



Published in final edited form as:

*Econ Hum Biol.* 2012 March ; 10(2): 113–126. doi:10.1016/j.ehb.2011.09.002.

## Smoking and Body Weight: Evidence using Genetic Instruments

**George Wehby, Ph.D.[Assistant Professor],**

Dept. of Health Management and Policy, College of Public Health University of Iowa, 200 Hawkins Drive, E205 GH Iowa City, IA 52242 USA

**Jeffrey C. Murray, MD,**

University of Iowa, Iowa City, Iowa, USA

**Allen Wilcox, MD, PhD, and**

National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

**Rolv T. Lie, PhD**

University of Bergen, Bergen, Norway

### Abstract

Several studies have evaluated whether the high and rising obesity rates over the past three decades may be due to the declining smoking rates. There is mixed evidence across studies – some find negative smoking effects and positive cigarette cost effects on body weight, while others find opposite effects. This study applies a unique approach to identify the smoking effects on body weight and to evaluate the heterogeneity in these effects across the body mass index (BMI) distribution by utilizing genetic instruments for smoking. Using a data sample of 1,057 mothers from Norway, the study finds heterogeneous effects of cigarette smoking on BMI – smoking increases BMI at low/moderate BMI levels and decreases BMI at high BMI levels. The study highlights the potential advantages and challenges of employing genetic instrumental variables to identify behavior effects including the importance of qualifying the instruments and the need for large samples.

### Keywords

Smoking; obesity; body mass index; genetic instrumental variables; quantile regression; mendelian randomization

### I. Introduction

Obesity has become one of the most prevalent population health risks in several developed countries over the past 20 years. Obesity significantly increases cardiovascular disease, diabetes, other chronic health conditions, mortality, and health care costs (Finkelstein et al., 2009). Also, obesity and overweight may reduce employment (Han et al., 2009) and wage rates (Baum and Ford, 2004), in part due to increasing insurance premiums (Bhattacharya and Bundorf, 2009).

© 2011 Elsevier B.V. All rights reserved.

Phone: 319- 384-5133; Fax: 319-384-5125; george-wehby@uiowa.edu.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Body weight and obesity are complex traits, likely with several determinants and contributors. Lifestyle including high caloric and fast food consumption and physical inactivity, food advertisements, restaurant availability, and changes in economic wellbeing and security may play a role (Chou et al., 2004; Smith et al., 2009). Furthermore, body weight and obesity may have a strong genetic etiologic component. Twin studies estimate an 80% heritability of the body mass index (BMI) (Hjelmborg et al., 2008), 73% heritability of obesity (Watson et al., 2006) and up to 45-60% heritability of eating behaviors (Keskitalo et al., 2008; Tholin et al., 2005). Several genes may contribute to obesity, some of which have been identified and found to be significantly correlated to obesity across multiple studies (Martinez et al., 2007).<sup>1</sup>

Risk behaviors may also contribute to obesity. There has been a wide interest in understanding the effects of smoking on body weight and several economic and epidemiological studies have evaluated these effects. However, the direction and magnitude of effects vary across studies. There is a general perception that smoking may decrease body weight by decreasing appetite and caloric intake, enhancing metabolism, and reducing fat accumulation. This may occur through the effects of nicotine on brain regulation of appetite and energy expenditure. Studies in mice have shown appetite decrease, weight and fat loss, and changes in the expression of uncoupling proteins in the fat tissue and of brain neuropeptide Y levels, which are involved in metabolism-related processes (Chen et al., 2008; Chen et al., 2005). However, smoking may also decrease exercise by constraining respiratory functioning, which may counteract the previously mentioned effects on appetite and metabolism resulting in an overall no effect or increase in body weight. Therefore, the biologic pathways suggest a rather ambiguous net effect of smoking on body weight.

This paper reports an application of genetic instrumental variables to identify the effects of smoking on BMI. Furthermore, we evaluate whether these effects vary at different locations of the BMI distribution. The paper discusses the advantages and challenges of employing genetic instrumental variables particularly when studying behavioral effects. The paper proceeds as follows: Section II summarizes the existing knowledge and the contributions of this paper; section III describes the data and methods; section IV reports the results; section V includes the discussion and conclusions.

## II. Background and Contributions

Researchers have questioned the extent to which decreases in smoking rates over the past two decades may have contributed to rising population obesity rates. Figure 1 shows the smoking and obesity rates among adults between 1971 and 2006 in the United States (US). These rates were changing in opposite directions, with smoking rates decreasing as obesity rates were climbing. About 34% of adults were obese in 2007-2008 in the US, and another 34% were overweight (Flegal et al., 2010). In 1971-1974, only about 15% of the US adult population were obese and another 32% were overweight (NCHS, 2010). Meanwhile, adult smoking rates decreased from about 42% in 1965 to 20% in 2007 (NCHS, 2010).<sup>2</sup> These opposing trends suggest that smoking rates may have contributed to the increase in obesity rates over the past 30 years. However, recent evidence suggests that the BMI rise started much earlier than the decline of smoking (Komlos and Brabec, 2010). Therefore, the real contribution of changes in smoking behavior to the high and increasing obesity prevalence over the last 30 years remains unclear.

---

<sup>1</sup>Several of the candidate genes are involved in regulating food intake, energy expenditure and fat accumulation (Martinez et al., 2007).

<sup>2</sup>Smoking rates among men and women were 51.2 and 33.7%, respectively, in 1965, and 22% and 17.5%, respectively, in 2007 (NCHS, 2010).

One challenge in studying the smoking effects is the potential simultaneous selection of smoking and body weight based on the individual's preferences for health and risk taking. These preferences are typically unobserved and ignoring their role may bias the estimation of smoking effects. Furthermore, body weight may have reverse effects on smoking if individuals select to smoke in order to control their weight, or if they quit smoking due to health problems that are related to obesity. Cawley et al (2004) and Rees and Sabia (2010) find that overweight (including obese) status may increase the risk of smoking initiation among female adolescents. However, these studies do not evaluate whether obesity has different effects from overweight.<sup>3</sup> Unobservables and reverse weight effects on smoking make the direction of bias in smoking effects on body weight using classical estimations theoretically ambiguous. For instance, unobservable health and risk-taking preferences may result in a positive bias in smoking effects on BMI if individuals who smoke tend to eat more high-calorie or unhealthy food and exercise less. In contrast, the reverse effects of body weight on smoking may result in either a positive bias if those who are overweight or obese are more likely to initiate and continue smoking and to smoke more or a negative bias if individuals with lower body weight tend to smoke more as an approach to prevent weight gain and control weight.

Economic studies have accounted for the endogenous selection of smoking using multiple approaches that included studying the effects of cigarette taxes and prices on body weight or using taxes or prices as instruments for smoking. The results of these studies are mixed and generally sensitive to the analytical approach. Using a reduced-form function for BMI and obesity that includes cigarette prices, indicators for policies on public smoking, alcohol/food prices, income and demographic characteristics and data from the Behavioral Risk Factor Surveillance System (BRFSS) for years 1984-1999, Chou et al (2004) report large positive effects of cigarette prices on BMI and obesity. This suggests that increases in the price of cigarettes (due to increasing taxes and lawsuits against smoking companies) may have contributed to the rising obesity rates.

Gruber and Frakes (2006) revisit this question using the BRFSS for years 1984-2002 and cigarette taxes instead of prices as well as year indicators. They find that higher cigarette taxes reduce BMI and obesity. The authors use an instrumental-variables (IV) model with taxes as instruments, and conclude that the estimated positive effects of smoking on BMI and obesity are unbelievably large. However, Chou, Grossman, and Saffer (2006) argue that prices are preferred to tax rates as measures for cigarette cost by reflecting differences in transportation costs and competition and that year indicators in a long panel may over-control for unmeasured variables compared to the non-linear time trends employed in their study. Nonnemaker et al. (2009) also evaluates the effects of cigarette prices on BMI using the BRFSS data for years 1984-2004 and tax rates as instruments for prices. The study includes state and time fixed effects and state-specific linear time trends and stratifies the model by smoking status (never smoker, former smoker, current smoker). They find overall an insignificant price effect in the pooled sample, but a significant positive price effect on the BMI of former smokers, which explains about 17.8% of their BMI increase. An opposite and insignificant effect is found for current smokers and never smokers. In their specification without using tax rate as an instrument, Nonnemaker et al. (2009) do not replicate the finding from Chou et al. (2004). Further using the specification of Chou et al. (2004) but stratifying the model by smoking status, Nonnemaker et al. (2009) find that the largest positive price effect is for never smokers.

Rashad (2006) applies an IV model to estimate the smoking effects on BMI using the National Health and Nutrition Examination Surveys (NHANES) I, II and III (years

---

<sup>3</sup>Also at least one paper (Rees and Sabia, 2010) seems to combine underweight and normal weight in the reference category.

1971-1994), and finds that the 2SLS smoking effects are insignificant, except among Hispanics, for whom the negative 2SLS effect is larger in absolute value than OLS. The OLS effects are negative and significant, both pooled and stratified by race, indicating a 1.3-point BMI decrease with smoking.

Baum (2009) evaluates the effects of cigarette costs (prices and taxes) on obesity and BMI using a difference-in-difference approach by subtracting the effect for individuals who smoked less than 100 cigarettes in their life from the effect for those who smoked 100 cigarettes or more, using the National Longitudinal survey of Youth (NLSY) for years 1979-2002. The study finds that increasing cigarette costs increase BMI and obesity. Courtemanche (2009) assesses the effect of lagged prices and taxes on BMI and obesity using the BRFSS for years 1984-2005 and the NLSY for years 1979-2004, and finds negative effects. The effect direction in that study is generally insensitive to the alternative model specifications that produce different results in the Chou et al. (2004) and Gruber and Frakes (2006). Finally, Fang et al. (2009) use data from China and instrumental-variables analysis (with community-level cigarette prices and average cigarette numbers as instruments) and find that smoking reduces BMI.

The mixed results from these studies highlight the analytical complexity of identifying the smoking effects on BMI and obesity. One common approach among these studies is the utilization of area-level measures for cigarette costs, either as instruments for smoking or as explanatory variables in reduced-form models. Area-level variables ignore variation between individuals in the same area and may be correlated with other area-level characteristics that affect BMI, such as healthcare quality and availability, food prices and social and cultural factors.

Another limitation of previous studies is that most of them ignore the potential heterogeneity in smoking effects on BMI and estimate effects only at the mean BMI. Heterogeneous effects may occur at various levels of BMI if preferences for body appearance and weight and risk factors for obesity (which cluster at different parts of the BMI distribution) modify the smoking effects on BMI. Furthermore, smoking effects on metabolism, appetite and exercise may vary by BMI levels. The study by Fang et al (2009) is the only one that evaluates the heterogeneity in smoking effects on the BMI distribution. The study finds, using quantile regression, that the negative smoking effects on BMI are highest between the 25<sup>th</sup> and 75<sup>th</sup> percentiles of BMI. However, their approach to account for the endogenous selection of smoking in the quantile regression does not provide consistent estimates (Chernozhukov and Hansen, 2004; Chernozhukov and Hansen, 2005).

Our study applies a novel approach to identify the smoking effects on BMI by utilizing genetic instruments for smoking. The instruments are genetic variants correlated with smoking but otherwise unrelated to BMI. Another contribution of this study is the assessment of heterogeneity of smoking effects across the BMI distribution. Our study addresses some of the earlier study limitations and provides a new analytical approach that integrates econometric approaches with genetic data to study the “causal effects” of smoking on BMI. Also, the paper discusses the requirements, strengths, and challenges of employing genetic instruments to identify behavior effects.

### III. Data and Methods

#### III.A. Analytical Approach

Body weight ( $W$ ) is studied as a function of smoking ( $S$ ) and socioeconomic and demographic factors that may relate to preferences for body weight, information about health risks and economic wellbeing ( $E$ ) as follows:

$$W_i = \alpha_0 + \gamma S_i + \mathbf{E}_i \lambda + e_i \quad (1),$$

where  $i$  represents an individual and  $\gamma$  represents the smoking effect. As discussed above, smoking is endogenous to body weight, suggesting that direct estimation of equation (1) will result in biased estimates of  $\gamma$ . A common approach is to utilize instruments for smoking that are unrelated to the error term  $e$  and are not directly related to  $W$ . The smoking function is modeled as follows:

$$S_i = \alpha_1 + \mathbf{G}_i \beta + \mathbf{E}_i \omega + u_i \quad (2),$$

where  $\mathbf{G}$  includes the instruments for smoking and  $u$  is the error term.

**III.A.1 Smoking Genetics and Genetic Instruments**—As discussed above, previous studies have used cigarette prices/taxes and other area-level instruments for smoking. In this study, we use a new source of variation in smoking that is due to genetic predisposition. There is a convergent evidence of strong links between specific genetic factors and cigarette smoking. Smoking behaviors are well known to involve a complex etiology of genetic, social and economic factors (Li, 2006; Tyndale, 2003). Many twin and adoption studies have demonstrated that genetic heritability is at least 50% for both smoking initiation and smoking persistence (Carmelli et al., 1992; Heath and Martin, 1993; Lessov et al., 2004; Maes et al., 2004). Researchers have identified several variants (SNPs) in nicotine, detoxification and neurotransmitter genes that are significantly correlated with several smoking behaviors including smoking initiation, smoking intensity such as cigarettes per day and other intensity measures, and smoking cessation.<sup>4</sup> Of course, non-genetic factors including economic, social, human capital and environmental factors such as prices, income, education, availability/supply of cigarettes, cigarette advertising, peer behavior and social network effects, and culture also affect smoking decisions. Nonetheless, there is strong and consistent evidence that various types of smoking behaviors are influenced by genetic factors. This clearly does not negate the importance of the economic, social and other non-genetic effects, but rather highlights the complex etiology that underlies smoking behaviors and that likely involves several genetic and non-genetic factors acting independently or interactively with each other. Furthermore, typically evaluated economic, social and demographic factors generally explain 10% or less of smoking behaviors (e.g., Cheng and Kenkel, 2010). The large percentage of variation in smoking that is unexplained by these factors is consistent with the suggested involvement of genetic factors in smoking behaviors.

The increasingly available data on specific genetic contributors to risk behaviors such as smoking provides a wealth of information for economists and other social scientists who wish to study the effects of these risk behaviors on outcomes (Fletcher, 2011; Wehby et al., 2008). Specifically, genetic variants may be utilized as instruments in order to account for the endogenous selection of risk behaviors. Other disciplines such as epidemiology have also begun to embrace this approach, referred to as “Mendelian randomization” in the epidemiology literature (Lawlor et al., 2008).

<sup>4</sup>Some of these genes are dopamine 2 receptor (*DRD2*) (Comings DE, 1996; Noble EP, 1994; Spitz et al., 1998), dopamine beta hydroxylase (*DBH*) (McKinney et al., 2000), *DOPA* decarboxylase (*DC*) (Ma et al., 2005; Yu et al., 2006), Cholecystokinin (*CCK*) (Comings et al., 2001), Tryptophan Hydroxylase (*TPH*) (Lerman et al., 2001; Sullivan, 2001), gamma-aminobutyric acid type B receptor subunit 2 gene (*GABBR2*) (Agrawal et al., 2009; Beuten, 2005), nicotinic acetylcholine receptor a4 subunit (*CHRNA4*) (Li et al., 2005) and a3 subunit (rs1051730 in *CHRNA3*, (Thorgeirsson et al., 2008) and Cytochrome P450 2A6 (*CYP2A6*; (Pianezza et al., 1998; Sellers et al., 2000). The variant rs1051730 is also related to quitting and intensity of smoking during pregnancy (Freathy et al., 2009). Recent studies have found significant associations between smoking and SNPs in *GABRA4*, *GABRA2* and *GABRE* (Agrawal et al., 2008). Two recent Genome Wide Association Studies (GWAS) have identified correlations between various nicotine dependence measures and SNPs in *CHRNA3* and *CHRNA5* (Berrettini et al., 2008; Caporaso et al., 2009), *MAOA* and *ACTN1* (Caporaso et al., 2009).



Genetic variants provide several advantages over traditional instruments for behaviors, which have ranged from individual-level enabling factors (such as income or employment) to area-level characteristics related to prices or policies. There are different types of DNA variants, but the most commonly-studied types are DNA base-pair variants, called single nucleotide polymorphisms (SNPs). Each SNP has two variants, called alleles (allele A or B). For each SNP, an individual may have a homozygote combination of either allele (AA or BB) or a heterozygote combination (AB). At the time of germ cell generation – a process called meiosis – each allele has an equal 50% chance of being transferred. In other words, the chance of inheriting either allele from a heterozygote parent is 50%.

Note that this “random” allele inheritance process does not imply that all individuals have the same probability of inheriting a certain allele. For example, a child born to AA-homozygote parents has a 100% chance of inheriting the AA genotype.<sup>5</sup> Nonetheless, the process of random assignment of alleles from each parent to their child over time suggests that genetic variants that are related to the risk behavior of interest (smoking in this study) are unlikely to be correlated with unobserved behaviors that are correlated with this risk behavior through preferences for health and risk taking, if these variants do not directly affect these unobserved behaviors or the unobserved preferences. Studies have found evidence of this randomizing effect of genetic instruments (Smith et al., 2007). Another advantage of genetic instruments is that they precede all behaviors in time and cannot be affected by behaviors. In contrast, other instruments (including area-level instruments such as cigarette prices or public smoking policies) may be affected by behaviors if individuals choose where to live based in part on their behaviors and area characteristics, and if population behaviors affect these characteristics (such as if smoking rates are a factor in policy decisions on cigarette taxes and policies).

Of course, violations of the exogeneity of genetic instruments may occur due to certain factors, as has been previously highlighted (von Hinke Kessler Scholder, 2011). First, genetic variants in physical proximity on the same chromosome may be correlated to each other, a process called linkage disequilibrium. If genetic instruments of interest are correlated with other genetic variants that are associated with unobserved confounders, such as other risk behaviors correlated with smoking, then the genetic instruments become endogenous (Lawlor et al., 2008). This threat to instrument exogeneity may be evaluated when suspected by studying the correlations between the genetic instruments and neighboring SNPs that are suspected of being related to unobserved confounders. Also, some genes have multiple functions and may affect several relevant behaviors and risk factors, which if unobserved, may violate the instrument exogeneity. Therefore, it is important to understand the gene functions and their implications for the instrument validity. Other violations may occur due to compensatory effects for genetic risks, although this is unlikely to be relevant for most applications aimed at studying behavior effects (von Hinke Kessler Scholder, 2011).

Another potential violation of instrument exogeneity may occur due to non-random matching within couples based *simultaneously* on genetic and non-genetic factors (behavioral, economic, and social factors) that may contribute to their children’s behaviors. If smokers are more likely to marry smokers due to shared preferences (including preferences for risk taking, future discounting and health importance), and if parental smoking increases children’s propensity to smoke (due for example to the availability of cigarettes in the household and the “acceptability” of smoking in the family), then this does not violate the exogeneity of genetic instruments by itself. In other words, as long as such preferences are not caused by the genetic instruments, the instruments remain exogenous

---

<sup>5</sup>Violations of this may occur when genetic (de-novo) mutations occur during meiosis or after pregnancy, which are very rare.

even if matching occurs based on these preferences. However, if smokers are more likely to intermarry due to the genetic risk factors that predispose to smoking, which is less expected compared with matching based on preferences that affect smoking, then this would violate the exogeneity of the genetic instruments. A standard test for Hardy-Weinberg Equilibrium (HWE) can be employed to evaluate whether the proportions of SNP genotypes satisfy the expected distribution based on random mating (Wigginton and Abecasis, 2005), which we employ as described below.

Another potential limitation of genetic instruments that may be common in several applications and that needs to be acknowledged is that the instruments may have significant but “weak” effects on the endogenous behavior – i.e. do not explain a large percentage of the variation in behavior – based on the commonly used thresholds for non-weak instruments such as an F-statistic above 10 or specific F-statistic thresholds based on 2SLS bias relative to OLS (Stock and Yogo, 2005).<sup>6</sup> This is expected because of the complex etiologies of behaviors that involve several genetic and non-genetic risk factors. Both, the current knowledge of biological pathways and empirical evidence to date suggest that most genetic factors are likely to explain individually only a small portion of the variation of complex traits such as behaviors. Weak instruments, even when exogenous as they should be, may bias the endogenous variable effect in the direction of the estimate assuming exogeneity and its variance downward, which may result in wrong inference of significant effects (Hahn and Hausman, 2003). Appropriate inference that is robust for weak instruments is needed in this case in order to avoid any bias with standard inference approaches. Therefore, there may be a tradeoff for several applications with genetic instruments between the theoretical advantage of instrument exogeneity compared to other non-genetic individual- or area-level instruments, and the empirical challenge of identifying the behavioral effect based on limited exogenous variation. However, with the availability of weak-instrument robust inference methods, and given the limitations and mixed evidence of previous studies of smoking and body weight employing other identification approaches, evaluating the smoking effects using genetic instruments likely adds significant knowledge to this area.

Our study uses the following three SNPs in the gene *GABBR2* as instruments for smoking: rs1435252, rs3780422 and rs1930139. *GABBR2* is on chromosome 9 and is involved in neuronal activity inhibition by coding for a protein for a GABA-B receptor involved in neurotransmitter release. This gene is considered to be a high priority candidate gene for nicotine dependence (NICSNP, 2007). Two studies using the same data source but different sample sizes found variants in *GABBR2* to be significantly involved in several smoking measures including the Fagerström Test for Nicotine Dependence (FTND), cigarettes per day, and the Heaviness of Smoking index (Beuten, 2005; Li et al., 2009). A candidate gene study of a sample of pregnant women from the Danish National Birth Cohort that was included in a GWAS prematurity study found several SNPs in *GABBR2* to be significantly associated with the average number of cigarettes smoked per day during pregnancy, although the associations were not very strong, with F-statistics ranging from 3 to 7 (Prater et al., 2011). Furthermore, a study found that nicotine significantly modifies the expression of *GABBR2* in the brain of rats both at the messenger RNA and protein levels, providing support for the involvement of this gene in smoking behaviors (Sun et al., 2007).

---

<sup>6</sup>F-statistics in several models of the few health economics studies that use genetic variants as instruments for obesity, ADHD, depression and smoking are less than 10 (e.g., Ding et al., 2009; Norton and Han, 2008). The F-statistics for the instruments for smoking are less than 10 (Ding et al, 2009). Some of the models in these studies for the behaviors other than smoking have F-statistics greater than 10. However, it is somewhat unclear to what extent the exclusion restrictions are theoretically justified in some of these models given that instruments might not be related only to the instrumented behaviors (obesity, ADHD, depression). Indeed, the p-values of the over-identification restrictions are lower in these studies than those we observe in this study.

The evidence from the above studies suggests a role for *GABBR2* in smoking and provides some support for considering it as an instrument. However, not all studies find significant association between *GABBR2* and smoking behaviors. A candidate-gene study of smoking that evaluated several *GABBR2* variants did not find significant associations between these variants and smoking (Agrawal et al., 2008). The inconsistency in results between that study and the first two studies finding significant association with smoking introduces some uncertainty regarding whether *GABBR2* truly affects smoking behaviors. However, there are major differences between that study and those other two studies in design and smoking measures. The first two studies reporting significant association use a family-based association design where probands – the nuclear participants who define the family structure for the genetic association analysis and determine the family’s eligibility to participate in the study – are defined based on consuming a minimum of 20 cigarettes per day for the last 12 months (Beuten, 2005; Li et al., 2009). In contrast, the study finding no significant association for *GABBR2* used a case-control design with cases (affected individuals) having an FTND equal to or greater than 4, and controls (unaffected individuals) having an FTND equal to 0 (Agrawal et al., 2008). Therefore, the first two studies finding significant association between *GABBR2* and smoking were more enriched with intensive smokers relative to the study finding no association.

Two genome-wide association studies (GWAS) did not report *GABBR2* to have significant effects on smoking (Berrettini et al., 2008; Caporaso et al., 2009). However, none of the SNPs in these studies was significant at the GWAS significance threshold corrected for multiple testing; some SNPs in *CHRNA3* and *CHRNA5* had low p-values but did not pass the p-value threshold. The samples for these two GWAS were also less enriched in intensive smokers relative to the two family-based studies finding significant association with *GABBR2*. Recently, two meta-analyses of GWAS data reported some variants to be significantly related to smoking at GWAS-appropriate significance thresholds with SNPs in *CHRNA3* having the largest effect and strongest statistical significance for cigarettes per day (Liu et al., 2010; The Tobacco and Genetics Consortium, 2010). These meta-analyses did not find evidence for *GABBR2* effects on smoking. The *CHRNA3* and other gene variants found to be significant in these GWAS meta-analyses have not been genotyped in our study sample.

Ideally, the selected genetic instruments should have unequivocal evidence of being related to the endogenous variable of interest (smoking behavior in this case). An increase in the number of previous studies replicating the associations of the genes or specific genetic variants being considered as instruments with the endogenous variable certainly provides stronger support for choosing these as instruments. However realistically, replications of genetic effects are not observed in many cases of complex phenotypes. The lack of replication is a rather common phenomenon for phenotypes and behaviors such as smoking that have complex genetic etiologies involving multiple genes and alleles that may interact amongst themselves and with environmental factors. The genetic/allelic heterogeneity may contribute to differences in results especially when comparing the results of studies that differ in their sample characteristics, phenotype measures, and designs. Further, the effects of genetic variants may be modified by environmental, behavioral and demographic characteristics which may result in differences in genetic effects across studies employing samples that vary in these characteristics.

GWAS provide several advantages over candidate gene association studies for detecting small to moderate genetic effects. However, they also face some limitations that should be considered when evaluating the replication of results. These include very low significance thresholds that may mask some important genetic effects as well as grouping heterogeneous samples from multiple data sources that may introduce noise into the phenotype measures



and increase the variance of the genetic effect estimates. The smoking GWAS described above included samples from several studies of various health conditions (diabetes, cardiovascular outcomes, depression/anxiety, others) which may increase measurement error in the smoking phenotypes.

In any application of genetic instrumental variables, it is important to evaluate the evidence both for and against employing the genetic variants as instruments and the implications for interpreting the study findings. While there is mixed evidence for *GABBR2*, the previous findings of significant effects of *GABBR2* variants on smoking, the knowledge of the gene's functions including the study results for gene expression in response to nicotine summarize above, and the significant relationship that we find in our study between *GABBR2* and number of cigarettes provides support for considering this gene as an instrument for smoking.

The second condition that instruments should satisfy is that they do not affect the outcome (body weight) either directly or through unobserved confounders. In other words, the instruments should affect body weight only indirectly through their effects on smoking. The support for this condition should be primarily obtained from the current knowledge of the gene functions based on previous studies. After searching the literature on *GABBR2*, we are unable to find any published studies that directly associate this gene with other behaviors or pathways that can affect body weight besides smoking.<sup>7</sup> Also, there is no evidence that this particular gene is involved in general addiction or in addictive personalities, and no studies have found evidence of it being linked to other addictions besides nicotine dependence. One study has reported that patients with autism have lower protein levels of GABA(B) receptor 2 units in the cerebellum compared to unaffected individuals (Fatemi et al., 2009). Another study found protein-level reductions for this receptor in the lateral cerebella of patients with schizophrenia, depression, and bipolar disorder compared to individuals without these conditions (Fatemi et al., 2011). While important, these studies are based on very small samples (less than 15 cases for each disorder) and do not provide evidence that *GABBR2* affects these health conditions or affects body weight through them. Furthermore, the fact that our sample is limited to pregnant and mostly healthy women reduces any potential threats of the instrument being correlated with body weight through these health conditions.

As would be the case for most genes involved in behaviors, it is important to recognize that the lack of evidence to date for a role of *GABBR2* in other behaviors besides smoking that may also affect body weight does not provide complete assurance against identifying such a role in the future. The continuously expanding and increasing capacity of genetic studies and gene-function assays for behavioral traits is expected to further characterize the functions of *GABBR2* and other candidate genes for smoking, which will allow for a more comprehensive evaluation of the fit of these genes as instruments in future studies. However, based on the current knowledge, there is a reasonable support for *GABBR2* to be unrelated to body weight other than through smoking.

The excludability of the instruments from the body weight function cannot be fully tested due to the role of unobservable factors but can be partially evaluated by testing the over-identification restrictions (Basman, 1960; Hausman, 1983; Wooldridge, 2002). These tests evaluate the excludability of the additional instruments from the body-weight function after identifying the smoking effect using one instrument. We employ these tests in order to provide a partial statistical evaluation of the instrument exogeneity. As described below, the instruments pass the over-identification restriction test at very high p-values (0.998), which

---

<sup>7</sup>We searched pubmed for published studies as well as the GENE database on pubmed which includes links to studies providing information on the genes' functions.

provides some assurance for the exogeneity of the instruments and is consistent with the current knowledge of the gene effects. As an additional check of whether the instruments may violate the assumption of being unrelated to BMI other than through smoking, we correlate the instruments with observed socioeconomic, demographic, and behavioral characteristics. While this does not fully evaluate the excludability condition, it reveals whether the instruments are systematically correlated with other observable confounders which would suggest that the instruments may not be exogenous. As described below, we find that the instruments are not related to these characteristics, which provides some further assurance that the instruments may be exogenous.

Two of the three SNPs that we employ as instruments, rs1435252 and rs3780422, have been found to be significantly related to smoking behaviors in European-American populations both on their own and as part of haplotypes (groups of alleles across multiple SNPs) involving other SNPs (Beuten, 2005). We include direct indicators for the genotypes of the three SNPs as instruments instead of haplotypes for two reasons. First, we are unable to find significant effects of haplotypes that have an adequate frequency to be used as instruments. Second, including indicators for multiple SNPs simultaneously in the model captures differences in the effects of various combinations of genotypes across these SNPs and some of the haplotype effects.<sup>8</sup> For each SNP, we include two indicators for the minor allele homozygote and heterozygote genotypes, with the major allele homozygote as the reference genotype. This avoids the assumption that the minor allele has similar effects in its homozygote and heterozygote forms.

The three SNPs we employ as instruments are significantly correlated with each other; the hypothesis of no linkage disequilibrium is rejected at  $p < 0.0001$ . However, this has no effect on the IV analysis which depends on the statistical strength of the instruments as a group in predicting the endogenous variable. Therefore, significant correlations between the instruments that increase the standard errors for the individual instrument effects on smoking do not reduce the significance of the joint instrument effects as a group. We find that the instruments have significant joint effects on smoking after adjusting for all the model exogenous variables, but that they are “weak” using the standard thresholds, with an F-statistic of 3.4. Therefore, we employ weak-instrument robust inference approaches that are described below in detail. In addition, we evaluate HWE for these SNPs.

### III.B. Data and Measures

The data are from a population-level study of oral clefts in Norway between 1996 and 2001 (Lie et al., 2008; NIEHS., 2009). That study enrolled 88% of Norwegian infants born with oral clefts in 1996-2001 and 76% of a randomly-selected sample of live-born infants born in Norway during the same period. The study obtained DNA samples from parents and children and obtained data from the mothers about 3-4 months after delivery on prenatal, behavioral, demographic, and socioeconomic characteristics. The majority of the mothers (94%) were born in Norway. In alternative models, we include an indicator for whether the mother was born in Norway or not and find similar results to the models that do not include this indicator.<sup>9</sup>

In this study, we evaluate the effects of number of cigarettes smoked per day (including 0 cigarettes for non-smokers) over the 12 months before becoming pregnant on BMI before pregnancy (BMI is based on the woman’s height and weight before pregnancy as reported shortly after delivery). Number of cigarettes is used because it reflects both smoking

<sup>8</sup>We recognize that this may not capture all haplotype effects which represent effects of combinations of alleles on the “same” chromosome of a certain chromosome pair.

<sup>9</sup>Detailed results are available from the authors.

participation and intensity. The use of smoking alone (yes or no) will mask variation in smoking intensity. Furthermore, the instruments are not significantly predictive of the probability of smoking participation (any smoking). This is not surprising given that the previous studies finding significant *GABBR2* effects used smoking quantity measures in the form of cigarettes per day and other intensity measures.

BMI is the standard measure of obesity (weight in kilograms divided by height in meters squared) and is employed in all the economic studies described above. We use BMI as a continuous measure in order to evaluate the smoking effect heterogeneity throughout the BMI distribution, and not just at the mean. In addition, we assess the smoking effects on a categorical BMI index of underweight, normal weight, overweight, and obesity given the common use of these categories as population-level indicators of body size and health.

The study sample includes 1,057 mothers who have complete data on the study measures and genetic instruments. The sample includes mothers of both affected and unaffected children with oral clefts. The inclusion of all mothers would not be expected to affect our analysis. We assess the effects of smoking 12 months before pregnancy on BMI before pregnancy, which suggests that the accuracy of such reports is unlikely to vary by the pregnancy outcome. Also, there is no a priori knowledge that the smoking effects on BMI may vary by whether the mother has genetic risk factors for having children with oral clefts. All regressions include the following covariates: age, marital status, education, and gross yearly income<sup>10</sup>. Table 1 reports the distribution of the study variables.

### III.C. Quantile Regression Model

One of our study objectives is to evaluate the heterogeneity of smoking effects throughout the BMI distribution. We use quantile regression (QR) to estimate these effects at various BMI distribution locations. QR provides a flexible semi-parametric model to evaluate effect heterogeneity. The model is estimated using the instrumental variables QR (IVQR) model of Chernozhukov and Hansen (2004; 2005; 2006)<sup>11</sup> and can be specified as follows:

$$W=Q(S, \mathbf{E}, U), \quad U \sim (0, 1) \quad (3)$$

where  $U$  is a uniformly distributed “unobserved” BMI technological factor that may represent the net value of all unobserved factors that are relevant to BMI, which may include preference, genetic, and social risk factors for body weight. Conditional on  $S$  and  $\mathbf{E}$ ,  $U$  determines each woman’s location or rank on the BMI distribution. Substituting the BMI quantile order  $q$ , which varies between 0 and 1, for  $U$ ,  $Q(S, \mathbf{E}, q)$  represents the conditional  $q$ th BMI quantile. The “ranking” factor  $U$  allows for interpreting the effects of  $S$  and  $\mathbf{E}$  on BMI quantiles as quantile treatment effects (QTEs) by unobserved BMI risk factors since  $U$  is held constant at  $q$ . The model estimates the smoking effects at each specified  $q$  (or  $U$ ):

$$W=Q(\alpha_{0q} + \gamma_q S + \mathbf{E}\lambda_q) \quad (4)$$

<sup>10</sup>We do not include pregnancy planning in the model because cigarette smoking is measured during the year before pregnancy which may affect pregnancy planning, making this variable endogenous, and its inclusion inappropriate. BMI may also affect pregnancy planning. One might argue that early pregnancy planning may in turn affect cigarette smoking and BMI. However, the instruments should account for any endogeneity due to early pregnancy planning. Indeed, the instruments have jointly insignificant effect on pregnancy planning when regressed on the instruments ( $p=0.9991$ ). Further, there is evidence that women who were planning their pregnancy in this sample and who smoked before pregnancy were more likely to quit cigarette smoking after pregnancy occurrence compared to those who were not planning their pregnancy (25.9% versus 17.2% quit smoking), suggesting that pregnancy planning has an effect on smoking after pregnancy occurrence.

<sup>11</sup>This model has been used in several applications (e.g., Wehby et al., 2009).

For quantile  $q$ , the IVQR model involves a grid search over the parameter space for  $\gamma$  and identifies the value that minimizes the coefficient  $\beta$  as close as possible to 0 in the following equation:

$$W - \gamma S = Q(\mathbf{E}\lambda_q(\gamma) + \beta_q(\gamma) Z) \tag{5}$$

where  $Z$  is the least squares projection of  $S$  on the identifying instruments  $\mathbf{G}$  and on  $\mathbf{E}$ , and where  $\beta$  and  $\lambda$  are functions of  $\gamma$ . The equivalence of  $\beta$  to 0 is the IV condition of the instruments affecting  $W$  only through  $S$  and not being related to  $W$  through unobserved confounders. To summarize, the IVQR estimate of  $\gamma$  is the value that makes  $\beta$  as close as possible to 0. We estimate the IVQR model for BMI quantiles ( $q$ ) 0.1, 0.25, 0.5, 0.75 and 0.9, which provide good coverage of the entire BMI distribution and are commonly used in QR applications. We avoid studying extreme quantiles as these may result in unstable estimates due to the small number of observations at these quantiles.

As described above, the instruments are considered “weak” with an F-statistic of 3.4 ( $p=0.0025$ ). Relying on asymptotic inference alone in this case may result in biased inference as the asymptotic standard errors may not be good estimates of those of the finite sample. Therefore, we estimate weak-instrument robust 95% confidence bounds for the smoking IVQR effects (Chernozhukov et al., 2007). There is no a-priori expectation for whether these confidence bounds are wider or tighter than the asymptotic confidence bounds (Chernozhukov and Hansen, 2008).

In addition to IVQR, equation (1) is also estimated by standard QR, assuming exogenous smoking for comparison purposes (Koenker and Bassett, 1978; Koenker and Hallock, 2001).<sup>12</sup> Note that quantile regression is estimated using the *whole sample*, and not by stratifying by the quantiles of the dependent variable. We also estimate the smoking effects on BMI mean for comparison purposes using OLS and two-stage least squares (2SLS) and estimate weak-instrument robust confidence bounds for the 2SLS effects (Chernozhukov and Hansen, 2008).

### III.D. Categorical BMI index

We assess the smoking effects on a categorical BMI index in order to better understand the smoking effects on the commonly-used weight indicators of underweight, normal weight, overweight, and obesity. A further motivation for using this index is that the BMI distribution of the study sample involves lower frequencies of high BMI values than the distribution for the US population. Therefore, evaluating the smoking effects on the weight categories complements the QTE estimation and allows for better extrapolation of the smoking effects to the US and other populations.

We assess the smoking effects on this index using multinomial logit. We use multinomial logit instead of an ordered logit model in order to evaluate the heterogeneity of the smoking effects on various weight categories and avoid the restrictive assumption of equal-proportionality of effects across weight categories under ordered logit. In order to account for the endogenous selection of cigarettes, we include the residual term from equation (2) as

<sup>12</sup>This model is estimated as follows:

$$\min \left[ q \sum_{W_i \geq Q_i} |W_i - Q_i| + (1 - q) \sum_{W_i < Q_i} |W_i - Q_i| \right] \tag{6}$$

The standard errors of this model are estimated by bootstrap with 2000 replications.

a regressor in the multinomial logit model of equation (1) (Terza et al., 2008). Testing the coefficient of this residual term provides a test of the endogenous selection of cigarettes (Wooldridge, 2002). Given that the standard variance estimates from this model are biased, we estimate the standard errors of the marginal cigarette effects on the weight index categories using bootstrap with 2,000 replications.

## IV. Results

### IV. A. Instrument Effects and Validity

Table 2 reports the coefficients of the cigarette function. The instruments have significant effects on cigarette number, with a joint F-statistic of 3.4 ( $p=0.0025$ ). Relative to the CC genotype, the TT genotype of rs1435252 decreases average daily cigarettes by 1.3 cigarettes, while the CT genotype increases smoking by 0.74 cigarettes per day. The AG of rs1930139 increases daily cigarettes by about one cigarette (relative to AA).

Some of these effects may seem unreasonably large as they exceed the GWAS-reported *CHRNA3* effects of one-cigarette increase per day with each risk allele (The Tobacco and Genetics Consortium, 2010). However, our study sample differs significantly from the GWAS samples on demographic and health characteristics. Mainly, our sample is limited to young mothers with measures of smoking during the 12 months before pregnancy while the GWAS include a wide range of samples of significantly older ages. Specifically, the average age among the various studies/samples included in the Tobacco and Genetics Consortium (2010) GWAS ranges from 39.6 to 72.3 years, which significantly exceeds the average age of 29 years in our sample. Given that differences in demographic, environmental, human capital and health factors may modify genetic effects on behaviors, differences in such characteristics between our sample and the GWAS samples and the pre-pregnancy period during which smoking is measured in our study may contribute to the observed large *GABBR2* effects in our sample. The incentive to quit smoking before pregnancy in order to avoid adverse effects on fetal health may intensify certain genetic effects on smoking.

The 2SLS over-identification restrictions evaluate the second IV assumption by testing the instrument excludability from the BMI function. The assumption cannot be rejected at a p-value of 0.998<sup>13</sup>. Exogenous but weak instruments are expected to over- not under-reject the over-identification restrictions (Hahn and Hausman, 2003). This, coupled to the high p-value of the over-identification test, provides some support for the exogeneity of the instruments, and suggests that the weak-instrument limitation is not affecting this result, although this cannot be fully verified. It is also important to note that the instruments are significantly correlated with each other, which generally weakens the value of the over-identification test. However, there is no evidence of collinearity issues between the instruments (the variance-inflation-factor ranges from 1.05 to 1.51), which suggests that there is still a large proportion of independent variation in each instrument. Further support for the instrument exogeneity is suggested by the lack of association between the instruments and selective socioeconomic, demographic and behavioral characteristics including education, income, age, marital status, pregnancy planning and alcohol consumption during the past few years before pregnancy (both number of drinks and binge drinking).<sup>14</sup> All instruments satisfy HWE at p-values ranging from 0.42 to 0.94, suggesting that there is no indication of non-random matching within couples based on these genetic variants that may result in bias when using them as instruments.<sup>15</sup>

<sup>13</sup>See Table A1 in the Appendix.

<sup>14</sup>These associations are summarized in Table A2 in the Appendix.

<sup>15</sup>The HWE test chi-square (1) for instruments rs1435252, rs3780422 and rs1930139 is 0.0077 ( $p=0.93$ ), 0.0049 ( $p=0.945$ ) and 0.416 ( $p=0.412$ ), respectively.



#### IV. B. Cigarette Effects on BMI Mean and Quantiles

Table 3 reports the cigarette effects on BMI mean and quantiles, estimated both in the standard and IV models.<sup>16</sup> Cigarettes have an insignificant effect on BMI mean. Under OLS, the effect is negative and insignificant. Under 2SLS the effect of cigarettes becomes positive and large but remains statistically insignificant ( $p=0.16$ ). The exogenous selection of smoking is rejected at  $p < 0.1$  based on the Durbin test.<sup>17</sup>

When treated as exogenous, cigarettes have overall insignificant negative effects at the evaluated BMI quantiles, except at the 0.9 quantile where cigarettes have an insignificant positive effect. The negative effects decrease (in absolute value) with the quantile order. The effect at the 0.1 quantile is marginally significant.

When treated as endogenous using the IVQR model, the cigarette effects become positive at BMI quantiles 0.1 through 0.75 and negative at the 0.9 quantile. The effects at the 0.1 and 0.9 quantiles are significant using both asymptotic standard errors and 95% weak-instrument robust confidence bounds. Cigarettes decrease the 0.9 BMI quantile by about 0.3 points per cigarette. The positive effects at BMI quantiles 0.25 and 0.5 are significant using the 95% weak-instrument robust confidence bounds, and the effect at quantile 0.25 is marginally significant based on the asymptotic standard errors. Cigarettes increase the 0.1 and 0.25 quantiles by about 0.17 points per cigarette. The effect is larger at the median, with an increase of about 0.9 points in BMI per cigarette. In the study sample, the unconditional 0.1, 0.5, and 0.9 BMI quantiles are 19.6, 22.7 and 28.7, respectively.

#### IV. C. Cigarette Effects on the Body Weight Index

Table 4 reports the marginal cigarette effects on the probabilities of underweight, overweight and obesity relative to normal weight as estimated from the multinomial logit model.<sup>18</sup> When treated as exogenous, cigarettes have a significant positive effect on underweight, but insignificant positive effects on overweight and obesity. When treated as endogenous, the cigarette effects on underweight become negative and marginally significant ( $p=0.06$ ). Cigarettes decrease underweight probability by 0.01 per cigarette. In this model, cigarettes have a marginally significant positive effect on overweight ( $p=0.055$ ), with a 0.03 point increase in overweight probability per cigarette. Finally, the cigarette effect on obesity switches to negative when treated as endogenous, but remains insignificant ( $p=0.3$ ). The exogenous selection of cigarettes is rejected in the underweight ( $p=0.046$ ) and overweight ( $p=0.065$ ) equations based on the significance of the cigarette function residual coefficient.

### V. Discussion and Conclusions

The study finds that smoking may increase BMI for women at low/moderate BMI levels, but may decrease BMI for those at high BMI levels. These heterogeneities are completely masked by evaluating the smoking effects at BMI mean. This effect heterogeneity implies that those with more risk factors for high BMI levels, who are at the extreme right margin of the BMI distribution, may experience weight loss due to smoking, while those with fewer risk factors for high BMI may experience weight gain. These results are overall consistent with those for body weight categories, which indicate that smoking reduces the probability of underweight, increases the probability of overweight and may reduce the probability of obesity.

<sup>16</sup>Tables A1, A3 and A4 in the Appendix report the full regression results.

<sup>17</sup>Durbin Chi-square (1) for exogeneity is 2.715 ( $p = 0.0994$ ).

<sup>18</sup>Table A5 in the Appendix reports the full multinomial logit results.

Several factors may contribute to this heterogeneity in smoking effects. First, smoking may affect certain weight-related biological processes differently at different BMI levels. For example, if smoking enhances the metabolism of individuals at low/moderate BMI levels more than at high BMI levels, and if faster metabolism increases food consumption because of a more frequent feeling of hunger, then smoking may increase BMI at low/moderate BMI levels but not at high BMI levels. Further, if smoking reduces the appetite of individuals at high BMI levels more than individuals at lower BMI levels, then this may also result in negative smoking effects at high BMI levels but not at low/moderate BMI levels. Furthermore, potential interactions between smoking and genetic risk factors for obesity may be more prevalent at high BMI levels and may contribute to this result.

The study suggests that ignoring self-selection into smoking may result in seriously biased estimates of smoking effects including both QTEs and effects on body weight categories. The study results suggest a negative bias in ordinary QTEs at quantiles 0.75 and lower and positive bias at quantile 0.9. The negative bias may be due to reverse BMI effects on smoking if individuals at lower BMI levels choose to smoke more if they perceive this to control their weight, and if those at higher BMI choose to smoke less, such as due to a higher perception of health risks (such as cardiovascular risks). The positive bias at the 0.9 quantile may be due to potentially strong preferences for risk-taking and weak preferences for body fitness at high BMI levels, which increase the positive correlation between smoking and unobserved confounders that increase BMI (such as unhealthy eating). Reverse weight effects on smoking and correlations with unobserved preferences may also explain the observed positive and negative bias in the cigarette effects on underweight and overweight status, respectively. If underweight women are more likely to smoke or to smoke more than normal-weight women in order to control their weight, and if overweight-women are less likely to smoke or to smoke less than normal-weight women because of concerns about health risks, then this may result in the observed bias. Obese individuals may also be less likely to smoke or smoke less than normal weight individuals because of such concerns, but may have a stronger positive correlation between smoking and other risk behaviors that increase obesity status, which may result in a net negative bias, given that the smoking effect switches to negative for obesity in the IV model. Of course, research is needed to test these hypotheses and the potential contributions of such pathways to the causal relationship between smoking and body weight.

The IVQR estimates in this study are opposite to those reported in the study by Fang et al. (2009), in which negative effects of cigarettes on BMI are reported at all quantiles, though they are significant or marginally significant only at quantiles 0.25-0.75. Several factors may contribute to the differences in the smoking effect heterogeneity between the two studies. First, the QR model employed in that study to account for endogenous smoking may result in inconsistent estimates. Second, that study utilizes area-level instruments that may be correlated with other excluded factors that may affect BMI on their own, such as food prices, quality, or nutritional value. Third, the study employs data from another population with potential differences in risk factors for BMI and in the BMI distribution compared with the Norwegian population.

The observed heterogeneity in this study may explain some of the conflicting results of previous studies. The majority of these studies focus on estimating smoking effects on BMI mean, which essentially represent the average of effects across the BMI distribution. Depending on the sample characteristics, model specification, and estimation, the observed differences in “mean effects” between these studies may be in-part due to the masked effect heterogeneity at other BMI distribution locations. This study highlights the importance of assessing the heterogeneity in smoking and cigarette-cost effects on BMI in future studies.

The study provides some evidence that genetic factors may be useful instruments for studying behavioral effects. This is of particular relevance to behaviors for which only area-level instruments have been feasible so far, and for which theoretically valid non-genetic individual-level instruments are generally unavailable, such as smoking. While the employed genetic instruments appear to satisfy the second IV assumption based on the over-identification tests, lack of associations with other socioeconomic and behavioral factors, and the current knowledge of the gene's functions, there is obviously a possibility that the employed genetic instruments may be correlated with other behaviors or health conditions besides smoking that may also affect body weight. Even though there is currently no consistent evidence from the literature about a pathway between *GABBR2* and body weight other than smoking, such a possibility is particularly relevant in the case of neurotransmitter genes, which may be involved in multiple behavioral pathways, as discussed previously for applications using instruments in other neurotransmitter genes (Cawley et al., 2011). As future studies further characterize genes' functions and their effects on health and behavior, it is important to reevaluate the extent to which genetic instruments satisfy the exclusion restrictions based on the new knowledge.

It is important to acknowledge the instruments' weakness and its implications. As discussed above, instrument weakness may bias IV estimates and result in wrong inference. We use an inference approach that is robust to weak instruments. This approach is expected to account for the instrument weakness, given the general support for the exogeneity of the employed instruments.<sup>19</sup> However, weak instruments reduce the power of estimating the effects of the endogenous variable and require large samples, perhaps tens of thousands of individuals for certain applications in order to have sufficient power to detect a reasonable effect size. Using weak instruments and relatively small samples such as in our application may prevent rejecting a false null hypothesis and may have resulted in our inability to reject the hypothesis of no effects of cigarette smoking on BMI mean. Weak instruments may be a common limitation for several applications employing genetic instruments given the complex etiologies of behavioral traits, suggesting a tradeoff between instrument weakness and exogeneity. Therefore, it is important to evaluate this tradeoff for each application. We believe that the instrument exogeneity gain outweighs the weak-instrument limitations for estimating the smoking effects on body weight given the challenges of employing other instruments to identify smoking effects. However, a more definitive evaluation of the relationship between smoking and BMI using genetic instruments likely requires a much larger sample than what is available to this study. This also highlights the importance of establishing consortia that pool multiple data sources with comparable data on behaviors, outcomes, and genetic instruments since no small to moderate size dataset may provide on its own sufficient power for genetic instrumental variable studies.

As mentioned above, it is imperative for studies using genes as instruments to provide both the evidence for and against the use of these specific genes as instruments. There is evidence from previous human and mice studies for a role of the gene employed in this study, *GABBR2*, in smoking behaviors, which is consistent with the observed association in this study. However, there is no evidence from GWAS and at least one other candidate gene study for an association between this gene and smoking. Furthermore, the genetic effects on smoking that we observe are larger than those reported in previous GWAS studies for other genes with more consistent evidence of being involved in smoking such as *CHRNA3*. We have discussed above several potential reasons for the lack of evidence in these other studies

---

<sup>19</sup>We cannot employ weak-instrument robust inference in the residual substitution IV model for the categorical BMI outcome using the bivariate probit model given that weak-instrument confidence bounds are not available for these models. However, given that inference is overall insensitive to the weak-instrument robust approach for the continuous BMI analysis, it is unlikely that the inference for this model is significantly biased by weak instruments.

for *GABBR2* and the large effects we observe in this sample such as genetic heterogeneity, gene-environment interactions, differences in demographic and socioeconomic characteristics between study samples, and the role of pregnancy planning in modifying genetic effects. Of course, we cannot confirm if any of these factors truly explains this inconsistency in the literature without additional studies and replications. As mentioned above, another candidate gene study using a Danish sample with GWAS data found statistically significant associations between several SNPs in *GABBR2* and number of cigarettes smoked during pregnancy, which provides further support for a role of this gene in smoking (Prater et al., 2011). However, further replications are needed in other samples for the evidence to be considered more consistent. It is also important to estimate the effects of smoking on BMI using as instruments other genetic variants that have been recently found to be strongly linked to smoking behaviors such as variants in *CHRNA3* and are now considered to have unequivocal evidence of being associated with smoking (Liu et al., 2010; The Tobacco and Genetics Consortium, 2010). These variants have not yet been typed in our study sample, but we hope to evaluate their utility as instruments for smoking in future studies.

Another limitation that should be considered is our inability to evaluate separately the effects of any smoking versus smoking intensity among smokers as the instruments are not predictive of the probability of any smoking participation. We leave it to future studies to evaluate the effects of alternative smoking measures on body weight using genetic instruments that are significantly predictive of these measures. Finally, it is possible that there are recall errors or biases in measuring body weight before pregnancy based on self-report after pregnancy. This may increase the standard errors and reduce the significance of the smoking effects.

In conclusion, the study employs a novel approach that integrates genetic information within an econometric framework to estimate the smoking effects on body weight and evaluate the effect heterogeneity. The application highlights the advantages and challenges of employing genetic instruments to identify behavior effects and is informative for related applications in health economics and other social science disciplines. The study results suggests tradeoffs in effects on BMI with decreasing smoking: a decrease in the weight of low/moderate weight women (*negative effect*), an increase in the underweight proportion (*negative effect*), a decrease in the overweight proportion (*positive effect*) and an increase in the weight of overweight/obese women (*negative effect*). This is not to say that policies that increase cigarette costs are suboptimal as this analysis considers only the smoking effects on body weight and for a specific population group. Furthermore, the results are subject to several qualifications related to the appropriateness of the employed instruments and adequacy of the sample size as discussed above. While there is some evidence supporting the use of this gene, the weakness of the instruments reduces the study power. Also, there is a possibility that the instruments may not be entirely exogenous which would bias the estimated effects of smoking on BMI in unknown directions. However, the application and results highlight the importance of evaluating this question in additional studies using larger and more diverse samples and stronger instruments.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Data analysis was supported in part by NIH/NIDCR grants R03 DE018394 and R01 DE20895. The authors thank research seminar participants and Robert Wallace at the University of Iowa and Shin-Yi Chou at Lehigh University for helpful comments.

## References

- Agrawal A, et al. Further evidence for an association between the gamma-aminobutyric acid receptor A, subunit 4 genes on chromosome 4 and Fagerstrom Test for Nicotine Dependence. *Addiction*. 2009; 104:471–477. [PubMed: 19207358]
- Agrawal A, et al. Gamma-aminobutyric acid receptor genes and nicotine dependence: evidence for association from a case-control study. *Addiction*. 2008; 103:1027–1038. [PubMed: 18482426]
- Basman RL. On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics. *Journal of the American Statistical Association*. 1960; 55:650–659.
- Baum CL. The effects of cigarette costs on BMI and obesity. *Health Econ*. 2009; 18:3–19. [PubMed: 18241088]
- Baum CL 2nd, Ford WF. The wage effects of obesity: a longitudinal study. *Health Econ*. 2004; 13:885–899. [PubMed: 15362180]
- Berrettini W, et al. Alpha-5/alpha-3 nicotinic receptor subunit alleles increase risk for heavy smoking. *Mol Psychiatry*. 2008; 13:368–373. [PubMed: 18227835]
- Beuten J, Ma JZ, Payne TJ, Dupont RT, Crews KM, Somes G, Williams NJ, Elston RC, Li MD. Single- and multilocus allelic variants within the GABA(B) receptor subunit 2 (GABAB2) gene are significantly associated with nicotine dependence. *Am J Hum Genet*. 2005; 76:6.
- Bhattacharya J, Bundorf MK. The incidence of the healthcare costs of obesity. *Journal of health economics*. 2009; 28:649–658. [PubMed: 19433210]
- Caporaso N, et al. Genome-wide and candidate gene association study of cigarette smoking behaviors. *PLoS One*. 2009; 4:e4653. [PubMed: 19247474]
- Carmelli D, et al. Genetic influence on smoking--a study of male twins. *N Engl J Med*. 1992; 327:829–833. [PubMed: 1508241]
- Cawley J, et al. The validity of genes related to neurotransmitters as instrumental variables. *Health Econ*. 2011; 20:884–888. [PubMed: 21612000]
- Cawley J, et al. Lighting up and slimming down: the effects of body weight and cigarette prices on adolescent smoking initiation. *Journal of Health Economics*. 2004; 23:293–311. [PubMed: 15019756]
- Chen H, et al. Long-term cigarette smoke exposure increases uncoupling protein expression but reduces energy intake. *Brain research*. 2008; 1228:81–88. [PubMed: 18619427]
- Chen H, et al. Effect of short-term cigarette smoke exposure on body weight, appetite and brain neuropeptide Y in mice. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*. 2005; 30:713–719. [PubMed: 15508020]
- Cheng K, Kenkel D. U.S. cigarette demand: 1944-2004. *Journal of Economic Analysis and Policy: Contributions to Economic Analysis and Policy*. 2010; 10
- Chernozhukov V, Hansen C. The effects of 401(k) participation on the wealth distribution: an instrumental quantile regression analysis. *Review of Economics and Statistics*. 2004; 86:735–751.
- Chernozhukov V, Hansen C. An IV Model of Quantile Treatment Effects. *Econometrica*. 2005; 73:245–261.
- Chernozhukov V, Hansen C. Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics*. 2006; 132:491–525.
- Chernozhukov V, Hansen C. The reduced form: a simple approach to inference with weak instruments. *Economics Letters*. 2008; 100:68–71.
- Chernozhukov V, et al. Inference approaches for instrumental variable quantile regression. *Economics Letters*. 2007; 95:272–277.
- Chou S, et al. Reply to Jonathan Gruber and Michael Frakes. *Journal of Health Economics*. 2006; 25:389–393.
- Chou SY, et al. An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System. *Journal of Health Economics*. 2004; 23:565–587. [PubMed: 15120471]
- Comings DE, F. L, Bradshaw-Robinson S, Burchette R, Chiu C, Muhleman D. The dopamine D2 receptor (DRD2) gene: a genetic risk factor in smoking. *Pharmacogenetics*. 1996; 6:73–79. [PubMed: 8845863]



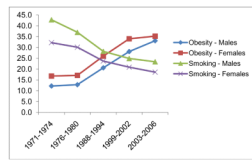
- Comings DE, et al. Cholecystokinin (CCK) Gene as a Possible Risk Factor for Smoking: A Replication in Two Independent Samples. *Molecular Genetics and Metabolism*. 2001; 73:349–353. [PubMed: 11509018]
- Courtemanche C. Rising Cigarette Prices and Rising Obesity: Coincidence or Unintended Consequence? *Journal of Health Economics*. 2009; 28:781–798. [PubMed: 19356816]
- Ding W, et al. The impact of poor health on academic performance: New evidence using genetic markers. *J Health Econ*. 2009; 28:578–597. [PubMed: 19217678]
- Fang H, et al. Does smoking affect body weight and obesity in China? *Econ Hum Biol*. 2009; 7:334–350. [PubMed: 19660996]
- Fatemi SH, et al. Expression of GABA(B) receptors is altered in brains of subjects with autism. *Cerebellum*. 2009; 8:64–69. [PubMed: 19002745]
- Fatemi SH, et al. Deficits in GABA(B) receptor system in schizophrenia and mood disorders: a postmortem study. *Schizophrenia research*. 2011; 128:37–43. [PubMed: 21303731]
- Finkelstein EA, et al. Annual medical spending attributable to obesity: payer-and service-specific estimates. *Health affairs (Project Hope)*. 2009; 28:w822–831. [PubMed: 19635784]
- Flegal KM, et al. Prevalence and trends in obesity among US adults, 1999–2008. *Jama*. 2010; 303:235–241. [PubMed: 20071471]
- Fletcher JM. The promise and pitfalls of combining genetic and economic research. *Health Econ*. 2011; 20:889–892. [PubMed: 21612001]
- Freathy RM, et al. A common genetic variant in the 15q24 nicotinic acetylcholine receptor gene cluster (CHRNA5-CHRNA3-CHRNA4) is associated with a reduced ability of women to quit smoking in pregnancy. *Hum. Mol. Genet*. 2009 ddp216.
- Gruber J, Frakes M. Does falling smoking lead to rising obesity? *Journal of health economics*. 2006; 25:183–197. discussion 389–193. [PubMed: 16165234]
- Hahn J, Hausman J. Weak instruments: Diagnosis and cures in empirical econometrics. *American Economic Review*. 2003; 93:118–125.
- Han E, et al. Weight and wages: fat versus lean paychecks. *Health Econ*. 2009; 18:535–548. [PubMed: 18677723]
- Hausman, JA. Specification and estimation of simultaneous equation models. 1983.
- Heath AC, Martin NG. Genetic models for the natural history of smoking: evidence for a genetic influence on smoking persistence. *Addict Behav*. 1993; 18:19–34. [PubMed: 8465673]
- Hjelmberg JB, et al. Genetic influences on growth traits of BMI: a longitudinal study of adult twins. *Obesity (Silver Spring)*. 2008; 16:847–852. [PubMed: 18239571]
- Keskitalo K, et al. The Three-Factor Eating Questionnaire, body mass index, and responses to sweet and salty fatty foods: a twin study of genetic and environmental associations. *Am J Clin Nutr*. 2008; 88:263–271. [PubMed: 18689360]
- Koenker R, Bassett G Jr. Regression Quantiles. *Econometrica*. 1978; 46:33–50.
- Koenker R, Hallock KF. Quantile Regression. *Journal of Economic Perspectives*. 2001; 15:143–156.
- Komlos, J.; Brabec, M. The Trend of Mean BMI Values of US Adults, Birth Cohorts 1882–1986 Indicates that the Obesity Epidemic Began Earlier than Hitherto Thought. 2010. National Bureau of Economic Research Working Paper Series No. 15862
- Lawlor DA, et al. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med*. 2008; 27:1133–1163. [PubMed: 17886233]
- Lerman C, et al. Tryptophan hydroxylase gene variant and smoking behavior. *Am J Med Genet*. 2001; 105:518–520. [PubMed: 11496367]
- Lessov CN, et al. Defining nicotine dependence for genetic research: evidence from Australian twins. *Psychol Med*. 2004; 34:865–879. [PubMed: 15500307]
- Li MD. The genetics of nicotine dependence. *Curr Psychiatry Rep*. 2006; 8:158–164. [PubMed: 16539894]
- Li MD, et al. Ethnic- and gender-specific association of the nicotinic acetylcholine receptor alpha4 subunit gene (CHRNA4) with nicotine dependence. *Hum Mol Genet*. 2005; 14:1211–1219. [PubMed: 15790597]

- Li MD, et al. Association and interaction analyses of GABBR1 and GABBR2 with nicotine dependence in European- and African-American populations. *PLoS One*. 2009; 4:e7055. [PubMed: 19763258]
- Lie RT, et al. Maternal smoking and oral clefts: the role of detoxification pathway genes. *Epidemiology*. 2008; 19:606–615. [PubMed: 18449058]
- Liu JZ, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet*. 2010; 42:436–440. [PubMed: 20418889]
- Ma JZ, et al. Haplotype analysis indicates an association between the DOPA decarboxylase (DDC) gene and nicotine dependence. *Hum. Mol. Genet*. 2005; 14:1691–1698. [PubMed: 15879433]
- Maes HH, et al. A twin study of genetic and environmental influences on tobacco initiation, regular tobacco use and nicotine dependence. *Psychol Med*. 2004; 34:1251–1261. [PubMed: 15697051]
- Martinez AJ, et al. Genetics of obesity. *Public Health Nutr*. 2007; 10:7. [PubMed: 17212837]
- McKinney EF, et al. Association between polymorphisms in dopamine metabolic enzymes and tobacco consumption in smokers. *Pharmacogenetics*. 2000; 10:483–491. [PubMed: 10975602]
- National Center for Health Statistics (NCHS). *Health, United States, 2009: With special feature on medical technology*. Library of Congress; Hyattsville, MD: 2010.
- NICSNP. The NICSNP Nicotine Project. National Institute on Drug Abuse; 2007.
- NIEHS. Norway Facial Cleft Study (NCL); 2009.
- Noble EP, S.J. S, Ritchie T, Sydulko K, St Jeor SC, Fitch RJ, Brunner RL, Sparkes RS. D2 dopamine receptor gene and cigarette smoking: a reward gene? *Med Hypotheses*. 1994; 42:257–260. [PubMed: 8072432]
- Nonemaker J, et al. Have efforts to reduce smoking really contributed to the obesity epidemic? *Economic Inquiry*. 2009; 47:366–376.
- Norton EC, Han E. Genetic Information, Obesity, and Labor Market Outcomes. *Health Econ*. 2008; 17:1089–1104. [PubMed: 18615836]
- Pianezza ML, et al. Nicotine metabolism defect reduces smoking. *Nature*. 1998; 393:750. [PubMed: 9655391]
- Prater, KN., et al. *Evaluating the Utility of Genetic Instruments for Studying Prenatal Behavior Effects Using Genome Wide Association Data*. University of Iowa, Department of Health Management and Policy; Iowa City, Iowa: 2011.
- Rashad I. Structural Estimation of Caloric Intake, Exercise, Smoking, and Obesity. *Quarterly Review of Economics and Finance*. 2006; 46:268–283.
- Rees DI, Sabia JJ. Body weight and smoking initiation: evidence from Add Health. *Journal of Health Economics*. 2010; 29:774–777. [PubMed: 20708283]
- Sellers EM, et al. Inhibition of cytochrome P450 2A6 increases nicotine's oral bioavailability and decreases smoking. *Clin Pharmacol Ther*. 2000; 68:35–43. [PubMed: 10945314]
- Smith GD, et al. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med*. 2007; 4:e352. [PubMed: 18076282]
- Smith, TG., et al. *Why the poor get fat: Weight gain and economic insecurity* Forum for Health Economics & Policy. 2009.
- Spitz MR, et al. Case-control study of the D2 dopamine receptor gene and smoking status in lung cancer patients. *J Natl Cancer Inst*. 1998; 90:358–363. [PubMed: 9498485]
- Stock, JH.; Yogo, M. *Testing for weak instruments in linear IV regression, Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*. Cambridge University Press; Cambridge and New York: 2005.
- Sullivan PF, Jiang Yuxin, Neale Michael C. Kendler Kenneth S. Straub Richard E. Association of the tryptophan hydroxylase gene with smoking initiation but not progression to nicotine dependence. *American Journal of Medical Genetics*. 2001; 105:479–484. [PubMed: 11449402]
- Sun D, et al. Regulation by nicotine of Gpr51 and Ntrk2 expression in various rat brain regions. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*. 2007; 32:110–116. [PubMed: 16794563]
- Terza JV, et al. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of health economics*. 2008; 27:531–543. [PubMed: 18192044]

- The Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet.* 2010; 42:441–447. [PubMed: 20418890]
- Tholin S, et al. Genetic and environmental influences on eating behavior: the Swedish Young Male Twins Study. *Am J Clin Nutr.* 2005; 81:564–569. [PubMed: 15755823]
- Thorgeirsson TE, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature.* 2008; 452:638–642. [PubMed: 18385739]
- Tyndale R. Genetics of alcohol and tobacco use in humans. *Annals of Medicine.* 2003; 35:94–122. [PubMed: 12795339]
- von Hinke Kessler Scholder S, et al. Mendelian randomization: the use of genes in instrumental variable analyses. *Health Economics.* 2011; 20:893–896. [PubMed: 21612002]
- Watson NF, et al. Genetic and environmental influences on insomnia, daytime sleepiness, and obesity in twins. *Sleep.* 2006; 29:645–649. [PubMed: 16774154]
- Wehby GL, et al. Quantile effects of prenatal care utilization on birth weight in Argentina. *Health Econ.* 2009; 18:1307–1321. [PubMed: 19142894]
- Wehby GL, et al. 'Mendelian randomization' equals instrumental variable analysis with genetic instruments. *Stat Med.* 2008; 27:2745–2749. [PubMed: 18344186]
- Wigginton JE, Abecasis GR. PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics.* 2005; 21:3445–3447. [PubMed: 15947021]
- Wooldridge, JM. *Econometric analysis of cross section and panel data.* MIT Press; Cambridge and London: 2002.
- Yu Y, et al. Intronic variants in the dopa decarboxylase (DDC) gene are associated with smoking behavior in European-Americans and African-Americans. *Hum Mol Genet.* 2006; 15:2192–2199. [PubMed: 16740595]

### Highlights

- Smoking may have heterogeneous effects on low versus high body weight quantiles.
- Genetic instruments may be useful for identifying behavioral effects.
- Instruments should be validated through literature review and statistical tests.
- Replication with large samples and strong instruments is needed.
- The study highlights the advantages and challenges of genetic instruments.



**Figure 1.**  
 Adult Smoking and Obesity Rates in 1971-2006 in the US  
 Note: Rates are from the “Health, United States, 2009: With Special Feature on Medical Technology “. National Center for Health Statistics, NCHS, (2010).



**Table 1**

Description of Study Variables

| Variable                          | Description  | Mean   | Std. Dev. | Min  | Max  |
|-----------------------------------|--|--------|-----------|------|------|
| BMI                               | Body Mass Index  | 23.6   | 4.0       | 14.3 | 47.3 |
| Underweight                       | Binary (0,1) indicator for BMI < 18.5  | 0.036  | 0.186     | 0    | 1    |
| Overweight                        | Binary (0,1) indicator for 25 ≤ BMI < 29.9   | 0.194  | 0.396     | 0    | 1    |
| Obesity                           | Binary (0,1) indicator for BMI > 29.9  | 0.077  | 0.266     | 0    | 1    |
| Cigarettes                        | Number of cigarettes per day (12 months before pregnancy)  | 4.2    | 6.4       | 0    | 40   |
| Age                               | Age in years   | 29.128 | 4.754     | 17   | 44   |
| Married                           | Binary (0,1) indicator for married mothers   | 0.516  | 0.500     | 0    | 1    |
| Less than high school<br><i>a</i> | Binary (0,1) indicator for maternal education of less than completed high school                         | 0.120  | 0.325     | 0    | 1    |
| High school<br><i>a</i>           | Binary (0,1) indicator for maternal education of completed high school                                   | 0.257  | 0.437     | 0    | 1    |
| Technical college<br><i>a</i>     | Binary (0,1) indicator for mother attending technical college  | 0.194  | 0.396     | 0    | 1    |
| University<br><i>a</i>            | Binary (0,1) indicator for mother completing university education  | 0.061  | 0.240     | 0    | 1    |
| Very low income<br><i>b</i>       | Binary (0,1) indicator for current gross maternal yearly income of less than 150,000 kr                  | 0.378  | 0.485     | 0    | 1    |
| Moderate income<br><i>b</i>       | Binary (0,1) indicator for current gross maternal yearly income between 201,000 and 250,000 kr inclusive | 0.235  | 0.424     | 0    | 1    |
| High income<br><i>b</i>           | Binary (0,1) indicator for current gross maternal yearly income of 251,000 kr or more                    | 0.135  | 0.342     | 0    | 1    |
| rs1435252_TT<br><i>c</i>          | Binary (0,1) indicator of TT genotype of rs1435252   | 0.083  | 0.276     | 0    | 1    |
| rs1435252_CT<br><i>c</i>          | Binary (0,1) indicator of CT genotype of rs1435252   | 0.412  | 0.493     | 0    | 1    |
| rs3780422_AA<br><i>c</i>          | Binary (0,1) indicator of AA genotype of rs3780422   | 0.201  | 0.401     | 0    | 1    |
| rs3780422_AG<br><i>c</i>          | Binary (0,1) indicator of AG genotype of rs3780422   | 0.496  | 0.5       | 0    | 1    |
| rs1930139_GG<br><i>c</i>          | Binary (0,1) indicator of GG genotype of rs1930139   | 0.052  | 0.222     | 0    | 1    |
| rs1930139_AG<br><i>c</i>          | Binary (0,1) indicator of AG genotype of rs1930139   | 0.333  | 0.472     | 0    | 1    |

*a*The reference category is 2-4 years of college

*b*The reference category is 151,000-200,000 kr

<sup>c</sup>The reference category for the SNP indicators is either CC, GG or AA.

**Table 2**

## Regression Coefficients of the Cigarette Function

|   |                                |
|---|--------------------------------|
| rs1435252_TT                            | -1.29 <sup>*</sup><br>(0.72)   |
| rs1435252_CT                            | 0.74 <sup>*</sup><br>(0.41)    |
| rs3780422_AA                            | 0.83<br>(0.57)                 |
| rs3780422_AG                            | -0.52<br>(0.44)                |
| rs1930139_GG                            | -0.01<br>(0.85)                |
| rs1930139_AG                            | 0.84 <sup>**</sup><br>(0.40)   |
| Maternal age                            | -0.06<br>(0.04)                |
| Married                                 | -2.36 <sup>***</sup><br>(0.39) |
| Less than high school                   | 4.28 <sup>***</sup><br>(0.66)  |
| High school                             | 0.81<br>(0.51)                 |
| Technical college                       | 2.67 <sup>***</sup><br>(0.55)  |
| University                              | -0.40<br>(0.84)                |
| Very low maternal income                | 0.09<br>(0.48)                 |
| Moderate maternal income                | 0.28<br>(0.56)                 |
| High maternal income                    | 0.06<br>(0.70)                 |
| Constant                                | 5.43 <sup>***</sup><br>(1.36)  |
| F Statistic for instruments<br>(6,1041) | 3.4 <sup>***</sup>             |

Note: This Table reports the regression coefficients of the cigarette function estimated by OLS. Standard errors are in parentheses.

\* indicate significance at  $p < 0.1$ , respectively.

\*\* indicate significance at  $p < 0.05$ , respectively.

\*\*\* indicate significance at  $p < 0.01$ , respectively.

**Table 3**

Cigarette Effects on BMI

|                    | Mean Effect                         | Quantile Effect / Quantiles            |                                     |                                     |                                       |  |
|--------------------|-------------------------------------|--|-------------------------------------|-------------------------------------|---------------------------------------|--|
|                    |                                     | 0.1                                    | 0.25                                | 0.5                                 | 0.75                                  | 0.9                                      |
| Smoking exogenous  | -0.020<br>(0.020)                   | -0.038*<br>(0.022)                     | -0.024<br>(0.018)                   | -0.015<br>(0.021)                   | -0.013<br>(0.037)                     | 0.043<br>(0.063)                         |
| Smoking endogenous | 0.218<br>(0.155)<br>[-0.324, 1.397] | 0.173**<br>(0.074)<br>[0.156, 0.208]** | 0.171*<br>(0.093)<br>[0.1, 0.243]** | 0.902<br>(0.62)<br>[0.002, 1.532]** | 0.103<br>(0.253)<br>[-0.177, 1.643]** | -0.307**<br>(0.133)<br>[-0.326, -0.15]** |

Note: This table presents the smoking effects on BMI means and quantiles. Standard errors of effects are in parentheses. 95% confidence bounds that are robust for weak instruments are in brackets.

\* indicate significance at  $p < 0.1$ , respectively.

\*\* indicate significance at  $p < 0.05$ , respectively.

**Table 4**

## Cigarette Effects on Weight Category Probabilities

|                          | <b>Underweight</b>   | <b>Overweight</b>   | <b>Obesity</b>      |
|--------------------------|----------------------|---------------------|---------------------|
| Cigarettes<br>exogenous  | 0.0015**<br>(0.0007) | 0.0007<br>(0.0021)  | 0.0004<br>(0.0012)  |
| Cigarettes<br>endogenous | -0.0108*<br>(0.0058) | 0.0317*<br>(0.0165) | -0.0089<br>(0.0088) |

Note: This table presents the marginal cigarette effects on the probabilities of weight categories relative to normal weight as estimated from the multinomial logit model.

\* indicate significance at  $p < 0.1$ , respectively.

\*\* indicate significance at  $p < 0.05$ , respectively.