HHS Public Access

Author manuscript

Rev Econ Househ. Author manuscript; available in PMC 2020 August 27.

Published in final edited form as:

Rev Econ Househ. 2014 ; 12(4): 737–770. doi:10.1007/s11150-012-9166-5.

The impact of early occupational choice on health behaviors

Inas Rashad Kelly,

Department of Economics, Queens College, CUNY, NBER, 65-30 Kissena Boulevard, Flushing, NY 11367, USA

Dhaval M. Dave,

Department of Economics, Bentley University, NBER, 175 Forest St., AAC 195, Waltham, MA 02452, USA

Jody L. Sindelar,

Yale School of Public Health, NBER, 60 College Street, P.O. Box 208034, New Haven, CT 06520-8034, USA

William T. Gallo

CUNY School of Public Health, Hunter College, CUNY, 425 E. 25th Street, New York, NY 10010, USA

Abstract

Occupational choice is a significant input into workers' health investments, operating in a manner that can be either health-promoting or health-depreciating. Recent studies have highlighted the potential importance of initial occupational choice on subsequent outcomes pertaining to morbidity. This study is the first to assess the existence and strength of a causal relationship between initial occupational choice at labor entry and subsequent health behaviors and habits. We utilize the Panel Study of Income Dynamics to analyze the effect of first occupation, as identified by industry category and blue collar work, on subsequent health outcomes relating to obesity, alcohol misuse, smoking, and physical activity in 2005. Our findings suggest blue collar work early in life is associated with increased probabilities of obesity, at-risk alcohol consumption, and smoking, and increased physical activity later in life, although effects may be masked by unobserved heterogeneity. The weight of the evidence bearing from various methodologies, which account for non-random unobserved selection, indicates that at least part of this effect is consistent with a causal interpretation. These estimates also underscore the potential durable impact of early labor market experiences on later health.

Keywords

Occupational choice; Obesity; Alcohol; Physical activity; PSID

JEL Classifications

I0; I12; J0; J24

Inas.Kelly@qc.cuny.edu.

1 Introduction

Working is an activity that occupies much of many people's lives. According to the 2005 American Time Use Survey, employed persons spend an average of 7.5 h a day working. A sizeable economics literature has established that early labor market history, including occupational choice, influences job mobility and income trajectories (Oreopoulos and von Wächter 2006; Light 2005). Economic resources have, in turn, been shown to affect health outcomes (Smith 1999). A possible compensating wage differential, wherein people trade off job safety and higher wages, and the presence of substantial heterogeneity in health insurance across occupation and industry classes, further implicate occupational choice as a significant input into health investments.[1]

This interplay of various reinforcing and competing mechanisms suggests that work-life can be both health-promoting and health-depreciating. Moreover, it may impact health investments and health outcomes directly, for instance through occupational hazards or job strain, as well as indirectly, through health care coverage, income, and peer influences. Given these numerous plausible pathways, a number of studies have examined the association between job conditions and health (Case and Deaton 2005; Theorell 2000), though most have been limited by potential selection bias and are thus unable to draw stronger conclusions regarding causality. One notable exception is a study by Gueorguieva et al. (2009), which conducts a more careful analysis of Health and Retirement Study data, uncovering significant differences in baseline health by occupation that persist over time.

Most of the extant literature in economics has also focused on contemporaneous effects rather than the cumulative, durable impact of early labor market choices. This is surprising given that the economic paradigm, which views health as a capital stock determined by lifetime investments, choices, and constraints (Grossman 2000), imparts a significant role to early investments and resources. Furthermore, the impact of occupational choice, which can establish future economic resources and other choices, may be most acute during early adulthood, when health levels and pathways are being established. Even if health utilization or health behaviors respond to current changes in circumstances, effects on health capital may not be realized until later in the working life-course, suggesting that the cumulative or durable impact of labor market choices may be more salient than contemporaneous effects.

This study is the first to assess the existence and strength of a causal relationship between initial occupational choice at labor entry and subsequent health behaviors and habits, such as smoking, drinking, physical activity, and obesity, all of which are important proximate inputs into later health status.[2] Health habits are often established relatively early in life (Fletcher and Sindelar 2009), and are therefore more likely to respond to early labor market choices. The focus on health behaviors also underscores potential pathways through which initial labor market choices may eventually have lasting effects on health; the identified

[1] See Viscusi and Aldy (2003) for a review of the literature on the value of a statistical life derived from the tradeoff between wages and job safety.

[2] We note that the chain linking occupational choice at initial labor market entry to subsequent health behaviors into late adulthood has several links, for instance associated with peer influences (from the acquisition of stable behavioral memes to their transmissibility to a subset of occupational entrants), health status (from occupational exposures to diagnosis to changes in health behaviors), health insurance (from coverage to access to shifts in information and incentives), and working conditions and income.

impact of initial occupation on later health outcomes (for instance, heart attacks, as studied in Sindelar et al. 2007) is more plausibly indicative of a causal link if effects on intermediate health behaviors and inputs are also evident. We also undertake an exploratory analysis of potential mediators and pathways, including income trajectories, health insurance, work hours, and other factors through which early occupational choice may have durable effects on subsequent health habits. Identifying these pathways can be important in targeting public policy interventions that may moderate potentially adverse effects on health behaviors and overall health status.

The empirical analysis uses data from the Panel Study of Income Dynamics (PSID), a nationally-representative longitudinal data set that contains information on over 65,000 individuals, spanning as much as 36 years of their lives. The PSID contains extensive information on pre-labor market conditions and family characteristics, as well as typically-unobserved measures such as individuals' risk tolerance, which allows us to account for potential selection bias. We further employ a series of methodologies to disentangle causal effects, including a standard instrumental variables-based strategy; a novel approach proposed by Lewbel (2012) that generates internal instrumental variables in the presence of heteroscedasticity; and an innovative approach proposed by Altonji et al. (2005) that permits causal inference without the need for exclusion restrictions or other restrictive assumptions. This direct focus on accounting for selection bias and sorting out causality is another advance to the existing literature.

2 Background

The objective of this study is to assess the extent to which first occupation impacts subsequent health behaviors among older adults. This question can be framed within the human capital model for the demand for health (Grossman 2000). Grossman combines the household production model of consumer behavior with the theory of human capital investment to analyze an individual's demand for health capital. Individuals invest in health up to the point where the marginal benefit equates the supply price of health capital at each age. The basic insight of this paradigm is that health is a capital stock and health behaviors and other inputs are investments in that stock. Today's health stock will be a function of the entire history of health investments, including current and past health behaviors, incomes, and health endowments.

Occupational choice can be an input into health production, affecting health behaviors and outcomes through a variety of channels. Aspects of work can have both direct and indirect effects on health, and these effects may be either health-promoting or health-depreciating.

2.1 Direct effects

Direct effects include occupational exposure to health and safety hazards, poor working conditions, and injury risks.[3] In theory, compensating wage differentials may also affect the trade-off between job safety and higher wages in unskilled jobs, but there is somewhat

[3]The Occupational Safety and Health Administration (OSHA) reports 4.1 non-fatal workplace injuries and illnesses among 100 equivalent full-time workers in 2008. This national all-industry average masks considerable heterogeneity; the rate in manufacturing

mixed evidence regarding this (see for instance, Brown 1980 and Rosen 1986). Thus, occupational choice can be an investment in health.[4] People may also initially select jobs with high injury rates and potentially risk-compensate by quitting smoking or engaging in more health behaviors external to the workplace, pointing to an a priori ambiguous net effect of occupational choice on subsequent health behaviors.

Job strain associated with working conditions may also have direct adverse effects on mental and physical health. Schnall et al. (1994) link working conditions and strain on the job to cardiovascular disease, and Kouvonen et al. (2009) link it to a reduced likelihood of smoking cessation. In a comprehensive review of the literature on job *stress*, Michie and Williams (2003) find that long hours worked, work overload and pressure, and the effects of these conditions on personal lives are key factors associated with psychological ill health and sickness absence. Depression and mental illness, in turn, have been found to causally impact participation in unhealthy behaviors such as smoking, drinking, and illicit drug use (Saffer and Dave 2005a, b). There is a large body of research on the direct health effects of occupation, but most of this occurs in fields outside of economics,[5] and is therefore beyond the scope of this paper. It is also well-evidenced that the burden of occupational disease is large and shared by workers and society at large. As an example, workers compensation helps to cover some, about 25 % of the burden of injury and illness (Leigh 2011).

2.2 Indirect effects

Indirect effects of occupational choice on health can occur via income, mobility and wealth constraints; health insurance coverage; time constraints; and influences through workplace peers. For instance, initial occupational choices affect occupational mobility, tenure and experience, and income trajectories over one's lifetime (Light 2005); these shifts in economic resources would in turn be expected to impact health. While the direction of causality is not well-established, a sizeable literature documents a strong association between income or wealth and a variety of health outcomes, including mortality and morbidity (Smith 1999). Ettner (1996), in an attempt to disentangle causality, applies an instrumental variables-based methodology to three large-scale nationally representative data sets. She estimates the structural impact of income on health and concludes that additions to

industries can vary non-linearly with firm size, with smaller and larger firms displaying lower rates (see for instance Leigh 1989), and the rate can also be higher among state and local government employers. Certain private industries such as crop and animal production, food/beverage and tobacco manufacturing, wood and primary metal manufacturing, hospitals, and nursing and residential care facilities also exhibit higher rates of occupational illness and injury. It should be noted that official data sources tend to underestimate occupational injury and that the reporting errors may systematically differ across industries with small firms more prone to under-reporting. See Leigh et al. (2004) for further discussion on these estimates and the under-counting.

[4]See Cropper (1977) for a formal introduction of occupational choice in Grossman's human capital framework for the demand for health capital.

[5]In particular, exposure to arsenic may lead to respiratory cancer (Lubin and Fraumeni 2000; Lubin et al. 2000), yet effects on other diseases may be masked by the healthy worker survivor effect (Lubin and Fraumeni 2000; Hertz-Picciotto et al. 2000). Coal miners, hard-rock miners, tunnel workers, concrete-manufacturing workers, and non-mining industrial workers have been shown to be at risk for developing COPD (chronic obstructive pulmonary disease) later in life (Boschetto et al. 2006; Govender et al. 2011). Thus, occupational exposure to dusts, chemicals, and gases has been shown to lead to COPD. Hip and knee injuries that occur at work have effects on osteoarthritis later in life. (See Aluoch and Wao 2009 for a comprehensive summary of this research. More information on occupational health can be found in Levy and Wegman 2000.) Many of these studies utilize small sample sizes and are correlational. Moreover, while they are key in identifying the mechanism through which blue collar work may affect subsequent health, they do not focus on the net effect, which is the focus of our study. We thank an anonymous reviewer for further highlighting this important mechanism.

income not only significantly improve mental and physical health, but also increase the prevalence of alcohol consumption.

The prevalence of uninsured individuals further varies substantially across occupation and industry classes, which in turn may mediate the impact of occupational choice on health (Newhouse 2004; McWilliams et al. 2007; Hadley 2003). For instance, among non-professional and non-managerial occupations, almost half of all non-elderly workers in agriculture are uninsured, 40 % of such workers in construction are uninsured, and 25 % of workers in the wholesale and retail trade lack insurance.[6]

Data from the American Time Use Surveys (ATUS) indicate that work-related physical activity (measured in equivalent metabolic units) varies substantially across occupations, being expectedly largest in mining, agriculture, construction, and manufacturing jobs, and lowest among management and administrative jobs. Leisure-time physical activity also varies across occupations, often in inverse relation to work-related activity, suggesting some substitution between the two types driven by time constraints (Saffer et al. 2011). This heterogeneity in physical activity across jobs, combined with differential effects of work-related versus leisure-time physical activity, may have lasting effects on subsequent health outcomes, *ceteris paribus*.

As much of life is spent working, social influences through workplace peers can further impact people's health behaviors. Bang and Kim (2001), for instance, estimate prevalence of cigarette smoking by occupation and industry in the US, using data from the National Health and Nutrition Examination Survey. They document considerable differences across occupations and industries. Smoking prevalence is highest among material movers, construction laborers, and vehicle mechanics and repairers, and lowest among teachers. Among industry groups, the construction industry had the highest prevalence of cigarette smoking. Using the California Occupational Mortality Study data set, Harford and Brooks (1992) find that cirrhosis mortality is highest among individuals with blue-collar type jobs (e.g., construction laborers and machinists) or jobs where alcohol was easily available (e.g., bartenders and waitresses). Using the 1995 Australian Health Survey, Burton and Turrell (2000) find that individuals in blue-collar occupations were more likely to be classified as insufficiently active, although this association could not be explained by hours worked. Powell et al. (2005) conclude that peer effects have a significant impact on youth smoking behavior and that there is a strong potential for social multiplier effects. Thus, with respect to initial occupation, health behaviors of young adults and youth may be especially susceptible to peer influences at the workplace. The presence of peer influences and social transmission in the workplace may be an indirect pathway through which initial occupational choice may exert a stable and persistent impact on health behaviors.

2.3 Prior studies on first occupation

Given these plausible mechanisms, numerous social scientists have studied the empirical relationship between work status, job characteristics, and health.[7] For instance, the

[6]See *Health Insurance Coverage in America, 2008*, accessed at The Kaiser Family Foundation website: http://facts.kff.org/chartbook.aspx?cb=57.

longitudinal Whitehall studies examine the health of civil servants in London, focusing on how occupation affects health (Marmot and Smith 1997; Marmot and Bobak 2000; Marmot 2001). In general, these studies find that occupational status, job insecurity, and stress, among other factors, affect various dimensions of health, including coronary heart disease, self-reported health status, various morbidities, and health behaviors.

This literature, however, has largely ignored any potential impact of occupational choice. Three recent studies address this gap and acknowledge the importance of early occupational choice; these studies are the first to empirically investigate how initial occupation and job characteristics may have a cumulative impact on subsequent health status. Fletcher et al. (2009) match job characteristics from the Dictionary of Occupational Titles to individual records from the Panel Study of Income Dynamics (PSID) to investigate the cumulative impact of job characteristics on self-rated health status. They construct 5-year cumulative measures of job characteristics, and find that individuals working in jobs with high physical demands or harsh conditions experience declines in their health, with stronger adverse effects for females and older workers. This is consistent with Fletcher and Sindelar (2009), who also find that a blue-collar occupation at labor force entry is associated with subsequent decrements in self-reported health status. Sindelar et al. (2007) aggregate three-digit occupational codes into ten broad categories and consider the effects of early occupation choice on self-rated health status and ever having a heart attack. They also confirm that first occupation has a durable impact on later health, though the impact varies by health measure and the degree of control for other observables.

2.4 Contributions

Our study adds to this emerging literature on the importance of early occupational choice on subsequent health and fits within the broader economics literature on enduring effects of early circumstances on health and labor outcomes. The studies noted above make a seminal contribution to this literature, though the focus thus far has been on self-rated health status and on the incidence of heart attacks. This study investigates the durable impact of first occupation on a host of subsequent health behaviors including smoking, drinking, physical activity, and obesity, all of which are important proximate inputs into health. The focus on health behaviors is warranted for at least two reasons. First, durable effects on health behaviors (that is, investments in health) may be relatively more apparent, and therefore easier to identify statistically, than effects on health outcomes, which can take a long time to materialize. Second, the focus on health behaviors also underscores potential pathways through which initial labor market choices may eventually have lasting effects on health, and in some cases capture longer-term effects on behaviors induced by shifts in health status. For instance, if a worker contracts cancer due to job-related exposures, he may quit smoking, lose weight, or perhaps exercise more. We also undertake a first step in directly investigating channels of effect, including shifts in income, hours worked, and other potential mediators through which initial occupational choice may influence health behaviors.

[7]See Theorell (2000) for a review of this literature.

In summary, this study provides the first empirical estimates for the lasting durable impact of first occupation on subsequent health behaviors for the general population and across demographic groups, while paying careful attention to potential bias from unobserved selection, potential channels of effect, and potential confounding between durable and contemporaneous effects.

3 Methodology

The above discussion suggests that early labor market choices can be a significant input into an individual's health production function. However, empirically identifying the causal effect of occupation on health behaviors is complicated by at least two issues. The first is what we refer to as structural endogeneity or reverse causality. Our focus on *initial* occupational choice and its impact on *subsequent* adult health investments bypasses this simultaneity concern. The second, what we refer to as statistical endogeneity, wherein an individual's early labor market choices and subsequent health investments may depend on a common set of unobserved factors (for instance, family history or risk tolerance), is a more relevant concern for this study.

Consider the following linear specifications of the structural production function for health behaviors (H_{it}) and initial occupational choice (O_{it-1})[8]:

$$H_{it} = \beta_1 O_{it-1} + \beta_2 X_i + \beta_3 \mu_i + \varepsilon_{it} \tag{1}$$

$$O_{it-1} = \alpha_1 X_i + \alpha_2 Z_{it-1} + \alpha_3 \mu_i + \upsilon_{it} \tag{2}$$

Equation (1) is a production function for health behaviors (H_{it}) at adulthood, which is a function of occupational choice at labor market entry (O_{it-1}) and observable characteristics such as age, gender, race, education, and years since initial labor market entry (X_i). Equation (2) postulates the determinants of occupational choice at labor market entry. The vector Z_{it-1} represents observed and unobserved variables specific to the occupation decision, such as parental occupation, initial labor market conditions, or private information regarding expected costs and benefits associated with the occupational choice, which may not directly impact the individual's subsequent health status (conditional on own and parental income or wealth, and other investments). The vector μ_i denotes common unobserved determinants of occupational choice that may also influence health behaviors, for instance family background, tolerance towards risk, and the rate of time preference. The subscripts refer to the ith individual in time period t, and $t-1$ denotes initial labor market entry or earlier periods.

Our objective is to estimate β_1 in order to assess the existence and strength of a possible causal relationship between first occupation and health behaviors. However, single equation methods applied to Eq. (1) may not yield causal information due to the presence of non-

[8]The health-investment production function is based on Grossman (2000), extended to include occupational choice as an input into health investment. The occupational choice model is based on the theory of human capital investment (for example, see Borjas 2004; Boskin 1974).

random selection into different occupations and investments in health—that is, correlation between μ_i and O_{it-1} ($\alpha_3 \neq 0$). Our estimation strategy proceeds in a stepwise fashion. Initially, we ignore the statistical endogeneity and estimate Eq. (1) using a standard regression model. We begin with a parsimonious set of covariates, and then estimate models with an expanded set of covariates, including fixed state effects based on both current residence as well as during labor market entry, family history, risk tolerance, and employment information, some of which are typically unobserved in other data sets. Estimating both the basic and the extended models allows us to evaluate how much of the association between early occupational choice and later health behaviors appears to be driven by omitted individual heterogeneity. If the magnitude of the marginal effect of first occupation is highly sensitive to the inclusion of the additional covariates and typically-unobserved factors (such as risk-aversion, family history, and state of residence at initial labor market entry), then it is likely that factors that remain unobserved also play some role in this relationship.[9] This assumption is reasonable if one is using a multi-purpose, secondary data set, where the information collected on respondents may not include all information relevant to the outcome under study (Altonji et al. 2005). This may not be the case, however, in a rich longitudinal data set such as the PSID, which includes measures of parental investments, family history, and residence at initial labor market entry, as well as measures of the respondent's tolerance towards risk.

We refer to this problem as selection on observables and selection on unobservables (Altonji et al. 2005). We use these terms to acknowledge that respondents are not randomly sorted into occupations and health behaviors. Selection on observables refers to observed factors (such as age, gender, and race) that are correlated with both initial occupation and subsequent health behaviors. Selection on unobservables refers to possible factors that are not available in our data set, and will therefore influence the marginal effect of initial occupation.[10]

We perform several robustness checks in order to add to the weight of the evidence on the issue of causality. Lewbel (2012) presents an instrumental variables (IV) technique that is useful when valid external instruments are weak or not available. This procedure relies on the presence of heteroscedasticity in the error term of the first-stage equation, which is tested using a Breusch and Pagan (1979) test. The Lewbel IV procedure uses the deviation from the mean of a vector of independent variables interacted with the residual from the first-stage (occupational choice) regression as the identifying instruments. Therefore, we also implement an instrumental variables analysis where we exploit the heteroscedastic nature of the residuals to generate internal instruments. In particular, the Lewbel IV procedure uses

[9]The direction and magnitude, however, is unknown, depending on the nature of the joint distribution of the observed and unobserved characteristics.

[10]To the extent that learning of job-related hazards may take time, workers may selectively quit their initial jobs when they do learn (see for instance Viscusi and O'Connor 1984), based on their preference for risk-taking and/or tradeoff between risk and wages. This would tend to understate any adverse effects of initial occupation on unhealthy behaviors since the first job may be a "mistake" for some individuals and these individuals are not exposed to their initial occupation for a lengthy period of time. In subsequent analyses we control for both the initial and current occupation to partially tease out such sorting. And, to the extent that the same set of unobservable (for instance, risk and time preference or alternate opportunities in the labor market) determine both sorting into initial occupation and quitting the initial occupation, our estimation strategy (constrained selection models and instrumental variables) would partly account for selective quitting.

$(X - \overline{X}) * \widehat{u_2}$ as the identifying instruments, where X is a vector of independent variables that may include all independent variables or a subset of them, and $\widehat{u_2}$ is the residual from the first-stage (occupational choice) regression. As validation for this technique, we consistently find evidence of heteroscedasticity in our samples, and the test of over identifying restrictions does not reject the exclusion restriction associated with the internal IVs.

In alternate models, we supplement these internal IVs with external IVs to potentially improve upon the strength of the instruments and maximize precision. We follow Fletcher and Sindelar (2009) in using parental occupation (conditional on parental income and education) and early state labor market conditions as instrumental variables (IV) for first occupational choice. Diagnostic tests are consistent with the identifying assumption that these measures have no direct impact on future health behaviors (outside of their impact through occupational choice), and that these measures are significant predictors of first occupation. These estimates should, nevertheless, be interpreted with caution, owing to the challenges in identifying plausible exclusion restrictions. This part of the analysis does, however, allow us to place our findings on health behaviors within the context of the sparse, but important, prior studies that have considered health outcomes. If these prior estimates on health outcomes are plausibly causal, then we should also observe commensurate effects on health behaviors, which are proximate inputs into later health.

The degree of selection on the observables can be gauged by comparing the estimated coefficients on first occupation from the parsimonious and extended models. The degree of selection on the unobserved characteristics cannot be measured directly with non-experimental data. However, we can potentially bound this latter effect, allowing us to draw some inferences regarding the unbiased relationship between first occupation and later health behaviors.

Thus, the next step in our empirical strategy relies on an innovative approach proposed by Altonji et al. (2005), comprising two parts. The first step involves obtaining estimates of the effect of first occupation on health behaviors from a bivariate probit regression model in which the correlation between unobserved variables is fixed at various levels. This part of the analysis allows us to assess how sensitive estimates of the effect of first occupation are to the potential problem of correlated unobservables. The second step computes the amount of sorting into first occupation and adult health behaviors on observed variables, and obtains estimates of the effect of first occupation under the assumption that the degree of sorting on unobserved variables is equal to the degree of sorting on observed variables.

Altonji et al. (2005) argue that if the observable determinants of an outcome are truly just a random subset of the complete set of determinants, then selection on observable characteristics must be equal to selection on unobservable characteristics. This assertion of equal selection is unlikely to be true, and in fact, given our specialized longitudinal data set, we would expect selection on observable factors to be greater than selection on unobservable factors. Thus, estimates obtained under the assumption of equal selection are likely biased downwards, and represent a lower-bound estimate. The upper bound effect is the estimate from the naïve single equation extended model that assumes no additional selection on unobservable variables.

The advantage of the Altonji et al. (2005) procedure is that it allows researchers to assess the possible existence and strength of a causal relationship without requiring the use of identifying assumptions that are often not credible—for example, the existence of valid instruments in an instrumental variables context or other ad hoc exclusion restrictions. As a result, without any other identifying assumptions, researchers can estimate the degree of sorting on unobservable factors using the observed data, and identify a lower bound on the causal parameter estimate.

As a final step, we implement an exploratory analysis of potential mediators to inform the strength of the specific mechanisms underlying the impact of first occupation on later health behaviors. The estimated specifications thus far only include exogenous socioeconomic and predetermined factors so as not to "over-control" for factors that may be potential pathways. In alternate analyses, we re-estimate specification (1) by incorporating household income, hours worked, and current occupational status to gauge the extent to which the estimated effect (if any) of first occupation on subsequent health behaviors can be explained by these mediators.

4 Data

The Panel Study of Income Dynamics was begun in 1968 and covers a representative sample of US. individuals (men, women, and children) and the family units in which they reside. By the end of the 2003 survey, the PSID had collected information from over 65,000 individuals, spanning as many as 36 years of their lives. Starting in 1997, the surveys were conducted biennially. Between 1968 and 1972, data collection took place through in-person interviews using paper and pencil questionnaires. Thereafter, most interviews were telephone interviews or, starting in 1993, computer assisted telephone interviews.[11] Comprehensive information on health behaviors, labor market characteristics, and demographic characteristics are readily available in the PSID. While data on health habits are reported in 1999–2005, we constrain our analysis to 2005 in order to focus on health behaviors later in life.[12] (The average age in our sample in 2005 is 54 years.) However, we exploit the longitudinal nature of the data set by utilizing information from prior years, particularly regarding labor market characteristics. Information on the head of the household and spouse are used due to the sparse information on health behaviors for other family members.

4.1 Health habits

4.1.1 Obesity—We consider obesity as the net effect of several health habits, specifically resulting from the caloric imbalance between eating (caloric intake) and physical activity (caloric expenditure). Self-reported weight and height are available in the PSID in the 1986,

[11] Source: http://psidonline.isr.umich.edu/Guide/Overview.html.

[12] We also estimated models based on 1999 outcomes (results available upon request); estimates were similar and conclusions remain unchanged.

and 1999–2005 waves, for the head of the household and the wife. The body mass index, or BMI, is computed as weight in kilograms divided by height in squared meters. Obesity is defined by the National Institutes of Health as having a BMI of 30 kg/m^2 or greater. According to data from the National Institutes of Health, the percentage of individuals 18 years of age or older classified as obese has risen in the United States from 12.7 % in the 1960s to 31.7 % in 2004. Obesity carries many risks for a host of disorders, including heart disease, hypertension, stroke, cancer, and depression (Must et al. 1999; Mokdad et al. 2003). A variety of economic causes have been explored, including reductions in job strenuousness (Philipson 2001; Lakdawalla and Philipson 2009), technological innovation in food processing and preparation (Cutler et al. 2003), the growing availability of restaurants and the increased labor force participation of females (Chou et al. 2004; Rashad et al. 2006), urban sprawl (Ewing et al. 2003), and time preference for the present (Komlos et al. 2004; Smith et al. 2005; Zhang and Rashad 2008). Note that due to differing response rates, the sample size for each health habit will vary, as can be seen in Table 2.

4.1.2 Alcohol consumption—For 1999–2005, the PSID asks the head of the household and spouse (if any) to report on the average number of drinks consumed per day. According to an NIAAA (National Institute on Alcohol Abuse and Alcoholism) screener for alcohol use disorders, consuming 5 or more drinks in a day (for men) or 4 or more drinks in a day (for women) at least once in the past year is indicative of an “at-risk drinker.” We therefore construct a dichotomous indicator for such at-risk drinking.[13] Alcohol consumption—and particularly abuse—can have adverse effects on labor market productivity, morbidity, mortality, and economic growth (Cesur 2009). Yet some studies have shown that moderate drinking has a positive effect on wages, largely operating through social networking channels (Berger and Leigh 1988; French and Zarkin 1995; Hamilton and Hamilton 1997; MacDonald and Shields 2001; Tekin 2004; Bray 2005). Other studies conclude that the positive relationship between moderate drinking and earnings mostly represents unobserved selection bias (Saffer and Dave 2005a, b; Dave and Kaestner 2002).

4.1.3 Smoking—The PSID asks questions on smoking by the household head and the spouse in 1986, and again in 1999–2005. We construct a dichotomous indicator for current smoking as the outcome measure. According to the Centers for Disease Control and Prevention, tobacco use, which can lead to lung, larynx, esophageal, and oral cancers, is the nation’s most preventable cause of morbidity and mortality.[14]

4.1.4 Physical activity—In 1999–2005, the PSID asks the head of the household and their spouse to report on the frequency of light and heavy physical activity. The questions are: “How often do you participate in light physical activity—such as walking, dancing, gardening, golfing, bowling, etc.?” and “How often do you participate in vigorous physical activity or sports—such as heavy housework, aerobics, running, swimming, or bicycling?” Individuals report on their frequency of participation and a reference time unit, which we standardize to an average weekly frequency. Physical activity has been shown to be an

[13]This is similar to the standard definition for binge drinking, though binge (or heavy episodic) drinking refers to consuming 5/4 or more drinks for males/females on a single occasion rather than a single day.

[14]See http://www.cdc.gov/tobacco/ for more information.

important factor in keeping morbidity and mortality at bay, and most Americans do not engage in sufficient amounts of physical activity (USDHHS 1996; Pratt et al. 1999).

4.2 First occupation variables

One challenge in identifying a relationship between occupational choice at labor market entry and health behaviors is in defining initial labor market entry. In this context, initial labor market entry ideally refers to initial employment in a full-time occupation. However, survey questions that ask about current employment do not necessarily restrict the response to full-time jobs. We thus employ two definitions of initial occupation: one based on a recent question about initial employment, which likely reflects first full-time employment, and one that exploits the longitudinal nature of the data and records the first employment observed in the data. Specifically, the first measure is based on the individual's own report in 1997–2005: "Thinking of your first full-time regular job, what kind of work did you do?" Since recall bias is likely minimized in earlier years, we use responses from 1997; if these are missing for the head of the household or the wife, we use 1999, then 2001 responses. These measures are based on 1970 Census, and three-digit occupation codes are provided. From these we derived 16 occupational categories: Craft, Operative, Transport, Labor, Farmer, Manager, Sales, Clerical, Craft, Operative, Transport, Labor, Farmer, Service, Private, and Professional. A dichotomous indicator is further defined as equal to 1 if the category is one of Craft, Operative, Transport, or Labor, and 0 otherwise, denoted as "blue-collar" occupation.[15] The advantage of this measure is that it is based on the individual's own report of the initial full-time, regular occupation, which may be the salient measure for their subsequent health habits; being based on three-digit occupation codes, it contains a greater degree of detail.

The second measure is generated based on the first occupation reported by the individual, which exploits the longitudinal nature of the PSID. The generated 'first occupation' code was created using 3-digit occupations for the head & wife starting from 1974; prior to 1974, however, only 1-digit occupation codes were initially coded[16] Based on these, a dichotomous indicator representing blue-collar work is defined to reflect the following occupations: Craft, Operative, or Labor. This measure, by using reported information at the time of first occupation, may minimize potential recall bias. However, a concern with this "generated" blue-collar first occupation measure is that, rather than coding the first occupation for some individuals, it will capture the occupation that was being held at the time the individual was first head of his or her household. This will likely not be until much later in the panel for individuals surviving to be heads of their households in 1999 and 2005. For this reason, the first measure based on the recalled first occupation variable is likely to be more reliable, though results are fairly consistent for key outcomes across both measures. Moreover, the first measure is also more likely to reflect the individual's first full-time occupation.

[15]Farmers denote farm owners and are therefore excluded from blue-collar classification. Farm workers are included in the laborer category and thus classified as blue-collar.

[16]A dummy variable is included in models indicating whether a 1-digit occupation code was used instead. This dummy variable will likely be correlated with age.

Both measures are highly consistent with each other, which adds a degree of validation. Table 1 indicates that approximately 33 % of respondents were initially blue collar workers based on recall in 2005, and approximately 30 % were initially blue collar workers based on their first reported occupation in the earliest PSID wave. Models also control for years since first occupation, defined as the difference between the current survey year and the year of first-reported occupation, to capture the duration effect of time since initial labor-market entry.

4.3 Individual characteristics

All models control for individual characteristics pertaining to gender, race/ethnicity, education, age, marital status, employment status, and fixed state effects for current residence.[17] Alternate models also control for parental characteristics, including the educational status of the mother and father and whether the family was poor, as well as fixed state effects for residence at time of initial labor market entry.[18] Note that controlling for marital status may also capture the potential effect of the partner's occupation on the respondent's outcomes (Mendolia 2012).

A module probing the individual's tolerance towards risk is administered to a subset of individuals in 1996. Measures of risk aversion are obtained from a series of questions involving the willingness to choose different levels of lifetime income with varying probabilities. Answers to the questionnaire separate individuals into four distinct categories of risk preference, ranging from the most risk tolerant to the most risk averse.[19] Since the PSID respondents only partake in the risk module once, the measure of risk tolerance is time-invariant. Some studies have shown that traits associated with risk tolerance are generally stable, may have a biogenic basis, and have some constancy across various situations (Howard et al. 1997; Menza et al. 1993), though it should also be noted that risk tolerance may evolve over the life cycle and may shift in response to changes in income and health. Individuals' propensity for risk-taking is typically unobserved in other datasets, and represents an important source of non-random selection into outcomes, since it may affect both occupational choice and participation in other risky and unhealthy behaviors. We therefore include measures of risk-tolerance in supplemental analyses and extended models to address this potential selection bias.

[17]In our initial specifications, we do not control for mediating factors such as household income and hours worked, which may represent mechanisms through which initial occupation affects health behaviors. Models reported in Table 4 assess the importance of these mediators. We also do not control for health insurance as well as health status and diagnoses such as hypertension, cancer, and other conditions in our main models for two reasons. First, these measures are potentially endogenous to occupation and health behaviors. Second, they represent potential pathways through which occupational choice may impact subsequent health behaviors. We do note however that in supplementary analyses (results available upon request), the inclusion of these covariates does not qualitatively alter our results or conclusions.

[18]Parental education is categorical: (1) grades 0–5, (2) grades 6–8, (3) grades 9–11, (4) grade 12, (5) 12 grades + non-academic training; R.N., (6) some college, no degree; associate's degree, (7) college baccalaureate degree and no advanced degree mentioned; normal school; RN with 3 years of college, and (8) college, advanced or professional degree; some graduate work.

[19]The categories can be ranked in order, without any functional form restrictions on the preference parameters or the utility function. Almost half (48.6 %) of the respondents can be classified in the most risk-averse category, with 31.8 % divided equally among the second and third most risk-averse groups, and 19.6 % comprising the least risk-averse categories. Barsky et al. (1997) provide a detailed analysis of the survey instrument and validate it by showing that it is related to behaviors (insurance, portfolio allocation, migration, risky health behaviors, self-employment) that would be expected to vary with an individual's propensity to take risks. See PSID documentation for the specific income gamble questions (http://psidonline.isr.umich.edu/data/Documentation/Cbks/Supp/rt.html).

4.4 Instrumental variables

In the instrumental variables (IV) models based on external instruments, we use information on the county unemployment rate in 1968 (the earliest year county unemployment is reported) when the average respondent is 18 years of age, and whether the respondent's father worked in a blue-collar occupation. These instruments reflect early economic conditions and likely affect occupational choice at labor market entry. Similar instruments (early labor market conditions and parental occupation) were also utilized by Fletcher and Sindelar (2009). We confirm that these measures are significant predictors of whether the respondent's first occupation was blue-collar in the expected direction. Higher county unemployment rates in the initial wave raise the probability of blue-collar work; similarly, a respondent is more likely to work in a blue-collar occupation if his or her father also did so. Conditional on parental education and family resources (which we control for in the extended models), these variables do not have any direct effects on health behaviors, as evidenced by the test of overidentification restrictions. Alternately, these instrumental variables are also statistically insignificant, with close-to-zero magnitudes when included in the extended specifications, again suggesting that they do not directly impact health behaviors. However, the instruments lack statistical power and the estimates should therefore be interpreted with caution. This underscores the difficulties of implementing a conventional IV-based strategy, particularly when analyzing the effects of early circumstances, since first occupation (at least 30 years prior to current adult outcomes) is difficult to predict with strong statistical power and in a way that is uncorrelated with subsequent inputs into health. Thus, the alternate approaches used in this paper add to the weight of the evidence bearing upon the research question.

Out of approximately 11,000 individuals who were either head of household or spouse in 2005, the sample size after deleting missing information on the aforementioned variables is 6303. Summary statistics are provided in Table 1.

5 Results

5.1 Baseline estimates

As shown in Table 1, there are significant differences in health behaviors between individuals whose initial labor market entry was in blue-collar occupations and those whose first jobs were in non-blue collar occupations. In general, initial blue-collar workers tend to engage in more unhealthy behaviors; they are more likely to be obese, have higher at-risk alcohol consumption, and be current smokers. However, initial blue-collar workers are also more physically active.

While these differences in health behaviors are suggestive, individuals are not randomly selected into initial blue-collar occupations. There are also significant differences with respect to other observable characteristics between blue-collar and non-blue collar workers. For instance, initial blue-collar workers are more likely to be male, low-educated, slightly older, married, and have low-educated and poor parents. Thus, the association between first occupation and subsequent unhealthy behaviors also reflects confounding due to such non-random selection on observables and potential selection on unobservables. The multivariate analyses address these concerns.

Table 2 presents estimates of the impact of first occupation on health behaviors. These models include the full set of covariates, including state-level indicators for both current residence and residence at time of labor market entry, parental characteristics, and indicators of risk tolerance. Baseline results pertaining to OLS and probit models will be discussed first. We primarily discuss estimates based on our preferred measure of recalled first-occupation, though for the most part estimates are consistent across both measures (recalled and generated first-occupation).

The first panel of Table 2 presents models for obesity. Marginal effects from probit estimates suggest that initial blue-collar occupation is associated with a 7.6 % point increase in obesity, based on recalled occupation[20] The reported estimates are derived from models with an extended, rich set of covariates. Full results for all coefficient estimates are reported in Appendix Table 6.[21] These estimates are similar to those derived by models with a limited set of covariates (results not shown).[22] This robustness is validating, and suggests that additional controls for parental history, risk tolerance, and indicators for state of residence at initial labor market entry, which are typically unobserved in other datasets, do not further lessen the impact of first occupation. This potentially diminishes the role of additional selection on other factors if these added controls are a random subset of all unobserved factors.

The second panel of Table 2 presents models for daily alcohol consumption. Individuals who enter the labor market in a blue-collar occupation have a higher probability of being an at-risk drinker by about 2.1 points (based on recalled first occupation). This is about a 10.6 % increase relative to the outcome mean. However, the coefficients are statistically insignificant. Other covariates affect alcohol consumption as expected and noted in the literature (Dave and Saffer 2008).[23] Notably, a higher degree of risk aversion is associated with lower levels of drinking.[24]

The third panel of Table 2 presents estimates of the impact of first occupation on the propensity of being a current smoker.[25] There is limited evidence that initial blue-collar work is associated with a slightly higher smoking rate (2.9 % points based on generated first occupation), though the effects are not significant. This may reflect decreased current

[20]The effect magnitude based on the generated first-occupation variable are smaller (4.7 % points) but still significant at the 5 % level (for a two-tailed test).

[21]The effects of the other covariates are consistent with the literature on obesity; obesity is higher among individuals who are black (relative to all other races), college-educated (though imprecisely estimated), and never-married, and individuals whose parents are low-educated. The coefficients on the indicators of risk tolerance (least risk-averse being the reference category) suggest that more risk-averse individuals have lower probabilities of being obese.

[22]Complete results for the parsimonious specifications for all outcomes can be found in Kelly et al. (2011).

[23]The prevalence of at-risk drinking is higher among males and non-Hispanic whites, and among never-married individuals.

[24]As a validation check, we also estimated models for moderate drinking (results not reported). There were generally no effects of initial occupation on moderate alcohol consumption, which is to be expected given that prior studies have generally found inconsistent effects between moderate alcohol consumption and labor market outcomes and health.

[25]Coefficient estimates for the other covariates are consistent with the literature (Chaloupka and Warner 2000) suggest that the likelihood of being a current smoker is higher among males, non-Hispanic whites, individuals with less than a high school education, those who are single, and those with low maternal education. Individuals with a high tolerance for risk are also more likely to be current smokers.

smoking prevalence among all groups. Our data show that current smoking prevalence declined from 24.9 % in 1999 to 20.7 % in 2005 among individuals whose initial occupation was blue-collar; this is a larger decrease than that experienced by individuals whose initial occupation was not blue-collar. Thus, there is some convergence in smoking rates between these two groups over time.[26]

The fourth panel of Table 2 presents models for physical activity. Initial blue-collar workers have a higher frequency of weekly physical activity by about 0.7 times, relative to those whose first occupation was not blue-collar, based on recalled first occupation. The higher value is about 70 % of the unadjusted difference based on the reported means in Table 1, which suggests that the increase in obesity reported among blue-collar workers reported above is likely due to higher caloric consumption. The effect based on generated first occupation is oppositely signed but close to zero in magnitude. The inconsistency in the estimates between the two measures of first occupation may partly be due to measurement error. The instrumental variables models (discussed below) address potential measurement error and resolve this inconsistency.

To summarize, single-equation estimates suggest two points. First, there is evidence that initial blue-collar work has some lasting effects on health behaviors. Second, selection on observed factors account for about 60–90 % of the unadjusted difference in health behaviors (as measured by obesity and physical activity) between the groups of workers; however, the effect magnitudes are not sensitive to additional controls for risk-tolerance, parental income and education, and indicators for state of residence at labor market entry.[27]

5.2 Instrumental variables

In order to bypass the issues with weak external instruments, we present estimates based on internal instruments as proposed in Lewbel (2012). These I Vs have stronger predictive power and are also plausibly excludable based on the tests of overidentification restrictions. Indeed, Lewbel (2012) recommends this methodology precisely to overcome issues with questionable and low-powered external instruments. These results indicate that initial blue-collar occupation leads to a higher probability of being obese (4.4 % points, based on recalled first occupation), a higher probability of being an at-risk drinker (5.8 % points), a higher probability of being a current smoker (6.2 % points), and a higher frequency of physical activity (5.8 times per week). The consistency of the estimates between both measures of first occupation is validating. Some of these estimates are imprecise due to limited sample sizes in the extended models and, while the internal IVs are stronger, the statistical power of these IVs may still not be adequate.

In order to improve statistical power, we also combine the internal IVs with external IVs. While the F-statistic on the joint significance of the set of IVs increases in most cases,

---

[26]We find some significant effects for smoking in 1999 (results available upon request).

[27]Studies have indicated that attrition in the PSID does not generally bias results (Fitzgerald et al. 1998; Zabel 1998). In Appendix Table 7, we show results from the baseline model where 2005 PSID longitudinal weights, which take attrition into account, are employed in the regressions. These weights take into account nonresponse and mortality; the reciprocal of the conditional probability that the individual responded given that the individual is alive is the factor used to adjust the weight for differential nonresponse and mortality. The procedure is described in further detail in Gouskova et al. (2008).

standard errors remain relatively large and the coefficients are imprecisely estimated in many cases. It is validating, however, that the coefficient magnitudes remain highly robust across all outcomes and measures with both sets of instruments (internal IVs versus internal + external IVs.) It is also validating that the IV-based estimates are generally consistent with the estimates from the extended baseline specifications.

5.3 Constrained selection models

Constrained selection models (reported in Table 3) allow us to assess the sensitivity of the estimates to additional amounts of selection on unobservable factors.[28] For obesity, at-risk drinking, and smoking, we find that even small amounts of additional positive selection on unobserved factors can wipe out the positive effects of initial blue-collar work on these outcomes.[29] Additional positive selection would lead to a negative effect of initial blue-collar occupation on obesity, at-risk drinking, and smoking. Additional negative selection, on the other hand, strengthens the magnitude of the effects reported in Table 2 for the extended specifications and suggests even stronger adverse effects on health behaviors.[30]

Altonji et al. (2005) note that selection on observable factors can be helpful in assessing selection on unobservable factors. Model 12 presents estimates based on the assumption that selection on unobservables is equal to the selection on observables; this assumption is appropriate in general datasets where the factors that we observe are a random subset of all determinants of the outcome. For the PSID, which is a specialized longitudinal dataset with extensive information on labor market history and other individual and family characteristics, the equal selection rule is likely to overestimate the amount of selection on unobservable factors. This is consistent with our earlier estimates, which showed that adding richer covariates to the specification do not lead to substantial changes in effect magnitudes. Thus, if the estimates from the single-equation extended models represent upper bound estimates, then the estimates from the models based on the equal selection constraint represent lower bound estimates.

Estimates from model 12 suggest that there may be slight additional positive selection on unobservables ($\rho = 0.1$) for obesity, in which case there would be no significant effect of initial blue-collar occupation on obesity. For smoking, model 12 suggests additional negative selection ($\rho = -0.3$) in which case initial blue-collar occupation is associated with a higher likelihood of being a smoker in 2005. Similarly, for at-risk drinking the equal selection constraint suggests additional negative selection ($\rho = -0.03$) and a higher likelihood of being an at-risk drinker associated with initial blue-collar work. Thus, if the estimates from the single-equation extended models represent upper bound estimates, then the estimates from the models based on the equal selection constraint represent lower bound estimates.

[28]See Kelly et al. (2011) for the full results for other years and specifications, which are qualitatively similar to those reported in Table 3.

[29]As physical activity is a continuous measure of weekly frequency, we are unable to estimate constrained-selection bivariate probit models for this outcome.

[30]Selection effects theoretically can be either negative or positive. For instance, individuals with a high rate of time preference (more present oriented) may be more likely to enter blue-collar occupations and also less likely to invest in their health leading to higher obesity; this would lead to positive selection bias. On the other hand, individuals with a taste for physical activity and manual labor may also be more likely to enter blue-collar occupations but would be less likely to be obese; this would lead to negative selection bias.

Note that these models suggest that *if* there is additional *positive* selection then the estimates are wiped out. The constrained selection models suggest that this may be the case for obesity, in which case the positive effect of initial blue collar work is essentially wiped out. However, with respect to smoking and at-risk drinking, the equal selection model (row 12) points to negative selection on unobservables, in which case initial blue-collar work is associated with a significantly higher smoking and at-risk drinking rate.

5.4 Heterogeneous effects

The estimates thus far represent an average population effect, which may mask considerable heterogeneity in responses across demographic groups. Table 4 presents estimates based on models stratified across socio-demographic characteristics. These models suggest that initial blue-collar work has larger positive effects on alcohol consumption and smoking among males, relative to females. Similarly, initial blue-collar work raises the probability of being obese and at-risk drinking somewhat more for Whites than for other races; however, the increase in smoking is larger among non-Whites. Initial blue-collar work is also associated with larger increases in physical activity among females (than among males) and among non-Whites (than among Whites); this is consistent with smaller increases in obesity among these groups. These patterns in effects on health behaviors across gender and race groups are also generally consistent with reported effects on health across these groups in Fletcher and Sindelar (2009). Some of these estimates are imprecise due to reduced cell sizes.

5.5 Exploratory analysis of potential mediators

Potential mechanisms through which initial occupational choice may impact health behaviors include income, hours worked, and current occupation. Estimates in Table 5 assess the importance of these potential mediators by alternately adding these measures to the baseline model and gauging the effect magnitudes. Comparing baseline estimates to those that include household income and hours worked, we find that the effect magnitudes are virtually unchanged.[31] This suggests that the effects of first occupation on health behaviors do not solely operate through income effects or work intensity. Interestingly, when models control for current occupation codes (model 4), positive effects of initial blue-collar work on obesity and frequency of physical activity become somewhat stronger. This is expected, and validating, since the correlation between initial blue-collar work and current blue-collar work is not particularly robust. When current occupation is not accounted for, initial occupation confounds two groups of individuals, those who shift from blue-collar to non-blue collar over time and those who do not. If the adverse effect of initial blue-collar work on healthy behaviors is attenuated when individuals are no longer currently working in blue-collar jobs (which is to be expected), then controlling for current occupation should make the estimated effects larger in magnitude. This latter effect is evidence of a dose-response relation; the impact of initial blue-collar occupational choice on health behaviors appears to be somewhat more pronounced if the individual continues in that occupation over their life.

[31]Household income is a computed variable, equal to the sum of: Taxable Income of Head and Wife, Transfer Income of Head and Wife, Taxable Income of Other Family Unit Members (OFUMs), Transfer Income of OFUMs, and Social Security Income.

6 Discussion

This study is the first to assess the existence and strength of a potential causal relationship between initial occupational choice at labor entry and subsequent health behaviors and habits. Unadjusted differences and single-equation models confirm that starting work in blue-collar occupations is subsequently associated with unhealthy behaviors (with the exception of physical activity) during later adulthood. In addition, we examine how much of this association is consistent with a causal mechanism and how much of it is being driven by non-random selection.

We utilize several methods to address potential confounding: (1) controlling for a rich set of individual characteristics and state fixed effects based on current residence as well as residence at time of labor market entry; (2) estimating instrumental variables models using external and internally-generated instruments; and (3) estimating constrained selection models. We estimate effects in 2005 in order to capture outcomes at older ages. Since our comprehensive measures on health behaviors are only available in more recent years, and the year of initial employment for the respondents in our data set varies, we cannot estimate effects for a set number of years (say, 30) between initial employment and subsequent health behaviors, though all models control for the number of years since initial employment.

Our results suggest that initial blue-collar work is associated with a higher probability of being obese, being an at-risk drinker, and being a current smoker later in life, as well as higher physical activity later in life. Specifically, results from the extended and IV specifications indicate that initial labor entry in blue-collar work raises obesity by about 4–8 % points (17–32 % relative to the baseline mean). There is also suggestive evidence that initial blue-collar work may raise smoking prevalence by about 6 % points (28 %), though the smoking effects are imprecise estimates. These results may explain the higher incidence of heart attacks found in Sindelar et al. (2007). There is also evidence of an increase in the likelihood of being an at-risk drinker by about 6 % points (30 %) associated with initial blue-collar occupation. We also find a suggestive increase in the frequency of physical activity by between 1 and 5 times weekly (10–40 %), which may be related to work-based physical activity. Studies have found some evidence of a substitution effect wherein individuals who have more physically-demanding jobs are less likely to be physically active outside of work (Saffer et al. 2011; Colman and Dave 2011). Even if total physical activity is higher among manual workers, the specific composition of physical activity has implications for health; specifically, leisure-based physical activity is found to be health promoting whereas work-based physical activity, especially repetitive or factory tasks, tend to have little positive health effect (Saffer et al. 2011).

We find that a substantial part of the observed difference (60–90 %) is due to non-random selection on observable factors, but that the effect magnitudes are sensitive (in terms of diminution) to additional positive selection on unobservable factors. If the additional selection is negative, then the estimated effect magnitudes become stronger. However, drawing upon the weight of the evidence from all of our various methodologies, a residual effect of first occupation on subsequent health behaviors remains, which is consistent with a causal behavioral framework.

Our results have implications for both policy and future research. That the impact of first occupation has such durable impacts on health habits suggests that disruptions of and improvements in the harmful health habits would be beneficial. Benefits would accrue to workers, their firms and society in general. Society would benefit due to possible increased productivity, and reduced costs of shared medical expenses (e.g. Medicaid and Medicare) and government programs (e.g. disability insurance). More indirect gains could include possible improved health habits through peer effects and parental patterning of health behaviors. Employers and governmental agencies such as OSHA could focus more on improving health habits. For example, Employee Assistance Programs, financial incentives to improve health habits, healthier food in cafeterias, and stricter bans on smoking in work environments (including outdoor work) could all help to improve health habits. Focus should be on workers in blue collar jobs and specific occupations with poor health habits. Future research should investigate the channels through which initial labor market experiences affect future health habits and why habits are tied stably to occupation.

Acknowledgments

The authors gratefully acknowledge funding from the National Institute of Aging (ROI AG027045). Maureen Canavan provided excellent research assistance. The authors acknowledge the use of Panel Study of Income Dynamics, Sensitive Data, produced and distributed by the Institute for Social Research, Survey Research Center, University of Michigan, Ann Arbor, MI (2007).

Appendix

See Tables 6 and 7

Table 6

Estimates of the effect of blue collar first occupation on health behaviors, 2005 OLS and probit models

Variables	Obese		Alcohol		Smoking		Exercise	
	First occupation recalled	First occupation generated	First occupation recalled	First occupation generated	First occupation recalled	First occupation generated	First occupation recalled	First occupation generated
Blue collar	0.0761*** (0.024)	0.0427* (0.024)	0.0208 (0.020)	0.0170 (0.020)	0.0004 (0.018)	0.0290 (0.019)	0.6846 (1.604)	−0.1654 (0.965)
Male	−0.1180*** (0.036)	−0.0975*** (0.035)	0.1790*** (0.030)	0.1727*** (0.029)	0.0735*** (0.027)	0.0679** (0.027)	2.7038 (1.749)	2.8057* (1.589)
Black	0.1309*** (0.028)	0.1289*** (0.027)	−0.0764*** (0.022)	−0.0715*** (0.022)	−0.0330 (0.020)	−0.0393** (0.020)	5.3877 (4.919)	5.0502 (4.714)
Hispanic	0.0164 (0.136)	0.0513 (0.128)	−0.1073 (0.066)	−0.0235 (0.083)	−0.1202** (0.051)	−0.0782 (0.064)	−2.9854 (2.543)	−1.4475 (2.316)
Other			−0.1473** (0.059)	−0.1452** (0.061)			−2.7397 (3.208)	−2.8551 (3.043)
Some high	0.0431 (0.072)	0.0125 (0.069)	0.0752 (0.083)	0.0859 (0.082)	0.0691 (0.062)	0.0645 (0.060)	2.7799 (2.753)	2.8869 (2.798)
High	0.0166 (0.065)	−0.0069 (0.063)	0.0267 (0.068)	0.0331 (0.067)	−0.0582 (0.046)	−0.0523 (0.046)	3.5063 (2.858)	3.3737 (2.844)
Some college	0.0966 (0.070)	0.0649 (0.068)	0.0211 (0.070)	0.0277 (0.069)	−0.1011** (0.039)	−0.0949** (0.040)	2.9339 (3.432)	2.6785 (3.423)
College	0.0180 (0.068)	−0.0188 (0.066)	0.0072 (0.069)	0.0203 (0.069)	−0.1698*** (0.035)	−0.1615*** (0.036)	1.2731 (2.190)	0.9367 (2.164)
Years since first occupation (generated)	−0.0005 (0.002)	−0.0001 (0.002)	0.0000 (0.001)	0.0006 (0.002)	−0.0003 (0.001)	0.0006 (0.001)	0.0321 (0.042)	0.0515 (0.045)
Dummy for years since first occupation missing	0.0783 (0.054)	0.0870 (0.056)	−0.0486 (0.040)	−0.0324 (0.042)	−0.0807*** (0.030)	−0.0627* (0.033)	1.5901 (2.059)	1.8278 (1.930)
Age	0.0057 (0.007)	0.0022 (0.007)	0.0029 (0.007)	0.0016 (0.007)	0.0186*** (0.006)	0.0159** (0.007)	−1.3881 (0.936)	−1.4574 (1.020)
Age squared	−0.0001 (0.000)	−0.0000 (0.000)	−0.0001 (0.000)	−0.0001 (0.000)	−0.0002*** (0.000)	−0.0002*** (0.000)	0.0154 (0.011)	0.0162 (0.012)
Married	0.0509 (0.041)	0.0592 (0.040)	−0.0690* (0.036)	−0.0671* (0.035)	−0.0999*** (0.035)	−0.1105*** (0.035)	1.3484 (1.449)	1.3622 (1.290)
Widowed	0.0229 (0.060)	0.0273 (0.060)	0.0376 (0.064)	0.0553 (0.065)	−0.0104 (0.046)	−0.0096 (0.046)	−3.6663 (3.288)	−3.7818 (3.334)
Divorced	−0.0192 (0.036)	−0.0051 (0.036)	0.0511 (0.034)	0.0551 (0.034)	0.0354 (0.030)	0.0302 (0.030)	−0.0487 (0.869)	0.0354 (0.855)
Work	−0.0433 (0.028)	−0.0395 (0.028)	0.0104 (0.024)	0.0136 (0.023)	−0.0445* (0.024)	−0.0407* (0.023)	−2.9411 (2.147)	−2.6851 (1.960)
Head	0.1446*** (0.038)	0.1419*** (0.038)	−0.0876** (0.041)	−0.0775* (0.040)	−0.0680* (0.038)	−0.0676* (0.037)	−0.8769 (1.227)	−0.6187 (1.280)
Mother’s educ.	−0.0003 (0.008)	0.0001 (0.008)	−0.0107 (0.007)	−0.0086 (0.007)	−0.0164** (0.007)	−0.0144** (0.007)	−0.6182 (0.389)	−0.5038 (0.374)
Father’s educ	−0.0191*** (0.007)	−0.0226*** (0.007)	0.0149** (0.006)	0.0157*** (0.006)	0.0052 (0.006)	0.0063 (0.006)	1.1278 (1.023)	1.0913 (0.999)
Parents poor	0.0266 (0.024)	0.0231 (0.023)	−0.0645*** (0.020)	−0.0616*** (0.019)	−0.0095 (0.018)	−0.0067 (0.018)	−1.2416 (1.600)	−0.9303 (1.586)
Dummy for 1-digit occupation used in blue coll def (gen)		−0.0350 (0.037)		−0.0125 (0.033)		−0.0196 (0.029)		−1.0047 (1.448)
Constant							33.3892** (15.141)	34.2538** (16.318)
Observations	2,486	2,583	2,530	2,639	2,559	2,659	1,499	1,554

Notes OLS coefficients (for physical activity) and marginal effects of probit models (for obese, binge drinking, and smoking) are reported for baseline models. State fixed effects pertaining to state of initial occupation and risk dummies are included in all models. Blue collar categories used are: cleaning/maintenance; construction; extraction; install/maintenance/repair; production; and transportation. Robust standard errors accounting for clustering by family are reported in parentheses. Asterisks denote statistical significance as follows:

*** p value < 0.01;

** 0.01 < p value ≤ 0.05;

$0.05 < p$ value ≤ 0.1

Table 7

Estimates of the Effect of blue collar first occupation on health behaviors, 2005 longitudinal PSID weights

Variables	Obese		Alcohol		Smoking		Exercise	
	First occupation recalled	First occupation generated	First occupation recalled	First occupation generated	First occupation recalled	First occupation generated	First occupation recalled	First occupation generated
Blue collar	0.0780*** (0.029)	0.0657** (0.029)	0.0153 (0.025)	0.0309 (0.026)	0.0052 (0.022)	0.0278 (0.023)	−0.3510 (1.501)	0.7566 (0.964)
Observations	2,077	2,161	2,119	2,213	2,137	2,228	1,237	1,287

Notes Each cell represents the coefficient of a separate regression of the effect of blue collar work on the health outcome listed. Individual characteristics listed in Table 1 are included in all regressions. OLS coefficients (for physical activity) and marginal effects of probit models (for obese, binge drinking, and smoking) are reported for baseline models. State fixed effects pertaining to state of initial occupation and risk dummies are included in all models. Blue collar categories used are: cleaning/maintenance; construction; extraction; install/maintenance/repair; production; and transportation. Robust standard errors accounting for clustering by family are reported in parentheses. Regressions are weighted using 2005 longitudinal weights accounting for attrition in PSID. Asterisks denote statistical significance as follows:

*** p value < 0.01;

** 0.01 < p value ≤ 0.05;

* 0.05 < p value ≤ 0.1

References

Altonji JG, Elder TE, & Taber C (2005). Selection on observable and unobservable variables: Assessing the effectiveness of Catholic schools. Journal of Political Economy, 113(1), 151–184.

Aluoch MA, & Wao HO (2009). Risk factors for occupational osteoarthritis: A literature review. AAOHN, 57(7), 283–290.

Bang KM, & Kim JH (2001). Prevalence of cigarette smoking by occupation and industry in the United States. American Journal of Industrial Medicine, 40, 233–239. [PubMed: 11598969]

Barsky R, Juster T, Kimball M, & Shapiro M (1997). Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study. The Quarterly Journal of Economics, 112(2), 537–579.

Berger MC, & Leigh JP (1988). The effect of alcohol use on wages. Applied Economics, 20(10), 1343–1351.

Borjas GJ (2004). Labor economics. New York, NY: McGraw-Hill.

Boschetto P, Quintavalle S, Miotto D, Lo Cascio N, Zeni E, & Mapp CE (2006). Chronic obstructive pulmonary disease (COPD) and occupational exposures. Journal of Occupational Medicine and Toxicology, 1, 11. [PubMed: 16756686]

Boskin MJ (1974). A conditional logit model of occupational choice. Journal of Political Economy, 82(2), 389–398.

Bray JW (2005). Alcohol use, human capital, and wages. Journal of Labor Economics, 23(2), 279–312.

Brown C (1980). Equalizing differences in the labor market. Quarterly Journal of Economics, 94(1), 113–134.

Breusch T, & Pagan A (1979). A simple test for heteroscedasticity and random coefficient variation. Econometrica, 47(5), 1287–1294.

Burton NW, & Turrell G (2000). Occupation, hours worked, and leisure-time physical activity. Preventive Medicine, 31(6), 673–681. [PubMed: 11133334]

Case A, & Deaton A (2005). Broken down by work and sex: How our health declines In Wise D (Ed.), Analyses in the economics of aging. Chicago: The University of Chicago Press.

Cesur R (2009). Essays on the aggregate burden of alcohol abuse, Ph.D Dissertation: Georgia State University.

Chaloupka F, & Warner K (2000). Economics of smoking In Newhouse J & Culyer A (Eds.), Handbook of health economics, volume IB. Amsterdam: North-Holland.

Chou S, Grossman M, & Saffer H (2004). An economic analysis of adult obesity: Results from the behavioral risk factor surveillance system. Journal of Health Economics, 23(3), 565–587. [PubMed: 15120471]

Colman G, & Dave D (2011). Exercise, physical activity, and exertion over the business cycle, National Bureau of Economic Research Working Paper 17406.

Cropper ML (1977). Health investment in health, and occupational choice. Journal of Political Economy, 85(6), 1273–1294.

Cutler D, Glaeser E, & Shapiro J (2003). Why have Americans become more obese? Journal of Economic Perspectives, 17(3), 93–118.

Dave D, & Kaestner R (2002). Alcohol taxes and labor market outcomes. Journal of Health Economics, 21(3), 357–371. [PubMed: 12022263]

Dave D, & Saffer H (2008). Alcohol demand and risk preference. Journal of Economic Psychology, 29(6), 810–831. [PubMed: 19956353]

Ettner S (1996). New evidence on the relationship between income and health. Journal of Health Economics, 15(1), 67–85. [PubMed: 10157429]

Ewing R, Schmid T, Killingsworth R, Zlot A, & Raudenbush S (2003). Relationship between urban sprawl and physical activity, obesity, and morbidity. American Journal of Health Promotion, 18(1), 47–57. [PubMed: 13677962]

Fitzgerald J, Gottschalk P, & Moffitt R (1998). The impact of attrition in the panel study of income dynamics on intergenerational analysis. Journal of Human Resources, 33(2), 300–344.

Fletcher JM, & Sindelar JL (2009). Estimating causal effects of early occupational choice on later health: Evidence using the PSID, National Bureau of Economic Research Working Paper 15256.

Fletcher JM, Sindelar JL, & Yamaguchi S (2009). Cumulative effects of job characteristics on health, National Bureau of Economic Research Working Paper 15121.

French MT, & Zarkin GA (1995). Is moderate alcohol use related to wages? Evidence from four worksites. Journal of Health Economics, 14(3), 319–344. [PubMed: 10145138]

Gouskova E, Heeringa SG, McGonagle K, & Schoenim RF (2008). Panel study of income dynamics technical report: Revised longitudinal weights 1993–2005. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan Available at: http://psidonline.isr.umich.edu/data/weights/Long-weights-doc.pdf.

Govender N, Lalloo UG, & Naidoo RN (2011). Occupational exposures and chronic obstructive pulmonary disease: A hospital based case-control study. Thorax, 66(7), 597–601. [PubMed: 21502099]

Grossman M (2000). The human capital model In Culyer AJ & Newhouse JP (Eds.), Handbook of health economics (Vol. 1A, pp. 348–408). Amsterdam: North-Holland.

Gueorguieva R, Sindelar JL, Falba TA, Fletcher JM, Keenan P, Wu R, & Gallo WT (2009). The impact of occupation on self-rated health: Cross-sectional and longitudinal evidence from the health and retirement survey. Journal of Gerontology: Social Sciences. 10.1093/geronb/gbn006.

Hadley J (2003). Sicker and poorer-the consequences of being uninsured: A review of the research on the relationship between health insurance, medical care use, health, work, and income. Medical Care Research Review, 60(2 Suppl), 3S–75S. [PubMed: 12800687]

Hamilton V, & Hamilton B (1997). Alcohol and earnings: Does drinking yield a wage premium? Canadian Journal of Economics, 30(1), 135–151.

Harford TC, & Brooks SD (1992). Cirrhosis mortality and occupation. Journal of Studies on Alcohol, 53(5), 463–468. [PubMed: 1405639]

Hertz-Picciotto I, Arrighi HM, & Hu SW (2000). Does arsenic exposure increase the risk for circulatory disease? American Journal of Epidemiology, 151(2), 174–181. [PubMed: 10645820]

Howard MO, Kivlahan D, & Walker RD (1997). Cloninger’s tridimensional theory of personality and psychopathology: Applications to substance use disorders. Journal of Studies on Alcohol, 58, 48–66. [PubMed: 8979213]

Kelly IR, Dave DM, Sindelar JL, & Gallo WT (2011). The impact of early occupational choice on health behaviors National Bureau of Economic Research Working Paper No. 16803, Cambridge, MA.

Komlos J, Smith P, & Bogin B (2004). Obesity and the rate of time preference: Is there a connection? Journal of Biosocial Science, 36(2), 209–219. [PubMed: 15030016]

Kouvonen A, Vahtera J, Väänänen A, De Vogli R, Heponiemi T, Elovainio M, et al. (2009). Relationship between job strain and smoking cessation: The finnish public sector study. Tobacco Control, 18, 108–114. [PubMed: 19052042]

Lakdawalla D, & Philipson T (2009). The growth of obesity and technological change. Economics and Human Biology, 7(3), 283–293. [PubMed: 19748839]

Leigh JP (1989). Firm size and occupational injury and illness incidence rates in manufacturing industries. Journal of Community Health, 14(1), 44–52. [PubMed: 2785537]

Leigh JP (2011). Economic burden of occupational injury and illness in the United States. Milbank Quarterly, 89(4), 728–772. [PubMed: 22188353]

Leigh JP, Marcin JP, & Miller TR (2004). An estimate of the U.S. government’s undercount of nonfatal occupational injuries. Journal of Occupational and Environmental Medicine, 46(1), 10–18. [PubMed: 14724473]

Levy BS, & Wegman DH (Eds.). (2000). Occupational health: Recognizing and preventing work-related disease and injury (4th ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Lewbel A (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. Journal of Business and Economic Statistics, 30(1), 67–80.

Light A (2005). Job mobility and wage growth: Evidence from the NLSY79. Monthly Labor Review, 128(2), 33–39.

Lubin JH, & Fraumeni JF Jr. (2000). Re: Does arsenic exposure increase the risk for circulatory disease? American Journal of Epidemiology, 152(3), 290–293. [PubMed: 10933278]

Lubin JH, Pottern LM, Stone BJ, & Fraumeni JF Jr. (2000). Respiratory cancer in a cohort of copper smelter workers: Results from more than 50 years of follow-up. American Journal of Epidemiology, 151(6), 554–565. [PubMed: 10733037]

Marmot M (2001). Income inequality, social environment, and inequalities in health. Journal of Policy Analysis and Management, 20(1), 156–159. [PubMed: 11850881]

Marmot M, & Bobak N (2000). Psychosocial and biological mechanisms behind the recent mortality crisis in central and eastern Europe In Cornia GA & Paniccia R (Eds.), The mortality crisis in transitional economies. UNU/WIDER studies in development economics (pp. 127–148). Oxford, NY: Oxford University Press.

Marmot MG, & Smith GD (1997). Socio-economic differentials in health: The contribution of the Whitehall studies. Journal of Health Psychology, 2(3), 283–296. [PubMed: 22013023]

MacDonald Z, & Shields MA (2001). The impact of alcohol consumption on occupational attainment in England. Economica, London School of Economics and Political Science, 68(271), 427–453.

McWilliams JM, Meara E, Zaslavsky AM, & Ayanian JZ (2007). Health of previously uninsured adults after acquiring medicare coverage. Journal of the American Medical Association, 298(24), 2886–2894. [PubMed: 18159058]

Mendolia S 2012 The impact of husband’s job loss on partners’ mental health, Review of economics of the household, online first 1–18. doi:10.1007/s11150-012-9149-6.

Menza M, Golbe L, Cody R, & Forman N (1993). Dopamine-related personality traits. Neurology, 43, 505–598. [PubMed: 8450991]

Michie S, & Williams S (2003). Reducing work-related psychological ill health and sickness absence: A systematic literature review. Occupational and Environmental Medicine, 60, 3–9. [PubMed: 12499449]

Mokdad AH, Ford ES, Bowman BA, Dietz WH, Vinicor F, Bales VS, et al. (2003). Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. Journal of the American Medical Association, 289(1), 76–79. [PubMed: 12503980]

Must A, Spadano J, Coakley EH, Field AE, Colditz G, & Dietz WH (1999). The disease burden associated with overweight and obesity. Journal of the American Medical Association, 282(16), 1523–1529. [PubMed: 10546691]

Newhouse JP (2004). Consumer-directed health plans and the RAND health insurance experiment. Health Affairs, 23(6), 107–113.

Oreopoulos P, & von Wächter T (2006). Short- and long-term career effects of graduating in a recession: Hysteresis and heterogeneity in the market for college graduates. National Bureau of Economic Research Working Paper 12159.

Philipson T (2001). The world-wide growth in obesity: An economic research agenda. Health Economics, 10(1), 1–7. [PubMed: 11180565]

Powell LM, Tauras JA, & Ross H (2005). The importance of peer effects, cigarette prices and tobacco control policies for youth smoking behavior. Journal of Health Economics, 24(5), 950–968. [PubMed: 15990184]

Pratt M, Macera CA, & Blanton C (1999). Levels of physical activity and inactivity in children and adults in the United States: Current evidence and research issues. Medicine and Science in Sports and Exercise, 31(S11), S526–S533. [PubMed: 10593523]

Rashad I, Grossman M, & Chou S (2006). The super size of America: An economic estimation of body mass index and obesity in adults. Eastern Economic Journal, 32(1), 133–148.

Rosen S (1986). The theory of equalizing differences In Ashenfelter O & Layard R (Eds.), Handbook of labor economics (pp. 641–692). Amsterdam: North Holland.

Saffer H, & Dave D (2005a). Mental Illness and the demand for alcohol cocaine and cigarettes. Economic Inquiry, 43(2), 229–246.

Saffer H, & Dave D (2005). The effect of alcohol consumption on the earnings of older workers In Lindgren Bjorn & Grossman Michael (Eds.), Advances in health economics and health services research (Vol. 16, pp. 61–90). [PubMed: 17867235]

Saffer H, Dave D, & Grossman M (2011). Racial, ethnic, and gender differences in physical activity. National Bureau of Economic Research Working Paper 17413.

Schnall PL, Landsbergis PA, & Baker D (1994). Annual Review of Public Health, 15, 381–411.

Sindelar JL, Fletcher J, Falba T, Keenan P, & Gallo WT (2007). Impact of first occupation on health at older ages. National Bureau of Economic Research Working Paper 13715.

Smith J (1999). Healthy bodies and thick wallets: The dual relation between health and SES. Journal of Economic Perspectives, 13(2), 145–166.

Smith PK, Bogin B, & Bishai D (2005). Are time preference and body mass index associated? Evidence from the national longitudinal survey of youth. Economics and Human Biology, 3(2), 259–270. [PubMed: 15964787]

Tekin E (2004). Employment, wages, and alcohol consumption in Russia. Southern Economic Journal, 71(2), 397–417.

Theorell T (2000). Working conditions and health In: Berkman L, & Kawachi L (Eds.). Social epidemiology. New York, NY:Oxford University Press, Inc.

US Department of Health and Human Services. (1996). Physical activity and health: A report of the surgeon general. Atlanta, GA: Centers for Disease Control and Prevention.

Viscusi WK, & Aldy JE (2003). The value of a statistical life: A critical review of market estimates throughout the world. Journal of Risk and Uncertainty, 27(1), 5–76.

Viscusi WK, & O'Connor CJ (1984). Adaptive responses to chemical labeling: Are workers bayesian decision makers? American Economic Review, 74(5), 942–956.

Zabel J (1998). An analysis of attrition in the panel study of income dynamics and the survey of income and program participation with an application to a model of labor market behavior. Journal of Human Resources, 33(2), 479–506.

Zhang L, & Rashad I (2008). Time preference and obesity: The health consequences of discounting the future. Journal of Biosocial Science, 40(1), 97–113. [PubMed: 17517169]

Table 1

Weighted summary statistics

Variable	Description	2005	Bluecoll = 0	Bluecoll = 1
Outcomes				
Obese	Dichotomous variable equal to 1 if BMI is ≥30 kg/m^2	0.241 *** (0.428)	0.215 (0.411)	0.292 (0.455)
Alcohol	Dichotomous variable equal to 1 consumed ≥5 drinks (males) or ≥4 drinks (females)	0.199 *** (0.399)	0.191 (0.393)	0.229 (0.420)
Smoke	Dichotomous variable equal to 1 if respondent smokes	0.167 *** (0.373)	0.140 (0.347)	0.211 (0.408)
Physical activity	Weekly frequency of participation in light or heavy physical activity	9.793 * (15.332)	9.342 (16.933)	10.249 (11.382)
First occupation variables				
Blue collar	First occupation blue collar (recall)	0.328 *** (0.469)	0.000 (0.000)	1.000 (0.000)
Blue collar	First occupation blue collar (generated)	0.297 *** (0.457)	0.131 (0.337)	0.601 (0.490)
Occ_years	Years since first generated occupation (2005)	18.661 *** (13.948)	16.432 (14.309)	23.374 (11.864)
Individual characteristics				
Male	Male respondent	0.459 *** (0.498)	0.299 (0.458)	0.779 (0.415)
White	Dichotomous variable equal to 1 if respondent is non-Hispanic White	0.906 (0.292)	0.908 (0.290)	0.914 (0.280)
Black	Dichotomous variable equal to 1 if respondent is non-Hispanic black	0.079 (0.270)	0.077 (0.266)	0.075 (0.263)
Hispanic	Dichotomous variable equal to 1 if respondent is Hispanic	0.010 (0.101)	0.010 (0.097)	0.009 (0.092)
Other	Dichotomous variable equal to 1 if respondent's race is other than above	0.005 ** (0.069)	0.006 (0.079)	0.002 (0.045)
Elementary	Dichotomous variable equal to 1 if elementary school educ	0.034 *** (0.182)	0.018 (0.133)	0.050 (0.218)
Some high	Dichotomous variable equal to 1 if some high school educ	0.103 *** (0.304)	0.064 (0.244)	0.161 (0.367)
High	Dichotomous variable equal to 1 if high school education	0.366 *** (0.482)	0.334 (0.472)	0.419 (0.493)
Some college	Dichotomous variable equal to 1 if some college education	0.223 (0.417)	0.229 (0.420)	0.228 (0.420)
College	Dichotomous variable equal to 1 if college education	0.273 *** (0.446)	0.355 (0.479)	0.142 (0.349)
Age	Age of respondent (in years)	54.355 *** (13.731)	53.738 (13.378)	54.998 (13.733)
Single	Dichotomous variable equal to 1 if respondent is single	0.047 *** (0.213)	0.055 (0.227)	0.032 (0.175)
Married	Dichotomous variable equal to 1 if respondent is married	0.751 * (0.433)	0.747 (0.435)	0.770 (0.421)
Widowed	Dichotomous variable equal to 1 if respondent is widowed	0.077 * (0.267)	0.079 (0.270)	0.064 (0.244)
Divorced	Dichotomous variable equal to 1 if respondent is divorced or separated	0.125 (0.331)	0.119 (0.324)	0.134 (0.341)
Employed	Dichotomous variable equal to 1 if respondent is employed	0.670 (0.470)	0.680 (0.467)	0.682 (0.466)

Variable	Description	2005	Bluecoll = 0	Bluecoll = 1
Head	Dichotomous variable equal to 1 if respondent is head of household	0.632*** (0.482)	0.507 (0.500)	0.871 (0.336)
Household income	Household income (in thousands of dollars)	89,523.040*** (144,597.000)	97,188.850 (126,737.800)	79,894.890 (184,952.500)
Mother's educ.	Mother's education (category) (1 = grades 0–5, 8 = college)	4.018*** (1.637)	4.251 (1.638)	3.659 (1.540)
Father's educ	Father's education (category) (1 = grades 0–5, 8 = college)	3.994*** (1.952)	4.299 (1.982)	3.488 (1.768)
Parents poor	Dichotomous variable equal to 1 if respondent answered that growing up, parents were poor	0.261*** (0.439)	0.227 (0.419)	0.312 (0.463)
Instrumental variables				
County unemp. (1968)	County unemployment rate in 1968	4.053*** (2.335)	3.953 (2.118)	4.216 (2.635)
Father blue coll	Dichotomous indicator for whether father's main occupation was blue-collar	0.463*** (0.499)	0.420 (0.494)	0.572 (0.495)

Notes Standard deviations are reported in parentheses. Sample size is 6,439. Asterisks denote that the difference in means by "blue collar" (based on recall) is statistically significant at the following levels:

*** p value ≤ 0.01;

** 0.01 < p value ≤ 0.05;

* 0.05 < p value ≤ 0.1

Table 2

Estimates of the effect of blue collar first occupation on health behaviors, 2005

	Baseline probit and OLS		Internal instruments		Internal + external instruments	
	First occupation recalled	**First occupation generated**	**First occupation recalled**	**First occupation generated**	**First occupation recalled**	**First occupation generated**
Panel 1: obese						
Blue collar	0.0761*** (0.024)	0.0427* (0.024)	0.0438 (0.060)	0.0288 (0.052)	0.0410 (0.065)	0.0663* (0.040)
Observations	2,486	2,583	2,519	2,615	2,380	2,466
Overidentification *p* value	-	-	0.72056	0.34193	0.86392	0.41891
F statistic	-	-	5.53	11.83	18.00	70.75
Panel 2: alcohol						
Blue collar	0.0208 (0.020)	0.0170 (0.020)	0.0578 (0.057)	0.0698 (0.045)	0.0784 (0.051)	0.0715 (0.059)
Observations	2,530	2,639	2,557	2,654	2,420	2,503
Overidentification *p* value	-	-	0.35128	0.78835	0.26933	0.47020
F statistic	-	-	4.97	11.63	13.65	74.28
PANEL 3: smoking						
Blue collar	0.0004 (0.018)	0.0290 (0.019)	0.0619 (0.054)	0.0715 (0.044)	0.0840 (0.061)	0.0642* (0.036)
Observations	2,559	2,659	2,585	2,682	2,444	2,531
Overidentification *p* value	-	-	0.31841	0.96825	0.67064	0.54840
F statistic	-	-	5.18	11.98	14.03	78.87
Panel 4: physical activity						
Blue collar	0.6846 (1.604)	−0.1654 (0.965)	5.7945** (2.506)	4.4770* (2.309)	4.9124*** (1.678)	3.7896** (1.784)
Observations	1,499	1,554	1,499	1,554	1,436	1,480
Overidentification *p* value	-	-	0.89848	0.44673	0.30466	0.33622
F statistic	-	-	8.96	12.31	42.77	98.61

Notes Each cell represents the coefficient of a separate regression of the effect of blue collar work on the health outcome listed. OLS coefficients (for physical activity) and marginal effects of probit models (for obese, binge drinking, and smoking) are reported for baseline models. Individual characteristics listed in Table 1 are included in all regressions. State fixed effects pertaining to state of initial occupation and risk dummies are included in all models. Blue collar categories used are: cleaning/maintenance; construction; extraction; install/maintenance/repair; production; and transportation. Robust standard errors accounting for clustering by family are reported in parentheses. Asterisks denote statistical significance as follows:

*** p value < 0.01;

** 0.01 < p value ≤ 0.05;

*$0.05 < p$ value ≤ 0.1.

Excluded instruments pertain to county unemployment (1968) and father's blue collar occupation. Standard errors in IV models are adjusted for arbitrary correlation within state of residence at initial labor market entry

Table 3

Constrained bivariate probit estimates of blue collar recalled first occupation on obesity, alcohol and smoking in 2005

Model	Constraint	Obesity	Alcohol	Smoking
1	Probit (ρ = 0)	0.0644 ***	−0.0001	0.0061
		(3.46)	(−0.01)	(0.52)
2	ρ = 0.1	0.0091	−0.0401 ***	−0.0331 ***
		(0.50)	(−3.52)	(−2.99)
3	ρ = 0.2	−0.0451 **	−0.0783 ***	−0.0704 ***
		(−2.53)	(−7.10)	(−6.57)
4	ρ = 0.3	−0.0979 ***	−0.1154 ***	−0.1066 ***
		(−5.66)	(−10.81)	(−10.28)
5	ρ = 0.4	−0.1497 ***	−0.1520 ***	−0.1424 ***
		(−8.98)	(−14.74)	(−14.21)
6	ρ = 0.5	−0.2005 ***	−0.1889 ***	−0.1781 ***
		(−12.58)	(−18.97)	(−18.44)
7	ρ = −0.1	0.1211 ***	0.0422 ***	0.0473 ***
		(6.42)	(3.49)	(4.01)
8	ρ = −0.2	0.1789 ***	0.0874 ***	0.0917 ***
		(9.44)	(7.04)	(7.55)
9	ρ = −0.3	0.2377 ***	0.1359 ***	0.1394 ***
		(12.56)	(10.69)	(11.19)
10	ρ = −0.4	0.2971 ***	0.1881 ***	0.1909 ***
		(15.85)	(14.53)	(15.01)
11	ρ = −0.5	0.3568 ***	0.2443 ***	0.2458 ***
		(19.41)	(18.63)	(19.09)
12	Sel on obs =	0.0109	0.0574 ***	0.1983 ***
	Sel on unobs	(0.70)	(5.33)	(17.71)
		[ρ = 0.0740]	[= −0.0350]	[ρ = −0.2697]
Observations		4,000	5,712	5,794

Notes Marginal effects of initial blue-collar occupation are reported. Physical activity (a continuous outcome) is excluded since only dichotomous outcomes are employed in constrained-selection bivariate probit models. Robust t statistics are reported in parentheses. Asterisks denote statistical significance as follows:

*** p value < 0.01;

** 0.01 < p value ≤ 0.05;

* 0.05 < p value ≤ 0.1

Table 4

Stratified samples effects of initial blue-collar occupation controlling for current occupation, 2005

Outcome	Obese	Alcohol	Smoking	Physical	Activity
1	Full sample	0.080***	0.012	0.008	1.79
		(0.028)	(0.024)	(0.021)	(1.362)
2	Male	0.064*	0.031	0.023	0.644
		(0.035)	(0.036)	(0.027)	(0.835)
3	Females	0.074	0.003	−0.035	4.381
		(0.051)	(0.034)	(0.032)	(4.09)
4	White	0.124***	0.017	0.002	0.346
		(0.033)	(0.032)	(0.025)	(0.799)
5	Non-white	−0.017	0.014	0.062	7.079
		(0.06)	(0.034)	(0.045)	(4.501)

Notes Each cell represents a separate regression model and shows coefficients on initial blue collar (based on self-report). Robust standard errors are reported in parentheses. All models include the controls used in Table 2. Asterisks denote statistical significance as follows:

*** p value ≤ 0.01;

** 0.01 < p value ≤ 0.05;

* 0.05 < p value ≤ 0.1. Sample sizes range from 2,043 to 6,599

Table 5

Effects of initial blue-collar occupation potential mediators 2005

Model	Outcome	Obese	Alcohol	Smoking	Physical activity
		First occupation recalled	**First occupation recalled**	**First occupation recalled**	**First occupation recalled**
1	Baseline	0.057*** (0.015)	0.004 (0.012)	0.004 (0.012)	−0.324 (1.069)
2	Baseline with household income	0.057*** (0.015)	0.004 (0.012)	0.003 (0.012)	−0.322 (1.071)
3	Baseline with work hours	0.057*** (0.015)	0.004 (0.012)	0.004 (0.012)	−0.324 (1.069)
4	Baseline with current occupation	0.066*** (0.02)	0.005 (0.017)	0.006 (0.016)	0.93 (0.84)

Notes Each cell represents a separate regression model. Coefficients (and marginal effects for dichotomous outcomes) of initial blue-collar occupation are presented. Robust standard errors are reported in parentheses. All models include the controls used in Table 2, with the exception of parental history and risk tolerance. Asterisks denote statistical significance as follows:

*** p value ≤ 0.01;

** 0.01 < p value ≤ 0.05;

* 0.05 < p value ≤ 0.1. Sample sizes range from 2,043 to 6,599