HEDG Health, Econometrics and Data Group
THE UNIVERSITY of York

HEDG Working Paper 07/15

Genetic Information, Obesity, and Labor Market Outcomes

Edward C. Norton

Euna Han

July 2007

ISSN 1751-1976

york.ac.uk/res/herc/hedgwp

Genetic Information, Obesity, and Labor Market Outcomes

Edward C. Norton
Professor, Department of Health Policy and Administration, and
Professor, Department of Economics
University of North Carolina at Chapel Hill

Euna Han
Post-doctoral Fellow, Institute for Health Research and Policy
University of Illinois at Chicago

July 2007

Please address correspondence to:

Edward C. Norton
Department of Health Policy and Administration
School of Public Health
CB #7411, McGavran-Greenberg Building
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-7411
USA

Tel: 919-966-8930
Fax: 919-966-6961
E-mail: edward_norton@unc.edu

SUMMARY

Economists have argued that obesity may lead to worse labor market outcomes, especially for women. Empirical methods to test this hypothesis have not thus far adequately controlled for the endogeneity of obesity. We use variation in genotype to predict variation in phenotype (obesity). Genetic information from specific genes linked to obesity in the biomedical literature provide strong exogenous variation in the body mass index, and thus can be used as instrumental variables. These genes predict swings in weight of between 5 and 20 pounds for persons between five and six feet tall. We use additional genetic information to control for omitted variables correlated with both obesity and labor market outcomes. We analyzed data from the third wave of the Add Health data set, when respondents are in their mid-twenties. Results from our preferred models show no effect of obesity on the probability of employment or on wages, for either men or women. This paper shows the potential of using genetic information in social sciences.

Keywords: Obesity, genetics, labor market outcomes, employment, wages, Add Health

Acknowledgements

This project was funded by the Demography and Economics of Aging Research Center grant from the National Institute on Aging (P30 AG04001). We thank participants in seminars at the Triangle Health Economics Workshop, Harvard University, Michigan University, the Southeastern Health Economics Working Group, and the International Health Economics Association World Congress for their helpful comments. David Bishai, David Blau, and Phil DeCicca also provided thoughtful suggestions.

This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu).

INTRODUCTION

Over the last several decades, the obesity rate among American adults has risen steadily. The majority are now overweight. The economic consequences of obesity, beyond worsening morbidity and mortality, may include worse probability of employment and lower wages. Obesity may affect these labor market outcomes because employers may want to avoid obese employees if obese employees have lower productivity due to their obesity, higher health care costs, or a higher discount rate manifested by lower investment in human capital (Baum and Ford, 2004). On the demand side, consumers may shun obese employees, particularly in service industries.

For all these reasons, economists have tried to measure empirically how obesity affects wages. The results vary widely, in part due to different econometric methods to deal with endogeneity (e.g., Averett and Korenman, 1996; Cawley, 2000; Behrman and Rosenzweig, 2001; Baum and Ford, 2004; Cawley, 2004; Conley and Glauber, 2005). Some studies find a wage penalty for obesity, others find no effect. The early studies of obesity on labor market outcomes ignored the endogeneity of obesity, making causal inference impossible. More recent studies have used weak instruments or just-identified models. Therefore, finding either good instruments that are correlated strongly with obesity and can be shown to be validly excluded from the main equation or other variables that predict both and have traditionally been excluded would advance the field.

This study uses genetic information to identify the effect of obesity on labor market outcomes. Genes that change the propensity of a person to be obese provide natural variation, as long as those genes are not correlated with other behavioral factors. Instrumental variables created from this genetic information allow us to control for the

endogeneity of obesity and obtain consistent estimates of the causal effect of obesity on labor market outcomes. In addition, genetic information that is correlated with obesity—but also is correlated with the main outcome—still helps reduce omitted variable bias.

Genetic information holds great promise for social science research. Many large surveys now routinely collect biomarkers, including genetic information. In the past, economists have used information on siblings and twins as controls or instruments by appealing to the argument that biological siblings and twins share many genes. Using genetic information can greatly improve on this approach if the genes are targeted to the endogenous variable, as they are in our study. Furthermore, genes are plausibly exogenous, being determined at conception. As one of the first economic studies to use genetic information, we also discuss the limitations of genetic information, including the importance of testing the exogeneity assumption rigorously.

We analyze data from the National Longitudinal Study of Adolescent Health (Add Health). In Wave III of Add Health, a subset of 2,612 of the Add Health respondents (total of about 17,000) contributed DNA samples. Six genes identified from the DNA samples were chosen specifically because they are believed to be related to obesity and to have a relatively high prevalence in the population. Polymorphisms in these six genes have been linked to obesity through behavior (Blundell 1977; Hoebel et al., 1989). All affect how the central nervous system regulates satiation and eating behavior. Several pass all specification tests for good instruments; others were used to control for omitted variable bias.

Results from our preferred models show no statistically significant effect of obesity either the probability of employment or wages conditional on employment, for either men or women. The genetic information provides strong instrumental variables, but in our sample of adults in their mid-twenties, lagged measures of obesity are generally not endogenous.

BACKGROUND

Obesity

The dramatic growth in obesity (or the state of being overweight) has been an important concern for policymakers and the public over the last several decades. Annual trends show that the average body mass index (BMI) increased by 9% during 1984–1998, and that the number of obese adults more than doubled during the same period (Chou, Grossman, and Saffer, 2002). The increase in the proportion of being overweight and obese affects all ages, racial and ethnic groups, and both genders (U.S. Department of Health and Human Services, 2001). The prevalence of obesity is higher than that of smoking, use of illegal drugs, or other risk factors for most of the highly prevalent chronic diseases, including heart disease, diabetes, and cancer (Philipson, 2001).

Several economic factors contribute to the increasing rate of obesity. First, low income is related to being overweight or obese due to a price effect (Stunkard, 1996; Sobal and Stunkard 1989). Fast food and convenience foods are both inexpensive and are higher in calories compared to other healthier foods. People with lower income are more likely to consume fattening foods (Chou, Grossman, and Saffer, 2002). Second, technological change has dramatically reduced both the fixed and variable costs of mass

preparation of foods. Furthermore, it has reduced the time delay before actual consumption of food, which particularly affects people with self-control problems (Cutler, Glaeser, and Shapiro, 2003). Third, trends in energy intake have also changed since the 1970s. Between 1977 and 2001, energy intake from sweetened beverage consumption increased by 135%, while energy intake from milk consumption fell by 38% for samples aged 2 to 60 years. This corresponds to an increase of 278 calories per day during the same period. Food portions have increased both inside and outside the home (Nielsen and Popkin, 2004; Popkin, et al., 1996). Fourth, income growth also encourages a sedentary lifestyle by increasing the cost of physical activity during leisure time. This leads to greater consumption of convenience or fast food (Lakdawalla and Philipson, 2002; Chou, Grossman, and Saffer, 2002).

Baum and Ford (2004) identified four ways that obesity affect labor market outcomes: health problems caused by obesity; myopia of obese individuals; consumer-based discrimination against obese workers in the labor market; and employer-based distaste for obese workers with regard to high health care cost for obese people or for other factors associated with weight and job productivity. First, low productivity of obese people may result from their health problems directly associated with obesity. The previous literature has consistently reported health problems and a high health care cost caused by obesity.

Second, individuals may discount the near future less heavily than the long-term future when they make decisions over time intervals (Becker and Murphy, 1988). If the individuals do not realize that their time preferences are inconsistent, then they become myopic with inconsistent or imprudent planning (Strotz, 1955). If individuals are

myopic, they ignore future effects when they make decisions about current consumption (Becker, Grossman, and Murphy, 1994). Food consumption brings immediate gratification, while costs of over-consumption of food occur in the future (Cutler, Glaeser, and Shapiro, 2003). Therefore, myopic workers are less likely to be concerned about the long-term adverse health effects of consuming fattening foods than non-myopic workers, and accordingly, more likely to be obese (Cawley, 2000). Like the choice for food consumption, those people will ignore future return to the investments on their human capital, such as on-the-job training, when they make decisions about current consumption of those investments.

Third, there may be some occupations where non-obese workers are more productive than obese workers due to consumer-based discrimination against obese workers. For example, consumers may prefer a slim sales representative in a beauty shop to an obese one. Results from an experiment demonstrated that employers perceived obese persons as unfit for public sales positions and as more appropriate for telephone sales involving little face-to-face contact (Puhl and Brownell, 2001). In another experiment, participants rated obese job applicants as lacking self-discipline, having low supervisory potential, and having poor personal hygiene and professional appearance (Martin, 1990). Consumers value goods based on the attributes or characteristics of those goods affecting their utility (Rosen, 1974). If those economic agents have a propensity for slim sellers, products sold by obese sellers may impose more constraint on their utility maximization due to high perceived price (Buffum and Whaples, 1995). Under the assumption that some individuals have a propensity for discrimination against obese sellers, while obese sellers are indifferent about the sliminess of the buyers, a consumer

with a distaste for obese sellers will perceive the price as being $P(1+d)$, where d is the discrimination coefficient and P is the monetary price. Discrimination coefficient d will measure the intensity of the propensity for discrimination against the obese seller (Becker, 1971).

Fourth, employers may have a distaste for obese employees for reasons including their own preferences for lean employees, consideration of their consumers' distaste for obese workers, belief of different ability to do jobs between obese and non-obese employees, or their concerns about rising employer-provided health insurance costs.

Collectively, these argue that obesity should reduce wages and the probability of employment, if there is any effect at all. Furthermore, the marginal effect of an increase in BMI may differ at different ends of the spectrum, being perhaps positive for waifs and negative for the vast majority who are in the normal to high weight range. We are not able to test the relative strengths of the contributions of each factor in our model. But we do acknowledge that in our data set of young men and women, the long-term health costs are probably least important. Like the prior empirical work, we are interested in estimating the magnitude of the effect of obesity on labor market outcomes.

Prior Empirical Work

Several studies have linked obesity to labor market outcomes, mostly wages. Even though all of those studies essentially used the same data, the NLSY79, their results differ markedly. These inconsistent trends in the previous literature may be attributed to the lack of valid control for the endogeneity of obesity (Register and Williams, 1992; Loh, 1993; Pagan and Davila, 1997; Gortmaker et al., 1993; Sargent and Blanchflower, 1994).

Recently, a few studies reported statistically significant negative effects of obesity on wages when they tried to control for the endogeneity of obesity. However, they used weak or just-identified instruments. Averett and Korenman (1996) replaced current body weight with a lagged body weight and controlled sister-fixed effects when estimating the effect of body weight on wages using the 1988 survey of the NLSY79. Women were estimated to suffer obesity penalties for their earnings. Similarly, Conley and Glauber (2005) took the lag of a 13 and 15 years of BMI as instruments for the current BMI, and used the sibling fixed-effects model. Using the Panel Study of Income Dynamics (PSID) 1986, 1999, and 2001 data, they estimated that obesity penalizes women for not only their own earnings, family income, and occupational prestige, but also spouse's earnings and spouse's occupational prestige. Although the sister or siblings fixed -effects sweep out the unobserved permanent endowment factors at the family level belonging to the error term, individually heterogeneous endowment factors will remain unobserved in the error term.

Behrman and Rosenzweig (2001) did not find a statistically significant effect of BMI on wages in the labor market with a survey on identical female twins from a sample from the Minnesota Twins Registry. In this study, the identical twin fixed-effects model

was used. In addition, the authors used lagged consumption or the lagged physical characteristics as an instrument for current BMI and height to control for the contemporaneous wage shocks in the error term. The results of this specification showed near zero effect of BMI on wages, while height has a statistically significant and strong positive effect on wages.

Cawley (2000) estimated the effect of obesity on women's employment disability, which was measured by limitations on the amount of paid work and limitations on types of paid work using 12 years of data from the NLSY79. When the sample women's own child's body weight was used as an instrument for the sample women's body weight, obesity did not have a statistically significant effect on a limitation on the amount of paid work nor a limitation on the type of paid work. However, the validity of the instrument remains untested, and children's body weight will not be a valid instrument for mother's body weight if there are unobserved heterogeneity for mothers in the wage residual, which affects both children's body weight and the mother's employment disability. For example, smoking or alcohol consumption during pregnancy may reflect the pregnant women's inconsistent discount rate for the future, which will affect both their performance in their job and their children's health at birth.

In another study, Cawley (2004) estimated the effect of obesity on the wage rate using 12 years of data from the NLSY79. Siblings' body weight was used as an instrument for sample persons' body weight, and the individual and sibling fixed-effects model were estimated. Cawley found obesity penalties in the labor market in terms of wages only for white women. However, validity of the instrument could not be tested.

In a study using 12 years of the NLSY79, Baum and Ford (2004) tested the four potential pathways linking obesity to labor market outcomes. Their empirical evidence suggested that those pathways might mediate the effect of obesity on labor market outcomes. However, the authors did not control for the endogeneity of obesity other than using the individual and family fixed-effects model. If any explanatory variables are not strictly exogenous, the fixed-effects estimators are inconsistent.

Bhattacharya and Bundorf (2004) studied on health care cost differentials as an explanation for wage differential between obese and non-obese people. They compared wage differentials between obese and non-obese individuals in employment with employer-provided health insurance to the wage differentials between obese and non-obese individuals in employment without employer-provided health insurance. Their difference-in-difference estimator using the NLSY79 data during 1989-1998 showed no statistically different wage differential between obese and non-obese individuals. However, they did not control for the endogeneity of either obesity or employer-provided health insurance. Obese individuals may choose a job with employer-provided health insurance over a job without employer-provided health insurance, recognizing that they have high need for health care services due to obesity. Jobs with employer-provided health insurance also may have different unobserved characteristics from the jobs without employer-provided health insurance, such as maternal leave or pension programs.

Genetics

To understand this study, we must explain a few key concepts of genetics. A *gene* is a unit of heredity, consisting of a string of DNA. Functionally, genes regulate the

production of proteins. Because human chromosomes, which are strings of genetic material divided into genes, come in pairs, each gene has two copies. *Alleles* are variations of the same gene. Humans have two alleles of each gene. The *genotype* is the specific genetic makeup of an individual. The genotype combined with environmental factors produce observable characteristics called *phenotypes*, such as whether or not a person has blue eyes or is obese. An individual's two alleles may be the same, but often differ. When two alleles differ, one may be dominant and one recessive, such as for eye or hair color. Or the phenotype may depend on the combined alleles in another manner, possibly also depending on the environment.

A number of genes have been linked to obesity in the biomedical literature (Snyder et al., 2004). However, it is not as simple as there being a "fat" gene. Instead there is a complex relationship between neurotransmitters in the brain, genes, and obesity. Neurotransmitters, including dopamine and serotonin, regulate food intake, and are thus related to obesity (Guo, North, and Choi, 2006). Certain genes interact with these neurotransmitters. Furthermore, the interaction depends on the exact genotype, with certain polymorphisms of genes related to high obesity and others to low obesity.

The Add Health data set has information on six relevant genes chosen because they have a high prevalence in the population and a direct role in either dopamine or serotonin. Furthermore, each gene has been shown in epidemiologic literature to be related to obesity (Guo, North, and Choi, 2006). In the Add Health data, we created genetic variables based on whether an individual had a particular genotype in either allele. Several of these variables are highly correlated to obesity, as described later. Because genes related to neurotransmitters may also affect other behaviors besides those

leading directly to obesity, we were not surprised to find that some of our genetic variables were directly related to labor market outcomes.

Value Added of Study

Compared to prior studies analyzing the effect of obesity on labor market outcomes, ours has several important advantages. We use genetic information to control for otherwise some unobservable hereditary traits that affect either obesity or labor market outcomes, or both. The variables that affect only obesity serve as instrumental variables; the others are exogenous controls. By having multiple instruments, we can also test the over-identifying restrictions. Through these methods we can better answer the empirical question that many economists have tried to answer—what is the marginal effect of BMI on the probability of employment and on wages?

We are aware of only one other economics study that uses genetic information as instrumental variables (Ding et al., 2006), in a study of how health affects education. Otherwise, economists have only used knowledge of family history. For example, Brown, Pagan, and Bastida (2005) use information on whether a person's family has a history of diabetes to instrument for diabetes in a model predicting employment. As mentioned in the literature review, other studies have used information on siblings or parents, usually as a proxy for actual genetic information.

Besides addressing whether obesity affects labor market outcomes in young adults, and demonstrating the value of using genetic information in social sciences, we also test whether sibling's BMI, which has frequently been used as a lone identifying

instrument, validly passes the over-identifying restrictions, conditional on the genetic information variables being valid.

METHODS

We investigate the role of obesity in two labor market outcomes. One is the probability that the person is employed, defined as working at least 10 hours per week at the time of the interview. The other outcome is wages, conditional on working. For both models, we are primarily interested in the effect of a change in body mass index (BMI) on the outcomes, holding other demographic factors constant. The literature review suggests that the relationship is negative. A person with higher BMI is less likely to be employed, and those that are employed are expected to have lower wages.

For the probability of employment we estimated a linear probability model. This makes the interpretation of the estimated coefficients easy, and does not substantively change the conclusions when compared to results from a logit model. Following the convention in labor economics, we transform wages by taking its natural logarithm. The distribution of logged wages is roughly normal in our data. This model assumes that the independent variables have a multiplicative (proportional) effect, not additive effect, on wages measured in dollars.

The two equations for labor market outcomes are therefore

$$\Pr(employment_i) = \alpha^E BMI_i + Gen_i\gamma^E + X_i\beta^E + \varepsilon_i^E$$

$$\ln(wage_i) = \alpha^W BMI_i + Gen_i\gamma^W + X_i\beta^W + \varepsilon_i^W \quad \text{if employed}$$

where BMI_i is individual *i*'s body mass index, the vector *Gen* includes genetic information that affects wages directly, the vector *X* includes exogenous explanatory

variables, ε is the iid error term, and the parameters to be estimated are α, β, and γ. The superscripts *E* and *W* refer to employment and wages. The goal is to estimate consistent estimates of α^E and α^W. The right-hand side of the equation for the probability of employment is the same as the equation for log hourly wages.

Endogeneity

The primary econometric concern is how to control for the endogeneity of obesity, which causes bias in the estimate of α. There are two primary stories for why obesity is endogenous. Each explains a different way that employment or income affects obesity. Ruhm argues in a series of articles (e.g., Ruhm, 2000; Ruhm 2003) that many individual health conditions are countercyclical with the macroeconomic conditions because unemployed persons have more time to exercise and eat well, and less income with which to purchase alcohol. Unemployed persons may also have more time to seek medical care. This reasoning implies that the estimated coefficient in a simple labor outcome regression would be biased upwards, compared to the true coefficient.

Other stories suggest that the bias is downwards. Continuing first with Ruhm's macroeconomic example, employed adults are more likely to have health insurance than unemployed adults, which lowers the out-of-pocket cost. As for individual characteristics, lower discount rates make a person invest more in all facets of human capital, creating a negative correlation between economic outcomes and obesity. Persons with low discount rates are more likely to invest in education, yielding higher average wages. They are also more likely to invest in health inputs, such as diet and exercise, which help to lower obesity. Each of these explanations implies that the estimated

coefficient in a simple labor outcome regression would be biased, specifically, more negative than the true coefficient.

Given the concern about endogeneity bias, one way that we use genetic information to control for endogeneity is as instrumental variables. First, some genes provide natural experiments because they directly affect weight, but not labor market outcomes. As explained in the background, at conception some people are naturally predisposed to have a higher BMI than other people. Because this happens at conception, this genetic information is exogenous to determining BMI. We estimate two-stage instrumental variable estimation, with multiple instruments that are correlated with obesity but unrelated to labor market outcomes. In our main models, the explanatory variable of interest, BMI, is continuous. Because we estimate both main outcomes with OLS, we use two-stage least squares.

The first-stage equation is

$$BMI_i = GenIV_i\theta + W\delta + Gen_i\gamma^{BMI} + X_i\beta^{BMI} + \upsilon_i$$

where *GenIV* is a vector of genetic information that are valid instruments, *W* includes additional valid instruments, ν is the error term, and the other Greek letters are parameters to be estimated. In some of our specifications, we follow the prior literature and also include sibling's BMI as an instrumental variable.

Obesity is also endogenous if there are unobserved variables related to both obesity and labor market outcomes. Variables measuring satiation and gratification are the most obvious examples of these. Therefore, we also try to mitigate the extent of endogeneity by including as many of these confounding variables as possible. Again we turn to genetic information. Some genetic information may be correlated with both

obesity and labor market outcomes. Such variables cannot be used as instruments because they cannot be excluded from the main equation. Therefore, we include these variables in both the main equation and the first-stage equation to predict BMI. Including those variables in both equations may help reduce omitted variable bias. These genetic variables may be correlated with satiation, self-control, and discounting of the future, all of which may affect labor market outcomes as well as body mass index.

Risky behaviors, such as smoking and drinking, may also reflect a lack of interest in investment in human capital. We control for four self-reported risky behaviors in *X*: ever smoked, like wild parties, drink alcohol or smoke marijuana, and like new experiences even if they are illegal, frightening or unconventional. Each of these may also be correlated with obesity, so omitting them would lead to further bias of the effect of obesity.

Another way that we try to break the endogeneity problem is to use lagged BMI, in most cases from the first wave of Add Health when respondents were in their late teens. Because part of the endogeneity problem is reverse causality of employment and wages affecting BMI through diet and exercise, lagged BMI is free of that problem. BMI in late teens is primarily dependent on genes, diet from home (typically a mother's cooking), and exercise. We found in preliminary analysis that the genes predicted lagged BMI (and other measures of obesity) much better than current BMI for women and about the same for men. Therefore, we used lagged BMI (or other measures of obesity) in all empirical models. The interpretation of the results is slightly different than in the prior literature. We estimate the effect of lagged BMI—when the respondents were in their late teens—on labor market outcomes when the respondents are in their mid-twenties.

Other econometric issues

Some have criticized the use of continuous BMI as a measure of obesity, and argued instead that categorical measures (e.g., is a person obese or not) are more appropriate to pick up non-linear effects. Therefore we also estimated models in which continuous BMI is replaced by a dummy variable indicating either that the person is overweight (BMI $\geq$ 25) or that the person is obese (BMI $\geq$ 30).

Another econometric issue is selection into the labor market. Although many studies of women's wages in the labor force estimate Heckman selection models to control for the unconditional effect of independent variables on wages, we do not. Therefore, the results in the wage equation are all conditional on employment and do not generalize to all women. The two reasons for not estimating selection models are closely related. There are few, if any, plausible variables in Add Health that determine whether a woman works but are unrelated to her wages conditional on working. Models that have no such identifying variables rely solely on function form for identification, and are notoriously unstable (Dow and Norton, 2003). Preliminary results confirmed that our results were quite unstable when identified through functional form. Therefore, we report a wage equation conditional on employment.

DATA

The National Longitudinal Study of Adolescent Health (Add Health) is a nationally-representative study of how health-related behaviors in adolescents affect various outcomes in early adulthood. The first wave, which began in 1994, collected

individual-, school-, and community-level information on respondents in grades 7 through 12. By Wave III, in 2001–2002, respondents were between 18 and 26 years old. We analyze data from the third wave on respondents at least age 21, because labor market outcomes are most relevant for these older respondents.

Among the 15,197 participants in the Wave III of the Add Health data, DNA information was collected for 2,574. The final sample of 769 women and 714 men and was obtained among those 1,344 women and 1,230 men with DNA information after applying the following exclusion criteria:

1) not interviewed in Wave III (314 women and 291 men dropped);
2) younger than 21 years old (255 women, 196 men);
3) in active military service at the time of interview (6 women, 29 men).

Dependent variables

The probability of employment was estimated for the full sample. Persons were coded as *employed* if they reported working for pay for at least 10 hours a week, or were a student. Eighty-two percent of the 713 women and 88 percent of the 767 men were employed in the final sample, according to our definition (see Table 1). The reference category includes both unemployed and discouraged workers. Although Add Health does not explicitly identify discouraged workers, there are likely to be few if any of them, given the age of the study sample.

The logarithm of wages was estimated for the sub-sample with positive wages. The Add Health data has collected data on respondents' usual earnings excluding overtime pay, tips, and commissions at the current job. The unit of collected earnings varies from hourly to yearly. However, for this analysis, we converted all reported units

for earnings into hourly basis. On average, women ($10.54) who worked earned slightly less per hour than men ($11.09). Hourly wages increased with age in the employed sample.

BMI

The explanatory variable of primary interest is the extent of obesity, which was measured with lagged body mass index (BMI). BMI is defined as weight in kilograms divided by height in meters squared. In the Add Health data, both measured and self-reported height and weight information are available. We used the measured height and body weight information because of the known biases in self-reported data on weight. Lagged BMI was defined as BMI at wave 1 for sample persons aged > 24 years old at wave 3, and BMI at wave 2 for sample aged < 25 year old at wave3. BMI lag in our final sample ranges from 15 to 52 for women and 15 to 42 for women (an average of 22.8), and 15 to 47 for men, with an average 23.6. Less than a quarter of the sample is overweight or obese in lag structure across all ages for women (22%) and for men (29%). Average lag BMI increases with age. The average height of women is five feet four inches; for men it is five feet eight inches.

Other control variables

Age ranges from 21 to 26, by construction, in our sample, with an average age of 22.6 years (see Table 1). About 15 percent of the sample was non-Hispanic Black and another 16 percent were Hispanic. One-fifth of the women are married, but only 15.5 percent of the men are married. Nearly two-thirds of women have a college education or

higher. The fraction is lower for men (54.4 percent). The vast majority of the sample report being in excellent or very good health status. In terms of geographic distribution, about 10 percent live in the Northeast, while the other three regional areas were represented fairly evenly.

Instruments and specification tests

All the genetic variables were based on whether either of the two alleles for each gene had a specific genotype. All genetic variables are dummy variables. A value of one indicates that either one of the alleles (or both) showed a specific polymorphism. Because multiply polymorphisms of one gene may be related to obesity, we sometimes created more than one variable per gene. For this study, our genetic instruments come from two of the genes (see Table 2). We use the 48-bp repeat polymorphism of the Dopamine D4 Receptor (DRD4 gene) and the Dopamine Transporter (DAT1 gene) as instruments for BMI. Specifically, for women we construct two instruments: the 379 frequency (or genotype) in allele A and other frequencies in allele B in the DRD4 gene; and the 427 frequency in either allele A or allele B in the DRD4 gene. For men, we construct three slightly different instruments: the 427 frequency in allele B and other frequencies in allele A in the DRD4 gene; the 427 frequency in both alleles A and B in the DRD4 gene; and the 440 frequency in either allele A or allele B in the DAT1 gene.

There is considerable variation in the fraction of the sample with the genetic variables equal to 1, ranging from 15 percent to 95 percent. Therefore, the variation in genetic information is considerable, and not confined to a small subset of the data.

The siblings' BMI lag averaged 24.3, just like for the study sample, and had similar range. The average of the siblings' BMI lag increases for the heavier BMI group. The correlation coefficient between BMI lag and siblings' BMI lag is 0.377. Nearly 20 percent of the sample did not report a BMI for any sibling, possibly because they have no sibling. For those sample that have missing data for siblings' BMI, we filled it with average siblings' BMI.

Specification tests confirm that our instruments are strong and validly excluded from the main equations. The full set of instruments—genetic information and siblings' BMI—have strong explanatory power in the first-stage regression. The instrumental variables are jointly statistically significant at the 1% level in the first-stage equation (see Table 3). The *F*-statistic for testing the null hypothesis that the first-stage coefficients on the instruments are jointly equal to zero is 14.83, which exceeds the minimum *F* statistic of 10 suggested by Staiger and Stock (1997). However, the F-statistics for men is 4.98. The marginal increase in R^2 by adding the instruments was 0.089 for women and 0.148 for men, after controlling for the other exogenous variables. These results confirm that the genetic information and siblings' BMI together are good instruments for BMI. The three genetic variables on their own also perform well on the specification tests, being jointly statistically significant and having a joint *F*-statistic in the first stage nearly 10 for women. Siblings' BMI lag also does well on its own; this is not surprising given its use as an instrument for obesity in the published literature.

The exogeneity of BMI lag was tested by including both the actual linear BMI lag and the predicted error term (from the first stage) in the second-stage estimation (Bollen, Guilkey, and Mroz, 1995). The statistical significance of the predicted error term tests

the null hypothesis of the exogeneity of BMI lag. For the probability of employment, the null hypothesis of the exogeneity of BMI lag was not rejected at the 5% level (see Table 3) for women. We could not reject the null hypothesis of exogeneity in the log wage equations for women, and could not reject exogeneity for either the employment or log wage equations for men.

The exclusion restriction of the over-identifying instruments was tested by the Lagrange Multiplier (LM) test, which examines whether all instruments are jointly excluded from the second-stage labor market outcomes equation. The null hypothesis to be tested is that the genetic information variables are validly excluded from the main equation, assuming that the sibling's BMI is a valid instrument. The test of the exclusion restriction did not reject the null hypothesis for both labor market outcomes (see Table 3).

Through the process of subjecting variables to the specification tests, we also found genetic variables that both predicted BMI and failed the over-identification test. Therefore, these variables are included in both the first stage and the main equations as controls to help control for omitted variables. Those variables are different between men and women. We created dummy variables based on frequencies in either or both alleles of four different genes (see Table 2 for details). The four genes are the Serotonin Transporter 6A4 (SLC), the Dopamine D2 Receptor TaqIA (DRD2), Monoamine Oxidase A-uVNTR (MAOA), and Cytochrome P450 2A6 (CYP).

RESULTS

First-stage results

The first-stage regression results provide further evidence that the instrumental variables are substantively related to obesity, as measured by BMI (see first and fourth

columns of Table 4). The genetic variables have estimated coefficients of –1.27 and –2.36 for women, and –0.90, –2.09, and –3.12 for men. To put this in perspective, a one-unit increase in BMI for a six-foot tall person corresponds to an increase of more than seven pounds. For a five-foot person a one-unit increase in BMI corresponds to an increase of over five pounds. Therefore, genetic variation alone leads to considerable exogenous changes in weight for a person of average height. A change in certain genes leads to exogenous changes in weight of at least 5 pounds for most people, and for some more than 20 pounds. These coefficients are of meaningful magnitude, even after controlling for other covariates. The genetic instrumental variables also strongly predict the probability of being overweight or obese, at least for women (see the second and third columns of Table 4). In sum, the instruments are strongly related to BMI in the first-stage regression.

As a further check, we ran simple regressions of BMI on each instrument and genetic variable by itself (with a constant) to show that simple correlations are both statistically strong and high in magnitude (see the third and fourth columns of Table 2). Four of the five coefficients are greater than 1.85 in absolute value and statistically significant at the one-percent level. Furthermore, adding the other exogenous covariates, but not the other instruments, does not change the simple correlation much (results not shown). Therefore, the large magnitudes found in the first-stage regression are not due to multicollinearity or other statistical problems.

Several other coefficients in the first-stage results have signs that are consistent with the literature and are statistically significant. Higher BMI is found for African-American and Hispanic women, taller women, and women not attending school.

African-American men have lower BMI on average, as do taller men. These additional results are reassuring. There is often a penalty for being older, unmarried, and in poor self-reported health, although those estimated coefficients are mostly not statistically significant at conventional levels.

Main results

The main results show that BMI has essentially no effect on labor market outcomes for women (see Tables 5 and 6). Five different models control for genetic information and endogeneity in different ways. In a simple regression, BMI has no effect on the probability of employment (see the first column of Table 5). In progressively richer models controlling for more covariates, the coefficient on BMI rises slightly, but it is never statistically significant at conventional levels. Adding the other genetic variables as controls for omitted variables does not seem to affect the coefficient on BMI. This may be because the extent of omitted variable bias is small. The 2SLS results are slightly higher in magnitude, but again are not statistically significant (see the fourth column of Table 5). A one-unit increase in BMI leads to a less than one percentage point increase in the predicted probability of employment for women.

Coefficients on other variables have the expected signs. The probability of employment is higher for women who are older, white, unmarried, and with higher education.

The overall results for the wage equation for women are not statistically significant (see Table 6). In the log wage equation, a higher BMI has no statistically significant effect on wages in any of the models, including the preferred OLS model. As

with the probability of employment, adding genetic information moved the coefficient from more negative to more positive. To the extent that controlling for endogeneity raises the estimated coefficient on BMI, previous studies may have been biased.

For men, BMI is never statistically significant in any of the models (see Tables 7 and 8). The coefficient on BMI is slightly positive (but insignificant) in a simple regression on the probability of employment. Adding more controls lowers this coefficient to be negative. For log wages, the trend for men is again the opposite of women. The simple correlation between BMI and log wages is positive (although not significant). Adding further controls shrinks the coefficient further towards zero.

The variable most often used in the prior obesity-labor market literature as an instrument was sibling's BMI. Because it was used by itself, justification was theoretical but not empirical. Prior researchers could not conduct the over-id test of whether the instruments could be excluded from the main equation. Having multiple genetic variables in our data set allows us to then test the hypothesis that sibling's BMI is a valid instrument. It is; see the lower half of Tables 3A and 3B for the specification test results. Including lagged sibling's BMI greatly improves the strength of the combined instruments in the first stage, at least for women. However, the 2SLS results that include lagged sibling's BMI are not substantially different than results with just the genetic information (results not reported).

Robustness checks

We have tried numerous other specifications to test the robustness of the results. For example, we tried various combinations of instruments, and different definitions of

instruments. We have also run the 2SLS models with either just the genetic variables or just the siblings' BMI (just identified models). The estimated effect of BMI remains stable and small across different sets of instruments.

To test for a possible nonlinear effect of BMI, we tried categorizing BMI into normal weight, overweight, and obese instead of using a continuous measure. The cutoff points are BMI of 25 and 30, the standard thresholds for being overweight and obese. The results of the specification tests are shown in Tables 3A and 3B. For women, the instruments predict being either overweight or being obese for the full sample, but for the subset of employed women, the instruments are weak. The effects of the categorical measures of obesity on labor market outcomes are statistically insignificant. We also estimated models with three linear splines. We have tried excluding variables such as education and marital status that are arguably endogenous too. None of these changes altered the results substantively.

DISCUSSION

Our study has one policy conclusion and two methodological contributions. Limitations include the relatively small sample size (although prior economic studies have had roughly the same sample size) and a focus on labor market outcomes restricted to persons in their mid-twenties. The policy conclusion is that an increase in BMI as a late teen has no statistically significant effect on either employment or wages for a person in their mid-twenties. Our study is by no means the first to find no significant effect of obesity measures on labor market outcome. We feel that our study is better able to combat the endogeneity of obesity in several ways—through the use of lagged BMI,

actual (not self-reported) BMI, and strong instrumental variables. Our findings, while a null finding, are important because of rise in obesity in the United States and the concern about the effect of obesity on labor market outcomes.

The main methodological contribution is the novel use of genetic information, both as instrumental variables and as extra exogenous variables to reduce omitted variable bias. We demonstrated that genetic information is highly predictive of BMI and the probability of being overweight or obese. The relationship is strong in both magnitude and statistical significance. Thus, this is one of the first social science papers to use genetic information as a natural experiment. We also were able to build on our work to test the over-identifying restrictions for sibling's BMI, and show that it does indeed pass these tests.

REFERENCES

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

Baum CL, and Ford WF. 2004. The wage effects of obesity: a longitudinal study. *Health Economics* 13:885-899.

Becker GS. 1971. The economics of discrimination. Second edition. University of Chicago Press.

Becker GS, Grossman M, Murphy KM. 1994. An empirical analysis of cigarette addiction. *The American Economic Review* 84(3): 396-418.

Becker GS, Murphy KM. 1988. A theory of rational addiction. *Journal of Political Economy* 96(4): 675-700.

Behrman JR, Rosenzweig MR. 2001. The Returns to Increasing Body Weight. PIER Working Paper 01-052.

Bhattacharya J, Bundorf MK. 2005. The incidence of the health care costs of obesity. NBER working paper No. 11303.

Blundell JE. 1977. Is there a role for serotonin (5-hydroxytryptamine) in feeding? *International Journal of Obesity* 1(1):15-42.

Bollen KA, Guilkey DK, 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.

Brown III HS, Pagan JA , Bastida E. 2005. The impact of diabetes on employment: genetic IVs in a bivariate probit. *Health Economics* 14(5):537-544.

Buffum D, Whaples R. 1995. Fear and lathing in the Michigan furniture industry: employee-based discrimination a century ago. *Economic Inquiry* 33: 234-252.

Cawley J. 2000. An instrumental variables approach to measuring the effect of body weight on employment disability. *Health Services Research* 35(5):1159-1179.

Cawley J. 2004. The impact of obesity on wages. *Journal of Human Resources* 39(2):451-74.

Chou SY, Grossman M, Saffer H. 2002. An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillances System. NBER working paper 9247.

Chou SY, Grossman M, Saffer H. 2004. An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillances System. *Journal of Health Economics* 23: 565-587.

Conley D, Glauber R. Gender, body mass and economic status. NBER working paper 11343. May 2005.

Cutler DM, Glaeser EL, Shapiro JM. 2003. Why have Americans become more obese? *Journal of Economic Perspectives* 17(3): 93-118.

Ding W, Lehrer SF, Rosenquist JN, Audrain-McGovern J. 2006. The impact of poor health on education: New evidence using genetic markers. NBER working paper 12304.

Dow WH, Norton EC. 2003. Choosing Between and Interpreting the Heckit and Two-Part Models for Corner Solutions. *Health Services & Outcomes Research Methodology* 4(1): 5–18.

Gortmaker SL, Must A, Perrin JM, Sobol AM, Dietz WH. 1993. Social and Economic Consequences of Overweight in Adolescence and Young Adulthood. *New England Journal of Medicine* 329:1008-1012.

Guo G, North, Choi. 2006. DRD4 gene variant associated with body mass: The National Longitudinal Study of Adolescent Health. *Human Mutation* 27(3), 236-241.

Hoebel BG, Hernandez L, Schwartz DH, Mark GP, Hunter GA. 1989. Microdialysis studies of brain norepinephrine, serotonin, and dopamine release during ingestive behavior. Theoretical and clinical implications. *Annals of the New York Academy of Sciences* 575:171-191.

Lakdawalla D, Philipson TJ. 2002. Technological change and the growth of obesity. NBER Working Paper 8946, Cambridge MA; National Bureau of Economic Research.

Loh ES. 1993. The economic effects of physical appearances. *Social Science uarterly* 74: 420-438.

Martin CJ. 1994. Protecting overweight workers against discrimination: is disability or appearance the real issue? *Employee Relations Law Journal* 20(1):133-142.

Nielsen SJ, Popkin BM. 2004. Changes in beverage intake between 1977 and 2001. *American Journal of Preventive Medicine* 27(3): 205-210.

Pagan JA, Davila A. 1997. Obesity, occupational attainment, and earnings. *Social Science Quarterly* 78(3).

Philipson T. 2001. The world-wide growth in obesity: an economic research agenda. *Health Economics* 10: 1-7.

Popkin BM, Siega-Riz AM, Haines PS. 1996. A Comparison of dietary trends among racial and socioeconomic groups in the US. *The New England Journal of Medicine* 335(10): 716-722.

Puhl R, Brownell KD. 2001. Bias, discrimination, and obesity. *Obesity Research* 9:788-805.

Register CA, Williams DR. 1992. Labor market effects of Marijuana and cocaine use among young men. *Industrial Labor Relations Review* 45: 435-448.

Rosen S. 1974. Hedonic prices and implicit markets: product differentiation in pure competition. *The Journal of Political Economy* 82(1): 34-55.

Ruhm CJ. 2000. Are recessions good for your health? *Quarterly Journal of Economics* 115(2):617-650.

Ruhm CJ. 2003. Good times make you sick. *Journal of Health Economics* 22(4):637-658.

Sargent JD, Blanchflower DG. 1994. Obesity and stature in adolescence and earnings in young adulthood. Analysis of a British birth cohort. *Archives of Pediatrics & Adolescent Medicine* 148(7): 681-687.

Snyder EE, Walts B, Perusse L, Chagnon YC, Weisnagel SJ, Rankinen T, Bouchard C. 2004. The human obesity gene map: the 2003 update. *Obesity Research* 12:369–439.

Sobal J, Stunkard AJ. 1989. Socioeconomic Status and Obesity: A Review of the Literature. *Psychological Bulletin* 105:260–275.

Staiger D, Stock JH. 1997. Instrumental variables regression with weak instruments. *Econometrica* 65(3): 557-586.

Strotz RH. 1955-1956. Myopia and Inconsistency in Dynamic Utility Maximization. *The Review of Economic Studies* 23(3): 165-180.

Stunkard AJ. 1996. Socioeconomic Status and Obesity. *Origins and Consequences of Obesity* 201: 174-187.

U.S. Department of Health and Human Services. 2001. Office of the Surgeon General. Public Health Service. The Surgeon General's call to action to prevent and decrease overweight and obesity. Accessed at Feb 28, 2005 from: http://www.surgeongeneral.gov/topics/obesity/calltoaction/CalltoAction.pdf.

Table 1. Summary statistics of non-genetic variables, Add Health data, Wave III

	Means			
Variables	**Women**	**Men**	**Min.**	**Max.**
Dependent Variables				
Employed	0.82	0.88	0	1
ln(Wages) (*N*=524; 540)	2.18	2.23	0	3.76
Wages ($) (*N*=524; 540)	10.54	11.09	1	43
Variable of Interest				
BMI, lagged	22.79	23.62	14.61	47.35
Sibling data				
Siblings' BMI, lagged	24.30	24.703	14.61	47.35
Demographic Variables				
Age – 20	2.610	2.669	1	6
African-American	0.162	0.141	0	1
Hispanic	0.148	0.169	0	1
Married	0.256	0.150	0	1
Education				
> High school	0.618	0.556	0	1
Attending school	0.369	0.328	0	1
Perceived Health Status				
Excellent or very good	0.727	0.762	0	1
Measured height (m)	1.629	1.775	1.22	2.13
Risky behaviors				
Ever smoked	0.343	0.451	0	1
Like wild parties	0.238	0.370	0	1
Drink, smoke marijuana	0.321	0.550	0	1
Like new experiences	0.328	0.561	0	1
Regional Variables				
West	0.228	0.249	0	1
Midwest	0.342	0.325	0	1
South	0.329	0.329	0	1
N	769	714	1,483	1,483

Table 2. Summary statistics of genetic variables, Add Health data, Wave III

	Means		Coefficients	
	Women	**Men**	**Women**	**Men**
Genetic variables used as IVs				
DRD4: Dopamine D4 Receptor				
DRD4_379_other	0.134		–1.85 **	
Allele A has 379 frequency, B has other				
DRD4_427either	0.048		–2.09 **	
Either Allele A or B has 427 frequency				
DRD4_other_427		0.010		–3.76 **
Allele A has other, B has 427 frequency				
DRD4_427_427		0.003		–2.16 **
Both alleles have 427 frequency				
DAT1: Dopamine Transporter				
DAT1_440either		0.385		–0.78
Either Allele A or B has 440 frequency				
Genetic variables used as controls				
SLC: Serotonin Transporter 6A4				
SLC_484_other	0.915	0.917	–0.05	0.33
Allele A has 484 frequency, B has other				
SLC_484_484	0.196	0.213	0.14	0.49
Both alleles have 484 frequency				
DRD2: Dopamine D2 Receptor TaqIA				
DRD2_178either	0.915	0.917	0.14	–0.77
Either Allele A or B has 178 frequency				
DRD2_304either	0.489	0.465	0.84	–0.39
Either Allele A or B has 304 frequency				
MAOA: Monoamine Oxidase A-uVNTR				
MAOA_351_other	0.017		0.003	
Allele A has 351 frequency, B has other				
MAOA_other_351	0.471		0.13	
Allele A has other, B has 351 frequency				
MAOA_351_351	0.334		–0.02	
Both alleles have 351 frequency				
MAOA_321either		0.622		0.58
Either Allele A or B has 321 frequency				
MAOA_351either		0.822		–0.50
Either Allele A or B has 351 frequency				
CYP: Cytochrome P450 2A6				
CYP_1either	0.957	0.948	–2.16	–2.08 *
Either Allele A or B has 1 frequency				
N	769	714	769	714

The coefficients reported in the last two columns are from 19 simple regressions of BMI on the genetic dummy variable and a constant term. Robust standard errors are in parentheses. *p*-value <0.01: **, <0.05 *.

Table 3A. Specification tests of the instrumental variables, Women

		Specification Tests			
Dependent Variables	***N***	**IV Strength**	**Over-id *p*-value**	**Exogeneity *p*-value**	**Conclusions**
No Sibling BMI as IV					
Continuous BMI (lagged)					
Pr(employed)	769	$F = 10.07$ $p = .0001$	0.223	0.893	Good IVs, exog.
ln(wages)	524	$F = 7.10$ $p = .0014$	0.590	0.612	Good IVs, exog.
If Overweight or Obese (lagged)					
Pr(employed)	769	$p = .0021$	0.927	0.213	Good IVs, exog.
ln(wages)	524	$p = .0007$	0.047	0.046	Poor IVs, endog.
If Obese (lagged)					
Pr(employed)	756	$p = .0004$	0.081	0.610	Good IVs, exog.
ln(wages)	490	$p = .0095$	0.745	0.714	Weak IVs, exog.
Sibling BMI included as IV					
Continuous BMI (lagged)					
Pr(employed)	769	$F = 14.83$ $p < .0001$	0.450	0.828	Good IVs, exog.
ln(wages)	524	$F = 18.38$ $p < .0001$	0.808	0.638	Good IVs, exog.
If Overweight or Obese (lagged)					
Pr(employed)	769	$p < .0001$	0.448	0.942	Good IVs, exog.
ln(wages)	524	$p < .0001$	0.486	0.484	Good IVs, endog.
If Obese (lagged)					
Pr(employed)	756	$p = .0001$	0.231	0.691	Good IVs, exog.
ln(wages)	490	$p = .0015$	0.964	0.261	Good IVs, exog.

The null hypothesis that the over-identifying instruments can be excluded from the main equation was tested with an LM test. The null hypothesis that BMI is exogenous was tested with an NR^2 test.

Table 3B. Specification tests of the instrumental variables, Men

		Specification Tests			
Dependent Variables	***N***	**IV Strength**	**Over-id *p*-value**	**Exogeneity *p*-value**	**Conclusions**
No Sibling BMI as IV					
Continuous BMI (lagged)					
Pr(employed)	714	$F = 4.43$ $p = .0057$	0.415	0.682	Weak IVs, exog.
ln(wages)	540	$F = 4.12$ $p = .0085$	0.865	0.716	Weak IVs, exog.
If Overweight or Obese (lagged)					
Pr(employed)	712	$p = .0608$	0.141	0.660	Weak IVs, exog.
ln(wages)	538	$p = .0698$	0.204	0.799	Weak IVs, endog.
If Obese (lagged)					
Pr(employed)	705	$p = .0794$	0.896	0.617	Weak IVs, exog.
ln(wages)	533	$p = .0551$	0.099	0.378	Weak IVs, exog.
Sibling BMI included as IV					
Continuous BMI (lagged)					
Pr(employed)	714	$F = 4.98$ $p = .0010$	0.471	0.253	Weak IVs, exog.
ln(wages)	540	$F = 4.25$ $p = .0032$	0.844	0.893	Weak IVs, exog.
If Overweight or Obese (lagged)					
Pr(employed)	712	$p < .0001$	0.303	0.224	Good IVs, exog.
ln(wages)	538	$p = .0045$	0.643	0.822	Weak IVs, exog.
If Obese (lagged)					
Pr(employed)	705	$p < .0001$	0.892	0.057	Good IVs, exog.
ln(wages)	533	$p = .0040$	0.709	0.759	Weak IVs, exog.

The null hypothesis that the over-identifying instruments are excluded from the main equations was tested using an LM test. The null hypothesis that BMI is exogenous was tested using an NR^2 test.

Table 4. First-stage regression results to predict BMI

	Women			Men		
Variables	**BMI**	**OW or Obese**	**Obese**	**BMI**	**OW or Obese**	**Obese**
Constant	9.95	–7.57 *	–16.06 **	25.28 **	–1.19	3.74
	(6.23)	(3.12)	(4.55)	(6.55)	(3.51)	(5.72)
Instrumental Variables						
DRD4_379_other	–1.27 **	–1.17 **	–2.86 **			
	(0.46)	(0.35)	(0.78)			
DRD4_427either	–2.36 **	–1.86 **	–3.48 *			
	(0.80)	(0.76)	(1.37)			
DAT1_440either				–0.90	–0.44	–0.62
				(0.52)	(0.29)	(0.35)
DRD4_427_427				–2.09		
				(1.09)		
DRD4_other_427				–3.12 **	–1.51	
				(1.18)	(0.96)	
Genetic Controls						
SLC_484_other	0.38	0.32	0.04	0.56	0.25	–0.24
	(0.50)	(0.24)	(0.41)	(0.58)	(0.27)	(0.40)
SLC_484_484	0.74	0.30	0.48	0.70	0.54	–0.40
	(0.69)	(0.41)	(0.48)	(0.79)	(0.30)	(0.56)
DRD2_178either	0.58	–0.04	0.74	–1.21	–0.55	0.04
	(0.90)	(0.41)	(0.90)	(0.87)	(0.45)	(0.79)
DRD2_304either	0.70	0.53 *	0.33	–0.20	–0.20	–0.07
	(0.52)	(0.21)	(0.33)	(0.53)	(0.26)	(0.43)
MAOA_351_other	0.43	1.39				
	(1.34)	(0.77)				
MAOA_other_351	0.18	–0.11	0.66			
	(0.62)	(0.29)	(0.43)			
MAOA_351_351	0.17	0.10	0.76			
	(0.69)	(0.34)	(0.48)			
MAOA_321either				1.02	–0.91	–0.98
				(1.33)	(0.85)	(1.18)
MAOA_351either				0.31	–0.57	–0.51
				(1.33)	(0.92)	(1.12)
CYP_1either	–1.83	–1.17 *	0.33	–1.83	–0.86 *	3.74
	(1.03)	(0.50)	(1.05)	(1.16)	(0.40)	(5.72)
Demographics						
Age – 20	0.20	–0.057	0.01	0.17	0.096	–0.03
	(0.20)	(0.091)	(0.13)	(0.19)	(0.083)	(0.13)
African-American	2.50 **	1.06 **	0.78	–1.46 *	–0.54	–1.87 *

	(0.96)	(0.36)	(0.50)	(0.64)	(0.39)	(0.94)
Hispanic	2.16 *	0.83	0.97	0.29	0.18	0.38
	(0.86)	(0.54)	(0.64)	(0.76)	(0.39)	(0.55)
Married	–0.68	–0.21	–0.59	0.21	0.44	–0.68
	(0.60)	(0.27)	(0.36)	(0.72)	(0.28)	(0.58)
Education						
> High school	0.14	–0.30	–0.63	0.51	0.12	0.49
	(0.55)	(0.28)	(0.38)	(0.61)	(0.22)	(0.33)
Attending school	–1.27 **	–0.54	–1.17 *	–0.81	–0.29	–1.0 *
	(0.49)	(0.30)	(0.48)	(0.63)	(0.21)	(0.48)
Health status	–0.83	–0.62 *	0.21	–1.18	–0.59 *	–0.81 *
Excellent/very good	(0.52)	(0.29)	(0.53)	(0.73)	(0.28)	(0.36)
Measured height	9.04 *	4.83 **	7.50 **	0.70	0.86	–2.14
	(3.62)	(1.76)	(2.35)	(3.46)	(1.65)	(2.94)
Risky behaviors						
Ever smoked	1.07 *	0.65 *	0.29	–0.52	–0.22	–0.30
	(0.50)	(0.25)	(0.29)	(0.53)	(0.29)	(0.36)
Like wild parties	–0.31	–0.43	–0.10	0.75	–0.1	0.65
	(0.60)	(0.50)	(0.64)	(0.59)	(0.23)	(0.35)
Drink, smoke	0.69	0.73	0.50	–0.62	–0.20	–0.76
Marijuana	(0.61)	(0.39)	(0.34)	(0.65)	(0.31)	(0.41)
Like new	–0.92	–0.48	–0.57	–1.27 *	–0.37	–0.58
experiences	(0.52)	(0.26)	(0.34)	(0.64)	(0.35)	(0.49)
N	769	769	756	714	712	705

Robust standard errors are in parentheses. *p*-value <0.01: **, <0.05 *. Regressions also include three regional dummy variables.

Table 5. Linear probability model predictions of employment for women

Variables	LPM	LPM	LPM	LPM	2SLS
Constant	0.81 **	0.56	0.72	0.70	0.68
	(0.10)	(0.42)	(0.42)	(0.42)	(0.50)
BMI (lagged)	0.0010	0.0049	0.0043	0.0042	0.007
	(0.0045)	(0.0036)	(0.0033)	(0.0032)	(0.025)
Instrumental Variables					
DRD4_379_other				0.030	
				(0.052)	
DRD4_427either				-0.047	
				(0.059)	
Genetic Controls					
SLC_484_other			0.033	0.037	0.032
			(0.034)	(0.034)	(0.034)
SLC_484_484			0.077	0.0080	0.0052
			(0.045)	(0.045)	(0.052)
DRD2_178either			-0.0045	-0.0041	-0.0060
			(0.057)	(0.056)	(0.059)
DRD2_304either			0.013	0.016	0.011
			(0.031)	(0.031)	(0.031)
MAOA_351_other			-0.17	-0.15	-0.17
			(0.23)	(0.24)	(0.24)
MAOA_other_351			-0.039	-0.039	-0.040
			(0.043)	(0.043)	(0.043)
MAOA_351_351			-0.093 *	-0.090 *	-0.094 *
			(0.041)	(0.041)	(0.040)
CYP_1either			-0.104 *	-0.102 *	-0.097
			(0.049)	(0.049)	(0.063)
Demographics					
Age – 20		0.018	0.018	0.017	0.018
		(0.011)	(0.012)	(0.012)	(0.013)
African-American		-0.121 *	-0.124 *	-0.123 *	-0.13
		(0.057)	(0.056)	(0.059)	(0.10)
Hispanic		0.0071	0.012	0.013	0.0047
		(0.046)	(0.045)	(0.047)	(0.084)
Married		-0.025	-0.014	-0.0099	-0.012
		(0.041)	(0.046)	(0.044)	(0.052)
Education					
> High school		0.177 **	0.176 **	0.179 **	0.175 **
		(0.048)	(0.047)	(0.052)	(0.053)
Attending school		0.207 **	0.209 **	0.205 **	0.214 **
			(0.027)	(0.026)	(0.037)

		(0.028)			
Perceived Health Status					
Excellent or very good		0.046	0.044	0.044	0.046
		(0.033)	(0.032)	(0.032)	(0.037)
Measured height		-0.082	-0.088	-0.081	-0.11
		(0.24)	(0.23)	(0.23)	(0.29)
Risky behaviors					
Ever smoked		0.037	0.045	0.048	0.041
		(0.033)	(0.035)	(0.037)	(0.052)
Like wild parties		0.003	-0.010	-0.0076	-0.0095
		(0.039)	(0.035)	(0.034)	(0.037)
Drink, smoke		-0.0033	-0.050	-0.0023	-0.0070
Marijuana		(0.039)	(0.036)	(0.036)	(0.040)
Like new experiences		0.011	0.012	0.0095	0.015
		(0.036)	(0.035)	(0.034)	(0.044)
N	769	769	769	769	769

Robust standard errors are in parentheses. *p*-value <0.01: **, <0.05 *. Regressions also include three regional dummy variables.

Table 6. Regressions to predict the logarithm of wages for women

Variables	OLS	OLS	OLS	OLS	2SLS
Constant	2.18 **	1.74 **	1.19 *	1.20 *	1.49
	(0.11)	(0.53)	(0.56)	(0.56)	(0.84)
BMI (lagged)	–0.0008	0.0014	0.0046	0.0050	–0.018
	(0.0045)	(0.0052)	(0.0046)	(0.0048)	(0.043)
Instrumental Variables					
DRD4_379_other				0.053	
				(0.082)	
DRD4_427either				0.03	
				(0.13)	
Genetic Controls					
SLC_484_other			0.098	0.097	0.095
			(0.066)	(0.070)	(0.068)
SLC_484_484			0.044	0.043	0.040
			(0.075)	(0.076)	(0.081)
DRD2_178either			-0.17	-0.17	-0.15
			(0.094)	(0.094)	(0.10)
DRD2_304either			0.052	0.052	0.069
			(0.074)	(0.069)	(0.086)
MAOA_351_other			-0.26 *	-0.26 *	-0.28 *
			(0.11)	(0.12)	(0.13)
MAOA_other_351			-0.008	-0.0067	0.0078
			(0.071)	(0.070)	(0.094)
MAOA_351_351			0.062	0.066	0.068
			(0.099)	(0.099)	(0.11)
CYP_1either			0.45 *	0.44 *	0.39
			(0.18)	(0.18)	(0.26)
Demographics					
Age – 20		0.105 **	0.103 **	0.103 **	0.108 **
		(0.023)	(0.023)	(0.023)	(0.025)
African-American		-0.24	-0.30 *	-0.29	-0.20
		(0.15)	(0.14)	(0.15)	(0.26)
Hispanic		0.030	-0.032	-0.026	0.034
		(0.10)	(0.10)	(0.10)	(0.16)
Married		0.097 *	0.088	0.091	0.068
		(0.042)	(0.049)	(0.048)	(0.061)
Education					
> High school		0.0098	0.0069	0.011	0.0030
		(0.061)	(0.057)	(0.058)	(0.056)
Attending school		-0.041	-0.075	-0.079	-0.11

		(0.070)	(0.069)	(0.069)	(0.11)
Perceived Health Status					
Excellent or very good		-0.073	-0.063	-0.063	-0.075
		(0.051)	(0.051)	(0.051)	(0.056)
Measured height		0.13	0.24	0.23	0.41
		(0.31)	(0.31)	(0.31)	(0.40)
Risky behaviors					
Ever smoked		-0.094	-0.13	-0.12	-0.096
		(0.073)	(0.073)	(0.076)	(0.067)
Like wild parties		-0.173 *	-0.15	-0.15	-0.18
		(0.086)	(0.084)	(0.084)	(0.11)
Drink, smoke marijuana		0.005	-0.016	-0.015	0.014
		(0.084)	(0.11)	(0.10)	(0.14)
Like new experiences		0.002	0.0075	0.0084	-0.015
		(0.089)	(0.071)	(0.075)	(0.078)
N	524	524	524	524	524

Robust standard errors are in parentheses. *p*-value <0.01: **, <0.05 *. Regressions also include three regional dummy variables.

Table 7. Linear probability model predictions of employment for men

Variables	LPM	LPM	LPM	LPM	2SLS
Constant	0.891 **	0.56	0.64 *	0.66 *	0.82
	(0.091)	(0.35)	(0.32)	(0.31)	(0.62)
BMI (lagged)	0.0004	0.0006	–0.0009	–0.0008	–0.008
	(0.0035)	(0.0031)	(0.0030)	(0.0030)	(0.019)
Instrumental Variables					
DAT1_440either				-0.012	
				(0.027)	
DRD4_427_427				0.076	
				(0.071)	
DRD4_other_427				0.135	
				(0.071)	
Genetic Controls					
SLC_484_other			0.0073	0.0040	0.011
			(0.039)	(0.040)	(0.041)
SLC_484_484			0.056	0.055	0.061
			(0.046)	(0.047)	(0.051)
DRD2_178either			-0.13 *	-0.131 *	-0.138 *
			(0.051)	(0.051)	(0.060)
DRD2_304either			0.0033	0.0050	0.0014
			(0.025)	(0.025)	(0.025)
MAOA_321either			0.116		
			(0.074)		
MAOA_351either			0.028		
			(0.073)		
Cyp2a6b_1			-0.089 *	-0.089 *	-0.10
			(0.038)	(0.038)	(0.056)
Demographic Variables					
Age – 20		-0.013	-0.013	-0.012	-0.012
		(0.015)	(0.014)	(0.014)	(0.015)
African-American		-0.161 **	-0.172 **	-0.173 **	-0.181 **
		(0.061)	(0.061)	(0.062)	(0.063)
Hispanic		-0.0056	-0.029	-0.027	-0.027
		(0.051)	(0.051)	(0.052)	(0.051)
Married		0.120 **	0.113 **	0.113 **	0.115 **
		(0.031)	(0.028)	(0.028)	(0.030)
Education					
> High school		0.111 *	0.114 *	0.114 *	0.118 *
		(0.049)	(0.047)	(0.048)	(0.050)
Attending school		0.101 **	0.101 **	0.103 **	0.094 **
		(0.028)	(0.028)	(0.029)	(0.033)

Perceived Health Status					
Excellent or very good		0.038	0.039	0.039	0.031
		(0.045)	(0.046)	(0.045)	(0.048)
Measured height		0.15	0.20	0.19	0.21
		(0.18)	(0.18)	(0.18)	(0.17)
Risky behaviors					
Ever smoked		0.065	0.070 *	0.070 *	0.067 *
		(0.035)	(0.033)	(0.033)	(0.034)
Like wild parties		0.011	0.019	0.018	0.024
		(0.044)	(0.043)	(0.042)	(0.049)
Drink, smoke marijuana		-0.034	-0.032	-0.033	-0.037
		(0.032)	(0.031)	(0.032)	(0.033)
Like new experiences		0.004	0.001	0.001	-0.008
		(0.033)	(0.031)	(0.031)	(0.036)
N	714	714	714	714	714

Robust standard errors are in parentheses. *p*-value <0.01: **, <0.05 *. Regressions also include three regional dummy variables.

Table 8. Regressions to predict the logarithm of wages for men

Variables	OLS	OLS	OLS	OLS	2SLS
Constant	2.15 **	1.55 *	1.15	1.14	0.64
	(0.12)	(0.75)	(1.05)	(1.02)	(2.14)
BMI (lagged)	0.0037	0.0008	0.0012	0.0008	0.023
	(0.0043)	(0.0048)	(0.0045)	(0.0040)	(0.059)
Instrumental Variables					
DAT1_440either				-0.024	
				(0.095)	
DRD4_427_427				0.20 *	
				(0.10)	
DRD4_other_427				-0.10	
				(0.13)	
Genetic Controls					
S484other			0.091	0.095	0.082
			(0.12)	(0.13)	(0.10)
S484484			-0.12	-0.12	-0.13
			(0.087)	(0.088)	(0.083)
D2178either			-0.086	-0.086	-0.063
			(0.16)	(0.16)	(0.19)
D2304either			-0.0025	-0.0011	0.0021
			(0.081)	(0.083)	(0.084)
Cyp2a6b_1			0.21	0.21	0.25
			(0.14)	(0.14)	(0.22)
Demographic Variables					
Age – 20		0.028	0.033	0.032	0.032
		(0.022)	(0.022)	(0.023)	(0.024)
African-American		-0.0046	-0.035	-0.043	-0.013
		(0.090)	(0.097)	(0.090)	(0.13)
Hispanic		-0.041	-0.064	-0.063	-0.074
		(0.13)	(0.13)	(0.14)	(0.13)
Married		0.241 **	0.224 **	0.222 **	0.214 **
		(0.060)	(0.057)	(0.056)	(0.058)
Education					
> High school		0.088	0.10	0.098	0.082
		(0.053)	(0.053)	(0.053)	(0.083)
Attending school		-0.280 **	-0.274 **	-0.268 **	-0.251 **
		(0.067)	(0.068)	(0.060)	(0.055)
Perceived Health Status					
Excellent or very good		0.046	0.022	0.020	0.045
		(0.050)	(0.48)	(0.053)	(0.075)

Measured height		0.37	0.49	0.51	0.44
		(0.37)	(0.42)	(0.42)	(0.40)
Risky behaviors					
Ever smoked		0.0056	0.0022	-0.0032	0.11
		(0.061)	(0.053)	(0.049)	(0.069)
Like wild parties		0.0079	-0.010	-0.0086	-0.023
		(0.073)	(0.083)	(0.086)	(0.065)
Drink, smoke marijuana		-0.048	-0.054	-0.052	-0.041
		(0.053)	(0.059)	(0.058)	(0.056)
Like new experiences		0.054	0.070	0.069	0.11
		(0.053)	(0.051)	(0.049)	(0.14)
N	540	540	540	540	540

Robust standard errors are in parentheses. *p*-value <0.01: **, <0.05 *. Regressions also include three regional dummy variables.