

**NBER WORKING PAPER SERIES**

**EDUCATION AND HEALTH:  
WHERE THERE'S SMOKE THERE'S  
AN INSTRUMENT**

**William N. Evans  
Edward Montgomery**

**Working Paper No. 4949**

**NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
December 1994**

**We wish to thank Victor Fuchs, Michael Grossman, Daniel Hamermesh, William Carrington, Michael Cook, Matthew Farrelly, and Sheila Murray for a number of helpful comments and suggestions. All errors and opinions are the authors'. This paper is part of NBER's research program in Health Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.**

**© 1994 by William N. Evans and Edward Montgomery. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.**

EDUCATION AND HEALTH:  
WHERE THERE'S SMOKE THERE'S  
AN INSTRUMENT

ABSTRACT

Victor Fuchs has suggested that the persistent positive correlation between education and health habits can be explained by interpersonal differences in the discount rate. If Fuchs is correct, some health habits can be used as instruments for education in standard wage equations. We use whether an individual smoked at age 18 in such a fashion. The instrument is strongly correlated with years of education, and IV estimates of the return to schooling are 10 percent larger than the OLS estimates. We fail to reject tests of overidentifying restrictions, show how the smoking/education link varies systematically across age cohorts and income groups, and demonstrate that the instrument is correlated with other intertemporal decisions such as home ownership. The results are replicated in four additional data sets, and for both males and females.

William N. Evans  
Department of Economics  
University of Maryland  
College Park, MD 20742  
and NBER

Edward Montgomery  
Department of Economics  
University of Maryland  
College Park, MD 20742  
and NBER

## **I. Introduction**

Public health officials have engaged in numerous campaigns to change the health habits of the American public. Efforts have been made to provide information on a broad range of health or safety practices covering such activities as diet, exercise, seat belt usage, drunk driving, drug and alcohol abuse. The 1964 Surgeon General's report on the hazards of smoking and the requirement that cigarette packs contain warning labels marked the initiation of perhaps the best known and widespread of these campaigns. Work by Hamilton (1972), Lewit, Coate, and Grossman (1981), Schneider, Klein and Murphy (1981) and others suggests that the widespread dissemination of information on health risks did alter public behavior as seen by the fact that the fraction of the population that smokes has declined from a high of 42.6 percent in 1966 to 26.7 percent in 1992.<sup>1</sup>

Despite these public health campaigns, a significant percentage of the population still engages in risky or unhealthy behavior, accounting for as much as 50 percent of the mortality in a given year.<sup>2</sup> Further, the decision to engage in healthy practices is not randomly distributed in the population. As noted by Farrell and Fuchs (1982), one of the most persistent relationships in health economics is that more educated people have better health and better health habits.

In this paper we make novel use of the correlation between the decision to invest in additional schooling and the decision to engage in an unhealthy habit (smoking), to derive estimates of the return to schooling. In economic models of human capital accumulation, individuals invest in education until the rate of return on the last year of education equals the individual's discount rate. If, as Fuchs (1982) conjectures, the correlation between

health and education arises because of unobserved differences in time preference or discount rates across individuals, then some health habits can be used as an instrument for education in a human capital earnings equation. This instrumental variable (IV) estimate of the rate of return to education will be purged of the ability bias that been the focus of much recent work.<sup>3</sup> We find that variations in whether an individual smoked at age 18 provides an exogenous source of variation in educational outcomes and the resulting IV estimates of the return to school are both precisely estimated and higher than the ordinary least-square (OLS) estimates. This relationship was found to be particularly strong for workers who turned 18 after evidence of smoking's effect on health became widespread in the late 1950's and early 1960's.

The paper is organized as follows. In section II we review previous work on the correlation between education and health habits. In section III we develop the conceptual framework for using health habits as an instrument in an earnings equation to remove ability bias. In section IV we describe the data sets used in our analysis. In section V we present estimates of the return to schooling from a variety of data sets using OLS and instrumental variable techniques. We also provide information to support our conjecture that differences in smoking behavior reflect differences in time preference by looking at the relationship between smoking and other future orientated behavior such as home ownership, saving, and seat belt use.

## **II. Previous Research on Health and Education**

The link between education and health is well documented. More educated people have lower mortality rates (Kitagawa and Hauser, 1973; Grossman, 1976; Rosen and

Taubman, 1982), have lower incidence of most chronic diseases (Pincus et al., 1987), better physiological measures of health (Newhouse and Friedlander, 1979), and better self-reported health status (Grossman, 1976). Fuchs (1982) notes that the better health outcomes of more educated people is driven primarily by the fact that more education workers are also more likely to have better health habits.

To provide some indication of the persistent correlation between education and health habits, we ran a series of reduced form health habit probit models where the dependent variables are indicators for 12 health habits and practices contained in the 1985 Health Promotion and Disease Prevention supplement (HPDP) to the National Health Interview Survey (NHIS). The NHIS is an annual, nationally representative multistage probability sample of the civilian, noninstitutionalized population 18 years of age and older. The NHIS is designed to provide national estimates of the distribution of illnesses, and the kinds of health services people receive. Each year, the NHIS contains a set of core questions, plus questions from a special supplement. In 1985, the special supplement was the HPDP survey which was designed to generate information on health and fitness awareness and health habits. The survey contained data for 24,685 adults. In each model, the independent variables are years of education, log family income and a dummy variable that equaled 1 if income is missing, age and age squared, family size, and indicators for race, region of the country, urban/rural, and when appropriate, gender.<sup>4</sup> The normalized coefficients for education and log income are reported in Table 1. The coefficients on the education and log income variables represent the change in the probability of engaging in a health practice resulting from a one year increase in education and a 10 percent increase in

income, respectively. The results indicate that more educated respondents are more likely to use their seat belt, exercise regularly, have their blood pressure checked within the previous year, eat breakfast every morning, and own a working smoke detector. More educated people are also less likely to have ever smoked, to be heavy drinkers, or to have driven drunk. Females with more education are more likely to have had breast exams and pap smears in the previous year, while more educated parents of small children are more likely to have Ipecac syrup in the home. All 12 coefficients on the education variable are precisely estimated with the smallest absolute value of a t-statistic being 6. The magnitude of these marginal effects is also quite large. For example, college-educated people have about a 14 percentage point higher probability of using a seat belt than a similarly defined high school graduate.

Although the evidence establishing a link between health and education is overwhelming, the interpretation of this correlation is, however, still a subject of debate. A number of authors have interpreted the education/health link as a causal relationship. Grossman (1976) and Michael (1973) developed models of investment in health care in which schooling increases the efficiency of time and goods in the production of health. Consequently, more educated individuals will be observed having better health and practicing "better" health habits.<sup>5</sup>

Fuchs (1982) offers another potential explanation for the strong correlation between years of education and health. He argues unobserved differences in the rate of time preference determines both the level of schooling a person acquires and their investments in health. Investments in health capital, such as exercising or not smoking, are similar to

investments in education in that both decisions involve a tradeoff between current costs and the discounted value of future benefits. In many models of human capital accumulation, and Grossman's (1976) model of health care investment, individuals invest until the rate of return from additional schooling or health inputs equals their discount rate. Individuals with higher discount rates would, *ceteris paribus*, end formal education at a younger age, and be less likely to invest in health through good health habits.<sup>6</sup> The notion that time preference plays an important role in explaining the consumption of potentially addictive goods such as cigarettes, alcohol and jogging is also highlighted in the "rational addiction" model of Becker and Murphy (1988).

In a test of this hypothesis, Farrell and Fuchs (1982) found that education at age 24 is as an important determinant of smoking at age 17 -- when all individuals have essentially the same level of education -- as it is at age 24, when education levels vary considerably. In a subsequent test of the Fuchs hypothesis, Kenkel (1991) found that after controlling for an individual's knowledge of the health consequences of certain health habits, education is still an important determinant of the smoking, exercise, and heavy drinking decision.

It should be pointed out that these two explanations for the correlation between schooling and health are not mutually exclusive. Schooling can be a source of knowledge even if individuals' propensities to use that knowledge vary because of unobserved differences in discount rates. It is an empirical question which is the more important source of variation in the schooling/health relationship. One might expect, however, that the causal link between schooling and health habits will be more important when knowledge about the efficiency or risks associated with health habits and practices comes primarily from formal

schooling rather than from other sources such as parents, relatives, doctors, or the government.

To the degree public information has succeeded in making the risks of particular health habits known to most citizens, the correlation between schooling and that health habit is more likely to be due to variations in unobserved factors such as discount rates. This is particularly relevant for the case of smoking because the public's awareness of the risks of smoking has changed considerably over time. In a 1954 Gallop Poll opinion survey, only 41 percent of adults identified cigarettes as a cause of lung cancer. By 1981, this number had risen to 83 percent. In 1949, only 66 percent of nonsmokers and 52 percent of smokers identified cigarette smoking as harmful. By 1981, this number had increased to 90 percent for all adults.<sup>7</sup> Because of the vast information available about the risks of smoking, we suspect that compared to other health habits the correlation between smoking and education is more likely to reflect unobserved factors such as discount rates rather than knowledge effects. Kenkel (1991), for example, notes that the average respondent in the 1985 NHIS/HPDP supplement was three to five times more likely to know the adverse health effects of smoking than they were to know the health effects of excessive drinking or lack of exercise.

### **III. Smoked at age 18 as an Instrument For Education**

One of the most important and widely examined relationships in labor economics is the link between schooling and earnings. As many authors have noted, however, OLS estimates of the marginal return to education are potentially biased because individuals with



greater unobserved ability will not only acquire more schooling, but may also have greater earnings.

A potential solution to this problem is to identify an instrument that is correlated with schooling but orthogonal to the unobserved component of an earnings equation.<sup>8</sup> Since educational attainment is an outcome from a choice process where individuals maximize wealth or utility subject to constraints, the natural set of instruments are measures of taste for schooling, the discount rate, or elements of the budget constraint.<sup>9</sup> A recent string of papers have used measures of the costs of education such as distance to school (Card, 1993) or state tuition rates (Kane and Rouse, 1993), while others have used constraints such as the interaction between compulsory schooling and the quarter of birth (Angrist and Krueger, 1991a), draft status (Angrist and Krueger, 1991b), or family gender mix (Butcher and Case, 1994) as instruments for education in earnings equations.<sup>10</sup> Still others have used siblings or twin data (Ashenfelter and Zimmerman, 1993; Ashenfelter and Krueger, 1994) and assumptions about the structure of the unobserved component of earnings to remove the ability bias.

Card (1994) notes that the common thread through all the studies listed above is that the IV estimates of the returns to education are higher than the OLS estimates. Despite the consistency of findings in these studies, they are not without their detractors. Griliches (1977) and Neumark (1994) outlined the problems inherent in using twins or siblings data. Bound et al. (1993) and Staiger and Stock (1994) noted that some of the other studies have been plagued by a lack of precision in their first stage estimates, potentially making the IV estimates as biased as the OLS estimates.

If the cross-sectional correlation between health habits such as smoking and education is primarily due to a common unobserved factor (the discount rate), then some health habits could potentially be used as instruments for education in a human capital earnings equation. The use of health habits, which previous work has shown to be strongly correlated with education, offers a potential solution to the problems identified with twin or sibling studies as well as the precision issues found in the "natural experiment" studies reviewed by Card (1994).

It should be noted, however, that even if the Fuchs hypothesis were the only reason for the correlation between health habits and schooling, not all health habits could be used as instruments for education in a standard earnings equation. The choice of which health habit to use as an instrument is complicated by two factors. First, many health habits, such as heavy drinking, may directly effect a worker's current productivity and hence his/her wage. Second, some health habits have consumption and not just investment value. Consequently, they might be correlated with family income (and potentially the unobserved component of labor earnings) because they are normal or inferior goods.<sup>11</sup>

We use whether a person smoked at age 18 as the instrument for schooling in this paper because it offers a potential solution to these aforementioned issues. First, whether one smoked at age 18 is strongly correlated with years of education. Much of the variation in years of education, especially among younger workers, is generated by the education decisions made during the late teen years. Similarly, most people who have ever smoked began smoking during their late teens. Surveys from the later 1980s note that about 70 percent of all people who ever smoked began smoking by the age 18.<sup>12</sup> As we illustrate

below, the concurrence in the timing of these two decisions generates a statistically precise and quantitatively large correlation between years of education and smoking at age 18.

Secondly, whether one smoked at age 18 should not independently affect one's current weekly wage. Although there is some evidence that current smokers lose more work days than nonsmokers, there is no systematic evidence linking even current smoking to lower productivity or earnings.<sup>13</sup> The link between smoking at age 18 and current productivity would be even more tenuous than the link between current smoking and productivity because many people who smoked at age 18 have long since quit smoking while others began smoking at later ages. In our data set, 43.9 percent of males who smoked at age 18 were not smoking at the time of the survey, while 24.5 percent of current smokers did not smoke at age 18. This fluid movement of people into and out of the smoking ranks suggests that even with a link between current smoking and earnings, past smoking is unlikely to be correlated with current earnings via a productivity effect. Finally, by using a past rather than a concurrent measure of a health habit as an instrument for current levels of schooling, any bias caused by the pure income effects on consumption is mitigated.<sup>14</sup>

#### **IV. Data**

Estimation of our model requires contemporaneous data on earnings as well as retrospective data on smoking habits. Many data sets such as the Current Population Survey (CPS) have the necessary wage data, but few of these surveys have any information about health habits. Likewise, the primary data used to estimate national smoking trends, the annual National Health Interview Survey (NHIS), does not have

individual wage data.<sup>15</sup> To our knowledge, the largest and most complete data set with the necessary information for our purposes is the 1987 National Medical Expenditure Survey (NMES). NMES was a nationally representative sample of approximately 14,000 households and 38,000 individuals.<sup>16</sup> NMES was designed to provide national estimates of family and individual medical care use and financing. Families were surveyed quarterly and asked detailed questions about their medical care utilization and insurance status. In the first quarter, detailed information was collected about the last job held. In some survey rounds, respondents were asked about health habits like smoking, exercise, and diet. Blacks and hispanics were oversampled in this survey.

The labor earnings questions in the NMES differ from the standard earnings question in the CPS in that respondents are asked how much they earned from their primary job. Respondents provide an hourly or daily wage rate, or reported weekly, bi-weekly, monthly, or yearly earnings. Using responses to questions about usual hours worked per week and usual days per week, we translate the earnings responses into weekly earnings. Earnings are not top-coded in the NMES data. From the detailed questions about jobs and individuals, we construct the other variables used in the analysis. Age, union status, region of the country, and type of city were taken from the first quarterly survey.

In the health habits section of the NMES, respondents were asked whether they ever smoked 100 cigarettes. If the respondent answered yes, information about whether they currently smoked, at what age they started or quit smoking, and how many cigarettes they smoked on average each day, were also collected. From the smoking initiation/quit questions, we can determine whether a person smoked at age 18. Questions about average

daily consumption were asked of both current and former smokers.

To make our analysis comparable to previous studies, we restrict our sample to males aged 21 to 64 who worked at least 30 hours per week, earned at least \$2/hour, and who were not self employed.<sup>17</sup> Using this sample selection criteria, and deleting those individuals with missing smoking data, the final sample contains data for 3302 people. Since the NMES was not designed to provide a nationally representative sample of workers or smokers, it is useful to compare the responses to the work questions with those from a nationally representative data set such as the entire outgoing rotation (OR) group of the 1987 CPS.<sup>18</sup> Using the same selection criteria for the 1987 CPS/OR as for the NMES we end up with a sample of 80,300 males. All respondents in the out-going rotation group of the CPS report their usual weekly earnings but weekly earnings are top-coded at \$999. Thus to make 1987 NMES comparable to the 1987 CPS/OR, we artificially top-coded the NMES numbers to \$999/week.<sup>19</sup>

Similarly, to gauge the reliability of the smoking numbers from the NMES, we utilized data from the 1987 NHIS Cancer Control Supplement (NHIS/CCS). The 1987 NHIS/CCS was designed to collect data on cancer knowledge and attitudes and cancer screening knowledge and practices. As part of the supplement, respondents were asked about past and current smoking habits. The 1987 NHIS/CCS contained data for 22,043 adults. In most years, blacks are over sampled in the NHIS, and for the 1987 survey, hispanics are over sampled as well. Because the 1987 NHIS/CCS contained limited information about jobs, the sample selection criteria for this data set is somewhat different from the previous two. In our sample from the 1987 NHIS/CCS, we included males age 21

to 64 who were working full time and not self-employed. With this selection criteria, we have data for 5092 respondents.

As seen in Table 2, the correspondence between the 1987 NMES and the 1987 CPS/OR data sets is quite close.<sup>20</sup> Weekly earnings in the CPS/OR data are only 2.7% higher than in the NMES data. This discrepancy is due in part to the fact that NMES over-sampled black and hispanic workers, and because the CPS/OR labor earnings questions included salaries from all jobs not just the primary job. The two wage distributions are nearly identical through about the third quartile. The 10th, 25th, 50th, 75th, and 90th percentile for weekly earnings from the 1987 NMES (1987 CPS/OR) are \$209 (\$210); \$300 (\$300); \$436 (\$441); \$602 (\$615); and \$825 (\$875) respectively. The two data sets are similar in mean years of education, age, hours per week, and union status. The age, education, and smoking data in the NMES are also comparable to those in the 1987 NHIS/CSS. The percentages of people who ever smoked or who smoked at the time of the survey are slightly higher in the 1987 NHIS/CSS whereas the percentage smoking at age 18 is identical in the two data sets. Both data sets indicate that the average and median smoker smokes about a pack a day.<sup>21</sup>

## **V. Results**

In the first two columns of Table 3, we report ordinary least-square (OLS) estimates of standard human capital earnings functions using the 1987 NMES and the 1987 CPS/OR data sets. The dependent variable in both models is the natural log of weekly earnings.<sup>22</sup> The explanatory variables used in the model are years of education, quadratic terms in age,

a union status dummy, indicators for black and hispanic workers, plus 3 dummies for regions of the country and indicators for large and medium sized SMSAs. The model explains roughly the same percent of the variation in log weekly earnings in both data sets. There are some differences between the results for the black, hispanic, and union status variables but, in general, the coefficients estimates from the NMES sample are comparable to those found in other studies. The rate of return to education for the NMES data (6.6%) is only slightly lower than the value from the CPS data (6.9%). The growth of earnings over time is nearly identical in both data sets. The NMES results predicted that earnings for a 31 year old are 35.6 percent higher than earnings for a 21 year old whereas the 1987 CPS/OR sample predicted a 36.1 percent difference.

In the third column of Table 3, we report two-stage least-square (2SLS) results using smoked at age 18 as an instrument for years of education. In the lower half of Table 3, we report the OLS coefficients from the first-stage regression of the years of education on the instruments and all other exogenous variables. People who smoked at age 18 have 1.25 fewer years of education than those who did not. It should be noted that in contrast to first stage results in some other studies, this coefficient is precisely estimated: the absolute value of the t-statistic is about 13.

The 2SLS estimate of the return to schooling is higher than the OLS value. In the studies surveyed by Card (1994), recent IV estimates of the education parameter are 10 to 30 percent larger than the OLS estimate. Our 2SLS estimate of the rate of return to education is .7 percentage points or about 10.6 percent larger than the OLS estimate.

One shortcoming of our instrument is that using whether a worker smoked at age 18

just identifies the 2SLS model. Therefore, we are unable to calculate tests of overidentifying restrictions (Hansen, 1982; Newey, 1985) for this model to see whether the instruments can be excluded from the structural earnings equation. However, by using the detailed smoking data in the NMES survey, we can construct other instruments that produced an overidentified system.

A number of researchers have demonstrated that among those who smoke, the heaviest smokers have the highest risk of heart disease and lung cancer.<sup>23</sup> If this information is well known among smokers, and if variations in discount rates affect smoking on the intensive as well as extensive margins, those who smoked more at age 18 would have higher discount rates.<sup>24</sup> The NMES survey collected average daily consumption of cigarettes for all people who ever smoked. We use this information to determine how much people smoked among those who smoked at age 18. We then construct measures of whether a person was a light ( $\leq 15$  cigarettes per day), moderate ( $<15$  but  $\leq 35$  cigarettes per day), or heavy smoker ( $>35$  cigarettes per day) among age 18 smokers for use as instruments. In the segmented model, the fact that there are 3 instruments and only one endogenous variable allows us to construct a test of overidentifying restrictions. The test is distributed as a chi-squared statistic with degrees of freedom equal to the difference between the number of instruments and the number of endogenous right-hand-side variables in the structural equation.

The 2SLS results with these alternate instruments are reported in the final column of Table 3. The coefficients on the three indicators of smoking intensity increase in absolute value as cigarettes consumed per day at age 18 increase. Light smokers have one fewer



year of education and heavy smokers have almost 1.9 fewer years than non-smokers. As we found before, the 2SLS estimates of the rate of return to education is larger than the OLS value. Using the three-part instrument, the estimated return to education is 7.1 percent, which is 7.6 percent larger than the OLS value. More importantly, the test of overidentifying restrictions is well below the critical value of 5.99, indicating that our model is internally consistent.

Bound et al. (1993) and Staiger and Stock (1994) have noted that in some recent papers that attempt to identify the return to schooling through natural experiments, the instruments are weakly correlated with years of schooling. The authors demonstrate that in this situation, even a weak correlation between the instrument and the unobserved component of the log weekly earnings will produce a bias in the IV estimates that is larger than the bias in OLS estimates. This is not a concern in our case. A regression of education solely on smoked at age 18 on the light, moderate, and heavy indicators produces  $R^2$ 's of 0.041 and 0.043 respectively. In the first-stage regression with the light, moderate, and heavy smoking instruments, the F-test statistic that these instruments are jointly zero is 60.8 with 3 and 3288 degrees of freedom. This value is about 72 standard deviations above the 95 percent critical value of 2.6.

The results reported in Table 3 are robust to minor variations in the model. First, we checked to see if our results were sensitive to the choice of age 18 as the year to measure lagged smoking behavior. The magnitude and general tenor of the results and unchanged when we used smoked at ages 16, 17, or 19 as instruments for smoking. At these ages, smoking rates are 25.4, 31.7, and 43.8 percent respectively, and the 2SLS

estimates for the years of education coefficient (standard error) are 0.070 (0.012), 0.064 (0.012), and 0.079 (0.013), respectively. Secondly, we estimated models where we replaced age with potential experience (calculated as age minus education minus 6). In these models, the results are similar to those reported in Table 3 in that the 2SLS estimates of the rate of return to education are 10 to 20 percent larger than the OLS estimates, depending on model specification. In the overidentified models, the specification test statistic is always well below the 95 percent critical value.<sup>25</sup>

### **Analysis Using Other Data Sets**

We repeated our primary analysis with several other data sets to insure that the results are not a function of having used the NMES data. Unfortunately, there are very few data sets that have the necessary variables to perform the tests outlined above. We have identified four data sets that have similar information to that provided in the NMES: special supplements to the September 1985 and September 1989 CPS, the 1986 wave of the Panel Study of Income Dynamics (PSID), and the 1987 NHIS/CSS utilized above. In this section, we replicate the baseline results from Table 3 with these three data sets.

Questions concerning smoking and immunization were asked in a special supplement to the September 1985 CPS. Questions about veterans' status and cardiovascular risk factors, including questions about cigarette consumption, were asked in the September 1989 CPS. In the 1985 CPS, adults age 16 and older were asked whether they ever smoked, when they started smoking, and whether they smoked at the time of the survey. These variables allow us to construct a variable similar to our previous instrument -- whether a

person began smoking by age 18. In this survey, we have no indication when former smokers quit and therefore, we cannot identify people who began and quit smoking all before the age of 18. As a result, the percent of people that began smoking by age 18 will be slightly larger than those who actually smoked at age 18. In the 1989 CPS, we can construct whether an individual smoked at age 18. These surveys have some advantages over the NMES in that the wage data from the out-going rotation group produces sample sizes about twice the size as the NMES data. However, no questions about the quantity of cigarettes smoked were asked of former smokers and therefore, we are unable to construct our primary test of overidentification. In the monthly CPS surveys, usual weekly earnings and hours were asked of the one-quarter of the people who were part of the out-going rotation sample. From this group, we produced a sample of working males between the ages of 21 and 64, who were not self-employed, who worked at least 30 hours per week, and who earned at least \$60 per week. These restrictions yielded a samples of 6313 and 6645 observations from the 1985 and 1989 CPS samples, respectively. Since weekly earnings in the September 1985 and September 1989 CPS are top-coded at \$999 and \$1923, respectively, we used the procedure discussed above to generate correction factors for the censored weekly earnings. For the 1985 and 1989 data, we multiplied the censored values by 1.25 and 1.23 respectively.

The smoking questions in the 1986 wave of the PSID also allow us to construct our primary instruments: whether a person smoked at age 18 and whether these smokers were light, moderate, or heavy smokers. Unlike the other data sets we used, the PSID reports hourly wages. The hourly wage data reported in the 1986 PSID is based on labor earnings

and hours for 1985. We include in our extract from the PSID 21 to 64 year old male heads of households whose families were part of the original random sample of households. We deleted self-employed workers, those who worked less than 1500 hours a year, and those with wages below \$2/hour. Because of small numbers of observations, we also deleted hispanic workers from our extract. Although constructed hourly wages in the PSID were top-coded at \$99.99/hour, no observations in our extract were censored. The final data set from the PSID contained 1787 observations.

The 1987 NHIS/CSS asked individuals for family income and not labor earnings. However, the survey collected extensive data on smoking. Family income will be a close approximation to labor earnings in families with only one working spouse. Therefore, we selected male workers who were the only working adults in their household, aged 21 to 64, working full time, but not self-employed. Family income is reported in \$1000 intervals up to \$20,000 annually, and then in larger groups. We choose the midpoint of each range as the measure of income. For those who made \$50,000 or more per year, we set income equal to \$75,000. Because of the detailed smoking questions contained in the survey, we can construct our primary instruments: the smoked at age 18 indicator, plus the light, moderate, and heavy smoking indicators for those who smoked at age 18. This data set contained 4515 observations.

Average weekly earnings in the September 1985 CPS were \$479.59 and 45.9 percent of these workers began smoking by age 18. Corresponding values for the September 1989 CPS data are \$551.50 and 36 percent. The average hourly earnings in the PSID sample was \$12.50 and 41.5 percent of the workers smoked at age 18. Average family earnings in

the 1987 NHIS/CSS were \$35,750 and 43 percent of the workers smoked at age 18. In Table 4 we report results from the September 1985 CPS outgoing rotation group (columns (1)-(2)), the September 1989 CPS (columns (3)-(4)), the 1986 PSID (columns (5) - (7)) and the 1987 NHIS data (columns (8) - (10)) . The OLS estimates of the rate of return to education from the CPS data sets are similar in magnitude to the values reported in Table 3. The 2SLS estimate of the rate of return to education from the CPS data sets reproduced the NMES estimates. The first-stage estimates indicate that those who began smoking by age 18 had 1.0 to 1.3 fewer years of education than those who did not, a result similar in magnitude to results presented in Table 3.

The OLS estimate of the return to education from the PSID sample is rather large. Further, using *Smoked at age 18* as an instrument increases the rate of return to education by almost 39 percent. We find that those who smoked at age 18 had 1.5 fewer years of education and those who smoked more had fewer years of schooling. More importantly, we cannot reject the test of overidentifying restrictions. In models using the data from the NHIS, the 2SLS estimates of the rate of return to education are 8 to 10% larger than the OLS estimates, smokers had 1.3 fewer years of education than those who did not smoke at age 18, years of education is monotonically decreasing in the intensity of smoking at age 18, and the test of overidentifying restrictions is well below the 95 percent critical value of 5.99.

In summary, all four data sets produce results consistent with those generated with the 1987 NMES data. In all five data sets, we find that those who smoked at age 18 have significantly fewer years of schooling. The results with the CPS and NHIS data indicate

little ability bias whereas the PSID data suggests substantial negative ability bias.

### **Is *Smoked at age 18* a valid instrument?**

Our paper hinges critically on whether *Smoked at age 18* is a valid instrument for education in a log earnings equation. Mathematically, the variable is a valid instrument if it is correlated with years of education but uncorrelated with the unobserved determinants of earnings. As shown above, our instrument satisfies the first criteria: those who smoked at age 18 have significantly lower years of education than those who did not smoked at that age. This was verified with five data sets and the result is precisely estimated. The tests of overidentifying restrictions also indicate that our model is internally consistent. However, Newey (1985) demonstrated that the test statistic is not consistent in models with general misspecification. Consequently, to support our argument that *Smoked at Age 18* is a proxy for the discount rate, we present evidence to suggest that this variable does not capture some measure of omitted ability and is providing some indication of future-orientated behavior.

If those who smoked at age 18 have different labor market skills, we can potentially lessen the correlation between the instrument and earnings by adding variables to the model that may control for the type of ability being signaled by the smoked at age 18 variable. For example, differences between those who smoked at age 18 and those who did not may manifest itself in earnings through lower disability days or better health of current non-smokers. To control for this possibility, we include measures of disability days and self-reported health status in the basic earnings equation. As seen in rows 2 through 4 of Table

5, the addition of either variable did not change the qualitative nature of our results.<sup>26</sup> In each model the 2SLS estimate of the rate of return to education is extremely close to the OLS estimate and the test of overidentifying restrictions is substantially smaller than the 95 percent critical value.

### **Does *Smoked at age 18* Measure the Discount Rate?**

The results above are consistent with the Fuchs hypothesis that those with lower discount rates will be less likely to smoke and more likely to accumulate more education. The results are, however, also consistent with another hypothesis -- that those with lower unobserved ability accumulate fewer years of education and are more likely to smoke. Because it is unlikely that any single test can rule out either the ability or discount rate hypothesis, we present a series of tests which support our contention that smoking at age 18 is a measure of the discount rate and not an indicator of unobserved ability.

Work by other researchers indicates that "ability" alone cannot explain the smoking/education link. Kenkle (1991) for example, regressed health habits on education and health knowledge and found that the addition of the health knowledge variables only altered slightly the coefficient on education. For the smoking/education link to be explained by differences in "ability" it should be the case that those with higher education are more knowledgeable about the risks of smoking. If smoking at age 18 reflects knowledge differences, then the inclusion of the measures of knowledge in an education attainment equation should significantly alter the coefficient on the smoked at age 18 variable. Conversely, if the smoked at age 18 variable primarily reflect factors that are

orthogonal to knowledge or "ability", then the coefficient would be unaltered. To test this we used data from the 1987 NHIS/CSS on people's perception of the strength of the causal link between smoking and various diseases. From this data, we construct indicator variables for whether people believed there was a strong or moderate link between smoking and heart disease, lung cancer, chronic bronchitis, and emphysema. The reference groups are those people who believe there is no or an uncertain link between smoking and these diseases. In a regression of education on observed factors and the smoked at age 18 dummy variable for the sample of working males listed in Table 2, the OLS coefficient (standard error) on the smoked at age 18 variable is -1.190 (0.075). Once the 8 dummy variables measuring health risk perception are added to the model, the coefficient changes only marginally to -1.125 (0.074).<sup>27</sup>

If smokers at age 18 have different endowments of some unobserved factor (ability), it seems likely that this same factor will affect other labor market outcomes as well. In Table 6, we use five different labor outcome measures as dependent variables and included education, age, age squared, race, region, SMSA, and *Smoked at age 18* as independent variables. We also estimate models where we include the indicators for light, moderate, and heavy smokers among those who smoked at age 18. The appropriate F or Chi-squared test for whether these three variables are jointly zero is reported in the last column of the table. These dependent variables are hours of work per week, indicators for whether the workers are in unions or high wage industries,<sup>28</sup> log of family income, and finally, log weekly earnings itself. In each case, education (a measure of ability) is statistically significant, while the coefficient on *Smoked at age 18* is small and statistically insignificant.



The point estimates also suggest economically unimportant effects as those who smoked at age 18 work only 0.03 more hours per week, are 1.9 percentage points more likely to be a union member, have 3.8 percent lower family incomes, and have only 0.6 percent lower earnings. This last model is obviously not a clean test of whether *Smoked at age 18* is uncorrelated with the unobserved component of earnings, but this dirty wage regression is instructive nonetheless. For all five dependent variables, we cannot reject the null hypothesis that the coefficients on the light, moderate, and heavy smoking variables are jointly zero. In fact, none of the 20 coefficients reported in the table is statistically significant.

Although these econometric results indicate that our smoking measure is not proxying for unobserved ability that is rewarded in the labor market, this does not prove that it is a measure of the discount rate. An alternative hypothesis is that the education/smoking relationship is causal via the knowledge effects of schooling on smoking. We devised three tests to analyze the potential importance of this hypothesis. First, following Farrell and Fuchs (1982), we delete those workers with less than 12 years of education. Therefore, at the age when we are recording the instrumental variable, most people have the same years of schooling. These results are reported in line 5 of Table 5. In this model, use of the smoked at age 18 instrument suggests no omitted ability bias. In the first-stage regression for this model, the coefficient on smoked at age 18 is still large in absolute value and precisely estimated.

A second way to eliminate the possible knowledge effect of schooling on smoking is to exploit the fact that although most individuals at age 18 have fairly similar education, 40

percent of them still smoke. Consequently, if there is a knowledge effect of education it will show up through generating changes in the smoking behavior among those who attain additional schooling after age 18. For those who acquire more schooling but do not change their smoking behavior, the knowledge effect of education must be limited. In a sample restricted to individuals who do not change their smoking behavior between age 18 and the time of the survey, there should be little or no correlation between schooling and smoking at age 18 due to knowledge or other time varying effects induced by the additional schooling people acquired.<sup>29</sup> However, if a time invariant factor like the discount rate were driving the correlation between smoked at 18 and schooling, the results would be unchanged by this sample restriction. On line 6 of Table 5 we report the OLS and 2SLS results from estimating our earnings equations on this restricted sample. The estimated return to schooling in the 2SLS results is still highly significant. Interestingly, the estimated return to schooling is even higher in the 2SLS results in the restricted sample. By eliminating people who began or quit smoking after age 18 from the sample, we removed those with moderate discount rates, leaving only the high (continuous smokers) and low (never smoked) discount rate individuals in the sample. To the degree this generates an increase in the mean discount rate in the sample, it should result in a higher estimated return to education since high discount rate individuals require a greater return to invest in schooling.

Finally, widespread news stories of the negative health effects of smoking were first reported in 1953. In 1964, the first Surgeon General's report documenting the health effects of smoking was published.<sup>30</sup> Given the dramatic changes in the public awareness

of the risks of cigarette smoking over the last forty years, one might expect that more recent cohorts who smoked at age 18 were more knowledgeable of the risks than older cohorts. If there has been a secular rise in the availability of information on, or perceptions of, the risks of smoking (related to the Surgeon General's and other reports), then we would expect that the decision to smoke at age 18 for older cohorts is less likely to reflect the discount rate considerations that are central to the Fuchs' hypothesis. If Fuchs is correct then as general knowledge of the risks improved over time, differences in smoking behavior across individuals will increasingly reflect differences in discount rates. This will generate an increasing negative correlation between smoking and education. Conversely, if the education/smoking link is due to knowledge effects, then as successive cohorts have greater knowledge of the dangers of smoking from sources other than schools the correlation should diminish. To test this hypothesis, we first interact our smoked at age 18 variables with cohort or decade of birth.<sup>31</sup> We also distributed the respondents into three categories: those who turned 18 before 1952, those who turned 18 in 1953-1963, and those who turned 18 during or after 1964. These results are presented in Table 7.

In column (2) the coefficient on smoking at age 18 in the first stage education equation shows almost a monotonic increase (in absolute value) with decade of birth. In the third column, there is a monotonically increasing relationship across the three groups. These results are clearly inconsistent with the causal or knowledge models of the education/smoking link. Consistent with the Fuchs's hypothesis, we find that among those reaching age 18 after the Surgeon General report becomes widely known, smoking is even more negatively related to educational attainment. In these models, the IV estimate of the

rate of return to education is 12 percent larger than the OLS estimate and the test of overidentifying restrictions is substantially smaller than the 95% critical value.

Given this lack of support for either the causal or unobserved ability explanations of the schooling/smoking link, we turn to some evidence which supports the discount rate explanation. Willis (1986) notes that when individuals in a human capital model face a rising supply of finance function, the discount rate will affect the schooling investment decision because individuals care about the timing of their earnings (consumption) stream and not just its present value. Those with high discount rates will acquire less schooling at any given borrowing rate. Similarly, it will take a lower borrowing rate to induce those with high discount rates to make the same level of educational investment. This suggests that there will be an interaction between borrowing rates and discount rates that affect the level of educational attainment. If our instrument -- smoked at age 18 -- is an indicator of the discount rate, then we might expect to see an interaction effect between it and measures of the borrowing rate. Since the higher the borrowing rate the less schooling an individual with a given discount rate would acquire, the coefficients on these interactions should be monotonically decreasing in the borrowing rate.

To test this we construct proxies for borrowing rates. Following Becker (1975) we assume that parental resources are correlated with the costs of financing education. Thus, we use interactions between measures of family resources (at age 18) and our smoking variable to test this hypothesis. Our family resource measures are based on survey responses from the 1986 PSID. In that survey year, respondents were asked to indicate whether they came from a poor, average income, or "well off" family when they were growing up.

Respondents were also asked the level of their parent's education. From the former question, we construct three indicators for poor, average, and well off families, and interact these indicators with smoked at age 18. Similarly, from the questions on parents' education, we construct three indicators for the level of parents' education (no high school degree, a high school degree or some college, and a college degree), and interact these variables with smoked at age 18. These results are reported in Table 8.<sup>32</sup>

In both cases, the coefficient on the smoked at age 18/family resources interactions decline in absolute value as the family resource measures increased. The 2SLS estimates of the return to education are larger than the OLS estimates and the tests of overidentifying restrictions are well below the critical values of 5.99. A potential problem with these tests is that they may be capturing the presence of differences in demand for education across wealth groups or the fact that children from wealthier families have greater ability. If this is strictly true the data should reject the null hypothesis in the tests of overidentifying restrictions. In fact, just using parents' education variable alone as instruments for children's education in the wage equation yields a test statistic of 7.54 (with 1 degree of freedom), which easily rejects the null hypothesis of internal model consistency. As this is not true when parent's education is interacted with our smoking measure, we appear to be capturing something other than ability with these interaction terms.

Finally, if our smoked at age 18 variable is primarily picking up discount rate considerations then one might expect it to correlated with other measures of the discount rate. Fuchs (1982) presented evidence that smoking is correlated with various measures of individuals' implicit interest rates derived from their answers to hypothetical money choice

questions. An alternate to calculating implicit interest rates from these type of surveys would be to see if smoking at age 18 is correlated with future oriented behavior besides educational attainment. In Table 9 we present models where we estimate the link between smoking at age 18 and several other measures of future orientated behavior.

An obvious candidate for future orientated behavior is saving. Unfortunately, all of the data sets that contain our smoking variable have only partial measures of household or individual saving. Consequently, we use the probability of home ownership and whether you received interest or dividend income (and how much) as proxies for saving. Since these are the primary asset in most household's portfolios, they should be correlated with saving behavior. The coefficient on smoking is the correct sign in each of the estimated equations. There is a strong negative correlation with home ownership even after controlling for education, income, and demographics. The marginal impact of smoking is quite large reducing the probability of home ownership by 6 percentage points. The size of this effect on home ownership is equivalent to a 6 year drop in education or a 31 percent drop in income.

Smoking at age 18 and other health habits that represent investment behavior should also be correlated if they share common discount rate considerations. Fuchs (1982) presents evidence that there is a significant raw correlation between smoking and other health habits. This correlation could, however, be due to common income effects or because of the health effects of smoking. For example, smokers may exercise less because it is difficult for them to do so, not necessarily because of time preferences. Therefore, we choose two health habits that share a common time preference component, but are uncorrelated with the health

effects of smoking -- whether a person wears a seat belt and whether a person flosses. In the last two rows of Table 9 we present equations for these two measures that control for income and other taste factors. In both cases smoking is negatively and significantly related to investments in other health habits. Taken as a whole these results are strongly supportive of the discount rate interpretation of our smoking instrument.

### **Results for Females**

So far our focus has been on estimating human capital earnings functions for samples of male workers. Although the underlying model of health and schooling investments are not gender specific, there are a number of reasons for looking at male and female behavior separately. First, the cross-sectional patterns of smoking and school attainment vary by gender. Historically, males have had both higher smoking participation rates and means years of education. Further, the time series or cohort patterns of smoking differ across these groups. Using data from the entire 1987 NHIS/CSS, we calculate the percent of people who smoked at age 18 by gender and by the year they turned 18. The percent of males that smoked at age 18 has declined steadily since the early 1960's, whereas the rate for females increased from the early 1960's to the late 1970's before declining. Farrell and Fuchs (1982) argue that while male behavior was affected by increase health risk knowledge, female smoking trends were confounded by other broader social trends. If this is true then we might expect that our smoking measure might not perform the same in male and female earnings equations.<sup>33</sup>

In Table 10 we repeat our basic OLS and 2SLS estimates using a sample of women

from the NMES. The first stage results again show a significant negative correlation between smoked at 18 and schooling. Both the incidence and quantity of smoking measures have the correct signs. Interestingly, the coefficients on smoking in the female education equations are only about half the size of those in the male equations. This suggests, as Fuchs argued, that smoking is more weakly related to discount rate differences among women than among men.

In all models, the 2SLS estimates of the return to schooling are again larger than the OLS values and are all precisely estimated. Further, we cannot reject the test of overidentifying restrictions. The point estimates on the cohort interaction terms again suggest that for older women smoking at age 18 serves as a weaker discount rate signal than among younger women. However, while the cohort effects are fairly monotonic for men they are not for women. Indeed, women smokers born in the 1920's had more education than nonsmokers suggesting that smoking at age 18 was correlated with other class or social factors.

Interestingly, the ability bias corrected return to schooling is much larger than the corresponding male estimate (12.2 vs 7.8 percent) despite more modest differences in the OLS estimates (7.9 vs 6.6). The large difference between the OLS and 2SLS estimates in the female sample is consistent with the findings of Butcher and Case (1994). Despite the larger increase in the coefficient, it should be noted that the IV estimate for women remains within the range of estimates found in previous samples for males and while the Butcher and Case IV estimate of the return to schooling is fairly imprecise, our second stage equations have t-statistics in excess of 4.



## **VI. Conclusions**

In this paper we exploit the strong link between health habits and education to develop an alternate econometric estimator of the return to schooling. Fuchs has suggested that variations in individual discount rates explained both differences in educational attainment and differences in the propensity to engage in various health habits. Consequently, health habits or practices that primarily reflect these discount rate considerations could potentially serve as valid instruments for education in deriving an estimate of the return to schooling purged of ability bias.

Our empirical results strongly support the hypothesis that smoking behavior at age 18 (the age that most decisions concerning years of schooling occur) is a statistically valid instrument for education. As found in other recent studies, the IV estimates of the return to schooling are about 10 percent higher than the OLS estimates. In contrast to many recent studies, however, our instrument is strongly correlated with education and hence devoid of the econometric problem associated with a weak instrument. To the degree ability bias is present it seems to reduce OLS estimates of the return to schooling.

We present evidence in support of the tests for overidentifying restrictions to show that the smoking decision does not reflect unobserved differences in knowledge or ability that are rewarded in the labor market. Smoking at age 18 has no independent effect on a variety of other labor market outcomes (controlling for schooling) and its correlation with education is insensitive to explicit controls for health risk knowledge. Despite a secular diminution in the need for formal education to acquire knowledge of the health risks of smoking, the link between schooling and smoking is stronger among younger cohorts of

men and women. Thus, this variable appears to be an increasingly useful indicator of discount rate differences. This conclusion is reinforced by the strong independent effect of smoking behavior on measures of saving and other future oriented activities.

Some readers might consider our use of health habits as an instrument to be an anomaly with no particular use in other econometric models. We would suggest, however, that the general class of problems where this instrument can be utilized is rather large. In many models where the variation in the potentially endogenous variable of interest are driven by heterogeneity in the time horizon, smoked at age 18 or other health habits may be considered as potential instruments. To give some indication how this type of variable can be used as an instrument or as a variable to identify hypothesis in other problems, consider the following four examples:

**The health costs of smoking:** A number of researchers have attempted to estimate the medical costs attributable to smokers. A recent report by Bartlett *et al.* (1994) places these costs at \$50 billion in 1993. These calculations are critically important in models such as those in Manning *et al.* (1989) that seek to determine whether current cigarette tax revenues pay for the external costs of smoking. These numbers are typically generated from models where total health expenditures per person are regressed on a vector of demographic variables and indicators of smoking. The estimates on the smoking variables are potentially biased because of omitted factors. Smokers are more likely to have a whole host of other bad health habits such as heavy drinking and poor diet that may also generate differences in medical care expenditures. Because of data limitations, these factors are typically omitted

from these models. Given the high degree of correlation between health habits, one could use some habits that may have no direct effect on health care expenses as instruments for smoking. For example, as demonstrated above, smokers are less likely to wear their seatbelt and floss regularly. We would not expect either of these health habits to directly effect health care use. Therefore, these variables can potentially be used as instruments for smoking in these models.

**The intergenerational link in smoking:** In a recent article that has received widespread media attention, Kandel *et al.* (1994) found that daughters of women who smoked during pregnancy have a substantially higher probability of smoking than other women. Although the authors are careful to point out that this result does not prove a causal relationship between the two variables, they do suggest a few biological avenues that would facilitate this intergenerational transmission. A test that could potentially falsify the biological hypothesis would be to estimate a model of the probability that a daughter smoked using only those women whose mothers did not smoke during pregnancy and include the smoked at age 18 variable. There is no biological reason why controlling for pre-natal behavior smoked at age 18 should affect the probability that a daughter would smoke. If, however, the smoked at age 18 coefficient is positive in this model, a biological reason for the linkage is ruled out and shared discount rate or taste factors are more likely.

**The health consequences of heavy drinking:** A number of authors have found that heavy drinkers have lower earnings than other workers, although this results is sensitive to data

sets and model specification. The possibility exists however, that heavy drinking is correlated with the unobserved determinants of earnings. As a number of authors have noted, however, the list of potential instruments for heavy drinking is rather limited. In most papers, therefore, heavy drinking has been assumed to be exogenous. One could use the interstate variation in taxes as an instrument for heavy drinking. However, with individual-level data, one cannot reject the null hypothesis that heavy drinkers have perfectly inelastic demand (Manning *et al.*, forthcoming). Using taxes as an instrument in this instance is therefore open to the criticism of Bound *et al.* (1993). Given the fact that in the 1985 NHIS/HPDP survey, those who ever smoked are substantially more likely to be heavy drinkers, one could potentially use smoked at age 18 as an instrument for current heavy drinking.

**The labor market consequences of teenage motherhood.** Women who give birth during their teen years have substantially lower incomes and years of education than non-teen mothers. A number of authors have questioned whether this is a causal effect. Geronimus and Korenman (1992) for example suggest that teen motherhood may signal other observed factors that are correlated with such labor market outcomes as low years of education or low income. To test their hypothesis, Geronimus and Korenman use within-family variation in the timing of births to estimate family fixed-effects models. In a family income equation with an indicator for teenage motherhood, one could potentially use teenage health habits such as whether one smoked at age 18 as an instrument for teenage motherhood. Those teens with a longer time horizon would be less likely to smoke, and less likely to

have a child during their teen years. In this instance, the raw correlation between teen motherhood and whether one smoked at age 18 is rather strong.

The work presented above and these four additional topics all suggest that certain health habits could potentially be used to identify other human capital type models where unobserved heterogeneity pose problems for econometric estimation.

## ENDNOTES

1. These trends are reviewed in Harris (1994).
2. These were the conclusions of the Surgeon General. See U.S. Department of Health, Education and Welfare (1979).
3. See Card (1994) for a survey of much of the recent literature on the returns to schooling and attempts to correct to ability bias.
4. In about 10 percent of all observations, family income was not reported. Rather than delete these observations, we set the value of log family income to zero for these observations and constructed a variable that equaled 1 if income was missing.
5. Numerous econometric tests of these different hypothesis have been conducted over the past 20 years. For a review of the empirical evidence concerning the various models, see Grossman and Kaestner (1994).
6. There are a number of ways in which discount rates will affect the decision to acquire additional schooling. If imperfect access to capital requires individuals to alter consumption patterns over time in order to finance education, or if the nonmonetary costs to education vary across individuals then discount rates will affect the investment decision. These models are discussed in detail in Willis (1986).
7. These surveys are review by Viscusi (1992).
8. Card (1994) provides a review of recent attempts to deal with this bias through instrumental variables as well as through the use of variants of a fixed effect estimator to control for unobserved ability. By assuming that unobserved ability is either invariant overtime, within families, or within-twins, differences across groups or time can be used to derive unbiased estimates of the effect of schooling on earnings.
9. See Lang (1993) for a discussion of the choice of instruments in an earnings equations,.
10. See Card (1994) for a review of these studies.
11. For example, results from Table 1 indicate that owning a smoke detector or using seat belts are normal goods, whereas smoking is an inferior good.
12. Authors' calculations from the 1987 National Health Interview Survey, Cancer Control Supplement.

13. See U.S. Public Health Service (1990), and Allen (1981). In a review of these studies, Ault *et al.* (1991) argues that once one controls for the fact that smokers are not randomly assigned in the population, the statistical relationship between smoking and absenteeism disappears.
14. The lagged values of health would be uncorrelated with the time variant components of earnings but still might be correlated with unobserved time invariant factors.
15. In the 1990 NHIS, labor earnings were included in the public use files for the first time. However, the smoking questions did not allow us to construct a measure of whether a person smoked at age 18.
16. In the NMES, black and hispanic families were over sampled.
17. Our results are not sensitive to these sample selection rules.
18. These data were taken from the NBER CD-ROM containing the outgoing rotation groups for the years 1978-1991.
19. In our regression work, we do not use this artificially top-coded data from the NMES.
20. Although the sample average for such variables as weekly earnings and education are quite similar in the two data sets, tests of means will reject the null hypothesis of identical means because of the large number of observations in both data sets. For example, holding the mean and variance of weekly earnings in the 87 CPS/OR constant as well as the variance of weekly earnings in the 87 NMES data, weekly earnings differences of as little as 1% would cause one to reject the null of identical means.
21. Warner (1987) noted that most self-reported studies of smoking tend to understate daily consumption.
22. As noted previously, weekly earnings in the 1987 CPS were top coded at \$999. To eliminate the bias associated with the censoring, we used the following procedure. First, we assumed log weekly earnings are normally distributed. Then using the definition of the expected value of a censored variable and the sample mean of the censored log weekly earnings, we computed the sample estimates for the mean and standard deviation of the uncensored log weekly wage distribution. We then use these values to calculate the expected value of the censored portion of log weekly earnings. We then use this expected value in place of all top coded earnings. For this data set, the expected value of the censored log weekly earnings is 1.26 times the top coded value of \$999. The weekly earnings in the NMES data are not top coded.
23. These results are reviewed in various Surgeon General's reports on smoking. See in particular 1982 report *The Health Consequences of Smoking: Cancer. A Report of the Surgeon General*, the 1983 report *The Health Consequences of Smoking: Cardiovascular Disease. A Report of the Surgeon General*, and the 1990 report *The Health Benefits of*

*Smoking Cessation. A Report of the Surgeon General.*

24. A number of public and private studies have tried to assess the public's general knowledge of smoking risks. These studies are discussed in Viscusi (1992). We have, however, been unable to locate any survey that addresses whether people perceive higher risks to heavier smokers.

25. We also consider a more limited set of right-hand-side variables in the log earnings equation where we include only education, age and age squared, plus the black and hispanic indicators. The OLS estimate (standard error) for the education coefficient is 0.070 (0.003) and the 2SLS estimates of the parameter with smoked at age 18 and the light, moderate, and heavy smoking breakdown are 0.075 (0.012) and 0.072 (0.012) respectively. For the latter model, the test of overidentifying restriction is 2.31 which is again well below the 95 percent critical value for a Chi-squared with 2 degrees of freedom.

26. Total disability days is entered linearly in the models. Self-reported health status is categorized into three dummy variables for excellent, good, and fair health, with the reference group being poor health.

27. We performed the same exercise with the light, moderate, and heavy smoking coefficients in the first-stage estimate of the education equation. Without the 8 risk perception variables, the coefficients on these three indicators are -0.573 (0.118), -1.107 (0.091), and -1.536 (0.133) respectively. Adding the risk perception variables, the coefficients are -0.535 (0.115), -1.050 (0.088), -1.487 (0.129).

28. High wage industries are mining, construction, transportation, communications, and utilities.

29. Differences in schooling and smoking at age 18 may still account for some of the correlation but, as noted above, most of the variation in schooling occurs because of decisions made near the age of 18.

30. Viscusi (1991) provides a detailed history of the health information concerning smoking.

31. We used 1987 NMES public use file contained on tapes 9, 13, and 18. Year of birth was not reported on any of these tapes. Therefore, we can only approximate year of birth as 87 minus age.

32. Because of missing values for parents' education and family income, the number of observations is lower than that reported in Table 4.

33. In addition to these issue, OLS estimates of wage equations for women are also thought to potentially suffer from sample selection and endogeneity bias that are not important for male wage equation.



## Bibliography

- Allen, S.G. (1981): "An Empirical Model of Work Attendance," *Review of Economics and Statistics*, 63, 77-87.
- Angrist, J.D., and A.B. Krueger (1991a): "Does Compulsory Schooling Affect Schooling and Earnings," *The Quarterly Journal of Economics*, 106, 979-1014.
- Angrist, J.D., and A.B. Krueger (1991b): "Estimating the Payoff to Schooling Using the Vietnam-Era Draft Lottery," Industrial Relations Section, Princeton University, Working Paper #290.
- Ashenfelter, O., and D. Zimmerman (1993): "Estimates of the Returns to Schooling from Siblings: Fathers, Sons, and Brothers," National Bureau of Economic Research Working Paper #4491.
- Ashenfelter, O., and A.B. Krueger (1994): "Estimates of the Economic Return to Schooling for a New Sample of Twins," Industrial Relations Section, Princeton University, Working Paper #304.
- Ault, R.W., et al. (1991): "Smoking and Absenteeism," *Applied Economics*, 23, 743-54.
- Bartlett, J.C., et al., (1994): "Medical-Care Expenditures Attributable to Cigarette Smoking - United States, 1993," *Morbidity and Mortality Weekly Report*, 43, 469-472.
- Becker, G.S. (1967): *Human Capital and the Personal Distribution of Income*, Ann Arbor, Michigan: University of Michigan Press.
- Becker, G.S. (1975): *Human Capital, 2nd. Ed.* New York: National Bureau of Economic Research.
- Becker, G.S. and K.M. Murphy (1988): "A Theory of Rational Addiction," *Journal of Political Economy*, 96, 675-700.
- Bound, J., D.A. Jaeger, and R. Baker (1993): "The Cure Can Be Worse Than the Disease: A Cautionary Tale Regarding Instrumental Variables," National Bureau of Economic Research Technical Working Paper #137.
- Butcher, K.F., and A. Case (1994): "The Effect of Sibling Composition on Women's Education and Earnings," *The Quarterly Journal of Economics*, 109, 531-564.

- Card, D. (1993): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," National Bureau of Economic Research Working Paper #4483.
- Card, D. (1994): "Earnings, Schooling, and Ability Revisited," National Bureau of Economic Research Working Paper #4832.
- Farrell, P., and V.R. Fuchs (1982): "Schooling and Health: The Cigarette Connection," *Journal of Health Economics*, 1, 217-30.
- Fuchs, V.R. (1982): "Time Preferences and Health: An Exploratory Study," in *Economic Aspects of Health*, edited by V.R. Fuchs, Chicago: University of Chicago Press.
- Geronimus, A.T., and S. Korenman, (1992): "The Socioeconomic Consequences of Teen Childbearing Reconsidered," *Quarterly Journal of Economics*, 107, 1187-1214.
- Griliches, Z., (1977): "Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica*, 45, 1-22.
- Grossman, M. (1976): "The Demand for Health: A Theoretical and Empirical Investigation," in *Household Production and Consumption*, edited by N.E. Terlecky, New York: Columbia University Press.
- Grossman, M., and R. Kaestner, (1994): "Social Benefits of Education: Health," working paper, City College of New York Graduate School.
- Hansen, L.P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-54.
- Hamilton, J.L. (1972): "The Demand for Cigarettes: Advertising, the Health Scare, and the Cigarette Advertising Ban," *Review of Economics and Statistics*, 54, 401-11.
- Harris, J.R. (1994): "A Working Model for Predicting the Consumption and Revenue Impacts of Large Increases in the U.S. Federal Excise Cigarette Tax," National Bureau of Economic Research Working Paper #4803.
- Kandel, D.B., P. Wu., and M. Davies, (1994): "Maternal Smoking During Pregnancy and Smoking by Adolescent Daughters," *American Journal of Public Health*, 84, 1407-1413.
- Kane, T.J., and C.E. Rouse (1993): "Labor Market Returns to Two-and Four-Year Colleges: Is a Credit a Credit and do Degrees Matter?" Industrial Relations Section, Princeton University, Working Paper #311.

- Kenkel, D.S. (1991): "Health Behavior, Health Knowledge, and Schooling," *Journal of Political Economy*, 99, 287-305.
- Kitagawa, E., and P. Hauser (1973): *Differential Mortality in the U.S.*, Cambridge, Massachusetts: Harvard University Press.
- Lang, K., (1993): "Ability Bias, Discount Rate Bias, and the Return to Education," Boston University Department of Economics, Working Paper.
- Lewit, E.M., D. Coate, and M. Grossman (1981): "The Effects of Government Regulations on Teenage Smoking," *Journal of Law and Economics*, 24, 545-69.
- Manning, W.G, *et al.*, (1989): "The Taxes of Sin: Do Smokers and Drinkers Pay Their Way?" *Journal of the American Medical Association*, 261, 1604-1609.
- Manning, W.G., L. Blumberg, and L.H. Moulton, (forthcoming): "The Demand for Alcohol: The Differential Response to Price," *Journal of Health Economics*.
- Michael, R.T. (1973): "Education in Nonmarket Production," *Journal of Political Economy*, 81, 306-27.
- Newey, W. (1985): "Generalized Methods of Moments Specification Testing," *Journal of Econometrics*, 29, 229-56.
- Neumark, D. (1994): "Biases in Twin Estimates of the Return to Schooling: A Note on Recent Research," National Bureau of Economic Research Technical Working Paper #158.
- Newhouse, J.P., and L.J. Friedlander (1979): "The Relationship Between Medical Resources and Measures of Health: Some Additional Evidence," *Journal of Human Resources*, 15, 200-18.
- Pincus, T., L.F. Callahan, and R.V. Burkhauser (1987): "Most Chronic Diseases are Reported More Frequently by Individuals with Fewer than 12 Years of Formal Education in the Age 18-64 United States Population," *Journal of Chronic Diseases*, 40, 865-874.
- Rosen, S. (1977). "Human Capital: A Review of Empirical Research," in R. Ehrenberg, ed., *Research in Labor Economics*, vol. 1, Greenwich, Connecticut: JAI Press.
- Rosen, S., and P. Taubman (1982): "Some Socioeconomic Determinants of Mortality," in *Economics of Health Care*, edited by J. van der Gagg, W.B. Neeman, and T.

Tsukahara, Jr., New York: Praeger Publishers.

Schneider, L., B. Klein, and K. M. Murphy (1981): "Governmental Regulation of Cigarette Health Information," *Journal of Law and Economics*, 24, 575-612.

Staiger, D., and J.H. Stock (1994): "Instrumental Variables Regressions with Weak Instruments," National Bureau of Economic Research Technical Working Paper 151.

U.S. Department of Health, Education, and Welfare (1979): *Healthy People: The Surgeon General's Report on Health Promotion and Disease Prevention*, Washinton, D.C., Government Printing Office.

U.S. Department of Health and Human Services (1982): *The Health Consequences of Smoking: Cancer. A Report of the Surgeon General*, U.S. Department of Health and Services, Public Health Service, Centers for Disease Control, Office of Smoking and Health, DHHS Publication Number (PHS) 82-50179.

U.S. Department of Health and Human Services (1982): *The Health Consequences of Smoking: Cardiovascular Disease. A Report of the Surgeon General*, U.S. Department of Health and Services, Public Health Service, Centers for Disease Control, Office of Smoking and Health, DHHS Publication Number (PHS) 84-50204.

U.S. Department of Health and Human Services (1990): *The Health Benefits of Smoking Cessation: A Report of the Surgeon General*, U.S. Department of Health and Services, Public Health Service, Centers for Disease Control, Office of Smoking and Health, DHHS Publication Number (CDC) 90-8416.

Viscusi, W. K. (1992): *Smoking: Making the Risky Decision*, New York: Oxford University Press.

Warner, Kenneth E. (1978): "Possible Increases in the Underreporting of Cigarette Consumption," *Journal of the American Statistical Association*, 73, 314-18.

Willis, R.J. (1986): "Wage Determination: A Survey and Reinterpretation of Human Capital Earnings Functions," in *Handbook of Labor Economics*, eds. O. Ashenfelter and R. Layard, New York: North Holland.

Table 1  
 Probit Estimates for Health Habit Indicators,  
 1985 NHIS/HPDP Supplement

Health Habit (1=yes, 0=no)	Sample *	Sample Size	Sample Mean	Normalized Coefficient and Asymptotic Standard Error on Years of education <sup>b</sup>	Normalized Coefficient and Asymptotic Standard Error on ln(Family income) <sup>c</sup>
Current smoker?	full sample	24,537	0.342	-0.029 (0.001)	-2.6E-3 (4.8E-4)
Ever smoked?	full sample	24,360	0.579	-0.025 (0.001)	-6.6E-4 (4.7E-4)
Wear safety belt most of the time?	full sample	24,351	0.371	0.034 (0.002)	5.0E-3 (4.9E-4)
Exercise regularly?	full sample	24,438	0.410	0.035 (0.001)	6.1E-3 (5.1E-4)
Had blood pressure checked in past year?	full sample	23,991	0.576	0.013 (0.002)	3.3E-4 (5.9E-3)
Heavy drinker (more than 4 drinks per day past 2 weeks)?	full sample	24,448	0.106	-0.009 (0.001)	-5.3E-4 (3.1E-4)
Eat breakfast every morning?	full sample	24,551	0.494	0.022 (0.001)	2.2E-5 (4.8E-4)
Have working smoke detector in home?	full sample	23,540	0.628	0.019 (0.001)	7.0E-3 (4.9E-4)
Ever drove drunk?	people who ever drank	16,787	0.212	-0.011 (0.001)	-7.2E-4 (4.9E-4)
Had pap smear in past year?	females	13,569	0.504	0.013 (0.002)	4.0E-3 (6.8E-4)
Had breast exam in past year?	females	13,595	0.533	0.012 (0.002)	5.7E-3 (6.8E-4)
Have Ipecac syrup in home?	families with young children	7,153	0.265	0.028 (0.003)	4.4E-3 (9.8E-4)

<sup>a</sup> There are 24,685 possible observations in the full sample, of which 13,833 are women. All models include the following covariates: Years of education, natural log of family income, an indicator for family income not reported, age and age squared, family size, and indicator variables for black, hispanic, married, never married, North East, Midwest, West, large SMSA, and medium-sized SMSA. Where appropriate, indicators for gender are also included.

<sup>b</sup> The normalized coefficient for Years of education represent the change in the probability of answering yes given a one year change in education. If  $p$  is the sample mean of the response variable and  $\beta_e$  is the probit coefficient for years of education, the marginal effect is calculated as  $\beta_e \phi(z)$ , where  $\Phi^{-1}(p)=z$ .

<sup>c</sup> The normalized coefficient for ln(Family income) represent the change in the probability of answering yes given a 10% increase in family income. If  $p$  is the sample mean of the response variable and  $\beta_i$  is the probit coefficient for ln(Family income), the marginal effect is calculated as  $.1 \beta_i \phi(z)$ , where  $\Phi^{-1}(p)=z$ .

Table 2  
Comparison of 1987 NMES, 1987 CPS/OR, and 1987 NHIS/CSS Survey Data

Variable Name	Sample Means and (Standard Deviations)		
	1987 NMES	1987 CPS/OR	1987 NHIS/CSS
Weekly earnings <sup>a</sup>	473.89 (228.29)	486.81 (237.68)	
Hours of work per week	43.58 (7.61)	43.27 (7.38)	
Years of education	13.02 (2.94)	13.12 (2.83)	13.22 (2.86)
Age	37.21 (10.90)	38.05 (11.16)	38.20 (11.01)
Black (1-0 indicator)	0.14 (0.35)	0.08 (0.27)	0.10 (0.30)
Hispanic (1-0 indicator)	0.11 (0.31)	0.06 (0.24)	0.08 (0.28)
Union (1-0 indicator)	0.24 (0.43)	0.23 (0.42)	
Ever Smoked 100 Cigarettes (1-0 indicator)	0.55 (0.50)		0.60 (0.49)
Current Smoker (1-0 indicator)	0.30 (0.46)		0.33 (0.47)
Smoked at age 18 (1-0 indicator)	0.41 (0.49)		0.41 (0.49)
Cigarettes per day -- current smokers only	21.00 (11.71)		21.88 (12.77)
Cigarettes per day -- age 18 smokers only	20.54 (11.66)		24.26 (14.75)
Observations	3302	80,300	5092

<sup>a</sup> To compare results across surveys, weekly earnings in the 1987 NMES survey were artificially top-coded at \$999/week.

Table 3  
Estimates of Log Weekly Earnings Equation.  
1987 CPS/OR and 1987 NMES Data

Parameter Estimates and Standard Errors

Variable Name <sup>a</sup>	1987 CPS/OR		1987 NMES	
	OLS	OLS	2SLS	2SLS
Intercept	3.422 (0.021)	3.558 (0.103)	3.492 (0.158)	3.511 (0.154)
Years of Education	0.069 (5.9E-4)	0.066 (0.003)	0.073 (0.013)	0.071 (0.012)
Age	0.076 (0.001)	0.075 (0.005)	0.073 (0.006)	0.074 (0.006)
Age <sup>2</sup>	-7.7E-4 (1.3E-5)	-7.5E-4 (6.3E-5)	-7.3E-4 (7.3E-5)	-7.4E-4 (7.2E-5)
Black (1-0 indicator)	-0.255 (0.006)	-0.191 (0.023)	-0.187 (0.024)	-0.188 (0.024)
Hispanic (1-0 indicator)	-0.195 (0.007)	-0.128 (0.027)	-0.109 (0.043)	-0.115 (0.042)
Union (1-0 indicator)	0.097 (0.004)	0.141 (0.019)	0.146 (0.021)	0.145 (0.020)
R <sup>2</sup>	0.334	0.335	0.256	0.257
Test of Overidentifying Restrictions <sup>b</sup>				1.408
			<b>First-Stage Estimates of Years of Education Equation</b>	
Smoked at Age 18 (1-0 indicator)			-1.245 (0.096)	
Light smoker (≤15 cigarettes per day)				-0.969 (0.138)
Moderate smoker (>15, ≤35 cigarettes per day)				-1.304 (0.119)
Heavy smoker (>35 cigarettes per day)				-1.873 (0.222)
R <sup>2</sup>			0.181	0.184

All models include 3 regional and 2 SMSA indicator variables.

<sup>b</sup> The test is distributed as a  $\chi^2$  with 2 degrees of freedom. The 95 percent critical value for  $\chi^2(2)$  is 5.99.

Table 4  
 Estimates of Log Earnings Model with  
 September 1985 CPS/OR, September 1989 CPS/OR  
 1986 PSID, and 1987 NHIS/CSS Data Sets

Parameter Estimates and Standard Errors

Variables <sup>a</sup>	9/85 CPS/OR		9/89 CPS/OR		1986 PSID			1987 NHIS/CSS		
	OLS	2SLS	OLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Years of Education	0.061 (0.002)	0.059 (0.008)	0.072 (0.002)	0.071 (0.012)	0.075 (0.004)	0.104 (0.014)	0.100 (0.014)	0.074 (0.003)	0.076 (0.014)	0.084 (0.014)
R <sup>2</sup>	0.324	0.261	0.335	0.250	0.385	0.322	0.323	0.248	0.176	0.177
Test of overidentifying restrictions (d.o.f.) <sup>b</sup>							1.232 (2)			0.346 (2)
First-Stage Estimates of Years of Education Equation										
Started smoking by age 18 (1-0 indicator)		-1.372 (0.065)								
Smoked at age 18 (1-0 indicator)				-0.998 (0.067)			-1.494 (0.108)		-1.223 (0.080)	
Light Smoker (1-0 indicator)							-1.086 (0.178)			-0.550 (0.125)
Moderate Smoker (1-0 indicator)							-1.512 (0.132)			-1.179 (0.097)
Heavy Smoker (1-0 indicator)							-2.009 (0.197)			-1.541 (0.139)
R <sup>2</sup>		0.196		0.151			0.218	0.225	0.160	0.160

The dependent variables are the log of weekly earnings for the September 1985 and 1989 CPS/OR data sets, the log of hourly earnings for the 1986 PSID data, and the log of yearly income in the 1987 NHIS/CSS. The 9/85 CPS, 9/89 CPS, PSID, and NHIS data sets samples contain 6312, 6645, 1787, and 4515 observations, respectively.

<sup>a</sup> All models contain the other exogenous variables listed in Table 3. The Hispanic indicator is not included in the PSID model and the UNION indicator is not included in the NHIS models.

<sup>b</sup> The test is distributed as a  $\chi^2$  with 2 degrees of freedom and the 95% critical value for a  $\chi^2(2)$  is 5.99.



Table 5  
Alternative Models of Log Weekly Earnings Equation,  
1987 NMES Data

Model	Comments <sup>a</sup>	Parameter Estimates and Standard Errors on Years of Education			
		OLS	Two-Stage Least Square Estimates		
			Smoke at Age 18 as Instrument	Light, Moderate, Heavy Smoker as Instruments	Test of overidentifying restrictions <sup>b</sup>
1	Base-line results (from Table 3)	0.066 (0.003)	0.073 (0.013)	0.071 (0.012)	1.408
2	Annual disability days included	0.066 (0.003)	0.072 (0.013)	0.070 (0.012)	1.488
3	3 indicators for self-reported health status included	0.064 (0.003)	0.066 (0.015)	0.064 (0.014)	1.462
4	Models (3) and (4)	0.063 (0.003)	0.065 (0.015)	0.063 (0.014)	1.506
5	Delete those with less than 12 years of education <sup>c</sup>	0.070 (0.004)	0.070 (0.020)	0.067 (0.018)	2.519
6	Include only those who do not change smoking habits between age 18 and time of survey <sup>d</sup>	0.069 (0.003)	0.099 (0.011)	0.097 (0.011)	2.641

<sup>a</sup> All models include the other exogenous variables listed in Table 3.

<sup>b</sup> The test is for the second instrument set. The test is distributed as a  $\chi^2$  with 2 degrees of freedom. The 95 percent critical value for  $\chi^2(2)$  is 5.99.

<sup>c</sup> There are 2749 observations in this model.

<sup>d</sup> There are 2481 observations in this model.

Table 6  
Labor Market Indicator Equations,  
1987 NMES Data

Parameter Estimates and Standard Error

Dependent Variable <sup>a</sup>	Estimation Method	Model (1)	Model (2)			F-test/ -2 log likelihood test <sup>d</sup>
		Smoked at Age 18 <sup>b</sup>	Light Smoker	Moderate Smoker	Heavy Smoker	
Hours of work per week	OLS	0.033 (0.279)	0.104 (0.395)	-0.107 (0.342)	0.486 (0.635)	0.29
Union Member (1-0 indicator)	Probit	0.019 (0.016)	0.001 (0.023)	0.025 (0.020)	0.029 (0.036)	1.89
Work in high wage industry (1-0 indicator) <sup>c</sup>	Probit	0.022 (0.015)	0.015 (0.021)	0.028 (0.018)	0.020 (0.030)	2.55
ln(Family Income)	OLS	-0.038 (0.029)	-0.051 (0.041)	-0.029 (0.036)	-0.042 (0.067)	0.65
ln(Weekly Earnings)	OLS	-0.006 (0.016)	-0.025 (0.023)	0.007 (0.020)	-0.010 (0.037)	0.53

<sup>a</sup> Other included explanatory variables are years of education, age, age squared, black and hispanic indicators, plus 3 regional and 2 SMSA indicator variables.

<sup>b</sup> The reported probit coefficients are normalized and the parameters measure the change in the probability given a change in the exogenous variable. If  $p$  is the sample mean of the response variable and  $\beta_{age18}$  is the probit coefficient for the smoked at age 18 variable, the marginal effect is calculated as  $\beta_{age18}\phi(z)$ , where  $\Phi^{-1}(p)=z$ .

<sup>c</sup> To determine high wage industries, we added 11 industry effects to the basic log weekly wage equation reported in Table 3 and defined as high wage industries those with statistically significant positive coefficients. Using this procedure, we defined high wage industries to include mining, construction, transportation, communications, and utilities.

<sup>d</sup> The test statistic is for the joint hypothesis that  $\beta_{light}=\beta_{moderate}=\beta_{heavy}=0$ . In the OLS models, the test is distributed as an F distribution with 3 and 3288 degrees of freedom. The 95% critical value for an  $F(3,3288)$  is 2.6. In the probit models, the test is distributed as a chi-squared with 3 degrees of freedom. The 95% critical value for a  $\chi^2(3)$  is 7.81.

Table 7  
Estimates of Log Weekly Earnings Equation  
1987 NMES Data

Parameter Estimates and Standard Errors

Variable Name <sup>a</sup>	OLS	2SLS	2SLS
Years of Education	0.066 (0.003)	0.075 (0.012)	0.074 (0.013)
R <sup>2</sup>	0.335	0.257	0.256
Test of overidentifying Restrictions (d.o.f.) <sup>b</sup>		3.354 (4)	0.670 (2)
		First-State Estimates of Years of Education Equation	
Smoked at Age 18 x Born in 20's		-0.979 (0.346)	
Smoked at Age 18 x Born in 30's		-0.873 (0.211)	
Smoked at Age 18 x Born in 40's		-1.419 (0.167)	
Smoked at Age 18 x Born in 50's		-1.114 (0.167)	
Smoked at Age 18 x Born in 60's		-1.557 (0.204)	
Smoked at Age 18 x Turned 18 by 1952			-0.767 (0.132)
Smoked at Age 18 x Turned 18 during 1953-1963.			-1.180 (0.171)
Smoked at Age 18 x Turned 18 in 1964 or after			-1.273 (0.115)
R <sup>2</sup>		0.183	0.177

<sup>a</sup> All models contain the other exogenous variables listed in Table 3.

<sup>b</sup> The test is distributed as a  $\chi^2$  and the 95% critical values for  $\chi^2(2)$  and  $\chi^2(4)$  are 5.99 and 9.48, respectively.

Table 8  
Estimates of ln(Hourly Wage) Equation,  
1986 PSID Sample

Parameter Estimates and Standard Errors

Variable <sup>a</sup>	OLS	2SLS	2SLS
Years of Education	0.076 (0.004)	0.109 (0.014)	0.106 (0.013)
R <sup>2</sup>	0.383	0.321	0.324
Test of overidentifying restrictions (d.o.f.) <sup>b</sup>		1.455 (2)	2.994 (2)
First-Stage Estimates of Years of Education Equation			
Smoked at Age 18 x Grew up in Poor Family		-2.093 (0.164)	
Smoked at Age 18 x Grew up in Avg. Income Family		-1.283 (0.146)	
Smoked at Age 18 x Grew up in Well-Off Family		-1.058 (0.187)	
Smoked at Age 18 x Parent's have less than HS Education			-2.304 (0.158)
Smoked at Age 18 x Parent's have HS Education			-1.224 (0.133)
Smoked at age 18 x Parent's have College Degree			-0.204 (0.262)
R <sup>2</sup>		0.228	0.244

There are 1703 observation in the data set.

<sup>a</sup> All models contain the other exogenous variables listed in Table 4.

<sup>b</sup> The test is distributed as a  $\chi^2$  with 2 degrees of freedom. The 95% critical values for a  $\chi^2(2)$  degrees of freedom is 5.99.

Table 9  
Correlates of Discount Rates,  
1987 NMES Data

Parameter Estimates and Standard Errors

Dependent Variable <sup>a</sup>	Estimation Method	Smoked at Age 18 <sup>b</sup>
Own Home (1-0 indicator)	Probit	-0.062 (0.019)
Have Interest or Dividend Income (1-0 indicator)	Probit	-0.010 (0.020)
Log of Interest and Dividend Income (if positive)	OLS	-0.040 (0.093)
Wear Seat Belt All the time (1-0 indicator)	Probit	-0.082 (0.019)
Used Dental Floss Yesterday (1-0 indicator)	Probit	-0.041 (0.015)

<sup>a</sup> Other included covariates are years of education, log of family income, family size, age and age squared, black and hispanic indicators, indicators for married, divorced/separated, plus 3 regional and 2 SMSA indicator variables

<sup>b</sup> The reported probit coefficients are normalized and the parameters measure the change in the probability given a change in the exogenous variable. If  $p$  is the sample mean of the response variable and  $\beta_{age18}$  is the probit coefficient for smoked at age 18 variable, the marginal effect is calculated as  $\beta_{age18}\phi(z)$ , where  $\Phi^{-1}(p)=z$ .

Table 10  
Estimates of Log Weekly Earnings Equation,  
1987 NMES Data,  
Female Sample

Parameter Estimates and Standard Errors

Variable*	OLS	2SLS	2SLS	2SLS
Years of Education	0.079 (0.003)	0.122 (0.030)	0.095 (0.025)	0.118 (0.024)
R <sup>2</sup>	0.318	0.172	0.180	0.175
Test of overidentifying restrictions (d.o.f.) <sup>b</sup>			4.801 (2)	0.434 (4)
	First-Stage Estimates of Years of Education Equation			
Smoked at age 18 (1-0 indicator)		-0.615 (0.107)		
Light Smoker ( $\leq 15$ cigarettes per day)			-0.363 (0.141)	
Moderate Smoker ( $>15, \leq 35$ cigarettes per day)			-0.849 (0.148)	
Heavy Smoker ( $> 35$ cigarettes per day)			-1.034 (0.401)	
Smoked at Age 18 x Born in 20's				0.899 (0.437)
Smoked at Age 18 x Born in 30's				-0.812 (0.278)
Smoked at Age 18 x Born in 40's				-0.747 (0.212)
Smoked at Age 18 x Born in 50's				-0.550 (0.164)
Smoked at Age 18 x Born in 60's				-0.845 (0.207)
R <sup>2</sup>		0.091	0.094	0.096

There are 3027 observation in the data set.

\* All models contain the other exogenous variables listed in Table 3.

<sup>b</sup> The test is distributed as a  $\chi^2$  and the 95% critical values for  $\chi^2(2)$  and  $\chi^2(4)$  are 5.99 and 9.49, respectively.