the results with the Heckman correction are ignored and not presented.

[5] Some panel data techniques are also employed in this study. See appendix for the results.

The first panel data technique to control for unobserved heterogeneity is fixed-effects model. Fixed-effects model is the main technique used for analysis of panel data. It lets us control for omitted variables that differ between cases but are constant over time. It does this by estimating a mean-differenced model for individuals across time periods. Let the error term be

$$\varepsilon_{it} = u_{it} + e_{it} \tag{17}$$

where e is a random error uncorrelated with explanatory variables in X and μ is unobserved individual characteristics. In this case, OLS estimates will not correctly characterize the effects of obesity, smoking, or drinking but some portion of the associations of health behavior with wages or unemployment will be due to this unobserved individual characteristic. In fixed effects, the wage equation becomes:

$$\left(ln(W_{it}) - ln(\overline{W}_i)\right) = \beta(X_{it} - \bar{X}_i) + (\mu_i - \bar{\mu}_i) + (e_{it} - \bar{e}_i) \tag{18}$$

where bars indicate individual specific means. It is assumed that the time-invariant unobserved individual-specific characteristics are constant among individuals, therefore $\mu_i = \bar{\mu}_i$ and these characteristics cancel out. Although the fixed-effects models produce consistent estimates, if μ contains unobserved individual characteristics that are correlated with X that vary over time, the estimates are not consistent. The other shortcomings of fixed-effects model are that the model takes out all covariates that are time-invariant, such as race, gender, marital status, etc., and the model ignores differences between individuals.

The second panel data technique is between-effects model. This model is used to control for omitted variables that change over time but are constant within cases; it provides estimates from changes across individuals. In the model, each covariate is assumed to be an individual's mean value across years, therefore individual-specific average wages are regressed on each covariate's individual-specific mean and the wage equation becomes:

$$ln\left(\overline{W_i}\right) = \beta\overline{X}_i + \overline{\mu}_i + \overline{e}_i \tag{19}$$

The shortcomings of the between-effects model are that the estimates of this model could also be biased if covariates are correlated with the error term, and this model ignores the within-person changes.

The third model is random-effects model. This model is used if we believe that some omitted variables may be constant over time but change between cases, and some other omitted variables may be constant between cases but vary over time. Therefore, this model is a combination of fixed and between-effects models. Stata is used to estimate the random-effects model and Stata's random effects model estimator is a weighted average of fixed and between-effects models. Although random-effects model estimates are efficient, its estimates could be biased if unobserved individual characteristics are correlated with the covariates. Hausman tests are used to compare fixed-effects and random-effects models under the that deals with unobserved individual heterogeneity is fixed-effects, which lets one control for omitted or unobserved variables that differ between cases but are constant over time by estimating the model as deviations from the means. However, the fixed-effects model does not eliminate bias in case of time-variant individual heterogeneity and takes out all the covariates that are time-invariant. Further, the model may worsen the bias caused by measurement error (Griliches and Hausman, 1986). Also, the model cannot identify the effects of individuals who do not change their behavior over time, and it assumes individuals who change their behavior over time are similar to the ones who do not change their behavior over time. However, (as found in this study) this is not always the case for individuals with health behaviors, e.g. permanent smokers are different than starters, quitters, or young experimenters.

To account for endogeneity, the Hausman-Taylor Instrumental Variable (HTIV) method is employed in wage analyses. This method is the most appropriate for the purpose and sample of this study. The Hausman-Taylor estimator has recently gained in popularity among many economists (Egger and Pfaffermayr, 2004) and econometric textbooks usually recommend the Hausman-Taylor method especially for panel data with time-invariant variables and correlated unit effects (Hausman and Taylor, 1981; Wooldridge, 2002). In the presence of correlation between some covariates and the unobserved individual characteristics, this method produces consistent estimates contrary to the OLS model. Furthermore, in contrast to the conventional IV methods, there is no need for finding external instruments. In addition, in contrast to the fixed effects model,

---

null hypothesis that the coefficients estimated by the two models are the same. If the test rejects, then the fixed effects must be used since random effects are biased.

the Hausman-Taylor IV model allows producing the estimates for time-invariant covariates. Let the wage model be:

$$ln(W_{it}) = \beta_1 X_{1it} + \beta_2 X_{2it} + \gamma_1 Z_{1i} + \gamma_2 Z_{2i} + \mu_i + \varepsilon_{it}, \tag{20}$$

where ε is a random error uncorrelated with explanatory variables, μ is unobserved individual characteristics, X_{1it} (X_{2it}) are variables that are time-varying and uncorrelated (correlated) with μ_i, and Z_{1i} (Z_{2i}) are variables that are time-invariant and uncorrelated (correlated) with μ_i. This model predicts potentially endogenous variables with a set of instruments obtained from within the model. Hausman and Taylor propose an instrumental variable approach using the following variables as instruments in the final GLS estimator: X_{1it}, Z_{1i}, $X_{2it} - \bar{X}_{2i}$, $\bar{X}_{1i}$. Particularly, this method estimates a random-effects model and uses exogenous time-varying variables as instruments for the endogenous time-varying variables, and exogenous time-invariant variables together with the unit means of the exogenous time-varying variables as instruments for the endogenous time-invariant variables (Plumper and Troeger, 2006). Obesity, smoking, and drinking variables are considered as endogenous variables and all others as exogenous variables.[6]

In order to account for the endogeneity of health behaviors in the unemployment analyses, the study estimates a multivariate probit model, which appears to be more appropriate and more efficient than conventional two-stage IV methods:

---

[6] I examined the endogeneity of several other variables, such as experience and marital status, but the results are unchanged.

$$\begin{gathered}U_{it}\text{=1 if } U_{it}^{*}\geq 0 \text{ and } U_{it}\text{=0 if } U_{it}^{*}\leq 0\\ O_{it}\text{=1 if } O_{it}^{*}\geq 0 \text{ and } O_{it}\text{=0 if } O_{it}^{*}\leq 0\\ S_{it}\text{=1 if } S_{it}^{*}\geq 0 \text{ and } S_{it}\text{=0 if } S_{it}^{*}\leq 0\\ BD_{it}\text{=1 if } BD_{it}^{*}\geq 0 \text{ and } BD_{it}\text{=0 if } BD_{it}^{*}\leq 0\end{gathered} \qquad (21)$$

where $U_{it} = 1$ means the individual is unemployed and $U_{it} = 0$ means the person is employed. Similarly, $O_{it} = 1$, $S_{it} = 1$, and $BD_{it} = 1$ indicate the individual is obese, smoker, and binge drinker, respectively. U_{it}^{*}, O_{it}^{*}, S_{it}^{*}, and BD_{it}^{*} are unobserved latent variables that determine being unemployed, obese, smoker, and binge drinker, respectively. The empirical specification for the multivariate model can be written as:

$$\begin{bmatrix} U_{it}^{*} \\ O_{it}^{*} \\ S_{it}^{*} \\ BD_{it}^{*} \end{bmatrix} = \begin{bmatrix} \beta_1 O_{it} + \beta_2 S_{it} + \beta_3 BD_{it} + \beta_4 X_{it} \\ \alpha_1 X_{it} + \alpha_2 Z_{it} \\ \gamma_1 X_{it} + \gamma_2 Z_{it} \\ \delta_1 X_{it} + \delta_2 Z_{it} \end{bmatrix} + \begin{bmatrix} u_{it} \\ o_{it} \\ s_{it} \\ bd_{it} \end{bmatrix}$$

$$\begin{gathered} E[u] = E[o] = E[s] = E[bd] = 0 \\ Var[u] = Var[o] = Var[s] = Var[bd] = 1 \\ Cov[u,o] = \rho_{12},\ Cov[u,s] = \rho_{13},\ Cov[u,bd] = \rho_{14} \end{gathered} \qquad (22)$$

where Z is a vector of instruments, β's, δ's, γ's, and α's are parameters to be estimated, β_1, β_2, β_3 are the coefficients of interest, u, o, s, and bd are error terms distributed multivariate normally with mean equal to 0 for each variable and a variance-covariance matrix that has the value of 1 on the principal diagonal and the correlation terms $\rho_{mn} = \rho_{nm}$ as the off-diagonal terms. ρ_{12}, ρ_{13} and ρ_{14} are the correlations between the error terms in employment and obesity, smoking, and binge drinking equation, respectively.

A combination of frequently used instruments is used in the multivariate probit model: several religion and religious attendance dummies, Rotter test score, youngest age started drinking dummies, alcoholic father in childhood (only satisfied for males), alcoholic mother in childhood (only satisfied for females) and cumulative years lived with alcoholic relative variables (Berger and Leigh, 1988; Bryant et al., 1996; Heien, 1996; Pagan and Davila, 1997; Hamilton and Hamilton, 1997).

This multinomial probit model is estimated according to the method of maximum likelihood using 'mvprobit' in Stata. Wald tests of significance of ρ's are used to test the endogeneity of unemployment and health behaviors (Wooldridge, 2002). If ρ's are significantly different from zero, then they are endogenous and probit/logit estimates are biased, hence the multinomial probit model should be used. If ρ's are zero, then the probit model of equation 11 would generate consistent estimates and there is no need for a multivariate model.

I also propose multinomial logit models with the following outcomes: 1) employed, unemployed, out of labor force, 2) part-time-employed, full-time-employed, unemployed, 3) self-employed, employed-not-self, unemployed.[7] These specifications are logical from a timing point of view in that the outcomes are observable conditioned on participating in the labor force which is broader than an employed-unemployed specification. Moreover, these types of specifications give one more opportunity to comprehensively examine the existence of discrimination, since not only the unemployed are more likely to be obese, smoker or binge drinker, but part-time employed, self-employed or out of labor force individuals may tend to be more obese, smoker, binge drinker than their counterparts (Grace, 2006). Therefore, the employment status, e.g. first

[7] The multinomial logit relies on the assumption of independence of irrelevant alternatives. Although this might be a limitation in this study, being the choices likely to be correlated based on unobservables, the results are reported for completeness.

case, can be represented by a vector $E = E[E1, E2, E3]$ which has the following binary elements:

$$E_1 = \begin{cases} 1 & \text{if employed} \\ 0 & \text{otherwise} \end{cases} \quad E_2 = \begin{cases} 1 & \text{if unemployed} \\ 0 & \text{otherwise} \end{cases} \quad E_3 = \begin{cases} 1 & \text{if out of labor force} \\ 0 & \text{otherwise} \end{cases}$$

Then, if the first category is the reference, for m = 2, 3:

$$Pr(E_{it} = m) = \frac{exp(\beta_1 O_{it} + \beta_2 S_{it} + \beta_3 BD_{it} + \beta_4 X_{it})}{1 + \sum_{m=2}^{3} exp(\beta_1 O_{it} + \beta_2 S_{it} + \beta_3 BD_{it} + \beta_4 X_{it})} \quad (23)$$

and for the first category:

$$Pr(E_{it} = 1) = \frac{1}{1 + \sum_{m=2}^{3} exp(\beta_1 O_{it} + \beta_2 S_{it} + \beta_3 BD_{it} + \beta_4 X_{it})} \quad (24)$$

The consistent parameter estimates, βs, can be obtained via maximum likelihood estimations.

Lastly, to examine whether the effects of these health behaviors are additive or interactive, interaction terms are added into wage and unemployment equations. These interaction terms would be statistically significant if dual and/or triple interactions exist.

In all estimations described in this study, NLSY sampling weights are used. The standard error for each reported coefficient is robust to heteroskedasticity and calculated with clustering by individual to account for correlations in the error terms of each individual over time via Stata command 'cluster'.[8]

[8] Formatting the data is one of the main steps in this study. NLSY79 data has longitudinal properties. It has more than 30,000 survey questions in each year for 21 years. Firstly only the questions and observations for use in the study are selected and the format of the data is changed to person-year observations. Then several covariates are created from these observations including the variables measuring the three health factors and some imputations are made for the missing values of the health factors which are explained later in the data section.

CHAPTER 4

Data and Descriptive Statistics

4.1.Data and Sample:

I use the National Longitudinal Survey of Youth (NLSY) data to study the effects of obesity, smoking and binge drinking on labor market outcomes. The NLSY is a comprehensive survey sponsored and directed by the Bureau of Labor Statistics of the U.S. Department of Labor. The main focus of the survey is labor force behavior. The NLSY began in 1979 with 12,686 respondents (6,403 are male) between the ages of 14 and 21. These individuals were chosen from both civilian and military populations, and were interviewed annually from 1979 until 1994. After the 1994 survey, the NLSY began interviewing biennially. In 2004, these individuals ranged in age from 39 to 46. In each survey, the NLSY gathers information on each respondent's employment status, salaries, age, personal characteristics, family characteristics, etc.

A combination of commonly used data specifications is employed to construct the sample from the data. All available waves of the survey are used from 1979 to 2004 for wage analyses and all waves from 1979 to 1998 are used for unemployment analyses since unemployment statuses of the individuals cannot be determined after 1998. Most of the existing studies exclude respondents in years during which their education has yet to be completed and person-year observations in which the respondent is less than 18 years old. The same exclusions are employed for my sample. I also exclude respondents who were in the armed forces and who did not have any employment or wage information.

Employing Cawley's (2004) specification, the females who were pregnant in the survey year are also excluded from the sample since weight (and obesity) may be affected by pregnancy. Additionally, for the wage analyses, the respondents working part-time or who received any wage in the survey year are included in my sample. For analyses of unemployment only, I exclude respondents who are out of the labor force, who worked as self-employed, without pay or in a family business as most other researchers did.

The sample for the wage analysis consists of 12,106 individuals (6,147 male, 5,959 female) after dropping 580 respondents, and the sample for unemployment analysis consists of 12,009 individuals (6,076 male, 5,933 female) after dropping 677 respondents. Person-year observations are used in the regression analyses and each person-year is included as a separate observation.[9] There are 125,264 person-year observations used in wage regressions, 65,364 of whom are male and 59,899 of whom are female (141,142 person-year observations are dropped). After dropping 109,315 person-year observations, the sample for unemployment regressions contains 119,033 person-year observations, 64,712 of whom are male and 54,321 of whom are female.

4.2.Descriptive Statistics:

4.2.1. Dependent Variables:

Hourly wages are calculated by dividing 'total income from wages and salary' to 'hours worked', and all wages are adjusted using the Consumer Price Index from the US

[9] Respondents provide multiple observations since person-year observations are employed and I use 'cluster' command in Stata to control for correlation among observations that come from the same individual. This command defines an error structure where only errors between observations from different people are independent.

Census Bureau with a 2003 base period. The average real wage, in 2003 dollars, for all person-year observations is $14.79 per hour. Real wages are increasing over time for both men and women, except the first five years when the respondents were still young or not employed yet as shown in Figure 2.1. Results indicate that there is a considerable wage differential between men and women in almost all years; for example the real wage for men in 2004 is $25.90 and it is $19.49 for women.

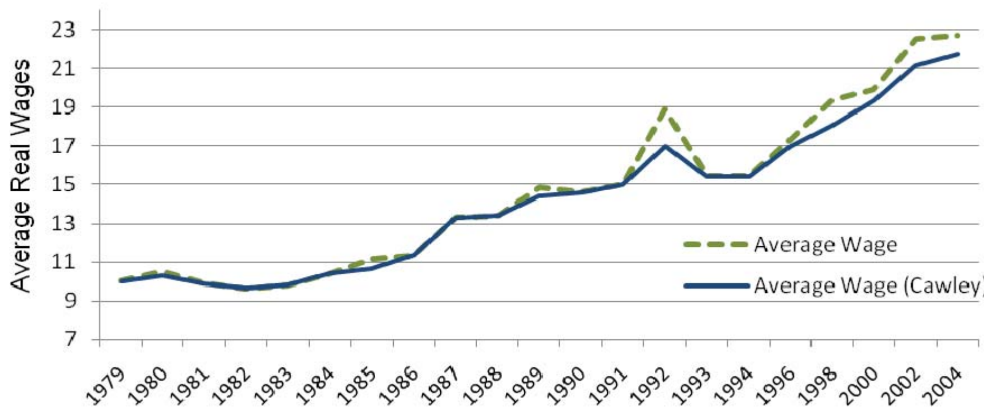

Figure 4.0.1. *Average real wages by survey years*

Because of the different question specifications in 1992, respondents' total net income from wages and salaries (and therefore average real wages) are not truncated, which explains the peak in this year. This problem is solved in Cawley's (2004) specifications. To get rid of the outliers, as Cawley (2004) did in his article, I recoded all the hourly wage information less than $1 as $1 and greater than $500 as $500. As a result, Cawley's specifications produce smoother wage distributions, and peaks in average wages, such as in 1992, disappeared.

As shown in previous research, there is a nonlinear relationship between wages and possible explanatory variables. For instance, wages tend to increase at an increasing (or decreasing) rate with years of schooling, age, years of tenure, or the effect of years of schooling on wages depends on gender. The standard approach to model this relationship is transformations, or more specifically using natural logarithms. Natural logarithm transformations are used to satisfy symmetry, homoscedasticity and linearity conditions. Because many statistical techniques work best when the data is single-peaked, when the variability is approximately the same within each group you are comparing and it is easier to describe the relationship between variables when it's approximately linear (Dallal, 2007).

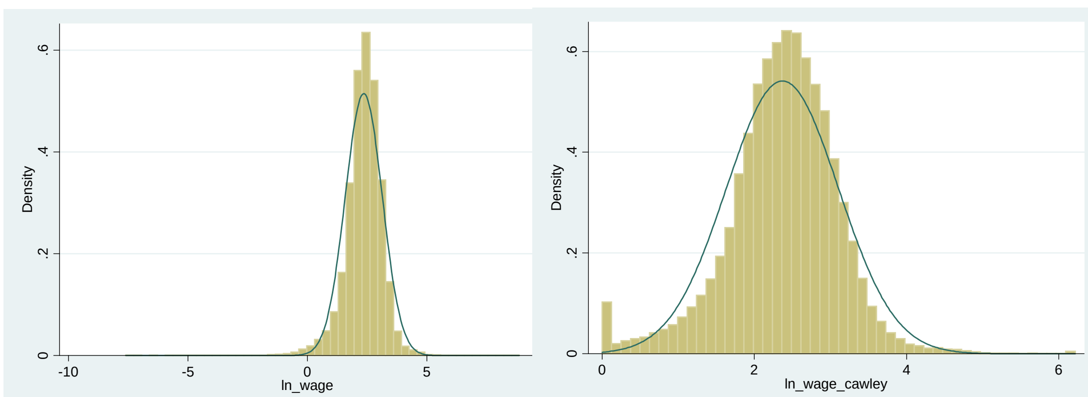

Figure 4.0.2. Histograms and normal density plots for natural logarithms of two different wage variables

Figure 4.2 shows the histograms of natural logarithms of three wage variables and normal density plots. Skewness is a measure of symmetry; it shows the lack of symmetry. The skewness for a normal distribution is zero. Kurtosis a measure of whether the data are peaked or flat relative to a normal distribution and the kurtosis for a standard normal distribution is three. Skewness values for the natural logarithm of three hourly real wage variables are -.83 and -.32 respectively for hourly wages and hourly wages with Cawley's specification. Kurtosis values are 9.37 and 4.17 respectively. Moreover, the Jarque-Bera (JB) test is a goodness of fit test that measures the departure from normality, taking skewness and kurtosis degrees into consideration. The JB statistic is lower for Cawley's specification. The histograms, these measures and the JB tests confirm that Cawley's wage specification is better and I will use this wage variable in my regression analyses.

The NLSY provides detailed information about the labor force status of respondents. The question: 'What were you doing most of last week?' is asked in all years of NLSY until 1998, thus only the waves between 1979 and 1998 are used in this study. The following seven categories of responses have been coded from each year's survey: (a) working, (b) with a job-not at work, (c) looking for work, (d) keeping house, (e) going to school, (f) unable to work, and (g) other. NLSY creates various employment statuses using the answers given to the question:

Employed: (1) All civilians who did any work as paid employees, or in their own business or farm, or who worked 15 hours or more as unpaid workers in an enterprise operated by a member of the family; and (2) persons who were temporarily absent from work because of various personal reasons, whether they were paid for the time off or were seeking other jobs. Persons whose only activity consisted of work around the house or volunteer work are excluded.

Unemployed: All civilians who had no employment during the survey week, were available for work, and (1) had made specific efforts to find employment some time during the prior four weeks, (2) were waiting to be recalled to a job from which they were laid off, or (3) were waiting to report to a new wage and salary job scheduled to start within 30 days.

Out of the Labor Force: All persons who are not classified as employed or unemployed or in the Armed Forces, who are engaged in own home housework, in school, unable to work because of long-term physical or mental illness, retired, seasonal workers in off season, who did not report looking for work, unpaid family workers and other.

Moreover, four other employment statuses are created for multinomial analyses purposes:

Full-time Employed: All employed persons working more than 35 hours or more per week.

Part-time Employed: All employed persons working less than 35 hours per week.

Self-Employed: All persons working for profit or fees in their own business, shop, office, or farm.

Employed-not-Self: All employed persons except self-employed.

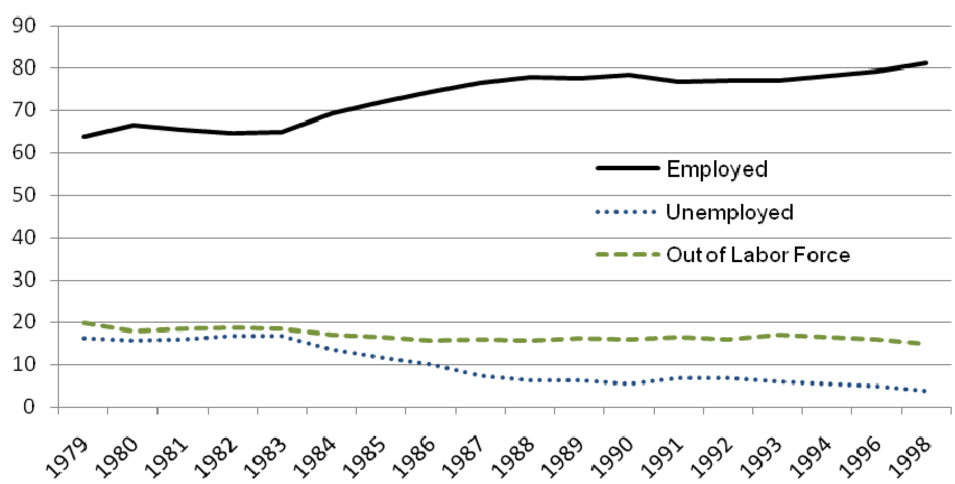

Figure 4.0.3. Employed/unemployed/out of labor force rates by survey years

Figure 4.3 displays the percentage of employed, unemployed and out of labor force individuals by survey year. The graph shows that as the respondents age, they are more likely to be employed and unemployment rate falls. Out of labor force percentages seem unlikely to change as individuals get older.

4.2.2. Obesity Variables:

Questions about weight were asked in every year except the 1979, 1980, 1983, 1984, 1988, and 1991 surveys and questions about height were asked only in the 1981, 1982, and 1985 survey years. I assumed that the heights of the respondents after the 1985 survey year are the same as the 1985 height observations since the youngest respondent was 21 years old in 1985. A linear increase (or decrease) of weight is assumed in the years, in which the respondent did not answer any weight question, and a linear increase of height is assumed in the 1983 and 1984 years since there is no height question asked in these years. Height information for 1981 is imputed for the first two years of the survey, 1979 and 1980. In 1981, respondents (between the ages of 16 and 23) weighed an average of 147.68 pounds. They weighed 185.44 pounds in 2004, when they were between the ages of 39 and 46.

Each respondent’s body mass index (BMI) is used to measure obesity. BMI is defined as weight in kilograms divided by height in meters squared. A BMI less than 18.5 is considered underweight, a BMI of 18.5-24.9 is normal and a BMI greater than 25 is overweight according to the Centers for Disease Control and Prevention (CDC). Obesity is defined as a BMI of 30 or more. The CDC’s current method for identifying obesity in those under age 21 is age- and gender-specific (CDC, 2006b) and only a few recent papers used this methodology. I use CDC's BMI growth charts to create obesity variables for the individuals under 21 years old.[10]

Health economics and medical studies demonstrate that there is a tendency for people to under-report their true weight but over-report their height (Cawley, 2004). Employing the third NHANES of 1988-1994, Cawley finds that on average men between the ages of 17 and 40 are inclined to over-report their true BMIs by 0.02% and woman under-report their BMIs by around 1.5%. Therefore, to predict the true BMIs of the respondents from their reported BMIs and to correct the measurement error, Cawley's information and specifications are used in this study.

The average BMI over all survey years is 25.34 which is accepted as an overweight measure. In 2004, 1,634 respondents were obese and 2,149 were overweight out of 5,013 respondents. Table 4.1 shows that 19,125 person-year observations out of 125,264 have obese characteristics, 10% of sample in all survey years are mildly obese (defined as BMI between 30 and 35) and 5% are morbid obese (defined as BMI 35 or more).

Table 4.1. *Person-year observations*

Underweight	7,263 (6%)
Normal	57,254 (46%)
Overweight	41,622 (33%)
Obese	19,125 (15%)
Mild Obese	13,152 (10%)
Morbid Obese	6,013 (5%)
	125,264

[10] This method leads 150 person-year observations for men under 21 years old to change from overweight to obese, but leads around 100 person-year observations for women under 21 years old to change from obese to overweight categories.

Figure 4.4 shows the average yearly BMI and obesity prevalence by year. In 1979, the average BMI was about 23.31, then average BMI increases monotonically and after 1989 the average BMI reaches 25 which means that respondents have an average BMI that is considered above the range currently considered best for health. The average wage for the full sample with a BMI of 30 or more is $14.65 and average wage with BMIs between 25 and 30 is $15.24.

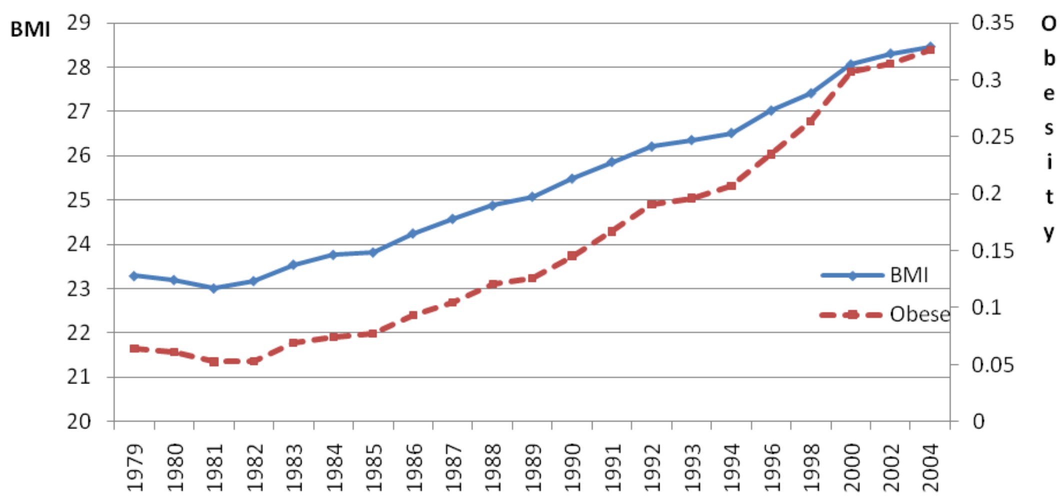

Figure 4.0.4. Body Mass Index (BMI) by survey years

4.2.3. Alcohol Consumption Variables:

Respondents were asked questions related to their alcohol consumption in the 1982, 1983, 1984, 1985, 1988, 1989, 1992, and 1994 survey years. However, the questions were not identical across the years. I employ Peters' (2004) and Keng and Huffman's (2007) methods to create alcohol consumption variables.

The 'current drinker' dummy variable is equal to one if the respondent answers positively to the question: "*did you drink any alcoholic beverages in last month*" or if the respondent gives an answer of at least one to the question in 1992: "*number of days drank alcohol in last month*". For the missing years, the earlier and later years' information about the drinking is imputed. For example, the 1986 and 1987 years do not have any questions about the drinking behavior of respondents. It is assumed that the 1986 drinking observations are the same as 1985, and the 1987 observations are the same as 1988. I also assume that the respondent is a current drinker in 1979, 1980 and 1981 if s/he is a current drinker in 1982, and that s/he is a current drinker in 2004 if s/he is a current drinker in 2002.

Binge drinking is recently defined by CDC as drinking 5 or more drinks during a single occasion for men and 4 or more drinks during a single occasion for women. However, there is too little information in NLSY surveys to create binge drinking variables in compliance with this definition. Instead, I use Peters' and Keng and Huffman's specifications to create dummy variables indicating excessive alcohol use of the individuals. It is supposed that if the respondent gives an answer of 4 or more to the question "*number of days had 6 or more drinks in last month*", then the 'binge drinker' variable is equal to one. If the respondent drinks 6 or more drinks on one to three occasions, the respondent is defined as a 'heavy drinker'.

Table 4.2 illustrates that 65% of person-year observations are current drinkers and around 12% are binge drinkers. Figure 4.5 shows that the prevalence of current drinkers and binge drinkers are decreasing over time.

Table 4.2. *Person-year observations*

Drinker	81,979 (65%)
Heavy Drinker	23,556 (19%)
Binge Drinker	14,820 (12%)
Permanent Abstainer	9,982 (8%)
Permanent Drinker	33,677 (27%)
	125,264

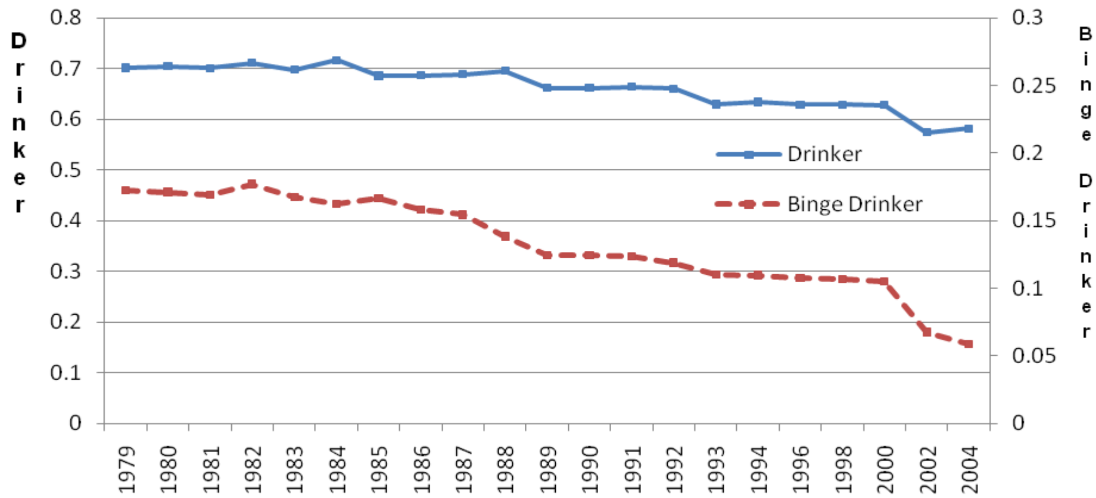

Figure 4.0.5. Percentage of drinkers by survey years

4.2.4. Smoking Variables:

Smoking questions were asked of respondents in the 1984, 1992, 1994 and 1998 survey years. The surveys were different across the years in the types of the questions asked. Therefore, I employ the similar methodologies that Levine et al. (1997) and Baum and Ford (2006) used in their papers.

'Daily smoker', 'heavy smoker' and 'light smoker' variables are created for all survey years. The 'daily smoker' variable is equal to one if the respondent answered positively to the question asked in 1992, 1994 and 1998: "*does respondent currently smoke daily*". In 1984, respondents were asked the average number of cigarettes smoked per day, and I assume that the variable is also equal to one if the respondent averaged one or more cigarettes per day. If the answer to the question "*number of cigarettes smoked per day*" asked in all years is 20 or more, then the respondent is assumed to be a 'heavy smoker'. If the answer is between one and 20, then the respondent is a 'light smoker'.

Respondents are also asked when they quit smoking. If they smoke in the year the questions were asked, the 'daily smoker' variables for the earlier (missing) years are assumed to be one. If they respond that they never quit and they do not smoke in the year the questions are asked, the individuals are presumed to be nonsmokers in previous years. Respondents who reported quitting within the previous year are recorded as smokers in the previous year and respondents are assigned as daily smokers from 1979 to the year they quit.

Not all the respondents answer the quitting smoking question. For the missing years, earlier and later year information, the average assumption, is imputed. For example, the smoking information for the 1985, 1986 and 1987 years is presumed to be the same in 1984, and smoking information for 1988-1991 is the same as 1992's smoking information.[11]

Levine et al. state that smokers often attempt to quit unsuccessfully. For example, using the 1991 NHIS survey, Levine estimates that over 40% of current smokers attempted to quit smoking in the previous year. Moreover, several retrospective questions

---

[11] If some of the respondents smoke heavily on some days and do not smoke on the other days, my definition of daily smoking may not match that implied by the 1992, 1994 and 1998 survey years and it may lead to some false transitions in smoking status between the survey years. However, using data from the 1991 National Health Interview Survey, Levine estimates that frequency of heavy smoking on some days and not smoking on other days is only 4% for individuals between the ages of 26 and 33, the same age as NLSY respondents in that year.

are used to create smoking questions in this study. However, Kenkel et al. (2003) state that contemporaneous and retrospective smoking status variables do not match very precisely in the NLSY79 data. Therefore, to observe the effects of smoking in longer periods and to minimize the measurement error of imputations, some additional smoking variables are created by considering the smoking behavior of the respondents for the years in which smoking questions are asked; smoker (in all four years), non-smoker (in all four years), quitter, starter, young-experimenter (only smoked during the year 1984) and unsuccessful quitter (quit before but started again).

Table 4.3. *Person-year observations*

Daily smoker	39,699 (32%)
Light smoker	18,756 (15%)
Heavy smoker	20,666 (17%)
Ever smoked daily	58,823 (47%)
Smoker	15,709 (13%)
Non-smoker	53,848 (43%)
Quitter	4,890 (4%)
Starter	4,301 (3%)
Young Experimenter	7,260 (6%)
Unsuccessful Quitter	10,154 (8%)
	125,264

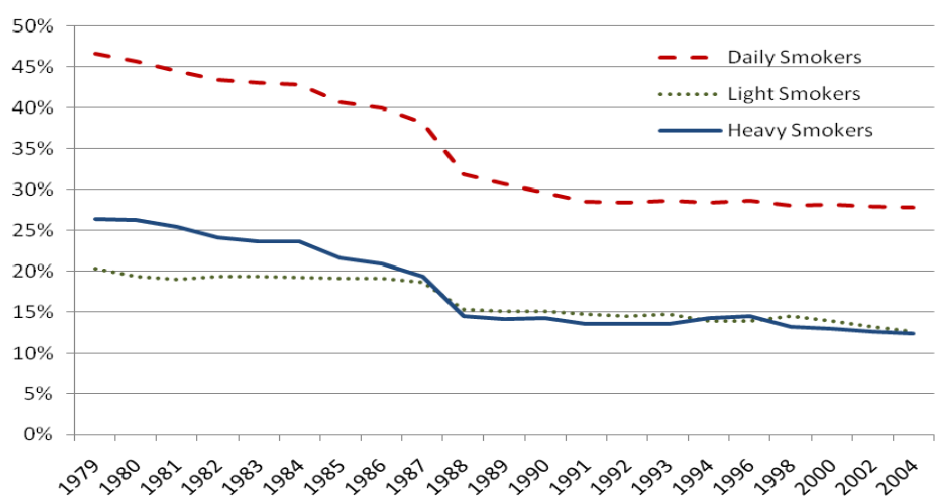

Figure 4.0.6. Percentage of smokers by survey years

Table 4.3 shows that 35% of the sample was daily smokers and 13% of the respondents smoked in all four years in which the smoking questions were asked. Figure 4.6 shows that the percentages of daily smokers, heavy smokers, and light smokers are decreasing over time. Percentages are high in first couple of years due to young experimenters. After 1990, although the number of respondents who answer the smoking questions in each survey year varies little, there are slight increases or decreases in the percentages because most of the current smokers attempt to quit smoking in these years. However, as Levine asserts, their attempts are often unsuccessful.

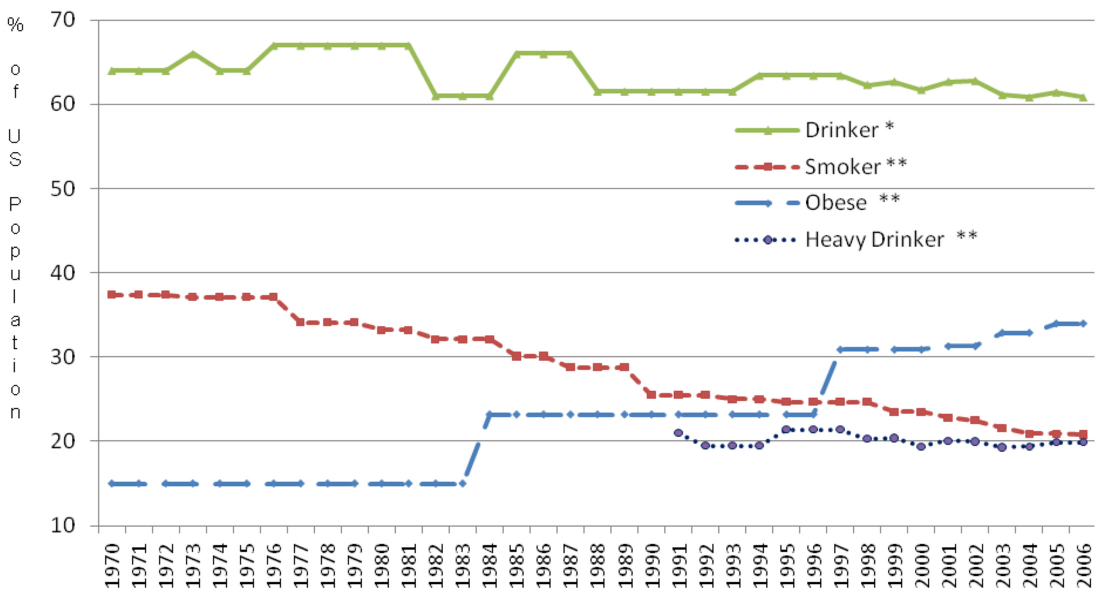

Figure 4.0.7. Percent of US Population: Obese, Smoker, Drinker, Heavy Drinker †
†: drinking 5 or more drinks more than 1 day in past month
SOURCES: * Center for Disease Control and Prevention: National Health and Nutrition Examination Survey (NHANES)

Figure 4.7 was prepared using CDC's NHANES and NHIS survey data. The percentage of US population that smokes has been decreasing from around 38% to about 21% during last 40 years. However, the prevalence of obesity has increased over the same period from 15% to 34%. The proportion of the US population that currently drinks and that heavily drinks (defined as drinking 5 or more drinks in more than one day in last 30 days) have stayed the same at about 65% and 20% respectively, though the proportions started to decline in the last decade.

The prevalence of these health behaviors in the sample are quite similar to the levels of these national surveys. For example, Table 4.2 and Figure 3 illustrate that around 63% of my sample are current drinkers, which is very close to the prevalence of drinkers in Figure 4.2. Moreover, Figure 4.7 shows that about 30% of the sample is obese in recent years, which is close to the most recent figure for obesity in Figure 4.4. The smoking prevalence of 21% in Figure 4.7 in recent years is also close to the most recent smoking figure shown in Figure 4.6.

4.2.5. Average Wages and Prevalence of being Employed by Health Behavior Characteristics:

Table 4.4 shows the average wages and prevalence of employed/unemployed individuals in the sample by single characteristics and demographic groups. According to Table 4.4, blacks and Hispanics are on average, more likely to be unemployed than whites. Only 67% of females are employed, which is around 15% less than employed males. Further, the average hourly wage of black workers is $2.26 less than white workers and $0.90 less than Hispanic workers. Females earn an average of $12.53, which is $3.10 less than for males.

Table 4.4. *Average Real Wages by Single Characteristics and Demographic Groups*

	Average Wages (After 1990)			Prevalence of Employment Variables		
	%	Wage	S.D.	%	Employed, %	Unempl., %
Full sample	*100*	17.65	23.331	*100*	74.16	9.22
By Single Characteristics						
Whites	*52.96*	19.47	25.524	*38.77*	90.73	3.97
Blacks	*28.41*	14.92	20.755	*27.19*	67.04	13.85
Hispanics	*18.63*	16.80	20.020	*17.53*	71.54	8.78
Males	*52.43*	19.59	25.393	*49.32*	81.99	9.92
Females	*47.57*	15.53	20.628	*50.68*	66.54	8.55
Smokers-all years	*13.54*	14.44	19.943	*12.84*	69.21	11.55
Nonsmokers-all years	*50.54*	19.33	25.277	*40.15*	79.26	6.67
Daily smokers (Smokers)	*26.01*	14.60	19.170	*33.22*	68.28	12.64
Non-daily smokers (Nonsmokers)	*70.07*	18.75	24.388	*61.38*	77.31	7.52
Obese	*22.69*	16.19	20.724	*13.21*	72.65	8.14
Nonobese	*75.32*	18.17	24.158	*84.16*	74.47	9.38
Drinker	*61.65*	18.81	24.595	*63.25*	78.68	9.19
Nondrinker	*36.14*	15.76	20.876	*36.75*	66.37	9.27
Binge drinker (B.Drinker)	*9.74*	15.63	18.772	*12.18*	76.80	12.22
By Combined Combined Characteristics						
Nonobese & Nonsmoker	*51.75*	19.49	25.307	*50.63*	77.86	7.57
Obese & Nonsmoker	*16.96*	16.77	21.438	*9.28*	77.73	7.40
Nonobese & Smoker	*20.54*	14.78	20.167	*29.04*	68.42	12.89
Obese & Smoker	*5.06*	13.96	15.337	*3.39*	66.92	10.36
Nondrinker & Nonsmoker	*27.28*	16.14	19.985	*25.03*	69.54	8.34
Nondrinker & Smoker	*7.60*	14.16	22.156	*9.41*	57.94	12.19
B.Drinker & Nonsmoker	*4.994*	17.13	19.851	*5.34*	81.18	9.85
B.Drinker & Smoker	*4.368*	13.95	17.514	*6.21*	72.87	14.46
Nondrinker & Nonobese	*25.22*	16.07	20.961	*29.48*	66.26	9.51
Nondrinker & Obese	*9.97*	15.01	20.024	*6.05*	66.59	8.38
B.Drinker & Nonobese	*7.48*	15.81	18.898	*10.43*	76.93	12.40
B.Drinker & Obese	*2.16*	15.15	18.725	*1.46*	76.75	10.05
Nonobese & Nonsmoker & Nondrinker	*18.53*	16.61	20.136	*19.67*	69.74	8.55
Nonobese & Nonsmoker & B.Drinker	*3.65*	17.54	18.639	*4.45*	81.45	10.02
Nonobese & Smoker & Nondrinker	*5.75*	14.29	23.247	*8.01*	57.62	12.42
Nonobese & Smoker & B.Drinker	*3.55*	14.11	19.101	*5.46*	73.03	14.59
Obese & Nonsmoker & Nondrinker	*7.98*	15.05	18.472	*4.60*	68.66	7.67
Obese & Nonsmoker & B.Drinker	*1.30*	16.23	22.918	*0.79*	79.80	8.71
Obese & Smoker & Nondrinker	*1.73*	13.83	19.054	*1.18*	58.76	11.22
Obese & Smoker & H.B.Drinker	*0.78*	13.29	8.282	*0.59*	71.96	11.90

The sample for wage analyses contains 125,264 person-year observations and the sample for unemployment analyses contains 119,033 person-year observations. Wages are in year-2003 dollars.

Table 4.4 also demonstrates the correlation between health behaviors and wages. For example, daily smokers in all years earn $14.40, which is $4.35 less than nonsmokers. Binge drinkers receive an average of $15.63, which is lower than average of nondrinkers or light to moderate drinkers. My sample consists of individuals from the early age of 18 to the oldest age of 46, and the obesity rates in early survey years are low as the real wage rates. Hence, descriptive analyses are done for the survey years after 1990 in order not to conclude wrongly that obesity has a positive wage effect. Table 4.4 shows that the average wage of obese workers is $16.19 which is almost $2.00 less than nonobese workers. Moreover, the table also reveals that daily smokers, binge drinkers, and obese individuals are on average, less likely to be employed than non-smokers, light to moderate drinkers, and the non-obese.

The descriptive statistics for the combined health factor groups support the importance of controlling for one health behavior when exploring the wage effect of another behavior since attaching another health behavior increases the wage penalty or the prevalence of being unemployed. For instance, the average wage for obese workers is $16.77 when they are also nonsmokers, but the average wage decreases to $13.96 for obese workers who are also smokers. Further, the prevalence of binge drinkers being unemployed is around 12.22%. The percentage of unemployment for binge drinkers rises to about 14.46% when they are also smokers, but the percentage decreases to 9.85% for binge drinkers when they are also nonsmokers.

The majority of the sample (51%) is nonobese nonsmokers, or nonobese drinkers (55%), or nonobese nonsmoker drinkers (31%). Consequently, one might think that these health factors will not have the expected effects on wages or unemployment prospects for the whole working population. However, around 5% of my sample is obese smokers, which is proportional to the US labor population, and this implies that roughly 5.5 million workers in the US are obese smokers whose wages or probability of being unemployed may be affected by both obesity and smoking behaviors. Approximately 0.6% of the sample are people who are obese, smoke and a binge drinker, and even this suggests that around 1 million US workers are affected by these three health behaviors at the same time.

Although the correlations between the health behaviors are low in my sample (highest is 0.25 between binge drinking and daily smoking, and the lowest is -0.09 between being obese and binge drinking) due to the panel data properties of the data, descriptive statistics still suggest that these health behaviors tend to cluster. For example, drinkers are less likely than nondrinkers to be obese (Z test=8.55). Additionally, obese people are less likely to be drinkers than non-obese people (Z test=7.78). Furthermore, smokers are less likely than nonsmokers to be obese, and the obese persons are less likely to than the nonobese persons to be smokers (Z test=10.02). Moreover, smokers are more likely to be drinkers, and drinkers are more likely to be smokers (Z test=15.45).

Consequently, these descriptive statistics, simple means and tabulations indicate that the prevalence of being unemployed (average wages) is lower (higher) for current drinkers but higher (lower) for obese, smokers or binge drinkers than their counterparts. On the other hand, these do not imply a causal relationship. As it is discussed, as a respondent gets older, the respondent’s human capital increases, thus his/her labor market success, wages and the likelihood of employment increase, but so does his/her weight. Hence, this relationship may be due to the increase in the age of the respondent. Similarly, some part of the effect of a health behavior on wages or on the probability of being unemployed might be due to another attached behavior of the respondent. As a result, to control for all the possible factors that may be correlated with health behaviors and wages/unemployment, multivariate regressions are employed. A set of standard (Table 4.5) and several supplemental covariates (Table 4.6) are included in multivariate regression analyses.

4.2.6. Other Explanatory Variables:

The NLSY enables one to learn respondents' educational backgrounds, household incomes, attitudes, residence information, demographic characteristics, family background characteristics, and a multiple of other categories. Most of the existing literature use similar individual, geographical and family background characteristics in their models and I use a combination of these variables in my estimations.

Table 4.5 provides the standard variables that are used in most of the human capital earnings regressions and employment/unemployment regressions in the literature and are used in this study. The table also provides each variable's sample mean and standard deviations of all the variables except the ones in percentage terms.

Table 4.5. *Descriptive statistics of standard variables*

	Variables	Mean	Std. Dev.
Dependent Variables			
Wage	Wage, *in 2003 $*	14.79	50.081
	Wage, *Averett, in 2003 $*	13.67	11.278
	Wage, *Cawley, in 2003 $*	14.15	18.444
	Log wage, *Cawley, in 2003 $*	2.37	0.737
Employment	Employed, *%*	74.16	
	Unemployed, *%*	9.22	
	Out of labor force, *%*	16.62	
	Part-time employed, *%*	7.14	
	Full-time employed, *%*	67.02	
	Self-employed, *%*	5.07	
	Employed-not-self, *%*	69.09	
Explanatory Variables			
Obesity Variables	Obese, *%*	15.27	
	Mild obese, *%*	10.51	
	Morbid obese, *%*	4.76	
	Body mass index, *kg/m^2*	25.34	4.965
	Overweight, *%*	33.23	
	Normal, *%*	45.71	
	Underweight, *%*	5.80	
Drinking Variables	Current drinker, *%*	66.47	
	Binge drinker, *%*	12.01	
	Heavy drinker, *%*	31.17	
Smoking Variables	Smoker, *%*	12.54	
	Nonsmoker, *%*	42.99	
	Quitter, *%*	3.90	
	Starter, *%*	3.43	
	Young experimenter, *%*	5.80	
	Unsuccessful quitter, *%*	8.11	
	Daily smoker, *%*	33.35	
	Light smoker, *%*	15.77	
	Heavy smoker, *%*	17.37	
Personal Characteristics			
Age	Age, *years*	28.66	6.648
Race-Ethnicity	Black, *%*	25.86	
	Hispanic, *%*	17.28	
	Non Black non Hispanic, %	56.86	
Gender	Male, *%*	52.18	
Marital Status	Married, *%*	44.82	
	Separated-divorced-widowed, *%*	12.80	
	Never married, *%*	42.38	

	Variables	Mean	Std. Dev.
Children	Children in household	0.86	1.111
Family Size	Family size	3.17	1.764
Health Status	Health limitation for work, *%*	6.62	
Acquired Human Capital			
Education Variables	Education Level, *years*	12.71	2.291
	Less than high school, *%*	3.28	
	Some high school, *%*	12.14	
	High school, *%*	48.13	
	More than high school, %	36.42	
	Some college, *%*	20.22	
	College or more, *%*	16.20	
Experience	Experience, *years*	7.41	4.948
Tenure	Tenure, *years*	3.39	4.084
Nonwage Income	Nonwage income, *in 2003 $*	17,308.98	46,257.853
Environmental Conditions			
Residence	South, *%*	39.49	
	Northeast, *%*	17.61	
	West, *%*	19.47	
	North central, %	23.43	
Urban-Rural	Urban, *%*	78.64	
	Rural, %	21.36	
Labor Market Conditions			
Unemployment Rate	Unemployment rate, *0-6%*	41.42	
	Unemployment rate, *6-9%*	36.16	
	Unemployment rate, *9%*	22.42	

Wage and log wage are hourly measures. Body mass index (BMI) is weight/(height^2). Separated-divorced-widowed dummy equals one if the respondent is separated or divorced or widowed. Children dummy is the number of children in household. Health limitation for work equals to one if the current health of respondent limits the kind or amount of work respondent can do. Education level is the highest grade in school respondent completed. South, northeast, west and north central dummies equal to one if the respondent's current residence is in south, northeast, west and north central, respectively. Urban and south are equal to one if the respondent's current residence is urban or south. Experience is cumulative hours worked since respondent completed schooling. Tenure is total years of tenure with employer. Nonwage income is total net family income minus total income from wages and salary.

Following Baum and Ford, and Peters, a set of supplemental covariates are included in the analyses to control more widely for individual background characteristics that might affect labor market outcomes and the respondent's attitudes as a way to limit the potential for unobserved heterogeneity bias. Table 4.6 shows descriptive statistics for these supplemental background variables and gives separate means and standard deviations.

Table 4.6. *Descriptive statistics of supplemental background variables*

	Variables	Mean	Std. Dev.
Supplemental Background Variables			
Personal Characteristics			
Native Born	Native born, *%*	93.21	
Foreign Language	Foreign language, *%*	1.57	
AFQT Score	AFQT score	37.60	27.562
Rotter External Control Score	Rotter score	8.75	2.409
Rosenberg Self-Esteem Score	Rosenberg score	33.17	3.977
Attitudes Towards Family Roles Score	Attitudes towards family roles score	16.52	3.316
Shyness	Shyness, *%*	29.09	
Anyone Present During Survey	Anyone present, *%*	20.47	
Substance Use			
Marijuana or Cocaine Use	Marijuana, 0, *%*	70.78	
	Marijuana, 1, *%*	16.07	
	Marijuana, 2-3, *%*	10.47	
	Marijuana, 4-5, *%*	2.69	
Childhood			
Foreign Language in Childhood	Foreign language in childhood, *%*	22.41	
Lived with Both Parents	Lived with parents *(at age 14)*, *%*	67.44	
Received Magazine	Magazine *(at age 14)*, *%*	53.71	
Received Newspaper	Newspaper *(at age 14)*, *%*	74.12	
Received Library Card	Library card *(at age 14)*, *%*	68.60	
Lived in South	South *(at age 14)*, *%*	36.35	
Lived in Urban	Urban *(at age 14)*, *%*	78.56	
Illegal Activity in 1980	Illegal activity, 1980, *%*	35.74	
Family			
Age of Youngest Child	No children, *%*	52.37	
	Youngest child, 0-1, *%*	8.67	
	Youngest child, 1-5, *%*	22.89	
	Youngest child, 5-, *%*	16.07	
Number of Siblings	No sibling, *%*	2.86	
	1 sibling, *%*	12.03	
	2-3 siblings, *%*	37.76	

	Variables	Mean	Std. Dev.
	4 or more siblings, *%*	47.35	
Older Sibling	Older sibling, %	76.93	
Family in Poverty	Family poverty status, %	12.14	
Family Information			
Mother's Work Information	Mother worked part-time *(at age 14)*, *%*	16.41	
	Mother worked full-time *(at age 14)*, *%*	40.88	
Father's Work Information	Father worked part-time *(at age 14)*, *%*	3.70	
	Father worked full-time *(at age 14)*, *%*	71.11	
Mother's Education	Mother, less than high school, *%*	19.48	
	Mother, some high school, *%*	23.04	
	Mother, high school, *%*	37.04	
	Mother, more than high school, *%*	13.70	
	Mother's education missing, *%*	6.74	
Father's Education	Father, less than high school, *%*	22.74	
	Father, some high school, *%*	15.66	
	Father, high school, *%*	28.91	
	Father, more than high school, *%*	17.59	
	Father's education missing, *%*	15.10	
Environmental Conditions			
SMSA Variables	Not in SMSA, *%*	23.12	
	SMSA, in central city, *%*	17.89	
	SMSA, not in central city, *%*	30.61	
	SMSA, central city unknown, *%*	28.37	
Work Information			
Experience	Experience, *years*	5.95	4.362
Tenure	Tenure, *years*	2.46	3.270
Nonwage Income	Nonwage income, *in 2003 $*	17,308.98	46,257.853

Native born is equal to one if respondent is born in the US. AFQT is the armed forces qualifications test given in 1980 (between 1-100). Rotter score (between 4 and 16): higher score means higher external control, lower score means higher internal control. Rosenberg self-esteem score (between 10 and 40): lower score denotes lower self-esteem. Attitudes towards family roles score (between 8 and 32): lower score indicates non-traditional views. Anyone present is if anyone was present during the survey. Marijuana questions are asked in 5 times, Marijuana, 0 is never used, Marijuana, 1 is answered positively only once, Marijuana, 2-3 means twice or three times, Marijuana, 4-5 is 4 or 5 times. Magazine and newspaper indicate if someone in respondent's household receives at least one magazine or newspaper subscription. Family poverty status is 1 if family is in poverty. White collar is 1 if respondent's occupation is professional, technical, manager, official, proprietor, sales worker or clerical. Experience is cumulative hours worked since respondent completed schooling. Tenure is total years of tenure with employer. Nonwage income is total net family income minus total income from wages and salary.

CHAPTER 5

Analyses of the Effects of Health Behaviors on Unemployment

5.1.Results when All Three Behaviors are Considered in Same Analysis:

The main aim of this study is to demonstrate that obesity, smoking, and binge drinking behaviors are correlated or tend to cluster, so their effects on unemployment may not be measured accurately in analyses that consider only one or two. No existing study has considered the effects of all three behaviors on unemployment at the same time and in the same analysis.

It is revealed by earlier health and economics literature that drinkers are more likely to be smokers than non-drinkers and smokers are more likely to be drinkers than non-smokers (Auld, 2005). The recent health journals argue that smokers are less likely to be obese than nonsmokers and a high percentage of people who quit smoking considerably gain weight (Baum et al., 2006). Moreover, recent research also shows that blacks, people with low income and low education are more likely to engage in all of these health behaviors.

As demonstrated in the descriptive statistics portion of the data section, similar to the results of earlier literature, obese individuals are found to be less likely than their non-obese counterparts to be a smoker or drinker, smokers are less likely than nonsmokers to be obese and more likely to be a drinker, and drinkers are less likely than abstainers to be obese and more likely to be a smoker.

In the light of these descriptive statistics and correlations, since the health factors are correlated, I argue that failing to control for one or more of these health behaviors in the employment/unemployment regression analyses may lead to underestimation or overestimation of the effects of the behaviors on the probability of being unemployed or employed.

Attention is on the employment/unemployment issue, because finding and holding a job is a prerequisite to the other relevant aspects of an individual's labor market experience, i.e. wages, labor force participation, occupation, or job quality. Furthermore, reduced employment propensities may constitute the greatest negative effects of health behaviors on labor market success (Mullahy and Sindelar, 1993) and such effects are ignored when analyzing samples consisting only of workers.

With the intention of observing the changes in the effects of these behaviors on the likelihood of being unemployed, four different regressions are run using the probit regression estimation model for unemployment analyses.

In the first three regression analyses, each behavior is included on the right-hand side of the unemployment equation in addition to standard and most commonly used explanatory variables, but without the other two behaviors. In the last model, all three health behaviors are included and controlled for each other. Table 5.1 shows the probit estimates of the behaviors on the probability of being unemployed for males and females separately.[12] Marginal effects are in brackets.[13]

[12] People who reported their employment status as out of labor force are excluded from the probit estimations.

Table 5.1. *The effects of obesity, daily smoking, and binge drinking on the likelihood of being unemployed when different specifications are used*

	Males				Females			
Probit results:	Model1	Model2	Model3	Model4	Model1	Model2	Model3	Model4
Obese	0.037 (0.044) [0.004]			0.078 * (0.041) [0.009]	0.052 (0.047) [0.006]			0.073 * (0.042) [0.009]
Daily smoker		0.212*** (0.027) [0.024]		0.217*** (0.027) [0.025]		0.185*** (0.028) [0.023]		0.187*** (0.028) [0.024]
Binge Drinker			0.056** (0.027) [0.007]	0.024 (0.026) [0.002]			0.126*** (0.046) [0.018]	0.085 * (0.048) [0.011]
Pseudo R^2	*0.203*	*0.207*	*0.202*	*0.209*	*0.209*	*0.213*	*0.206*	*0.215*
LR test with Model 1				LR chi2(2) = 268.81 P = 0.000				LR chi2(2) = 147.17 P = 0.000
LR test with Model 2				LR chi2(2) = 20.08 P = 0.001				LR chi2(2) = 15.13 P = 0.001
LR test with Model 3				LR chi2(2) = 252.79 P = 0.000				LR chi2(2) = 198.56 P = 0.000
Goodness of Fit Test with Model 1				Difference in BIC = 245.43				Difference in BIC = 183.33
Goodness of Fit Test with Model 2				Difference in BIC = 26.66				Difference in BIC = 19.01
Goodness of Fit Test with Model 3				Difference in BIC = 231.51				Difference in BIC = 186.15

Regression models contain 64,712 person-year observations for males and 54,321 person-year observations for females. Clustered robust standard errors are in parentheses. Marginal effects are in brackets. Sampling weights are controlled. *** Significant at the 0.01 level. ** Significant at the 0.05 level. * Significant at the 0.10 level. See Table 4.5 for other explanatory variables.

[13] Marginal effect of a health behavior on unemployment is calculated as the difference in expected probabilities of unemployment between having the behavior and not having it.

Table 5.1 illustrates that failing to include one or more of health behaviors in unemployment regressions would lead to underestimation of the impacts of being obese and overestimation of the effect of binge drinking for both males and females.

The table shows that the estimated parameters of the obesity for males and females increase and become statistically significantly different from zero at the 10% level when smoking and binge drinking behavior of individuals are controlled for (P values are 0.019 and 0.027 for males and females, respectively; obesity coefficients are different in the regression with single behavior and the regression with all behaviors). The marginal effect of obesity on unemployment increases from 0.4 to 0.9 for males and increases from 0.6 to 0.9 for females.

Further, the estimates of binge drinking and their significance levels decrease when obesity and smoking behaviors are included in the same regression for both genders (P values are 0.013 and 0.017 for males and females, respectively; binge drinking coefficients are different in regression with single behavior and regression with all behaviors). The marginal effect of binge drinking on unemployment falls from 0.7 to 0.2 for males and the estimated coefficient, which is statistically significant at the 1% level, becomes insignificant. The marginal effect falls from 1.8 to 1.1 for females and the estimated parameter becomes statistically significant only at the 10% level.

The table also reveals that the coefficient of smoking changes only slightly for males and females (changes are insignificant) when the other behaviors are included into unemployment regression. Also, the impact of smoking on the odds of being unemployed is greater than the impact of binge drinking and the impact of obesity in all different models for both genders, indicating that smoking has the highest impact on unemployment. The marginal effect of binge drinking is greater than the marginal effect of obesity (insignificant) for both genders when the behaviors are not controlled for. But, the marginal effect of obesity becomes greater than the marginal effect of binge drinking (insignificant) for males when the behaviors are controlled for.

To support my ideas, above models are compared using likelihood ratio tests and Hosmer and Lemeshow's goodness of fit tests. The likelihood ratio tests test the null hypothesis that removing some variables (namely obesity, smoking, or binge drinking) has no effect; it does not lead to a poorer fitting model. The chi square statistics for all comparisons are greater than 10 and the corresponding P values are less than 0.05, which are statistically significant. These results mean that the health behaviors that were removed to produce the reduced models resulted in a model that has a significantly poorer fit, and therefore all the variables should be included in the unemployment models.

The idea of goodness of fit tests is that if the predicted frequency and observed frequency match more closely, they fit better. AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) are calculated to compare maximum likelihood models. Difference in BIC and difference in AIC are all positive and large enough for each model specification, therefore the results strongly support the models with all three health behaviors over the models with only one health behavior.[14]

---

[14] The values of AIC are not shown on the table since they are very close to BIC.

Although Pseudo R^2 is not a commonly accepted measure of explanatory power in logistic regression models, I would claim that the basic contention is also supported since Pseudo R^2 increases every time an additional health behavior is controlled for.

5.2. Endogeneity of the Health Behaviors:

Providing unbiased estimates of obesity, smoking, and binge drinking on unemployment is not straightforward because the probit estimates do not correctly characterize the effects of health behaviors since health behaviors may be endogenous. The health behaviors and unemployment may simultaneously affect each other. In addition, an unobserved factor may be correlated with both health behaviors and unemployment, such as time preference, self-control, sociability, etc. Technically the possibilities mean that the error term in the unemployment equation is correlated with obesity, smoking, or binge drinking binary variables. In this case the coefficient estimates are biased.

In order to limit the potential for endogeneity bias or the issue of omitted variables and to control more widely for individual background characteristics that might affect labor market outcomes, the supplemental background variables are included into the unemployment equation (Table A1 in Appendix).[15] Once the supplemental

[15] See Table A1 in Appendix for the results. To correct the possible specification error that might be created by employing numerous supplemental covariates, the Stata command 'linktest' is used which detects if the model is not properly specified. Although the test results marginally reject the null hypothesis that specification error occurs due to omitted relevant variables or function is not correctly specified, Box-Tidwell model is employed to specify the model better. It transforms a predictor using power transformations and finds the best power for model fit based on maximum likelihood estimate. Further, to correct the possible multicollinearity problem, another Stata command 'vif' is employed. There seems to be no multicollinearity with health behaviors even when early years marijuana-cocaine usage is included in

background variables are controlled, the estimated parameters of smoking and binge drinking slightly decrease for males (the changes are not statistically significant). The coefficient of obesity increases and becomes statistically significant at the 1% level for males (P value is 0.015, the change is statistically significant), and the coefficients of obesity and binge drinking fall and become insignificant for females (P value are 0.029 and 0.023, the changes are statistically significant). The effects of smoking are unchanged. Hence, it can be argued that the effects of obesity and binge drinking are subject to omitted variable bias, which shows the importance of controlling for supplemental background variables.[16]

Nonlinear models such as logit or probit are often used to estimate the likelihood of being unemployed/employed, but these models do not allow us to control for the individual heterogeneity, so it may produce biased results due to a correlation between some unobservable characteristics and some explanatory variables. In order to account for this possible unobserved heterogeneity, panel data techniques are also employed in this study.[17] The fixed-effects estimates of obesity and smoking are smaller in magnitude than the cross-sectional estimates and the coefficients of binge drinking are greater for

---

unemployment model. To reduce the multicollinearity, age and years of tenure are centered by subtracting the means from their predictor values before generating the square terms.

[16] The likelihood ratio tests support the new model with supplemental covariates over the model without them for both males and females. Similarly, the goodness of fit tests argue that large and positive differences of BIC's between the two models strongly support the second model with supplemental covariates over the first model.

[17] The results from the fixed-effects models, random-effects models and Hausman test results are displayed in Appendix, Table A3. Firstly, fixed-effects regressions are used to control for unobserved variables that differ between individuals but constant over time. It is argued that an alternative to fixed-effects is random-effects, which assumes that some unobserved variables may vary between individuals but constant over time, and others vary over time but constant between individuals. For both genders, the Hausman specification tests reject the null hypothesis that fixed and random-effects models produce the same coefficients, with P-values less than 0.001. Therefore, these suggest that the differences between these two models are systematic and only fixed-effects models should be taken into account.

males and females. Moreover, the fixed-effects estimates of obesity become statistically insignificant for both genders. These indicate that the effects of being obese on the odds of being unemployed may be due to the unobserved time-invariant individual characteristics.

Fixed-effects model does not eliminate bias in case of time-variant individual heterogeneity and takes out all the covariates that are time-invariant. Further, the model may worsen the bias caused by measurement error. Also, the model cannot identify the effects of individuals who do not change their behavior over time, and it assumes individuals who change their behavior over time are similar to the ones who do not change their behavior over time. Therefore, the results must be analyzed with caution.

Existing literature has generally dealt with the endogeneity problem by implementing instrumental variables methods. The instrumental variables method is conceptually difficult and easily misused; nonetheless it is widely used in econometrics, as Cameron and Trivedi (2005) point out. The basic assumption of instrumental variables method is that the instruments should be correlated with and exogenous to obesity, daily smoking, and binge drinking, but should be uncorrelated with the employment/unemployment. To employ the instrumental variables regression method and to test for exogeneity, a two-stage equations approach in the probit model is applied by using the Stata command 'ivprobit'. In order to exploit the panel data properties of the sample, the IV method with fixed-effects is also employed. [18] [19]

---

[18] See Appendix, Table A6, for the results of conventional IV methods and IV methods with fixed-effects. The orthogonality of the instruments for the independence is examined and the ones that satisfy the basic assumptions of instrumental variables methods are included in the obesity, smoking, and drinking equations

The results for males show that when an instrumental variables approach is applied, the estimated coefficients and marginal effects of all health behaviors increase, but the estimate of obesity becomes statistically insignificant (the change is statistically significant) and the estimate of binge drinking stays statistically insignificant. For females, when instrumental variables approach is applied, similar to the results of the males, the estimated coefficients on all health behaviors increase, but the estimate of binge drinking becomes statistically insignificant (the change is statistically significant). When the panel data technique (fixed-effects) is employed together with the IV model, the magnitudes of all of the estimated parameters fall and only the estimated coefficients of daily smoking are statistically significant for males and females.

The achievement of unbiased estimates via the IV method depends essentially on the predictive power and validity of the instruments. If weak correlation exists between the instruments and the employment status of individuals, this can lead to bias in instrumental variables estimates. Also, mean square errors with an IV method may be large which implies a trade-off between bias and variance (Bollen et al., 1995; Norton et al., 1998). Moreover, instrumental variables method is less efficient since it does not fully

---

as instruments: the religious attendance dummy in 1982, religious affiliation dummies (non-religious, Baptist, Catholic, Jewish), Rotter test score, youngest age started drinking dummies, alcoholic father (only satisfied for males), alcoholic mother (only satisfied for females) and cumulative years lived with alcoholic relative variables. The correlations between obesity, daily smoking, and binge drinking and these instruments are tested by F-tests. All the exogenous variables and instruments are regressed on being obese, being daily smoker, and being binge drinker in three different equations to attain the reduced form residuals. Then, obesity, daily smoking and binge drinking are regressed on being unemployed by adding these reduced form residuals into the unemployment equation. Wald test of exogeneity is employed to test the correlation between the error terms in the first stage and second stage equations. Hansen J statistic for over-identification and Cragg-Donald Wald statistic for under-identifications are calculated and tested.

[19] However, this model needs instruments that vary over time, predict obesity, smoking, or binge drinking, and uncorrelated with employment/unemployment prospects of the individual. The only instruments I use that vary over time are religiosity dummies. Thus, the results of IV with fixed effects must be analyzed with caution.

make use of the non-linearity of the unemployment model. Therefore, a more appropriate non-linear model, i.e. multivariate probit model, specification is also applied, which jointly estimates several correlated binary outcomes using maximum likelihood estimation.

Table 5.2. *Multivariate probit estimation results of being unemployed for obese, smokers, and binge drinkers, using instrumental variables*

	Male		Female	
Multivariate probit results:	Probit	Mvprobit	Probit	Mvprobit
Obese	0.118 *** (0.039) [0.010]	0.155 (0.096) [0.016]	0.026 (0.038) [0.002]	0.191 * (0.099) [0.018]
Daily smoker	0.115 *** (0.026) [0.009]	0.271 *** (0.049) [0.025]	0.102 *** (0.027) [0.010]	0.207 *** (0.041) [0.018]
Binge Drinker	0.013 (0.026) [0.001]	0.039 (0.032) [0.002]	0.009 (0.042) [0.001]	0.102 (0.070) [0.008]
Instruments: *(included in obesity, smoking and drinking equations)*	Rotter test; Youngest Age Started Drinking, less than 13; Youngest Age Started Drinking, 13-16; Youngest Age Started Drinking, 16-18; Religious Attendance in 1982, at least once a week; Religious Attendance in 1982, at least once a week; Alcoholic Father; Alcoholic Mother; Alcoholic relative, years lived			
LR Test of all correlations are zero	Chi2(6)=65.10, P=0.000		Chi2(6)=43.33, P=0.000	

Regression models contain 64,712 male and 54,321 female person-year observations. People who reported their employment status as out of labor force are excluded from the analyses. Clustered robust standard errors are in parentheses. Adjusted predicted probabilities are in brackets. Sampling weights are controlled. *** Significant at the 0.01 level. **Significant at the 0.05 level. *Significant at the 0.10 level. See Table 4.6 for the list of all explanatory covariates.

Multivariate probit results are displayed in Table 5.2. Probit results of employment status (being unemployed) are compared to the results of multivariate probit models. In the multivariate probit estimation, unemployment, obesity, daily smoking, and binge drinking equations are jointly estimated. Obesity, daily smoking, and binge drinking equations include the instruments used in earlier IV models. Marginal effects are in brackets.

Table 5.2 shows that the multivariate probit estimates generally demonstrate different relationships than the probit models. When all four equations are jointly estimated and health behaviors are instrumented, the estimated coefficients of all health behaviors increase. Marginal effect of smoking rises from 0.9 to 2.5 (P value is 0.007, the change is statistically significant) for males and marginal effect of smoking rises from 1.0 to 1.8 (P value is 0.027, the change is statistically significant) for females. Although the marginal effect of obesity increases for males, the estimated coefficient becomes statistically insignificant (P value is 0.022, the change is statistically significant), and the marginal effect of obesity considerably rises from 0.2 to 1.8 (P value is 0.009, the change is statistically significant) for females and the coefficient becomes statistically significantly different from zero.

These results illustrate that the effects of the health behaviors on unemployment are associated with endogeneity bias associated with probit models. Being obese increases the probability of being unemployed by 1.8% for females only, smoking increases the likelihood of being unemployed by 2.5% and 1.8% for males and females, respectively, and binge drinking appears to have no effect for either gender.

The Wald test is used to reject the null hypothesis that the correlation parameter, rho, between two equations is zero. Explicitly, the test simply checks whether the error terms in the unemployment equation and the reduced-form equations for the endogenous variables, obesity, daily smoking, or binge drinking are correlated. As it turns out, the unemployment equation is not strongly associated with any of the health behavior equations (P values are 0.117, 0.103 and 0.212, respectively). But, the joint test of the covariances between the errors of three reduced form equations and the error of the unemployment equation reject the null that none of the equations are strongly associated, which demonstrates the importance of controlling for endogeneity.

Table 5.3 shows the results of the multivariate probit method when three health behaviors are considered individually and simultaneously for males and females separately.

Table 5.3. *The estimates of the multivariate probit models of the effects of obesity, daily smoking, and binge drinking on unemployment when different specifications are used*

	Males				Females			
	Model1	Model 2	Model3	Model4	Model1	Model2	Model3	Model4
Obese	0.106 (0.092) [0.009]			0.155 (0.096) [0.014]	0.148 * (0.088) [0.013]			0.191 * (0.099) [0.018]
Daily smoker		0.252*** (0.051) [0.024]		0.271*** (0.049) [0.025]		0.118*** (0.043) [0.015]		0.207*** (0.041) [0.018]
Binge Drinker			0.046 (0.033) [0.004]	0.039 (0.032) [0.002]			0.146 * (0.075) [0.012]	0.102 (0.070) [0.008]

Regression models contain 64,712 person-year observations for males and 54,321 person-year observations for females. People who reported their employment status as out of labor force are excluded from the analyses. Clustered robust standard errors are in parentheses. Marginal effects are in brackets. Sampling weights are controlled. *** Significant at the 0.01 level. ** Significant at the 0.05 level. * Significant at the 0.10 level. See Table 4.6 for other explanatory variables.

Table 5.3 shows that, although the estimated parameters of the health behaviors noticeably change, the changes are not statistically significant. When all the behaviors are controlled for each other, the marginal effects of obesity increase from 0.9 to 1.4 and from 1.3 to 1.8 for males and females, respectively, but the changes are not significantly different from zero (P values are 0.17 and 0.19, respectively). Also, the significance levels of the coefficients stay the same; the estimated parameter of obesity is still statistically insignificant for males and statistically significant at the 10% level for females. Moreover, the marginal effects of daily smoking only slightly increase for both males and females. Furthermore, the marginal effect of binge drinking falls from 0.4 to 0.2 for males and the change is not significant (P value is 0.34). The only statistically significant change occurs in the binge drinking coefficient for females; the marginal effect falls from 1.2 to 0.8 (P value is 0.045).

Consequently, in contrast to the earlier results, when endogeneity is accounted for, the estimated coefficients of the health behaviors are statistically the same whether these three behaviors are considered individually or all together for both genders. Therefore, the biases in the probit estimates are due to failure of probit to account for endogeneity.

5.3. Sensitivity and Robustness Analyses:

5.3.1. Interactive/Additive Characteristics of Health Behaviors:

Up to now it has been supposed that the impact of one health behavior is not correlated with the impact of another health behavior. The behaviors could be additive, for instance if health worsens when a binge drinker smokes or a smoker becomes obese; hence the likelihood of being unemployed may increase proportionally by poorer health. However, the health behaviors would not be additive, if for example, they are due to the same unobserved factors or if smokers already perceive a wage penalty or unemployment effect and being obese produces minor or no additional impact on wages or unemployment, or vice versa (Baum et al., 2006).

In order to support the earlier results and to investigate if the effects of these behaviors are additive or interactive, four interaction terms are added to unemployment equations: obese*daily smoker, obese*binge drinker, daily smoker*binge drinker, and obese*daily smoker*binge drinker. These terms will be statistically significant if dual or triple interactions exist. If the interaction terms are statistically insignificant, then one could conclude that the behaviors are not interactive; their total effects could be found by just adding the coefficients together. The results from multivariate probit estimates of three health behaviors and their interactions are displayed in Table 5.4.[20]

The results illustrate that almost all of the estimated coefficients of health behaviors are statistically the same as the earlier estimates of the models without the interaction terms for both genders. Moreover, none of the interaction terms are statistically significantly different from zero; hence the effects of the behaviors are not interactive.

But, when interaction terms are included into the multivariate probit models, binge drinking estimates become statistically significant for males and only marginally rejected at the 10% level for females, which shows the importance of including interaction terms. Therefore, if an individual is neither obese nor a smoker, being a binge

[20] Firstly, the interactions are included into the unemployment equation once at a time in order not to lead to any multicollinearity problem. Then, obesity, smoking, and binge drinking dummy variables are centered at their mean values and interaction terms are created using these centered variables. Results are largely unchanged when all centered interaction terms are included in the same unemployment equation, therefore only one estimate is displayed in table.

drinker results in an increase in the probability of being unemployed by 0.5%. Hence, the effect of binge drinking could be correlated with the obesity and/or smoking behavior of the individual, for instance, if an obese or smoker already faces an unemployment effect, e.g. due to discrimination, also becoming a binge drinker may not result in any additional impact on unemployment.

Table 5.4. *The estimates Multivariate Probit models of the effects of health behaviors and their interactions on unemployment, for males and females separately*

Multivariate Probit	Males	Females
	0.158	0.195 *
Obese	(0.097)	(0.100)
	[0.016]	[0.018]
	0.266 ***	0.199 ***
Daily smoker	(0.050)	(0.041)
	[0.025]	[0.018]
	0.051 *	0.114
Binge Drinker	(0.030)	(0.074)
	[0.005]	[0.010]
	-0.029	-0.075
Obese * Smoker	(0.097)	(0.075)
	[-0.002]	[-0.006]
	0.059	-0.016
Obese * Binge Drinker	(0.114)	(0.108)
	[0.005]	[-0.001]
	-0.028	0.042
Smoker * Binge Drinker	(0.053)	(0.085)
	[-0.002]	[0.003]
	0.063	0.088
Obese * Smoker * Binge Drinker	(0.011)	(0.092)
	[-0.005]	[0.008]

Unemployment regression models contain 64,712 male and 54,321 female person-year observations. People who reported their employment status as out of labor force are excluded from employment analyses. Marginal effects are in brackets.

5.3.2. More Detailed Measures of the Health Behaviors:

The effects of some additional measures of three health behaviors and their interactions on the probability of unemployment are also analyzed in order to observe whether they show consistent results with the earlier findings and to provide several sensitivity and robustness checks. The results are displayed in Table 5.5.

Table 5.5. *The estimates of Multivariate Probit models of the effects of more detailed measures of health behaviors on unemployment*

Multivariate Probit Results:	Males	Females
	0.289 ***	0.212 ***
Smoker	(0.045)	(0.039)
	[0.027]	[0.019]
	0.032	0.015
Quitter	(0.045)	(0.036)
	[0.003]	[0.001]
	0.106 **	0.085 **
Starter	(0.046)	(0.035)
	[0.009]	[0.007]
	0.075	0.009
Young Experimenter	(0.055)	(0.047)
	[0.006]	[0.001]
	0.178 **	0.089
Unsuccessful Quitter	(0.071)	(0.068)
	[0.015]	[0.007]
	0.077	0.126 *
Mild Obese	(0.096)	(0.078)
	[0.007]	[0.011]
	0.196 *	0.261 ***
Morbid Obese	(0.102)	(0.109)
	[0.018]	[0.024]
	0.168 ***	0.164 ***
Light Smoker	(0.038)	(0.044)
	[0.016]	[0.015]
	0.302 ***	0.233 ***
Heavy Smoker	(0.057)	(0.041)
	[0.028]	[0.021]
	-0.073	-0.085
Current Drinker	(0.053)	(0.055)

	[-0.006]	[-0.07]
Heavy Drinker	-0.016 (0.026) [-0.001]	-0.009 (0.037) [-0.001]

Unemployment regression models contain 64,712 male and 54,321 female person-year observations. People who reported their employment status as out of labor force are excluded from employment analyses.

'Morbid obesity' is defined as having a BMI of 35 or more and 'mild obesity' is defined as having a BMI of between 30 and 35. 'Smoker' is defined as being a daily smoker in all survey years when smoking questions are asked, i.e. 1984, 1992, 1994, and 1998. A 'young experimenter' is an individual who smokes only in the 1984 survey year and then quits smoking in later survey year, 1992. A 'heavy smoker' is defined as smoking more than 20 cigarettes per day and a 'light smoker' is someone who smokes one to 20 cigarettes per day. A 'current drinker' is equal to one if the respondent drinks any alcoholic beverage in last month. 'Heavy drinking' is defined as drinking 6 or more drinks in one to three occasions in past month of the survey year. [21]

Table 5.5 shows that the estimates of smoking variables are as expected; higher marginal effects of smokers, starters, and unsuccessful quitters than nonsmokers and the impacts are greater in magnitude for males. However, the results demonstrate that smokers are a heterogeneous group of people: the unemployment effects of (persistent)

[21] In order not to fall into overlapping problem, light and heavy smoking variables and heavy drinking variable are analyzed in different regressions than other smoking and drinking variables. Interaction terms are included in two additional different regressions and but the results are not displayed since the coefficients are largely unchanged.

smokers are different than the effect of starters and quitters, and young experimenters and quitters seem to have no effect on unemployment for both genders.[22]

Additional obesity variables help us to understand the earlier insignificant effect of being obese on unemployment for males. Even though mild or morbid obesity leads to a higher likelihood of unemployment for females, only morbid obesity appears to affect the likelihood of being unemployed for males (marginal effect of morbid obesity is 1.8 and it is significant at the 10% level). Hence, obesity affects the unemployment only at the extremes of obesity for males.

Moreover, contrary to the previous literature, drinking is found to have no effect on unemployment for either gender when endogeneity is accounted for. The results indicate that the marginal effect of current drinking or binge drinking is statistically significant in probit models, which supports the idea of positive effects of drinking on labor market outcomes. However, once endogeneity is addressed via multivariate probit models, all the coefficients become statistically insignificant for both genders, hence the effects are mostly due to failure of probit models to account for endogeneity.

Furthermore, the results show that daily smoking has impacts on unemployment even in different measures of daily smoking. Heavy smokers and light smokers have statistically significant effects on the likelihood of being unemployed of both genders, and the marginal effects are greater for males.

[22] The heterogeneity among smokers gives an insight of why the fixed-effects results might be biased. The results show that those who change their behavior over time, e.g. quitters, smokers, young experimenters, are different from the ones who do not change, and they have different effects on unemployment. However, the whole identification of fixed-effects models is based on the individuals who change their behavior.

5.3.3. Multinomial Analyses:

5.3.3.1. Employed – Unemployed – Out of Labor Force Analysis:

The analyses of the effects of health factors on the likelihood of unemployment have ignored the individuals who are out of labor force. In order to examine the effects of behaviors more broadly and to obtain more comprehensive measures for labor market relations, all possible states of employment are included in the analyses and multinomial logit models are employed.[23] The dependent variable has three outcomes: employed, unemployed, and out of labor force. This model shows the probability of being unemployed (or out of labor force) as opposed to being employed. Table 5.6 shows the multinomial logit estimates (base category is employed).

Table 5.6. *Multinomial logit estimates of being employed for obese, smoker, and binge drinker*

	Male		Female	
Multinomial Logit:	Unemployed	Out of Labor Force	Unemployed	Out of Labor Force
Obese	1.043 (0.073)	1.149 (.107)	1.132 * (0.077)	1.061 (0.066)
Daily smoker	1.240 *** (0.058)	1.066 (0.067)	1.172 *** (0.058)	0. 938 (0.041)
Binge Drinker	1.049 (0.050)	0. 716 *** (0.050)	1.037 (0.078)	0. 995 (0.077)

Regression models contain 70,409 male and 72,344 female person-year observations. Clustered robust standard errors are in parentheses. Sampling weights are controlled. *** Significant at the 0.01 level. **Significant at the 0.05 level. *Significant at the 0.10 level. See Table 4.6 for the list of all explanatory covariates.

[23] The multinomial logit relies on the assumption of independence of irrelevant alternatives. Although this might be a limitation in this study, being the choices likely to be correlated based on unobservables, the results are reported for completeness.

The table shows similar results as compared to earlier results; daily smoker males and females are more likely to be unemployed than nonsmokers, and obese females are more likely than nonobese females to be unemployed. Although most of the coefficient estimates of behaviors for out of labor force status are statistically insignificant, the binge drinking coefficient is statistically significantly different from zero at the 1% level for males. The estimate states that a significantly lower probability of being out of labor force compared to being employed exists for binge drinker males, in other words, binge drinker males are more likely to be employed than out of labor force.

One explanation for the result could be that a binge drinker needs work to assure the current and future alcohol consumption. Alternatively, binge drinker males could be less likely to be out of labor force, e.g. less likely to be inactive due to studying or parental leave, than their moderate drinker or abstainer counterparts. However, it should be noted that out of labor force individuals cannot be treated as a uniform group and the changes in health behaviors and in lifestyle as one enters into and out of the work force should be studied in the future.

5.3.3.2. Full-time – Part-time Employment Analysis:

Multinomial analysis is also done for full-time/part-time employment prospects. One may argue that since obesity, smoking, or binge drinking has adversa effects on health, people with these behaviors may observe health limitations to work full-time or may choose to work part-time. Alternatively, these three behaviors may affect the employment status of an individual through an evaluation by co-workers, customers or employers based on the workers’ smoking, drinking, or obesity status. Employers may discriminate against them in hiring or firing decisions, or co-workers or customers may dislike them, therefore these obese, smoker, or binge drinker individuals may tend to be in part-time jobs.

Multinomial logit estimates of the employment status for people with three health behaviors are displayed in Table 5.7. ‘Part-time employed’ is defined as working less than 35 hours per week in the survey year and ‘full-time employed’ is defined as working 35 or more hours per week in the survey year. We expect to see that obese, smoker, or binge drinkers are more likely to be employed part-time than employed full-time. The based category is being full-time employed.

Table 5.7. *Multinomial logit estimates of being full-time employed, part-time employed, base category is being full-time employed*

	Male			Female		
Mlogit results	Part-time	Unemployed	Out of Labor Force	Part-time	Unemployed	Out of Labor Force
Obese	1.088 (0.134)	1.047 (0.072)	1.158 (0.109)	0.879 (0.075)	1.130 * (0.078)	1.023 (0.071)
Daily smoker	0.886 (0.068)	1.227*** (0.059)	1.052 (0.067)	0.747*** (0.045)	1.111** (0.056)	0. 944 (0.040)
Binge Drinker	0 .800*** (0.062)	1.032 (0.050)	0.703*** (0.050)	1.078 (0.104)	1.047 (0.082)	1.007 (0.083)

Regression models contain 70,409 male and 72,344 female person-year observations. Clustered robust standard errors are in parentheses. Sampling weights are controlled. *** Significant at the 0.01 level. **Significant at the 0.05 level. *Significant at the 0.10 level. See Table 4.6 for the list of all explanatory covariates.

The table shows that some of the results contradict earlier assumptions. As expected, obese females, and daily smoker males and females are less likely to be full-time employed than unemployed. But daily smoker females and binge drinker males are less likely to be part-time employed than to be full-time employed, and all the other relationships are statistically insignificant. These results contradict our assumption that individuals having these behaviors are more likely to observe health limitations to work full-time or more likely to be discriminated; therefore they will choose to be part-time employed than full-time than their counterparts.

5.3.3.3. Self-employed – Employed by others Analysis:

Another conflicting result appears to be on the impacts of obesity, daily smoking, and binge drinking on being self-employed. One may think that people having these three behaviors may tend to choose self-employment as they are more likely to develop health limitations to work and they are more likely to be discriminated against by potential employers and co-workers. Table 5.8 shows the multivariate logit estimates of the employment status for obese, smoker, and binge drinkers. Being employed by others is the base category.

Table 5.8. *Multinomial logit estimates of being self-employed, employed by others, and unemployed, base category is being employed by others*

	Male		Female	
Mlogit results:	Unemployed	Self-employed	Unemployed	Self-employed
Obese	1.041 (0.070)	0.747 ** (0.094)	1.130 * (0.078)	0.944 (0.130)
Daily smoker	1.217 *** (0.056)	0.821 ** (0.040)	1.180 *** (0.058)	0.793 * (0.095)
Binge Drinker	1.024 (0.048)	0.908 (0.080)	1.012 (0.079)	0.983 (0.175)

Regression models contain 64,712 male and 54,321 female person-year observations. Clustered robust standard errors are in parentheses. Sampling weights are controlled. *** Significant at the 0.01 level. **Significant at the 0.05 level. *Significant at the 0.10 level. See Table 12 for the list of all explanatory covariates.

Being obese for females and being daily smoker for both genders are more likely to be unemployed than employed-not-self, as expected. However, daily smoker males and females are less likely to be self-employed than employed by others. Furthermore, obese males are also less likely than their counterparts to be self-employed than employed by others.

The results of the full-time/part-time analyses and self-employment/employed-not-self analyses contradict our earlier hypotheses. One possible explanation could be that people having these health behaviors who face wage discrimination may move to the sectors or full-time employment where the wage structure is fixed, even if they receive lower wages. Alternatively, the relationship could be the opposite: self-employed people (part-time workers) might be less likely to be obese or be a daily smoker than employed-not-self individuals (full-time workers). Maybe looks (i.e., not obese) and energy are more important for self-employment, so that people with poor health habits are less likely to be self-employed. On the other hand, we might be getting these results simply because part-time or self-employed people could be more likely to underreport their actual weights or to misreport their smoking and drinking behaviors.

Furthermore, the results must be assessed with caution since the multinomial logit model relies on the assumption of independence of irrelevant alternatives and it does not account for endogeneity or unobserved heterogeneity. The analysis requires further research.

CHAPTER 6

Analyses of the Effects of Health Behaviors on Wages

6.1.Results when All Three Behaviors are Considered in Same Analysis:

The first aim of this chapter is to show that obesity, smoking, and binge drinking behaviors are correlated or tend to cluster, so their effects on wages may not be measured accurately in analyses that consider only one or two. Since the issue of this correlation from Chapter 5 has given important results, the same issue is carried over for wage analyses. Although previously the combined effects of obesity and smoking, and the combined effects of smoking and drinking have been analyzed in the economics literature, no author has considered the effects of all three health factors on labor market outcomes at the same time in the same analysis.

Cigarette smoking, excessive alcohol use, and poor eating habits tend to reinforce each other (Betts, 2000), and smokers are less likely to be obese than nonsmokers (Baum et al., 2006). Hence, failing to control for all health behaviors in the same wage effects analyses may lead to biased results.

Similar to the unemployment analysis, in order to examine the variations in the effects of these behaviors on wages, four different wage regressions are run using ordinary least square estimation technique (OLS). In the first three wage regression estimations, the health behavior variables are included on the right-hand side of the wage level model alone together with most commonly used standard covariates, and the last wage regression controls for all health behaviors at the same time. Table 6.1 shows the effects of these three behaviors on log wages for males and females separately.

Table 6.1. *The effects of obesity, daily smoking and binge drinking on log wages when different specifications are used*

	Males				Females			
OLS results	Model1	Model2	Model3	Model4	Model1	Model2	Model3	Model4
Obese	-0.023 (0.017)			-0.038** (0.016)	-0.119*** (0.018)			-0.128*** (0.018)
Daily smoker		-0.079*** (0.012)		-0.085*** (0.012)		-0.026** (0.012)		-0.032*** (0.011)
Binge Drinker			0.009 (0.012)	0.019* (0.011)			-0.032* (0.018)	-0.022 (0.019)
Adj R^2	*0.266*	*0.267*	*0.266*	*0.273*	*0.258*	*0.257*	*0.257*	*0.262*
F test with Model 1				F (2) = 22.12 P = 0.000				F (2) = 18.47 P = 0.000
F test with Model 2				F (2) = 33.94 P = 0.000				F (2) = 29.76 P = 0.000
F test with Model 3				F (2) = 11.23 P = 0.001				F (2) = 9.22 P = 0.001

Regression models contain 65,365 person-year observations for males and 59,899 person-year observations for females. Clustered robust standard errors are in parentheses. Sampling weights are controlled. *** Significant at the 0.01 level. ** Significant at the 0.05 level. * Significant at the 0.10 level. See Table 4.5 for other explanatory variables.

Table 6.1 demonstrates that when the health behaviors are considered individually, only daily smoking statistically significantly affects the wages of males (wage penalty is 7.9%), but obesity, daily smoking, and binge drinking significantly affect the wages of females (wage penalties are 11.9%, 2.6%, and 3.2%, respectively).

Table 6.1 also illustrates that the estimates noticeably change when these health factors are controlled for each other. Obesity, daily smoking, and binge drinking statistically significantly impact the wages of males and obesity and daily smoking significantly impact the wages of females after controlling for all behaviors in the same regression analyses. The results reveal that the wage effect of obesity is underestimated if one fails to control for the smoking and drinking behaviors of males. The insignificant coefficient of obesity decreases from -0.023 to -0.038 and becomes statistically significantly different from zero at the 0.05% level (P value=0.032; the change is statistically significant).

Moreover, the wage effects of binge drinking are underestimated for males (estimate is positive) and overestimated for females if alcohol use of the respondents is considered individually. The positive but statistically insignificant estimated coefficient of binge drinking increases and becomes statistically significant at the 10% level (P value=0.024, the change is significant) for males, and the statistically significant binge drinking wage penalty decreases and becomes insignificant for females (P value=0.029; the change is significant).

Although the significant estimated parameter of daily smoking noticeably increases when obesity and binge drinking health behaviors of the individuals are controlled in wage regressions for both males and females, the estimated coefficients are not statistically significantly different from each other (P values are 0.39 and 0.44 for males and females, respectively; the changes are insignificant).

To test my ideas, the first three models are compared to the fourth model by using F-tests. All F-test values are greater than or very close to 10, thus, the collective contribution of the health behaviors is statistically significant and the data gives evidence of a statistically significant departure from the first model to the full model; therefore all three health variables should be included in the same analysis. Moreover, adjusted R^2 is a commonly accepted measure of the explanatory power of the linear regression models and it quantifies goodness of fit. It can be claimed that the earlier arguments are also supported since adjusted R^2 increases each time when an additional health behavior is controlled for, for males and for females.

6.2. Endogeneity of the Health Behaviors:

As discussed in Chapter 4 and Chapter 5, the wage effects of obesity, smoking, and binge drinking are difficult to examine since these variables could be endogenous; they could be correlated with the regression error term in the wage equation. The wages may simultaneously affect weight, smoking, and drinking decisions of the individuals, for instance people may demand more alcohol, cigarettes, and food when they earn more. Alternatively, some outside factor may also explain the relationship between the wages and these behaviors, such as motivation, self-control, or being economically myopic.

The OLS estimates in wage analyses do not correctly characterize the effects of obesity, smoking or binge drinking since health behaviors may be endogenous. In order to limit the potential for heterogeneity bias or the issue of omitted variables and to control more widely for individual background characteristics that might affect labor market outcomes, similar to Chapter 5, the supplemental background variables are included in the wage level equations.[24]

When the supplemental background variables are included into the wage level equations, the estimated coefficients of the health behaviors change and they all decrease. However, the changes are not statistically significant (most P values are all greater than 0.10). The obesity wage penalties decrease for both males and females, from 3.8% to 3.2% and from 12.8% to 9.6% respectively (P values are 0.41 and 0.38; the changes are statistically insignificant) and their significance level stays the same. The estimated parameter of daily smoking decreases for both males and females (from -0.085 to -0.074 for males and from -0.032 to -0.023 for females). The estimate of daily smoking for females becomes statistically significant at the 5% level and the change is statistically significant (P value is 0.088). Although the wage penalty for binge drinking falls from 1.9% to 1.7% and stays statistically significant at the 10% level for males (P value is 0.51; the change is insignificant), the estimate of binge drinking is never significant for females. Therefore, these results show that the effect of daily smoking is subject to omitted variable bias for females.

OLS is often used to estimate the wage models but it does not allow one to control for the unobserved individual heterogeneity, so it may produce biased results due to a correlation between some unobservable characteristics and some explanatory variables.

[24] See Table A2 in Appendix for the results. As it is done in fifth chapter, to correct the possible specification error that might be created by employing numerous supplemental explanatory variables, the Stata command 'linktest' is used which detects if the model is not properly specified. To correct the possible multicollinearity problem, Stata command 'vif' is also employed. There seems to be no multicollinearity with health behaviors. To reduce the multicollinearity in age and tenure variables, age and years of tenure are centered by subtracting the means from their predictor values before generating the square terms.

In order to account for this unobserved heterogeneity, panel data techniques are also employed in this chapter.[25] The fixed-effects estimates of all health behaviors are smaller in magnitude than the cross-sectional estimates for both males and females. Moreover, the fixed-effects estimates of obesity become statistically insignificant for males and become statistically significant only at the 10% level for females. Further, the positive effect of binge drinking on wages for males disappears once fixed-effects or random-effects models are employed. The results reveal that the wage effect of obesity and the wage effect of binge drinking for males may be due to the unobserved time-invariant individual characteristics.

However, as it is previously noted, fixed-effects estimates must be reviewed carefully. Although the identification of the fixed-effects models is largely based on individuals who change their behavior, it was reported in earlier chapter that there is a lot of heterogeneity among individuals who have these health behaviors, especially smokers. Additionally, the model does not eliminate bias in case of time-variant individual heterogeneity and takes out all the covariates that are time-invariant. Further, the model may worsen the bias caused by measurement error.

Earlier studies have generally dealt with endogeneity problem with twin studies (Behrman and Rosenzweig, 2001), with a measure of previous body weight (Averett and Korenman, 1996), or mostly with the instrumental variables methods (Pagan and Davila, 1997; Cawley, 2004; Baum and Ford, 2006; Morris, 2006). A similar conventional

[25] The results from the fixed-effects, random-effects models and Hausman test results are displayed in Appendix, Tables A4 and A5. For both genders, the Hausman specification tests reject the null hypothesis that fixed and random-effects models produce the same coefficients, with P-values less than 0.001. Therefore, these suggest that the differences between these two models are systematic and only fixed-effects models should be taken into account.

instrumental variable model is also employed in this study: a two-stage equations approach in the OLS model by using the Stata command 'ivreg2'.[26] A set of instruments that are commonly used by other studies is employed in a conventional instrumental variable method: the religious attendance dummy in 1982, religious affiliation dummies (non-religious, Baptist, Catholic, Jewish), times charged with illegal activity in 1980, cumulative years lived with alcoholic relative dummy, alcoholic father (only for males) and alcoholic mother (only for females) variables. Furthermore, some panel data models (fixed effects and random effects models) in which obesity, smoking and heavy binge drinking are assumed to be endogenous are employed.[27]

The instrumental variable results for males show that when the instrumental variables approach is applied, the wage effects of most of the health behaviors increase. However, the estimate of binge drinking becomes statistically insignificant (P value is 0.044; the change is statistically significant), the estimated parameter of daily smoking falls and becomes statistically significant only at the 10% level (P value is 0.089; the change is significant), and the estimate of obesity stays statistically insignificant. For females, when the instrumental variables method is employed, wage effects of all health behaviors rise (the changes are not significant) but their significance levels stay the same.

[26] In the first stage, the predicted values are obtained from regressing obesity, daily smoking, binge drinking separately on all exogenous variables and the instrumental variables. In the second stage, these predicted values are added to the OLS model on wages, in place of the actual values. To test the correlation between the error term in the first stage and the error term in the second stage, a Wald test is employed. To test the validity of the estimates, Sargan–Hansen test, which tests the null hypothesis that all instruments are uncorrelated with the error term of the second-stage, is used in addition to the F-tests of the first stage equations.

[27] See Appendix, A7, for the results of conventional IV methods, panel-data techniques and comparison of the OLS methods. Hausman tests reject the null hypothesis that the differences between the coefficients of fixed and random effects models are not systematic, for both males and females. Therefore, only the IV for fixed effects model is studied.

When the panel data technique (fixed-effects) is employed together with IV model, the magnitudes of all of the estimated parameters fall and only the estimated coefficient of daily smoking for males and the estimated parameter of obesity for females are statistically significant. [28]

To account for endogeneity, the Hausman-Taylor Instrumental Variable (HTIV) method is also employed in wage analyses. The method appears the most appropriate for the purpose and sample in this study. In the presence of correlation between some covariates and the unobserved individual characteristics, this method produces consistent estimates contrary to the OLS model. Furthermore, in contrast to the conventional IV methods, there is no need to find external instruments. Additionally, in contrast to the fixed effects model, the HTIV model produces the estimates for time-invariant covariates.

Table 6.2 displays the results of the OLS and HTIV models for males and females separately. Table 6.2 illustrates that the wage effects of health behaviors are associated with endogeneity biases. When HTIV model is employed, the estimate of obesity decreases and becomes statistically insignificant for males (P value is 0.023; the change is statistically significant). Further, the positive wage effect of binge drinking for males

---

[28] There is a large literature claiming that failing to control for selection into working could lead to sample selection bias. For instance, individuals whose reservation wage is greater than the current wage they face may choose not to work (Heckman, 1979). To account for this bias, a two-step Heckman correction method is employed for males and females separately. Non-wage family income is served as an instrument for the propensity of the individual to work or receive positive wage. The Mills Ratio correction terms which are employed to control for selection into working are not statistically significant in the wage equation either for males or females. Furthermore, the correction does not change the sizes and significance levels of the three health behaviors. Moreover, instrumenting the health behaviors do not lead to any significant change in either the Mills Ratio term or in the estimated parameters of the health behaviors. Hence, the results with the Heckman correction are ignored and not presented.

becomes negative and statistically insignificant (P value is 0.008; the change is statistically significant). Although the wage effect of obesity for females and wage effects of daily smoking for males and females all noticeably change, the changes are not statistically significantly different from zero (P values are all above 0.10).

Table 6.2. *The estimates of OLS and HTIV models of the effects of obesity, smoking, and binge drinking on log wages, for males and females separately*

	Males		Females	
	OLS	HTIV	OLS	HTIV
Obese	-0.032 ***	-0.017	-0.096 ***	-0.043 ***
	(0.014)	(0.012)	(0.016)	(0.013)
Daily smoker	-0.074 ***	-0.048 ***	-0.023 **	-0.025 **
	(0.011)	(0.009)	(0.011)	(0.011)
Binge Drinker	0.017 *	-0.006	-0.007	-0.007
	(0.010)	(0.007)	(0.016)	(0.013)

Regression models contain 65,365 person-year observations for males and 59,899 person-year observations for females. Clustered robust standard errors are in parentheses. Sampling weights are controlled. *** Significant at the 0.01 level. ** Significant at the 0.05 level. * Significant at the 0.10 level. See Table 4.6 for other explanatory variables.

Hence, once endogeneity is accounted for, obesity has negative effects only on the wages of females (a penalty of 4.3%), smoking wage penalties are 4.8% and 2.5% for males and females, respectively, and binge drinking has no effect on wages for either gender.

Table 6.3 shows the results of the HTIV method when three health behaviors are considered individually and simultaneously for males and females separately. For both genders, Model 4 includes all the health behaviors in the same wage regression estimation.

Table 6.3. *The estimates of HTIV models of the effects of obesity, daily smoking and binge drinking on log wages when different specifications are used*

	Males				Females			
HTIV results	Model1	Model2	Model3	Model4	Model1	Model2	Model3	Model4
Obese	-0.012 (0.015)			-0.017 (0.012)	-0.031** (0.014)			-0.043*** (0.013)
Daily smoker		-0.045*** (0.009)		-0.048*** (0.009)		-0.024** (0.011)		-0.025** (0.011)
Binge Drinker			-0.010 (0.007)	-0.006 (0.007)			-0.013 (0.012)	-0.007 (0.013)

Regression models contain 65,365 person-year observations for males and 59,899 person-year observations for females. Clustered robust standard errors are in parentheses. Sampling weights are controlled. *** Significant at the 0.01 level. ** Significant at the 0.05 level. * Significant at the 0.10 level. See Table 4.6 for other variables.

The table illustrates that, similar to the unemployment analyses, even though the estimated parameters of health behaviors visibly change, the changes are not statistically significantly different from zero. The wage effect of obesity for males and the wage effect of binge drinking for males and females are all statistically insignificant even after accounting for endogeneity (P values are greater than 0.10; the changes are insignificant). The estimates of daily smoking only slightly increase from -0.045 to -0.048 for males and from -0.024 to -0.025 for females. The wage effect of obesity rises from 3.1% to 4.3% for females but the change is insignificant (P value is 0.19).

Therefore, the results reveal that once endogeneity is accounted for via HTIV model, the estimated parameters of the health behaviors are not statistically significantly different whether these three behaviors are considered individually or simultaneously. Therefore, the biases in the OLS estimates are due to OLS's failure to account for endogeneity.

6.3. Sensitivity and Robustness Analyses:

6.3.1. Interactive/Additive Characteristics of Health Behaviors:

To support the results of wage analyses, interactive and additive characteristics of health behaviors are also examined. Previously, it has been assumed that the impact of one health behavior is the same regardless of another health behavior. The health factors could be additive, for instance if health worsens when an obese individual is also smokes; hence the wages may be decreased proportionally by poorer health. However, the health behaviors would not be additive if for example, they are due to the same unobserved factors (Baum et al., 2006).

With the aim of investigating whether the effects of these behaviors are additive or interactive, four interaction terms are added into the wage and the unemployment regressions: obese*daily smoker, obese*binge drinker, daily smoker*binge drinker, and obese*daily smoker*binge drinker.[29] These interaction terms will be statistically significant if dual or triple interactions exist. If the interaction terms are statistically insignificant, then one could conclude that the behaviors are not interactive; their total effects could be found by just adding the coefficients together.

Table 6.4 shows the effects of the health factors and their interactions on wages accounting for endogeneity by using the HTIV.

[29] Although no formal presence of multicollinearity is found with the health risk factors and their interactions, to reduce the unobserved potential multicollinearity, the health risk factor variables are centered by subtracting the means from their predictor values before generating the interaction terms. The results of the models of centered and non-centered variables are mostly the same, therefore only the results of the non-centered variables are displayed.

Table 6.4. *The estimates of HTIV models of the effects of obesity, smoking and binge drinking and their interactions on log wages*

HTIV Results:	Males	Females
Obese	-0.010 (0.014)	-0.051 *** (0.015)
Daily smoker	-0.051 *** (0.010)	-0.031 *** (0.011)
Binge Drinker	-0.017 * (0.010)	-0.034 (0.021)
Obese * Smoker	0.039 (0.025)	0.005 (0.057)
Obese * Binge Drinker	0.014 (0.015)	0.037 (0.027)
Smoker * Binge Drinker	-0.001 (0.022)	0.023 (0.023)
Obese * Smoker * Binge Drinker	0.001 (0.041)	0.070 (0.077)

Regression models of wage analyses for males contain 65,365 person-year observations, for females contain 59,899 person-year observations. People who reported their employment status as out of labor force are excluded from employment analyses. Marginal effects are in brackets.

The results illustrate that while nearly all of the estimates of the health behaviors in each model are similar to the results of earlier analyses without the interaction terms, none of the interaction term appears to be statistically significant in any model. Therefore, one may conclude that obesity, smoking, and binge drinking have no additional interactive effects; these behaviors do not bring any additional interaction to the overall multiplication model since the wage level models are in log-linear format and behaviors already have some interaction due to the exponential terms in the wage level model.

One variation in the results attracts attention. Similar to the unemployment analyses, when interaction terms are included into the wage models, binge drinking estimates for males become statistically significant and it becomes only marginally insignificant for females (P value is 0.108), which demonstrates the importance of including interaction terms. The results show that even though the interaction terms are insignificant, when all are added into the wage model, binge drinking leads to a negative effect on wages for males in HTIV model. In other words, if an individual is neither obese nor a smoker, being a binge drinker results in a 1.7% wage penalty. Therefore, the earlier positive effect of binge drinking on wages in OLS and insignificant effect in HTIV models could be related to the effects of obesity and/or smoking, for instance, if an obese or smoker already perceives a wage penalty due to obesity and/or smoking, deciding to be binge drinker may not result in any additional wage penalty.[30]

6.3.2. Wage Comparisons:

The impacts of three health factors on wages might vary across the wage structure with higher wage workers less likely to be affected, possibly because of the nature of their jobs, than lower wage workers or workers with less education, or vice versa. Partitioning the dependent variable, log wages, into different groups, e.g. high wage group, middle wage group, and low wage group, and regressing log wages on explanatory variables for these specific groups could produce incorrect results. Referring to this problem, Heckman (1979) argues that the method of truncation is open to selection bias

[30] Moreover, the likelihood ratio test comparing the two models with and without the interaction terms rejects the null hypothesis that removing the interaction terms has no effect and does not lead to a poorer-fitting model. Moreover, R^2 increases by more than 0.02 once the interaction terms are included into the wage regression, and goodness of fit test and the Stata command ‘linktest’ that is used to detect a specification error work better with the model including interaction terms. These findings suggest controlling the interaction terms while examining the impacts of health behaviors, especially binge drinking, on wages and a more detailed analysis is required for a clear assessment.

and produces both biased and inconsistent estimates, since one is truncating the full sample based on the dependent variable in the model.

Koenker and Hallock (2001) state that a more suitable econometric technique is quantile regression, in which one can focus on the conditional distribution of the dependent variable and avoid the selection bias associated with truncated regression. This is because with a quantile regression, one can choose the central tendency point around which to estimate a regression, for example, 25th decile rather than the mean without truncating the sample to exclude the upper 75% of data. OLS models the relationship between one or more covariates X and the conditional mean of a dependent variable Y given $X = x$. On the contrary, quantile regression models the relationship between X and the conditional quantiles of the dependent variable Y given $X = x$, so it is especially useful in applications where extremes are important or when both lower and upper or all quantiles are of interest, as in this study.

The problem of whether the impact of three health factors on wages might vary across the wage structure is addressed by comparing estimates for various quantiles.[31] Table 6.5 displays the differences in the effects of three health behaviors on log wages for males and females at different points of the conditional wage distribution using simultaneous-quantile regression. The quantiles are chosen as 0.10, 0.25, 0.50, 0.75 and 0.90. Standard errors of the quantile regressions are estimated by bootstrapping with 500

[31] In general the Stata command ‘qreg’ is used to fit quantile regression models. In this study, another Stata command ‘sqreg’ is employed which estimates the simultaneous-quantile regression. It produces the same coefficients as ‘qreg’ would produce for each quantile, but it can estimate the results for multiple quantiles simultaneously, thus it is possible to test the coefficients describing different quantiles. However, standard quantile regressions do not present estimates for the variances of the differences in the coefficients of separately estimated quantile regressions. Such estimates are attained by bootstrapping.

replications. Nevertheless, simultaneous-quantile regression does not allow controlling for either sampling weights or clustering. The equality of the coefficients across quantiles is tested with Wald tests.

Table 6.5. *The estimates of Simultaneous Quantiles Regression model of the impact of obesity, smoking and binge drinking on log wages, for different quantiles separately, for males and females*

Males	10th Quantile	25th Quantile	50th Quantile	75th Quantile	90th Quantile
Obese	-0.004 (0.016)	-0.026 ** (0.011)	-0.020 *** (0.007)	-0.008 (0.009)	-0.015 (0.011)
Daily smoker	-0.055 *** (0.010)	-0.052 *** (0.005)	-0.050 *** (0.004)	-0.050 *** (0.004)	-0.052 *** (0.011)
Binge Drinker	0.006 (0.013)	0.009 (0.007)	0.009 (0.006)	0.004 (0.007)	0.001 (0.008)
Females	**10th Quantile**	**25th Quantile**	**50th Quantile**	**75th Quantile**	**90th Quantile**
Obese	-0.091 *** (0.018)	-0.076 *** (0.010)	-0.071 *** (0.008)	-0.059 *** (0.006)	-0.055 *** (0.012)
Daily smoker	-0.010 (0.010)	-0.031 *** (0.006)	-0.027 *** (0.005)	-0.025 *** (0.006)	-0.024 *** (0.005)
Binge Drinker	-0.003 (0.031)	0.013 (0.017)	-0.001 (0.006)	-0.020 ** (0.009)	-0.010 (0.011)

Regression models contain 65,365 person-year observations for males and 59,899 person-year observations for females. Standard errors with bootstrapping with 500 replications are in parentheses.

The table reveals that the wage penalties for daily smoking are fairly constant over the wage distribution for both genders. The wage penalties for smoking range from 5.5% to 5.2% for males whereas the wage penalties range from 3.1% to 2.4% for females.[32]

The table illustrates that obesity affects the wages of females relatively more at lower quantiles. The wage penalty for obesity is around 9.1% at the 10th quantile, and the

[32] Hausman tests that compares estimated coefficients across different quantiles (except the 10th quantile for females) fail to reject the null hypothesis that the wage penalties are the same. Moreover, a Hausman test also fails to reject the hypothesis that the full set of daily smoking coefficients are equal at the five quantiles estimated for males and at the four quantiles estimated (except the 10th quantile) for females.

penalty decreases to 7.1% at the 50th quantile, and becomes 5.5% at the 90th quantile.[33] However, obesity affects the wages of males only at lower quantiles. Although the impact of obesity at the 10th percentile is statistically insignificant, the wage penalty for obesity is 2.6% at the 25th percentile, it is 2% at the 50th percentile, and there seems to be no wage penalty either at the 75th percentile or at the 90th percentile for males.[34]

The results demonstrate that there is no wage penalty for being a binge drinker for males at any quantile. The same conclusion is true for females, except the wage penalty of 2% at the 75th quantile. The insignificant wage effect of obesity at the 10th quantile for males and insignificant wage effect of daily smoking for females at the 10th quantile could possibly be due to the unstable working characteristics of the individuals that they might be working in different jobs as part-time or it could be because they may be working in jobs that offer low wages but less or no discrimination against obese or smoker in hiring or firing conditions. These results could also be because of the quirk of the sample.

6.3.3. More Detailed Measures of Health Behaviors:

The effects of some additional measures of these three health behaviors and their interactions on the wages are analyzed in order to observe whether they show consistent

[33] A Hausman test that tests the null hypothesis that all obesity coefficients are equal at the five quantiles rejects the null hypothesis, thus obesity wage penalties vary with wage structure with lower wage workers more likely to be affected than higher wage workers.
[34] Hausman test rejects the null hypothesis that the full set of obesity coefficients are equal at the last four quantiles estimated for males. Therefore, the impact of obesity varies with wage structure for males as well; obesity affects wages for males relatively more at the lower quantiles.

results with the earlier findings to provide several sensitivity and robustness checks. The results are displayed in Table 6.6.[35]

Table 6.6. *The estimates of HTIV model of the effects of more detailed measures of health behaviors on log wages, for males and females separately*

HTIV Results:	Males	Females
Smoker	-0.059 ***	-0.031 ***
	(0.008)	(0.009)
Quitter	0.031	-0.007
	(0.030)	(0.031)
Starter	-0.063 *	-0.024
	(0.032)	(0.030)
Young Experimenter	-0.019	-0.022
	(0.026)	(0.023)
Unsuccessful Quitter	-0.051 **	-0.014
	(0.021)	(0.021)
Mild Obese	0.007	-0.029 **
	(0.013)	(0.014)
Morbid Obese	-0.029 **	-0.075 ***
	(0.015)	(0.020)
Light Smoker	-0.061 ***	-0.018 *
	(0.010)	(0.011)
Heavy Smoker	-0.031 ***	-0.043 ***
	(0.011)	(0.014)
Current Drinker	0.007	0.013
	(0.006)	(0.009)
Heavy Drinker	0.005	0.004
	(0.006)	(0.007)

Regression models of wage analyses for males contain 65,365 person-year observations, for females contain 59,899 person-year observations. People who reported their employment status as out of labor force are excluded from employment analyses.

[35] In order not to fall into overlapping problem, light and heavy smoking variables and heavy drinking variable are analyzed in different regressions than other smoking and drinking variables. Interaction terms are included in two additional different regressions and but the results are not displayed since the coefficients are largely unchanged.

The obesity dummy variable is divided into two dummies: 'morbid obesity' is defined as having a BMI of 35 or more and 'mild obesity' is defined as having a BMI of between 30 and 35. 'Smoker' is defined as being a daily smoker in all years when smoking questions are asked. A 'young experimenter' is an individual who smokes only in the 1984 survey year, when smoking questions are asked for the first time, and then quits smoking in later survey years. A 'heavy smoker' is defined as smoking more than 20 cigarettes per day and a 'light smoker' is someone who smokes at least one but less than 20 cigarettes per day. The 'current drinker' dummy variable is equal to one if the respondent reported drinking any alcoholic beverage in the previous month. 'Heavy drinking' is defined as drinking 6 or more drinks in one to three occasions in the previous month.

The table demonstrates very similar results to the results of the unemployment analyses. The results show that the wage effects of permanent smokers, quitters and unsuccessful quitters are statistically significant for males and greater than the wage effects of quitters, young experimenters and nonsmokers. For females, only permanent smoking has a negative impact on wages. Therefore, these findings reveal that smokers are a heterogeneous group of people: the wage effects of smoking groups are different.[36]

The results of HTIV models show that obesity affects wages of only females; it has no impact on the wages of males. However, table 6.6 shows that mild and morbid obesity leads to wage penalties for females, and morbid obesity appears to affect the

---

[36] As it is discussed in fifth chapter, the heterogeneity among smokers explains why the fixed-effects results might be biased. The results show that those who change their behavior over time, e.g. quitters, smokers, young experimenters, are different from the ones who do not change, and their effects are different on wages. However, the identification of fixed-effects models is based on the individuals who change their behavior.

wages males (the wage penalty of being morbidly obese is 2.9% for males). Hence, obesity affects the wages of males only at the extremes of obesity.

Earlier studies and the OLS results of this study demonstrate that current drinking and heavy drinking have positive effects on wages for both males and females. However, these effects turn out to be insignificant in HTIV models when endogeneity is accounted for, thus the OLS results and the results of the previous literature are biased; these positive effects are due to the failure of OLS accounting for endogeneity.[37]

This study uses several retrospective measures and questions to create health behavior variables for the years missing health behavior questions in the survey. Kenkel et al. (2003) state that contemporaneous and retrospective smoking status variables do not match very precisely in the NLSY79 data. Hence, to minimize the measurement error due to imputations and problems with retrospective measures, besides using additional measures of health behaviors, this study runs alternative regression estimations to reveal that the results are robust to the choice of specification. For this reason, wage and unemployment regressions are re-run using only the survey years of 1984, 1992, and 1994 in which the questions about all health behaviors are asked to the respondents.[38]

The results of HTIV model for wage analysis and multivariate probit model for unemployment analysis show that the estimated coefficients of health behaviors in

[37] One last additional variable is drinks per week variable (the results are not shown in the table). Six dummy variables are created related to drinks per week: 0-2 drinks per week, 2-5 drinks per week, 5-10 drinks per week, 10-20 drinks per week, 20-30 drinks per week and 30-more drinks per week. OLS results show that drinkers who drink less than 2 drinks earn 3.4% more, respondents who drink between 2 and 5 drinks earn 5.7% more, and drinkers who drink more than 30 drinks per week earn 7.3% less than nondrinkers. The coefficients of other dummy variables are statistically insignificant. However, the coefficients of all dummy variables turn out to be statistically insignificant when HTIV model is employed to account for endogeneity.

[38] See Appendix, Table A8, to see the results of the robustness analysis.

regressions using all survey years are not statistically significantly different from the estimates of the behaviors in robustness regressions that use only 1984, 1992, and 1998 survey years (P values are all greater than 0.10). Thus, previous results of the study are robust and retrospective questions do not lead to serious bias.

6.3.4. Ethic/Racial Differences:

To conduct a richer examination of ethnic/racial differences of health behavior wage penalties, I separated whites, blacks and Hispanics into subsamples and examined the wage effects separately. For each demographic group, the impacts are analyzed using HTIV models. Table 6.7 shows the estimates of obesity, smoking, and binge drinking for white, black and Hispanic males and females, separately.

Table 6.7. *The estimates of Hausman-Taylor IV models of the impact of obesity, smoking, and binge drinking on log wages, for whites , blacks and Hispanics separately*

	Males			Females		
HTIV Results:	Whites	Blacks	Hispanics	Whites	Blacks	Hispanics
Obese	0.005 (0.016)	0.004 (0.023)	-0.001 (0.027)	-0.062 *** (0.018)	-0.024 (0.025)	-0.017 (0.031)
Daily smoker	-0.041 *** (0.012)	-0.070 *** (0.018)	-0.037 * (0.020)	-0.031 ** (0.013)	-0.006 (0.022)	-0.033 (0.027)
Binge Drinker	-0.018 ** (0.008)	0.002 (0.016)	0.025 (0.015)	-0.018 (0.015)	0.022 (0.032)	-0.008 (0.035)

Regression model for white males contains 33,103 person-year observations, model for black males contains 15,595 person-year observations and model for Hispanic males contains 10,616 person-year observations. Regression model for white females contains 30,652 person-year observations, model for black females contains 13,419 person-year observations and model for Hispanic females contains 8,595 person-year observations.

The table reveals that smoking has wage penalties for all subsamples of males but affects the wages of only white females. Further, there appears to be an obesity wage penalty only for white females, and surprisingly a binge drinking wage penalty only for white males.

However, the results must be assessed with caution since the person-year observations of blacks and Hispanics are relatively smaller than whites in the sample. The person-year observations of black males and Hispanic males are only around 12% and 8% of the whole sample respectively, and the person-year observations of black females and Hispanic females only 11% and 7% of the whole sample. As a result, the standard errors are high and the estimates might be doubtful.

6.3.5. Public-Private Sector Differences:

The effects of obesity, smoking, and binge drinking on wages are also examined for private and public sector workers separately. In the public sector, wages are generally predetermined for all groups of workers, thus the sector is a relatively compact and fixed wage sector. On the contrary, wages of the private sector workers are determined by the employer, so workers are more exposed to wage discrimination in a private sector. Moreover, obese, smoker, or binge drinker workers who faced discrimination in a private sector may move to public sector where the wage structure is fixed.

Descriptive statistics illustrate that the average BMI of male (female) workers working in a private sector is 25.1 (23.7) while the average BMI in public sector is 26.9 (25.4). However, the statistics do not show any statistically significant difference of average smokers or average heavy binge drinkers between the private and public sectors.

Table 6.8 displays the estimates of HTIV model of the three health behaviors on wages for males and females working in a private or public sector, separately.

Table 6.8. *The estimates of HTIV models of the effects of obesity, smoking and binge drinking on log wages, for public and private sector workers*

	Males		Females	
HTIV Results:	Private	Public	Private	Public
Obese	0.021	0.016	-0.036 **	-0.006
	(0.015)	(0.025)	(0.017)	(0.026)
Daily smoker	-0.045 ***	-0.039 *	-0.036 ***	-0.031
	(0.012)	(0.021)	(0.013)	(0.026)
Binge Drinker	-0.008	0.037	-0.020	0.005
	(0.008)	(0.023)	(0.015)	(0.040)

Regression model for private employment contains 39,809 person-year observations and model for public employment contains 11,490 person-year observations for males. Regression model for private employment contains 32,549 person-year observations and model for public employment contains 10,562 person-year observations for females.

The table shows evidence of wage penalties for being obese or a smoker for females and a smoking wage penalty for males in private sector jobs. However, in the public sector, only males face lower wages only due to smoking, and the estimate of smoking is statistically significant only at the 10% level, which lets one argue that private sector jobs are more prone to wage penalties than public sector jobs.

CHAPTER 7

Conclusions and Discussion

This dissertation aims to find the joint effects of obesity, smoking, and binge drinking on labor market outcomes, using the National Longitudinal Survey of Youth data. The main objective of this study is to show that the effects of obesity, smoking, and binge drinking on wages or on unemployment may not be measured accurately in analyses that consider only one or two health behaviors since these behaviors are correlated or tend to cluster.

The OLS results of wage analyses reveal that the wage effect of obesity is underestimated for males and the wage effect of binge drinking is underestimated for males (estimate is positive) and overestimated for females if one fails to control for other health behaviors in the same analyses. Similarly, the probit results of unemployment analyses illustrate that failing to include one or more of the health behaviors in unemployment regressions would lead to an underestimation of the impact of being obese and an overestimation of the effect of binge drinking for both genders.

However, the OLS and probit estimates do not correctly characterize the effects of obesity, smoking, or binge drinking since these behaviors may be endogenous. I address the potential endogeneity by employing the Hausman-Taylor instrumental variable (HTIV) method in wage analyses and the multivariate probit method in unemployment analyses. The results illustrate that the wage and unemployment effects of the behaviors are subject to endogeneity biases. Further, when endogeneity is controlled for, the estimated parameters of the health behaviors are not statistically significantly different whether these behaviors are considered individually or all together.

The Hausman-Taylor results show that obesity, defined as a body mass index of 30 or higher, has negative effects only on the wages of females (a penalty of 4.3%), and smoking wage penalties are 4.8% and 2.5% for males and females, respectively. Being obese increases the probability of being unemployed by 1.8% for females only, and smoking increases the likelihood of being unemployed by 2.5% and 1.8% for males and females, respectively. However, binge drinking appears to have no effect on wages or unemployment for either gender.

I have also conducted several sensitivity and robustness analyses on the labor market effects of obesity, smoking, and binge drinking. Firstly, this study examines whether the effects of the three health behaviors on wages or unemployment are interactive. None of the interaction terms are statistically significantly different from zero; hence the effects of the behaviors are not interactive. However, when interaction terms are included into the models, binge drinking estimates for males become statistically significant both in wage and unemployment analyses which shows the importance of including interaction terms in regressions. Hence, binge drinking negatively affects the wage of a male or probability of being unemployed if he is neither obese nor a smoker.

I examined whether the impacts of these behaviors on wages might vary across the wage structure using quantile regressions. The results reveal that the wage penalties for daily smoking are fairly constant over the wage distribution for both genders.

However, obesity affects the wages of males and females relatively more at lower quantiles, and there is no wage penalty for being a binge drinker for males and females at any quantile.

I also estimated wage and unemployment models using more detailed measures. When endogeneity is addressed, the results demonstrate that smokers are a heterogeneous group of people: the wage and unemployment effects of persistent smokers are different than starters, quitters or young experimenters. Furthermore, obesity appears to affect the wages and the likelihood of being unemployed of males only at the extremes: men who are morbidly obese (body mass index 35 or higher) suffer a wage penalty of 2.9% and are more likely to be unemployed. Contrary to the previous literature, drinking is found to have no effect on wages or on unemployment for either gender when endogeneity is accounted for.

I separated whites, blacks, and Hispanics into subsamples to examine ethnic/racial differences. I found that, although smoking has wage penalties for all subsamples, obesity affects the wages of only white females, and binge drinking affects wages of only white males. However, the results must be assessed with caution since the person-year observations of blacks and Hispanics are relatively smaller than whites in the sample.

Lastly, health behaviors may affect the wages of individuals working in private sector more than those working in the public sector. Because public sector is a relatively compact and fixed wage sector, but wages of the private sector workers are determined by the employer, so workers are more exposed to wage discrimination in a private sector.

I find evidence of wage penalties for being obese or a smoker in private sector jobs. However, in the public sector, only males face lower wages only due to smoking.

My results draw attention to methodological issues involved in estimating relationships between health behaviors and labor market outcomes, especially the need to account for unobserved heterogeneity or endogeneity. Failure to consider the endogeneity or unobserved factors that affect wages/unemployment and behaviors could lead to incorrect inferences, and I argue that the results of most of the earlier literature suffer from these problems.

My results are similar to Register and Williams (1990), Pagan and Davila (1997) and Baum et al. (2006), who also find obesity wage penalties for females but not for males. However, my results differ from Cawley (2004), who finds significant negative wage effects of weight for males. The results are also parallel to Leigh and Berger (1989), Viscusi and Hersh (2001), Auld (1998) and Levine et al. (1997), who find a wage penalty for smokers in both genders. But the results are different from a more recent study by Baum and Ford (2006), who fail to observe significant negative effects of smoking for males or females after controlling for obesity in the same analysis and after addressing the unobserved heterogeneity.

The literature of the wage effects of drinking has contrasting results due to the use of different drinking measures and a failure to control for endogeneity. My results diverge from the results of most of the earlier literature, such as Kenkel and Ribar (1994), Mullahy and Sindlear (1993), French and Zarkin (1995), etc., who find negative effects of excessive alcohol consumption on earnings. But, similar to Peters' more recent results, (2004), I find that the effect of binge drinking disappears after the unobserved heterogeneity is addressed.

The results of this dissertation are also similar to Sousa (2005), Garcia and Quintana-Domeque (2007), Morris (2007), and Greeve (2008), who find significant negative employment effects of obesity, and Kenkel and Ribar (1994), and Terza (2002) who assert that there is no effect of problem drinking on employment. However, my results differ from Sargent and Blanchflower (1994), Kenkel and Ribar (1994), Sarlio-Lahteenkorva and Lahelma (1999), and Cawley (2000a), who do not find any effect of obesity on unemployment, and Mullahy and Sindelar (1996), MacDonald and Shields (2004), and Tekin (2004) who find negative effects of alcoholism on employment.

My sample, measures and methods differ from those used by some of these authors. Firstly, none of the studies have used all three behaviors in the same regression analysis. Secondly, I used all years of the NLSY data set and did not rely only on early years of the longitudinal data. Additionally, I used panel data techniques to control for unobserved heterogeneity and instrumental variable methods to account for endogeneity of health behaviors. Lastly, the results may differ in the measures of employment/wages and health behaviors used, since the choice of the measures is usually dictated by data availability.

This study has some limitations. For instance, one extension would be to use semi-parametric estimators to avoid making strong functional form and/or distributional assumptions. Furthermore, relying on self-reported alcohol or cigarette consumption may lead to biases since respondents may give inaccurate or biased information. Also, defining obesity as having BMI of 30 or more could result in a bias since BMI is a poor predictor of body fat. I plan to re-define these health behaviors and re-examine the labor market effects using these new measures.

The findings have direct implications for policy:

Firstly, recently some health and public policy studies argue that health or public policy campaigns must target multiple lifestyle habits or risk behaviors instead of just one. It is stated that it has been the prevailing attitude to address one habit at a time, but those programs have met with limited success, because these behaviors are linked to each other and individuals may have difficulty getting rid of one risk behavior or habit as long as they are attached to the other behaviors. Hence, to minimize the health, labor or economic costs of these behaviors, campaigns must target multiple behaviors.

The results of this study state that estimates of health behaviors on labor market outcomes are not different whether they are considered individually or all together after controlling for endogeneity; the effects of these behaviors are not linked. Therefore, considering health behaviors individually on the effects of wages or unemployment does not lead to any biased results. One may also assert that the campaigns that aim to lessen the negative labor market effects of health behaviors can target single behaviors without leading to any incorrect results.

Secondly, the study suggests that policies that aim to decrease the negative economic and social effects of health behaviors must focus more on smoking. The methodological analysis in this dissertation shows that results of earlier studies may suffer from endogeneity. After accounting for endogeneity, this study demonstrates that out of the three health behaviors, smoking has the highest negative impact on wages and unemployment. Moreover, I found that the wage effect of smoking is fairly constant over the wage distribution; individuals at different wage levels face similar wage penalties due to smoking. Further, the smoking wage penalty appears both in public and private sector jobs. Therefore, all these results suggest the importance of smoking prevention policies or policies that aim to reduce the negative effects of smoking on labor market outcomes, e.g. preventing discrimination against smokers.

Thirdly, this study shows the possible effects of obesity prevention policies, education of low income people about the negative social and economic effects of obesity, and discrimination against obese individuals, i.e. especially obese females, prevention policies. As it is found in this study, obesity affects the wages and unemployment of females. Obese women face wage penalties in private sector jobs. Furthermore, it is shown that obesity effects appear only at the extremes of obesity for males, but the penalty for women is much larger in the same range. Additionally, this dissertation reveals that obesity wage penalties vary with wage structure; obesity affects the wages more at the lower quantiles of wage distribution, so low income people are more prone to the obesity wage penalties. These results offer supports to policies to educate specific populations, i.e. low income people, regarding food choices and negative effects of obesity. The results also support policies designed to reduce discrimination against obese individuals, particularly obese women. Higher taxes on unhealthy foods and beverages, and regulation of certain food advertising and fast foods, especially in districts where average household income is lower, would also be consistent with these results.

Lastly, this dissertation argues that results of the earlier literature may suffer from endogeneity in the form of simultaneity and unobserved heterogeneity. After accounting for endogeneity, it is found that obesity affects the labor market outcomes of only females, smoking affects the wages and unemployment of both genders, but drinking, whether moderate or binge drinking, has no impact on labor market outcomes. Thus, economists and health and public policy analysts may use these results about health behaviors to re-formulate and re-organize the policies, and insurers may use these results on health behaviors to better predict utilization and to adjust pricing of insurance accordingly.

The current study can also be expanded to give more insight for policy makers. For instance, I plan to combine the results of wage and unemployment effects of these three health behaviors to examine their total effects on life-time earnings of the individuals. Also, I want to focus on the impacts of health behaviors on spells of unemployment using duration regression specifications. In one line of work, I plan to analyze whether changes in these behaviors over time may lead to different results. For example, it may be examined to see if people who lose/gain weight, or who currently smoke, or who currently drink experience different outcomes from those who have or do not have those behaviors consistently throughout the time period. In another line of work, I plan to expand this work by considering additional outcomes such as educational attainment, occupational distribution or total number of jobs of the individuals.

Appendix

Table A.1. *Probit estimates of the effects of obesity, smoking and binge drinking on unemployment, with standard covariates only, and with standard and supplemental covariates*

	Males		Females	
Probit results:	Standard	Standard + Supplemental	Standard	Standard + Supplemental
Obese	0.078 *	0.118 ***	0.073 *	0.026
	(0.041)	(0.039)	(0.042)	(0.038)
	[0.009]	[0.010]	[0.009]	[0.002]
Daily smoker	0.217 ***	0.115 ***	0.187 ***	0.102 ***
	(0.027)	(0.026)	(0.028)	(0.027)
	[0.025]	[0.009]	[0.024]	[0.010]
Binge Drinker	0.024	0.013	0.085 *	0.009
	(0.026)	(0.026)	(0.048)	(0.042)
	[0.002]	[0.001]	[0.011]	[0.001]

Regression models contain 64,712 person-year observations for males and 54,321 person-year observations for females. People who reported their employment status as out of labor force are excluded from the analyses. Clustered robust standard errors are in parentheses. Marginal effects are in brackets. Sampling weights are controlled. *** Significant at the 0.01 level. ** Significant at the 0.05 level. * Significant at the 0.10 level. See Table 4.5 and Table 4.6 for standard and supplementary covariates.

Table A.2. *The OLS estimates of the effects of obesity, smoking and binge drinking on log wages, with standard covariates only, and with standard and supplemental covariates*

	Males		Females	
OLS results:	Standard	Standard + Supplemental	Standard	Standard + Supplemental
Obese	-0.038 **	-0.032 **	-0.128 ***	-0.096 ***
	(0.016)	(0.014)	(0.018)	(0.016)
Daily smoker	-0.085 ***	-0.074 ***	-0.032 ***	-0.023 **
	(0.012)	(0.011)	(0.011)	(0.011)
Binge Drinker	0.019 *	0.017 *	-0.022	-0.007
	(0.011)	(0.010)	(0.020)	(0.016)

Regression models contain 65,365 person-year observations for males and 59,899 person-year observations for females. Clustered robust standard errors are in parentheses. Sampling weights are controlled. *** Significant at the 0.01 level. ** Significant at the 0.05 level. * Significant at the 0.10 level. See Table 4.5 and Table 4.6 for standard and supplementary covariates.

Table A.3. *Fixed effects and random effects estimates of the impact of obesity, smoking and drinking measures on unemployment*

	Male			Female		
Logit results:	Logit	Fixed	Random	Logit	Fixed	Random
Obese	1.246 ***	1.052	1.135 **	1.045	0.877	1.018
	(0.094)	(0.098)	(0.067)	(0.073)	(0.084)	(0.060)
Daily smoker	1.238 ***	1.061	1.242 ***	1.199 ***	1.163 **	1.206 ***
	(0.058)	(0.068)	(0.046)	(0.059)	(0.090)	(0.049)
Binge Drinker	1.031	1.095 **	1.073*	1.021	1.138	1.030
	(0.049)	(0.052)	(0.041)	(0.078)	(0.099)	(0.070)

Regression models contain 64,712 male and 54,321 female person-year observations. People who reported their employment status as out of labor force are excluded from the analyses. Clustered robust standard errors are in parentheses. Sampling weights are controlled. *** Significant at the 0.01 level. **Significant at the 0.05 level. *Significant at the 0.10 level. See Table 4.6 for the list of all explanatory covariates.

Table A.4. *Fixed effects, between effects and random effects estimates of the impact of obesity, smoking and drinking measures on log wages, for "males" only*

Males:	OLS	Fixed Effects	Between Effects	Random Effects
Obese	-0.022	-0.013	-0.004	0.010
	(0.014)	(0.015)	(0.019)	(0.012)
Daily smoker	-0.074 ***	-0.037 ***	-0.058 ***	-0.050 ***
	(0.011)	(0.012)	(0.013)	(0.009)
Binge Drinker	0.017 *	0.006	0.042 **	-0.001
	(0.010)	(0.009)	(0.018)	(0.008)

Regression models for males contain 65,365 person-year observations. Standard errors are in parentheses for between effects. Clustered robust standard errors are in parentheses for OLS, fixed effects and random effects. Sampling weights are controlled in OLS. *** Significant at the 0.01 level. ** Significant at the 0.05 level. * Significant at the 0.10 level. See Table 4.6 for the list of explanatory variables.

Table A.5. *Fixed effects, between effects and random effects estimates of the impact of obesity, smoking and drinking measures on log wages, for "females" only*

Females:	OLS	Fixed Effects	Between Effects	Random Effects
Obese	-0.096 ***	-0.031 *	-0.076 ***	-0.057 ***
	(0.016)	(0.017)	(0.020)	(0.013)
Daily smoker	-0.023 **	-0.013	-0.005	-0.017 *
	(0.011)	(0.013)	(0.013)	(0.009)
Binge Drinker	-0.007	-0.005	-0.036	-0.009
	(0.016)	(0.015)	(0.032)	(0.013)

Regression models for females contain 59,899 person-year observations. Standard errors are in parentheses for between effects. Clustered robust standard errors are in parentheses for OLS, fixed effects and random effects. Sampling weights are controlled in OLS. *** Significant at the 0.01 level. ** Significant at the 0.05 level. * Significant at the 0.10 level. See Table 4.6 for the list of explanatory variables.

Table A.6. *The estimates of IV and IV for fixed effects models of the effects of obesity, smoking and binge drinking on unemployment, for males and females separately*

	Males			Females		
	OLS	IV	IV- Fixed Effects	OLS	IV	IV- Fixed Effects
Obese	0.118 *** (0.039) [0.010]	0.129 (0.097) [0.011]	0.066 (0.057) [0.005]	0.026 (0.038) [0.002]	0.138 * (0.072) [0.011]	0.076 (0.058) [0.006]
Daily smoker	0.115 *** (0.026) [0.009]	0.211 ** (0.101) [0.016]	0.178 * (0.095) [0.014]	0.102 *** (0.027) [0.010]	0.198 ** (0.99) [0.016]	0.155 * (0.089) [0.11]
Binge Drinker	0.013 (0.026) [0.001]	0.022 (0.033) [0.002]	0.017 (0.029) [0.001]	0.009 (0.042) [0.001]	0.077 (0.068) [0.006]	0.042 (0.048) [0.003]

Regression models contain 64,712 person-year observations for males and 54,321 person-year observations for females. People who reported their employment status as out of labor force are excluded from the analyses. Clustered robust standard errors are in parentheses. Marginal effects are in brackets. Sampling weights are controlled. See Table 4.6 for other explanatory variables.

Table A.7. *The estimates of IV and IV for fixed effects models of the effects of obesity, smoking and binge drinking on log wages, for males and females separately*

	Males			Females		
	OLS	IV	IV-Fixed Effects	OLS	IV	IV-Fixed Effects
Obese	-0.022 (0.014)	-0.045 (0.039)	-0.010 (0.012)	-0.096 *** (0.016)	-0.123 ** (0.071)	-0.026 * (0.013)
Daily smoker	-0.074 *** (0.011)	-0.046 * (0.024)	-0.037 *** (0.010)	-0.023 ** (0.011)	-0.035 ** (0.015)	-0.013 (0.010)
Binge Drinker	0.017 * (0.010)	0.022 (0.020)	0.005 (0.008)	-0.007 (0.016)	0.011 (0.019)	-0.005 (0.013)

Regression models for males contain 65,365 person-year observations. Regression models for females contain 59,899 person-year observations. Standard errors are in parentheses. Sampling weights are controlled. See Table 4.6 for the list of explanatory variables.

Table A.8. *HTIV estimates of the effects of obesity, smoking and binge drinking on log wages, when different survey years are used*

	Males		Females	
HTIV results:	All Years	1984, 1992, 1994	All Years	1984, 1992, 1994
Obese	-0.017 (0.012)	-0.011 (0.015)	-0.043 *** (0.013)	-0.036 ** (0.018)
Daily smoker	-0.048 *** (0.009)	-0.033 *** (0.013)	-0.025 ** (0.011)	-0.020 * (0.012)
Binge Drinker	-0.006 (0.007)	-0.011 (0.011)	-0.007 (0.013)	-0.010 (0.016)

Regression models with all years contain 65,365 person-year observations for males and 59,899 person-year observations for females. Regressions with only 1984, 1992, 1994 years contain 10,002 person-year observations for males, 9,203 person-year observations for females. Clustered robust standard errors are in parentheses. Sampling weights are controlled. See Table 4.6 for standard and supplementary covariates.

References

Auld MC(2000). Smoking, drinking, and income. *Journal of Human Resources* 40,505-518.

Auld MC, Sidhu N (2005). Schooling, cognitive ability and health. *Health Economics* 14-10, 1019-1034.

Averett S, Korenman S (1996). The economic reality of the beauty myth. *Journal of Human Resources* 31, 304-330.

Baum CL, Ford WF (2004). The wage effects of obesity: a longitudinal study. *Health Economics* 13, 885-889.

Baum CL, Ford WF, Hopper JD (2006). The Obese Smoker's Wage Penalty. *Social Science Quarterly* 87, 863–881.

Becker G (1971). *The Economics of Discrimination*, vol I, second ed. Chicago: The University of Chicago Press.

Becker GS, Grossman M, Murphy KM (1994). An Empirical Analysis of Cigarette Addiction. *American Economic Review* 84, 396-418.

Behrman J, Rosenzweig M (2001). The Returns to Increasing Body Weight. Penn Institute for Economic Research Working Paper No.01-052.

Benham L, Benham A (1982). Employment, earnings, and psychiatric diagnosis, in: V. Fuchs, ed., Economic Aspects of Health, Chicago: University of Chicago Press

Berger MC, Leigh JP (1988). The effect of alcohol use on wages. *Applied Economics* 20, 1343-1351.

Betts N (2000). Smoking, Heavy Drinking and Poor Nutrition Tend To Cluster. *The American Journal of Health Promotion.*

Bhattacharya J, Sood N (2005). Health Insurance and the Obesity Externality. NBER Working Papers 11529, National Bureau of Economic Research, Inc.

Bollen K, Guilkey D, Mroz TA (1995). Binary Outcomes and Endogenous Explanatory Variables: Tests and Solutions with an Application to the Demand for Contraceptive Use in Tunisia. *Demography* 32-1, 111-131.

Bove C, Olson C (2005). Factors Contributing to Obesity in Rural, Low-Income New York Women. Working paper from the conference Rural Women's Health.

Bryant R, Samaranayake V, Wilhite A (1992). Alcohol use and the wages of young men: whites vs. nonwhites. *International Review of Applied Economics* 6, 184–202.

Bryant R, Jayawardhana A, Samaranayake V, Wilhite A (1996). The Impact of Alcohol and Drug Use on Employment: A Labor Market Study Using the National Longitudinal Survey of Youth. Institute for Research on Poverty. Discussion Paper no. 1092-96.

Burkhauser R, Cawley J (2004). Obesity, Disability, and Movement Onto the Disability Insurance Roll. Working Paper 2004-089 University of Michigan, Retirement Research Center.

Cawley J (2000a). Body weight and women's labor market outcomes. Working paper no. 7841, National Bureau of Economic Research, Cambridge, MA.

Cawley J (2004). The Impact of Obesity on Wages. *Journal of Human Resources* 39-2, 451-74.

Centers for Disease Control and Prevention (CDC) (2006b). BMI—Body Mass Index: Home. Washington, DC: Centers for Disease Control and Prevention, Department of Health and Human Services. Available at http://www.cdc.gov/nccdphp/dnpa/bmi/index.htm

Centers for Disease Control and Prevention (CDC) (2007). Behavioral Risk Factor Surveillance System (BRFSS). Washington, DC: Centers for Disease Control and Prevention, Department of Health and Human Services. Available at http://www.cdc.gov/brfss/index.htm

Centers for Disease Control and Prevention (CDC) (2007). Obesity Among Adults in the United States--No Statistically Significant Change Since 2003-2004. Washington, DC: Centers for Disease Control and Prevention, Department of Health and Human Services. Available at http://www.cdc.gov/nchs/data/databriefs/db01.pdf

Conway TL, Cronan TA (1992). Smoking, Exercise, and Physical Fitness. *Preventive Medicine* 21, 723-34.

Dallal, GE, Wilkinson L (1986). An analytic approximation to the distribution of Lilliefors' test for normality. *Am. Stat.* 40, 294-296.

Egger P, Pfaffermayr M (2004). Distance, trade and fdi: A hausman-taylor sur approach. *Journal of Applied Econometrics 19*, 227-246.

Everett M (1990). Let an overweight person call on your best customers? Fat chance. *Sales and Marketing Management* 142, 66–70.

Feng W, Zhou W, Butler JS, Booth BM, French MT (2001). The impact of problem drinking on employment. *Health Econ* 10, 509–521.

Finkelstein EA, Fiebelkorn IC, Wang G (2003). National Medical Spending Attributable to Overweight and Obesity: How Much, and Who's Paying? *Health Affairs* Available at hhttp://www.asu.edu/educ/epsl/CERU/Articles/CERU-0305-71-OWI.pdf

Forcier MW (1985). Labor force behavior of alcoholics: A review. *The International Journal of Addictions* 20, 253-268.

French M, Roebuck C, Alexandre P (2001). Illicit drug use, employment, and labor force participation. *Southern Econ J* 68, 349–368.

Fuchs, VR (1974). *Who Shall Live? Health, Economics, and Social Choice*. New York: Basic Books. Gottschalk 168.

Garcia J, Quintana-Domeque C (2006). Obesity, Employment and Wages in Europe, In K. Bolin and J. Cawley (Eds.), Advances in Health and Health Services Research, Vol. 17: The Economics of obesity. New York: Elsevier.

Gortmaker SL, Must A, Perrin JM, Sobol AM, Dietz WH (1993). Social and economic consequences of overweight in adolescence and young adulthood. N *Engl J Med*. 329, 1008-1012

Grafova IB, Stafford FP (2009). The Wage Effects of Personal Smoking History. *Industrial and Labor Relations Review* 62, 381-393.

Greeve J (2007). Obesity and Labour Market Outcomes: New Danish Evidence. wp 07-13, Department of Economics, Aarhus School of Business, University of Aarhus. Statistical Discrimination of Disabled Workers? Working paper SOLE 2005 Available: http://gsbwww.uchicago.edu/labor/E.08.4.%20Greve%20&%20Tranaes%20&%20Jensen.pdf

Griliches Z, Hausman J (1986). Errors in variables in panel data. *Journal of Econometrics* 1986; 31, 93-118.

Hamilton V, Hamilton BH (1997). Alcohol and earnings: does drinking yield a wage premium? *Canadian Journal of Economics* 97, 135-151.

Harper B (2000). Beauty, stature and the labour market: a British cohort study. *Oxford Bulletin of Economics and Statistics* 62, 771–801.

Harris R, Zhou J, Youngblood BD, Rybkin I, Smagin G, Ryan D (1998). Effect of repeated stress on body weight and body composition of rats fed low- and high-fat diets. *Am J Physiol Regul Integr Comp Physiol* 275, R1928-R1938.

Harwood H (2000). *Updating Estimates of the Economic Costs of Alcohol Abuse in the United States: Estimates Update Methods and Data.* Report prepared by the Lewin Group for the National Institute on Alcohol Abuse and Alcoholism.

Hausman JA, Taylor WE (1981). Panel data and unobservable individual effects. *Econometrica* 49, 1377–1398.

Heien DM (1996). Do drinkers earn less? *Southern Economic Journal* 63(1), 60–8.

Hoad NA, Clay DN (1992). Smoking Impairs the Response to a Physical Training Regime, A Study of Officer Cadets. *Journal of the Royal Army Medical Corps* 238, 115-17.

Jahns L, Baturin A, Popkin BM (2003). Obesity, diet, and poverty: trends in the Russian transition to market economy. *European Journal of Clinical Nutrition* 57, 1295-1302.

Johansson E, Alho H, Kiiskinen U, Poikolainen K (2007). The association of alcohol dependency with employment probability: Evidence from the population survey 'Health 2000 in Finland'. *Health Economics* 16, 739-5

Jusot F, Khlat M, Rochereau T, Serme C (2008). Job loss from poor health, smoking and obesity: a national prospective survey in France. *J Epidemiol Community Health* 62; 332-337.

Kandel DB, Davies M (1990). Labor force experiences of a national sample of young adult men, the role of drug involvement. *Youth & Society* 21, 411-445.

Keeler T, Hu T, Barnett P, Manning W (1993). Taxation, Regulation, and Addiction, A Demand Function for Cigarettes Based on Time-Series Evidence. *Journal of Health Economics* 12, 1-18.

Keng S, Huffman WE (2007). Binge drinking and labor market success: a longitudinal study on young people. *J Popul Econ* 20, 35-54.

Kenkel DS, Ribar D (1994). Alcohol consumption and young adults' socioeconomic status. *Brookings Papers on Economic Activity-Micro* 119-161.

Kenkel D, Lillard R, Mathios A (2003). Smoke or fog? The usefulness of retrospectively reported information about smoking. *Addiction* 98, 1307-1313.

Lahelma E, Kangas R, Manderbacka K (1995). Drinking and unemployment: contrasting patterns among men and women. *Drug and Alcohol Dependence* 37, 71-82.

Lee YL (1999). Wage effects of drinking and smoking: an analysis using Australian twins data. Working paper, Department of Economics, University of Western Australia.

Leigh JP, Berger MC (1989). Effects of smoking and being overweight on current earnings. *American Journal of Preventive Medicine* 5, 8-14.

Levine PB, Gustafson TA, Velenchik AD (1997). More bad news for smokers? The effects of cigarette smoking on wages. *Industrial and Labor Relations Review* 50, 493-509.

Lewit EM, Coate D (1982). The potential for using excise taxes to reduce smoking. *Journal of Health Economics* 121-45

Loh ES (1993). The economic effects of physical appearance. *Soc Sci Q* 74, 420-438.

Lundborg P, Bolin K, Höjgård S, Lindgren B (2007). Obesity and Occupational Attainment among the 50+ of Europe, In K. Bolin and J. Cawley (Eds.), Advances in Health and Health Services Research, Vol. 17: *The Economics of obesity*. New York: Elsevier.

MacDonald Z, Shields MA (2001). The impact of alcohol consumption on occupational attainment in England. *Economica* 68,427-53.

Maranto CL, Stenoien AF (2000). Weight discrimination: a multidisciplinary analysis. *Emp Respons Rights J.* 12, 9–24.

McArthur LH, Holbert D, Pena M (2001). Obesity knowledge of adolescents from six Latin American Cities: Effects of socioeconomic status. *International Journal of Obesity and Related Metabolic Disorders* 25, 1262-1268.

Montgomery JD (1991). Social networks and labor-market outcomes: Toward an economic analysis. *American Economic Review* 81, 1408-18.

Moon M, McLean R (1980). Health, obesity and earnings. *American Journal of Public Health* 70, 1006–1009.

Morris S (2007). The impact of obesity on employment. *Labour Economics*. 14, 413-33.

Mullahy J, Sindelar JL (1993). Alcoholism, work, and income. *Journal of Labor Economics*. 11, 494-520.

Mullahy J, Sindelar JL (1996). Employment, unemployment, and problem drinking. *Journal of Health Economics* 15, 409-34.

National Longitudinal Survey of Youth Data, (2002) [Internet]. [Cited 2002]. Available from: http://www.bls.gov/nls/79guide/2002/nls79g0.pdf

NIAAA (1993), Eight special report to US Congress on alcohol and heath, Washington, DC, N.I.H. Publ 943699.

Norton EC, Lindrooth RC, Ennett ST (1998). Controlling for the Endogeneity of Peer Substance Use on Adolescent Alcohol and Tobacco Use. *Health Economics* 7, 439-53.

Obesity Action Coalition (OAC) (2005). Obesity Action Coalition (OAC) Calls on Wal-Mart to Renounce Discrimination against Obese Employees. Available at: http://www.obesityaction.org/news/2005/oaccallsonwalmart.php

Pagan JA, Davila A (1997). Obesity, occupational attainment, and earnings. *Soc Sci Q* 78, 756-770.

Peters BL (2004). Is there a wage bonus from drinking? Unobserved heterogeneity examined. *Applied Economics* 36, 2299-2315.

Petry NM (1999). Alcohol use in HIV patients: What we don't know may hurt us. *International Journal of STD and AIDS* 10(9), 561–570.

Plumber T, Troeger VE (2006). The Euro and Monetary Policy Autonomy in European Non-Euro Countries. *European Union Politics* 7, 213-234.

Puhl R, Brownell KD (2001). Bias, Discrimination and Obesity. *Obesity Research* 12, 788-805.

Register CA, Williams DR (1990). Wage effects of obesity among young workers. *Soc Sci Q* 71, 130-141.

Renna F (2008). Alcohol Abuse, Alcoholism and Labor Market Outcomes: Looking for the Missing Link. *Industrial and Labor Relations Review* 62, 92-103.

Rice, DP (1999). The economic impact of schizophrenia. *Journal of Clinical Psychiatry* 60 (Suppl. 1), 4-6.

Sargent, J. D., Blanchflower, D. G (1994). Obesity and stature in adolescence and earnings in young adulthood. *Archives of Pediatric Adolescent Medicine* 148, 681–687.

Sarlio-Lahteenkorva S, Lahelma E (1999). The Association of Body Mass Index with Social and Economic Disadvantage in Women and Men. *International Journal of Epidemiology* 28, 445-449.

Sesso HD (2001). Alcohol and cardiovascular health: recent findings. *American Journal of Cardiovascular Drugs* 1, 167–172.

Shaper AG (1998). Alcohol and mortality in British men: Explaining the U-shaped curve. *The Lancet* 332, 1267-1273

Sousa S (2005). Does size matter? A propensity score approach to the effect of BMI on labour market outcomes. Unpublished manuscript, European University Institute and University of Minho.

Sturm R (2002). The Effects of Obesity, Smoking, And Drinking on Medical Problems. *Health Affairs* 245-253.

Tekin E (2004). Employment, wages, and alcohol consumption in Russia. *Southern Economic Journal* 71, 397-417.

Terza J (2002). Alcohol abuse and employment: A second look. *Journal of Applied Econometrics* 393-404.

Viscusi WK, Hersch J (2001). Cigarette Smokers as Job Risk Takers. *Review of Economics and Statistics* 83, 269-80.

Wasserman J, Manning W, Newhouse J, Winkler J (1991). The Effects of Excise Taxes and Regulations on Cigarette Smoking. *Journal of Health Economics* 10, 43-64.

WHO (2002). The world health report 2002 - Reducing Risks, Promoting Healthy Life.

Wolf, AM., Colditz GA (1998). Current Estimates of the Economic Cost of Obesity in the United States. *Obesity Research* 6, 97–106.

Wolf A (2001). Personal communication. November 26.

Wooldridge JM (2002). Econometric Analysis of Cross-Section and Panel Data. MIT Press, Cambridge, MA.

Yamaguchi K, Kandel DB (1985). On the resolution of role incompatibility: Life event history analysis of family roles and marijuana use. *American Journal of Sociology* 90, 1284-1325.

Zarkin GA, French MT, Mroz T, Bray JW (1998). Alcohol use and wages: new results from the National Household Survey on Drug Abuse. *Journal of Health Economics* 17, 53-68.

Curriculum Vitae

Ilker Dastan

2010	Ph.D. in Economics, Rutgers University, New Brunswick, New Jersey
2005	M.A. in Economics, Rutgers University, New Brunswick, New Jersey
2003	B.A. in Economics, Bilkent University, Ankara, Turkey

2007-2010	Instructor, Department of Economics, Rutgers University
2004-2007	Teaching Assistant, Department of Economics, Rutgers University