Economics of Education Review 30 (2011) 792–812

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

The impact of education on health knowledge☆

Duha Altindag[a], Colin Cannonier[a], Naci Mocan[b,*]

[a] Auburn University, United States
[b] Louisiana State University, NBER and IZA, Baton Rouge, LA 70803-6306, United States

ARTICLE INFO

Article history:
Received 13 August 2010
Received in revised form 6 October 2010
Accepted 20 October 2010

JEL classification:
D83
I12
I21

Keywords:
Schooling
Health
Education
Teenagers
Heart disease
Smoking
Pregnancy
Drinking
Alcohol addiction
Health knowledge
NLSY

ABSTRACT

The theory on the demand for health suggests that schooling causes health because schooling increases the efficiency of health production. Alternatively, the allocative efficiency hypothesis argues that schooling alters the input mix chosen to produce health. This suggests that the more educated have more knowledge about the health production function and they have more health knowledge. This paper uses data from the 1997 and 2002 waves of the NLSY97 to conduct an investigation of the allocative efficiency hypothesis by analyzing whether education improves health knowledge. The survey design allows us to observe the increase in health knowledge of young adults after their level of schooling is increased by differential and plausibly exogenous amounts. Using nine different questions measuring health knowledge, we find weak evidence that an increase in education generates an improvement in health knowledge for those who ultimately attend college. For those with high school as the terminal degree, no relationship is found between education and health knowledge. These results imply that the allocative efficiency hypothesis may not be the primary reason for why schooling impacts health outcomes.

1. Introduction

The seminal work of Grossman (1972a, 1972b) created the theoretical framework of a human capital model for the demand for health, where health is both demanded and also produced. The model posits that the stock of health capital enters the utility function as a consumption good because better health increases utility. Health capital also determines the amount of time that can be devoted to work in the market, and to the production of nonmarket goods. Health capital depreciates over time, and gross investment in health can be produced by a household production function that uses the person's own time and such health inputs as medical care, diet, and cigarette and alcohol consumption.

The pure investment version of the Grossman model, where health does not provide direct utility, generates unambiguous predictions. For example, as long as the marginal product of health capital declines as the stock of health gets larger (which is reasonable because output produced by health capital has a finite upper limit, such as 8800 h per year) schooling should increase the quantity of health demanded (for the details of Grossman model, see Grossman, 2000). In Grossman's model schooling causes health because schooling increases the efficiency

☆ We thank Michael Grossman, Joe Price, Jason Fletcher, Partha Deb, and the seminar participants of the 2010 Western Economic Association Meetings in Portland, Oregon, and the Graduate Center of CUNY for helpful discussions and suggestions. Two anonymous referees provided helpful comments.

* Corresponding author.
E-mail addresses: altindag@auburn.edu (D. Altindag), ccanno5@lsu.edu (C. Cannonier), mocan@lsu.edu (N. Mocan).

0272-7757/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.econedurev.2010.10.003

of health production. Numerous studies provided evidence for the causal impact of schooling on health (Berger & Leigh, 1989; Chou, Liu, Grossman, & Joyce, 2010; Conti, Heckman, & Urzua, 2010; Currie & Moretti, 2003; Lleras-Muney, 2005).[1] However, it has also been suggested that schooling influences health mainly through its impact on allocative efficiency. In this hypothesis, schooling alters the input mix chosen to produce health. Specifically, it is presumed that the more educated choose a combination of inputs that produces more output than does the input mix chosen by the less educated (see Rosenzweig & Schultz, 1982 and papers discussed in Grossman, 2006).

A horse race between allocative efficiency and productive efficiency can be conducted by estimating health production functions which include education and all potential health inputs. Because it is difficult to measure all health inputs (i.e. health behavior) and also because health inputs themselves are functions of education, estimation of such production functions is plagued with empirical difficulties. An indirect method to determine the relative importance of allocative versus productive efficiency involves estimating the relationship between schooling and health inputs by controlling for the impact of health knowledge (Kenkel, 1991). If schooling improves allocative efficiency by increasing health knowledge and thereby altering the choice of health inputs, schooling should have little or no direct effect on health inputs in a model that controls for health knowledge. However, Kenkel (1991) finds that even though part of the impact of schooling on smoking and drinking alcohol is attributable to health knowledge, most of the impact of schooling on these (negative) health inputs remains even after controlling for health knowledge, suggesting that allocative efficiency is not a major factor. Price and Simon (2009) find that during the three-month period after the publication of an article in the *New England Journal of Medicine* about the risk of a vaginal birth after having a previous C-section birth (VBAC), the incidence of VBAC dropped more significantly among the more educated mothers. This finding suggests that more educated people absorb new information more quickly, which may then change their behavior. Similarly, Aizer and Stroud (2010) report that more educated mothers reduced their smoking after the release of the 1964 Surgeon General Report on smoking and health while the less-educated did not. These findings support the allocative efficiency hypothesis because they imply that education increases information about the true impact of the inputs on health.

Allocative efficiency hypothesis suggests that more educated individuals have more information about the structure of the production function. As explained by Grossman (2008), this implies that "...the more educated have more knowledge about the harmful effects of smoking or about what constitutes an appropriate diet." In this paper we conduct a direct investigation of the allocative efficiency hypothesis by analyzing the relationship between schooling and health knowledge. Using data from the 1997 and 2002 waves of the National Longitudinal Survey of Youth 1997 cohort (NLSY97) we investigate whether education improves health knowledge. It is plausible that more educated individuals have more health knowledge not because of education itself, but because of other attributes that allow them to acquire health knowledge and that these attributes are correlated with education. For example, if wealthier and knowledgeable parents transmit their health knowledge to their children at home and if children of such parents obtain more education, then the impact of the home environment may be attributed to education if home environment is not adequately controlled for. In empirical analyses we are able to adjust for the impact of household characteristics. Furthermore, the survey design allows us to observe the increase in the health knowledge of young adults after their level of schooling is increased at differential and plausibly exogenous amounts. Specifically, two observationally identical young adults who have the same level of education in the first wave (1997 wave) of the survey may have completed different levels of schooling in the second wave (2002 wave) because the surveys are administrated in different time periods. More specifically, the 1997 wave of the NLSY97 was administrated between January 1997 and May 1998, and the 2002 survey was administered between November 2002 and July 2003. This means, for example, that a 9th grader in the 1997 wave could have been interviewed 54 months later in the 2002 wave of the survey, while another 9th grader could have been interviewed 78 months after the first survey. Given that the timing of the surveys is not correlated with student or parent attributes, this design implies that the second student was exposed to 20 additional months of schooling in comparison to the first student.[2]

Using school attendance between the two surveys in Ordinary Least Squares (OLS) regressions and using the distance between surveys as an instrument for attendance in Instrumental Variables (IV) regressions, we find spotty evidence for the hypothesis that an increase in the level of education generates an improvement in health knowledge among those who ultimately attend college. The effect is even weaker for males. In the sample of individuals where high school is the terminal degree, there is no discernible impact of education on health knowledge. These results imply that the allocative efficiency hypothesis may not be the primary reason for why schooling impacts health outcomes.

Section 2 describes the empirical specification. Section 3 presents the data. Section 4 discusses the results and Section 5 is the conclusion.

2. Empirical specification

Consider Eq. (1) below, where K_{it} stands for the extent of the health knowledge of the person i at time t, approximated by the probability of answering correctly various questions about health production and disease prevention.

$$K_{it} = \alpha + \beta S_{it} + \mathbf{X_{it}} \mathbf{\Omega}' + \boldsymbol{\varepsilon}_{it}, \quad (1)$$

[1] Also see the papers cited in Chou et al. (2010).

[2] The difference in exposure to schooling in this example is 20 months rather than 24 because there is no schooling in Summer months. We provide evidence for the exogeneity of the timing of the survey to observable household attributes in the data section.

where S_{it} stands for the schooling level of the individual i at time t, and $\mathbf{X_{it}}$ stands for a vector of variables that are expected to influence the health knowledge of the person. There is no reasonable reverse causality in this specification because the extent of an individual's health knowledge should have no impact on his/her level of schooling. On the other hand, the estimated impact of schooling on health knowledge, represented by β, could be biased if the empirical analysis fails to control for such factors as family income and parent education that may influence both the level of schooling of the person and his/her health knowledge. Furthermore, it could also be the case that unobservable personal and family characteristics, which are part of the error term $\boldsymbol{\varepsilon}_{it}$, may be correlated with schooling S_{it}. If $\boldsymbol{\varepsilon}_{it}$ consists of two components, π_{it} and μ_i, where π_{it} is idiosyncratic white noise but $\mathrm{Cov}(S_{it}, \mu_i) \neq 0$, the estimate of β will be biased.

We exploit the fact that the NLSY97 respondents answered the same health knowledge questions both in the 1997 and 2002 waves. This allows us to investigate the link between the change in education and the change in health knowledge between these two waves. Taking the time-difference of Eq. (1) between the two survey years differences out the person-specific, time-invariant component μ_i, which generates

$$\Delta^k K_{it} = \beta \Delta^k S_{it} + \Delta^k \mathbf{X_{it}} \boldsymbol{\Omega}' + \Delta^k \pi_{it}, \tag{2}$$

where Δ^k stands for k-month difference. In Eq. (2), current health knowledge is a function of the change in the quantity of schooling over the last k months, as well as the change in observable family attributes such as family income and household size. We estimate regressions based on Eq. (2).

It is easy to show that under a distributed-lag representation, Eq. (2) can be re-written as

$$K_{it} = \beta \Delta^k S_{it} + \Delta^k \mathbf{X_{it}} \boldsymbol{\Omega}' + \gamma K_{it-k} + e_{it}, \tag{3}$$

where e_{it} is the error term, which is uncorrelated with schooling. As an alternative specification, we also estimate versions of Eq. (3).

Note that, as explained in the previous section, not each person is observed in exactly 5-year intervals. In fact, the average time between the first and second interview is 68 months (the shortest distance between the two interviews is 54 months, and the longest distance is 76 months). Therefore, the k-th difference in Eq. (3) stands for different lengths for different people. The distribution of the number of months between the two interviews is displayed in Fig. 1. The variation in the distance between the two interview dates translates into variation in schooling respondents received between the first and second interviews. The average number of months of school attendance between the surveys (*Attendance*) is 50.[3]

In principle, a higher value of *Attendance* captures both the effect of schooling and the effect of aging. Thus, age at the second interview can be added as a control variable to isolate the impact of schooling. However, a variety of analyses, described in the Section on Instrumental Variables demonstrate that aging, in the range that exists in this data set, has no direct impact on health knowledge, or on other types of knowledge that are plausibly acquired in school. Moreover, age was never significant in any regression estimated, and inclusion of age did not alter other coefficients (Altindag, Cannonier, & Mocan, 2010). Therefore, we do not control for age in the regressions. As explained below, this allows us to estimate the models also using instrumental variables, where school attendance is instrumented by the distance between interviews.

The correlation between household income (household size) in 1997 and the length of time in months between the two survey waves is −0.01 (0.02). This confirms that distance between the two surveys is exogenous to household attributes. Given that school attendance between the two survey years is directly related to the time passed between these two survey years, much of the variation in *Attendance* is due to the variation in time between interviews. However, part of the variation in *Attendance* is due to college attendance. For example, school attendance will differ between two students who have the same distance between the interviews if one of them has a terminal high school degree, but the second one decided to go to college. To isolate the impact of attendance, we followed the students until the survey year 2007 to identify those who eventually went to college.[4] Thus, we created a sample that consists of students who attended college at some point in their lives until they were 24 years of age on average (in year 2007).[5] We name this sample "College-bound students". We also created a second sample that consists of students whose terminal degrees are high school. This sample, which is called "Eventual High School Graduates," includes those students who graduated from high school, but have not attended college as of 2007.[6]

3. Data

The main data set is constructed using information from the 1997 and 2002 waves of the National Longitudinal Survey of Youth 1997 cohort (NLSY97). The NLSY97 contains a nationally representative sample of 8984 youths who were aged 12–16 as of December 31st 1996. The respondents have been followed annually since the survey was initiated. In the first year of the survey, biological and/or residential

[3] The number of months in school is smaller than the number of months between the interview dates because the former does not include those months where the school is not in session, whereas the latter does. However, in a handful of cases the number of months in school is greater than the number of months between the interview dates because in these cases the corresponding students had dual enrollments (e.g. simultaneous enrollment in high school and college), and we counted enrollment in each month in each type of institution towards total school attendance.

[4] These include those respondents who were still in high school in the second survey (2002–2003), but subsequently graduated and attended college. If schooling information was missing in 2007, we utilized information reported in previous waves.

[5] This sample is similar in design to that used by Farrell and Fuchs (1982). They selected a group of students with at least 12 years of education at age 17, who acquired additional schooling (college education) until the age of 24. Farrell and Fuchs (1982) analyzed the impact of additional schooling between the ages of 17 and 24 on the propensity to smoke.

[6] This includes students who were in high school during the second survey in 2002–2003, graduated subsequently, but have not enrolled in college as of 2007.

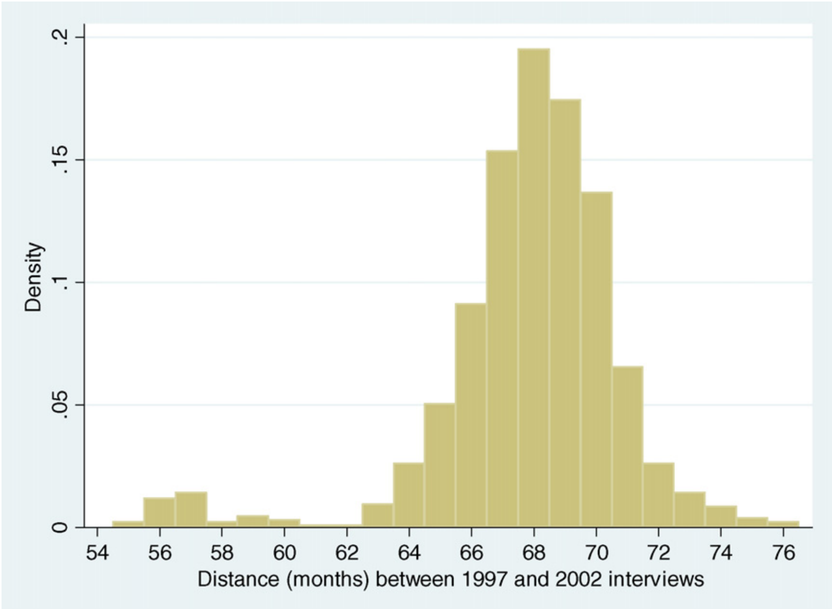


Fig. 1. Distribution of number of months between interview dates (college bound students and eventual high school graduates combined).

parents of the respondents were interviewed to provide supplemental information about their education and their knowledge of health-related issues.

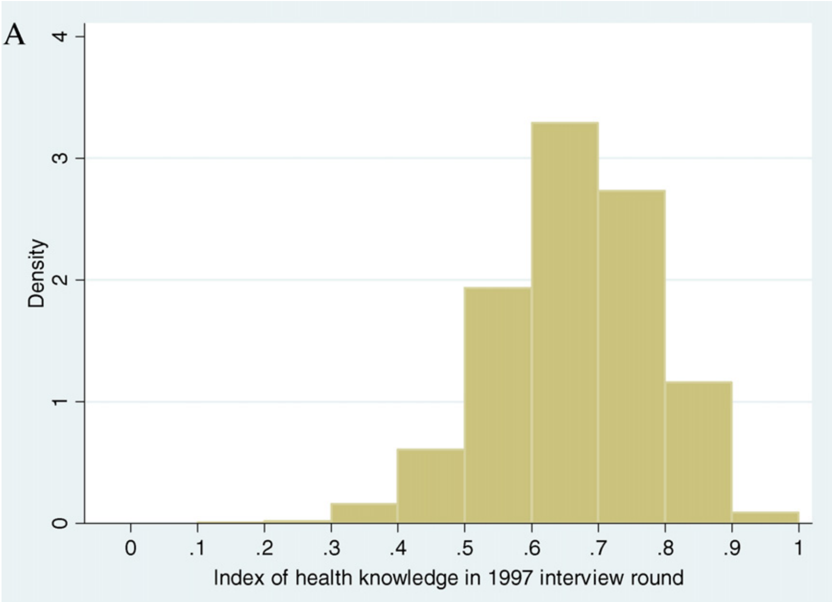
Key dependent variables that gauge the extent of individuals' health knowledge, are based on nine questions that were asked in both the 1997 and 2002 waves to the cohort born in 1983. These questions are listed in Appendix A. Table 1 presents the dependent variables that are created based on these nine health questions posed to the respondents in the college-bound sample. The descriptive statistics for the sample of eventual high school graduates are presented in Altindag et al. (2010). The variables take the value of one if the respondent answered the questions correctly, and zero if the answers were incorrect. For example, the respondents were asked if they believed that smoking one or more packs of cigarettes per day increased the risk of getting heart disease (*Smoking leads to Heart Disease*). Based on the Surgeon General's report, the correct answer to this question is yes. For each question, we referred to authoritative sources such as the American Heart Association, Surgeon General's Report, National Institute on Drug Abuse and American Liver Foundation to identify the correct answer. For the question regarding the timing of pregnancy, the correct answer is option *c* (about a week after the period begins), although we also considered both options *c* and *d* as correct. The latter variable is titled *Pregnancy Most Likely-2*.[7] As Table 1 reveals, some health questions, such as the impact of smoking on heart disease and the link between drinking and alcohol addiction have been answered correctly by most respondents. On the other hand, the proportion of correct answers is low for some other questions such as the link between drinking and heart disease. The proportion of correct responses is particularly low for pregnancy related questions.

The proportion of correct answers to the question about the relationship between drinking and heart disease is markedly lower in the 2002 survey in comparison to the 1997 survey. It is possible that such a decline in the rate of correct answers took place because some respondents started drinking between the survey years, and because of cognitive dissonance they denied in the second survey that drinking leads to heart disease. The NLSY97 contains a question about alcohol consumption. Because the health knowledge question pertains to drinking at least five drinks of alcohol once/twice a week, we identified individuals who consumed five or more drinks for at least four days a month (heavy drinkers). If cognitive dissonance is a major factor, we expect a strong correlation between the onset of heavy drinking and the propensity to change the answers between the two survey waves. Of the 250 college-bound individuals who answered correctly in the first wave but incorrectly in the second wave, 218 were not heavy drinkers in either wave. Only 32 were not heavy drinkers in the first wave, but became heavy drinkers in the second wave. In a more general context, we investigated the correlation between the change in the knowledge question and the change in heavy drinking between the survey years by running a regression where the dependent variable is the change in the correct response indicator to the drinking-heart disease question and the independent variables consist of three dichotomous indicators of whether the person was not a heavy drinker in either wave of the survey, whether he/she was a heavy drinker in the first

[7] We did not use the question that asks whether smoking one or more packs of cigarettes per day increases the risk of getting AIDS. This is because scientific evidence on the impact of smoking on faster progression to AIDS among HIV-infected individuals is mixed (Coates et al., 1990; Nieman, Fleming, Harris, & Mitchell, 1993). Furthermore, the question may have been misinterpreted as the impact of smoking on the risk of HIV infection, which is different from progressions to AIDS, conditional on HIV.

Table 1
Summary statistics of health knowledge questions. NLSY97 sample for 1997 and 2002 waves (college-bound students).

Variables	Variable definition	Means		
		Full sample ($N = 1969$)	1997 Sample ($N = 1003$)	2002 Sample ($N = 966$)
	(1)	(2)	(3)	(4)
Outcome variables:				
Smoking leads to heart disease	Correct response for whether smoking at least one pack of cigarette per day increases the risk of heart disease, and equal zero otherwise.	0.938	0.938	0.938
Drinking leads to heart disease	Correct response for whether having at least five drinks of alcohols once/twice per week increases the risk of heart disease and equal zero otherwise.	0.601	0.646	0.555
Drinking leads to liver damage	Correct response for whether having at least five drinks of alcohols once/twice per week increases the risk of liver damage and equal zero otherwise.	0.917	0.912	0.922
Drinking DOES NOT lead to arthritis	Correct response for whether having at least five drinks of alcohols once/twice per week DOES NOT increase the risk of getting arthritis, and equal zero otherwise.	0.778	0.742	0.815
Drinking leads to alcohol addiction	Correct response for whether having at least five drinks of alcohols once/twice per week increases the risk of alcohol addiction, and equal zero otherwise	0.928	0.940	0.916
Drinking harms unborn child	Correct response for whether having at least five drinks of alcohols once/twice per week increases the risk of harming an unborn child, and equal zero otherwise	0.972	0.979	0.964
Pill prevents pregnancy	Correct response for whether birth control pill is the most effective in preventing pregnancy, and equal zero otherwise	0.421	0.264	0.583
Condom prevents STD	Correct response for whether condom is the most effective in preventing STDs, and equal zero otherwise	0.772	0.639	0.910
Pregnancy most likely	Correct response for whether pregnancy is most likely to occur one week after the period begins, and equal zero otherwise	0.109	0.081	0.138
Pregnancy most likely-2	Correct response for whether pregnancy is most likely to occur one week after the period begins or two weeks after the period begins, and equal zero otherwise	0.255	0.160	0.352
Index of Health Knowledge	The proportion of correct answers in response to health knowledge questions (standard errors in parentheses)	0.716 (0.136)	0.684 (0.133)	0.749 (0.131)
Index of Health Knowledge – Alcohol	The proportion of correct answers in response to alcohol related health knowledge questions (standard errors in parentheses)	0.839 (0.165)	0.844 (0.160)	0.834 (0.171)
Index of Health Knowledge – Non alcohol	The proportion of correct answers in response to non-alcohol related health knowledge questions (standard errors in parentheses)	0.561 (0.216)	0.483 (0.212)	0.642 (0.188)



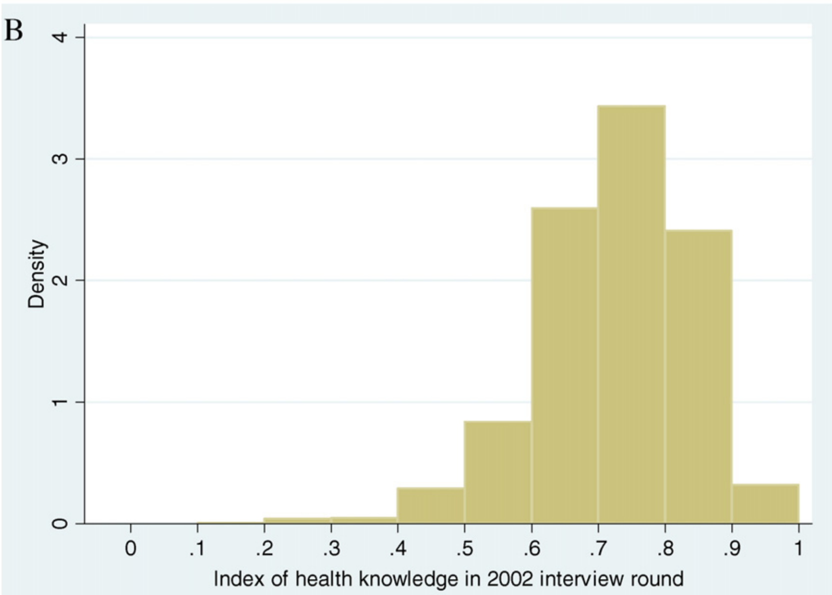


Fig. 2. (A) Distribution of the index of health knowledge during 1997 round of interviews (college-bound students). (B) Distribution of the index of health knowledge during 2002 round of interviews (college-bound students).

wave but not in the second wave, and whether the person was not a heavy drinker in the first wave but became a heavy drinker in the second wave. None of the dummy variables were statistically different from zero. This evidence suggests that there is no systematic correlation between the change in the drinking pattern and the propensity to answer this particular health question correctly.

We created an index of health knowledge questions, which is the percentage of correct answers to all nine questions.[8] This index, which is listed in Table 1, may represent the overall accuracy of health knowledge. The mean value of the index increases between the survey years, demonstrating an improvement in overall health knowledge. The distribution of the index is presented in Fig. 2A and B for both waves.

Although we have no evidence that the onset of heavy drinking between the two waves of the survey is correlated with the change in answers related to drinking, we cannot completely rule out this possibility. It is also possible that the amount of drinking specified in the questions is not considered as heavy drinking by the respondents, but instead it is considered as moderate drinking. If this is the case, given the information on the health benefits of moderate drink-

[8] We used *Pregnancy most likely* question as one of the nine questions.

ing (Mukamal et al., 2003; Rimm, Williams, Fosher, Criqui, & Stampfer, 1999) it is possible that the answers are noisy. Therefore, we also created an index of health knowledge, which excludes all drinking-related questions. This index, titled *Index of Health Knowledge – Non Alcohol*, consists of the four non-alcohol related questions: smoking and heart disease, pill and pregnancy, condom and STD, and timing of pregnancy. *Index of Health Knowledge – Alcohol* stands for the index that consists of five alcohol related questions.

Table 2 provides the descriptive statistics of the explanatory variables in the sample of individuals who eventually attended college. Schooling is measured using two different variables. *Highest Grade Completed* refers to the highest grade completed at the time of the survey. This variable may contain measurement error due to the fact that some respondents are interviewed after the end of the school year. Consider two students as an example. If the first student was interviewed in April of 1998 (when the school was in session) and the second one was interviewed in June of 1998 (after the semester is over), the latter student will report a value for "highest grade completed" which is one year larger than the former, although the difference in the amount of schooling is only two months. Such measurement error would generate attenuation bias in the estimated coefficient of *Highest grade completed*. To account for the timing of the survey, we created an indicator variable (*Aftersemester*) which takes the value of one if the individual was interviewed between June 1997 and May 1998 or between June and July 2003. Note that two students who are interviewed in different months of the same academic year would provide the same answer to the highest grade completed questions although the amount of schooling they have received (in months) would be different.[9] This second source of measurement error in the *Highest grade completed* variable cannot be controlled for.

As an alternative, and more accurate, measure of schooling we use *Attendance,* which is the cumulative number of months the individual has attended any type of school (Kindergarten to college) since the first interview in the 1997 wave.[10] This variable is created by using monthly schooling status information available in the schooling event history of each wave of the NSLY between 1997 and 2002. Consequently, *Attendance* measures schooling with more precision than *Highest grade completed.*

Five of the nine health knowledge questions were also posed to the parents of the respondents in the 1997 wave. These variables are also listed in Table 2. It is interesting to note that the percentage of correct answers is lower for parent answers.[11] In some specifications, we control for parent's health knowledge using these variables. Race, Hispanic ethnicity and gender indicators capture time-invariant characteristics of the respondents. They are included in Table 2 to provide information about the sample, although such characteristics will not be used in empirical analyses as they drop out due to time-differencing.

4. Results

Table 3A presents the results pertaining to the college-bound students sample. The coefficients are obtained from estimating models as depicted by Eq. (2), and robust standard errors are in parentheses. The variable $\Delta^k S_{it}$ in Eq. (2) measures the change in schooling between the two survey years. The primary measure of S_{it} is *Attendance*, which stands for the cumulative number of months the individual attended school since the first survey.[12] Thus, this variable identifies the number of months the respondent was in school between the two survey cycles.

In Table 3A schooling has an impact on health knowledge for the questions about the association between drinking and arthritis (column 4), and the questions about pregnancy listed in columns (7)–(10). An additional month of schooling increases the probability of answering these health questions correctly by 0.3–0.9 percentage points. Independent variables that measure the change in household income, household size, whether the respondent lives with at least one biological parent and whether the respondent lives in an urban area are not significant determinants of health knowledge. These variables may be noisy because they are constructed, like other independent variables, by taking the difference in relevant values between the two survey years. Transitory changes in 1997 or 2002, which may have temporarily altered the values of these variables from their permanent levels, would create measurement error in these variables. Thus, as an alternative measure we used the average values of these variables between the years 1997 and 2002. The results, which are not reported in the interest of space, were very similar to those reported in Table 3A.

For completeness, we also report the results obtained from the model where the change in schooling is measured by the change in *Highest Grade Completed*. The results, reported in Table 3B, show that in this specification schooling has no impact on health knowledge with the exception of the knowledge about the pill and the knowledge about the timing of pregnancy. Each additional grade completed is associated with an increase in the probability of answering the question on the pill (column 7) correctly by six percentage points, on the timing of pregnancy (column 10) by about seven percentage points. A comparison of these magnitudes with those reported in Table 3A shows that the effects associated with *Highest grade completed* are 7–11 times larger than those obtained from the models that employed *Attendance.* This is consistent with the fact that a typical school year is about 9 months long.

[9] An example is two students who have been interviewed in September and April of the same academic year.

[10] However, none of the individuals in our sample is in kindergarten or in primary school at the time of the first interview.

[11] Parents were asked five questions about the health impact of alcohol. The lower rate of accuracy of their responses to these questions may be a reflection of the fact that the parents were born in the 1950s and 1960s and grew up during a period where society's attitudes towards alcohol consumption were more tolerant. It is possible that early impressions about the impact of alcohol consumption persisted over time.

[12] In the 1997 wave of the survey, when this question was asked for the first time the possible answers are 0 months or 1 month.

Table 2
Summary statistics of personal attributes. NLSY97 sample for 1997 and 2002 waves (college-bound students).

Variables	Variable definition (1)	Mean (std. dev.) Full sample (N = 1969) (2)	Mean (std. dev.) 1997 Sample (N = 1003) (3)	Mean (std. dev.) 2002 Sample (N = 966) (4)
Highest grade completed	Years of schooling (highest grade completed).	9.577 (2.828)	6.900 (0.675)	12.356 (0.815)
Attendance	Months respondent attended high school or college since the first wave.	27.804 (27.931)	0.978 (0.147)	55.657 (8.151)
Age	Age in years.	16.140 (2.857)	13.366 (0.500)	19.021 (0.294)
Female	Equals one if female, zero otherwise.	0.542	0.542	0.542
Black	Equals one if Black and equal zero otherwise.	0.238	0.239	0.238
Hispanic	Equals one if Hispanic and equal zero otherwise.	0.177	0.177	0.177
Hours worked	Cumulative hours (in 000's) worked from age 14 at employee-type jobs.	1.219 (1.755)	0.004 (0.062)	2.485 (1.778)
Household size[a]	Number of individuals in respondent's household.	4.197 (1.554)	4.513 (1.428)	3.869 (1.611)
Household income[a]	Household income (in tens of thousands of US dollars) of respondent.	5.849 (4.825)	5.269 (3.892)	6.451 (5.571)
Urban[a]	Equals one if the respondent resides in an urban area, and equal zero otherwise.	0.750	0.756	0.744
Biological parent(s)[a]	Equals one if the respondent lives with at least one biological parent, and equal zero otherwise.	0.872	0.962	0.777
Household size[b]	Average number of people living in the respondent's household in the last five waves (1998–2002).			4.210 (1.290)
Household income[b]	Average income (in tens of thousands of US dollars) of the respondent's household in the last five waves (1998–2002).			6.549 (5.349)
Urban[b]	Proportion of the times (waves) the respondent has resided in an urban area during the last five waves (1998–2002).			0.762 (0.400)
Biological parent(s)[b]	Proportion of the times (waves) the respondent has lived with at least one biological parent during the last five waves (1998–2002).			0.907 (0.202)
Parent health knowledge – Drinking leads to heart disease	Correct response for whether respondent's parent thinks having at least five drinks of alcohols once/twice per week increases the risk of getting heart disease, and equal zero otherwise.		0.618	
Parent health knowledge – Drinking leads to liver damage	Correct response for whether respondent's parent thinks having at least five drinks of alcohols once/twice per week increases the risk of liver damage and equal zero otherwise.		0.784	
Parent health knowledge – Drinking DOES NOT lead to arthritis	Correct response for whether respondent's parent thinks having at least five drinks of alcohols once/twice per week DOES NOT increase the risk of getting arthritis, and equal zero otherwise.		0.589	
Parent health knowledge – Drinking leads to alcohol addiction	Correct response for whether respondent's parent thinks having at least five drinks of alcohol once/twice per week increases the risk of alcohol addiction and equal zero otherwise.		0.775	
Parent health knowledge – Drinking harms unborn child	Correct response for whether respondent's parent thinks having at least five drinks of alcohols once/twice per week increases the risk of harming an unborn child, and equal zero otherwise.		0.839	

Notes: Variables with missing values were converted to the mean value of each survey round. In some cases, the number of observations varies because of missing information.

[a] Descriptive statistics are calculated using the sample specified in the column header.

[b] Descriptive statistics are calculated using the last 5 waves.

Table 3A
The impact of school attendance on health knowledge: OLS regressions (college-bound students).

Variables	(1) Smoking leads to heart disease	(2) Drinking leads to heart disease	(3) Drinking leads to liver damage	(4) Drinking DOES NOT lead to arthritis	(5) Drinking leads to alcohol addiction	(6) Drinking harms the unborn child	(7) Pill prevents pregnancy	(8) Condom prevents STDs	(9) Pregnancy most likely	(10) Pregnancy most likely-2
Attendance	0.002	−0.004	−0.000	0.006***	−0.001	0.000	0.009***	0.005**	0.003*	0.006**
	(0.001)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.003)	(0.002)	(0.002)	(0.003)
Hours worked	0.000	−0.009	−0.006	0.004	0.004	−0.004	0.001	−0.004	−0.003	−0.020*
	(0.006)	(0.013)	(0.008)	(0.011)	(0.007)	(0.003)	(0.011)	(0.010)	(0.009)	(0.011)
Household income[a]	−0.000	−0.000	0.000	−0.000	−0.000	−0.000	0.000	0.000	0.000	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Household size[a]	0.007	−0.028*	0.010	−0.007	0.003	−0.004	−0.013	−0.012	0.010	0.014
	(0.007)	(0.016)	(0.009)	(0.014)	(0.008)	(0.006)	(0.015)	(0.013)	(0.011)	(0.014)
Urban[a]	0.028	−0.077	−0.002	−0.027	0.041	−0.005	0.043	−0.103*	0.049	0.077
	(0.043)	(0.075)	(0.038)	(0.053)	(0.035)	(0.023)	(0.069)	(0.062)	(0.052)	(0.069)
Biological parent(s)[a]	0.018	0.076	−0.032	−0.052	0.033	−0.006	−0.018	0.008	0.011	−0.033
	(0.027)	(0.058)	(0.031)	(0.049)	(0.030)	(0.022)	(0.053)	(0.045)	(0.036)	(0.053)
Observations	958	958	958	958	958	958	951	953	949	949
R-square	0.02	0.02	0.03	0.02	0.02	0.01	0.02	0.01	0.01	0.03

Notes: Robust standard errors are reported in parentheses.
[a] Household characteristics are measured as the change between the two survey years.
* Statistical level of significance: $p < 0.1$.
** Statistical level of significance: $p < 0.05$.
*** Statistical level of significance: $p < 0.01$.

Parents' health knowledge may directly influence the health knowledge of their offspring because parent health knowledge can be transmitted to the child at home. It is possible that such transmission of knowledge from parent to the child is stronger in households where parent supervision and parent-child contact is greater. If children from such households have better school attendance records, then the effect of schooling will be overestimated in models that do not control for parent health knowledge. However, as we argued earlier, because of the survey design, the potential change in schooling between the two survey years is independent of household attributes. Thus, controlling for parent health knowledge should not impact the estimated schooling coefficients in our models. Tables 4A and 4B display the results of the models where parent health knowledge is added as an additional control. This is a dichotomous indicator which takes the value of one if the parent who was asked the health question answered it correctly. As mentioned in the data section, only the five questions listed in Tables 4A and 4B were posed to the par-

Table 3B
The impact of highest grade completed on health knowledge: OLS regressions (college-bound students).

Variables	(1) Smoking leads to heart disease	(2) Drinking leads to heart disease	(3) Drinking leads to liver damage	(4) Drinking DOES NOT lead to arthritis	(5) Drinking leads to alcohol addiction	(6) Drinking harms the unborn child	(7) Pill prevents pregnancy	(8) Condom prevents STDs	(9) Pregnancy most likely	(10) Pregnancy most likely-2
Highest grade completed	0.017	−0.019	−0.006	0.032	−0.008	0.007	0.062**	0.033	0.030	0.068**
	(0.012)	(0.030)	(0.018)	(0.023)	(0.014)	(0.008)	(0.027)	(0.023)	(0.020)	(0.028)
Aftersemester	−0.003	0.031	−0.021	0.019	0.006	0.016	0.041	0.040	0.035	−0.006
	(0.026)	(0.054)	(0.028)	(0.045)	(0.026)	(0.016)	(0.050)	(0.043)	(0.034)	(0.050)
Hours worked	0.000	−0.008	−0.006	0.001	0.004	−0.004	−0.003	−0.006	−0.004	−0.021*
	(0.006)	(0.013)	(0.008)	(0.011)	(0.007)	(0.003)	(0.011)	(0.011)	(0.009)	(0.011)
Household income[a]	−0.000	−0.000	0.000	−0.000	−0.000	0.000	0.000	0.000	0.000	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Household size[a]	0.007	−0.028*	0.010	−0.008	0.003	−0.004	−0.013	−0.012	0.010	0.014
	(0.007)	(0.016)	(0.009)	(0.014)	(0.008)	(0.006)	(0.016)	(0.013)	(0.011)	(0.014)
Urban[a]	0.030	−0.083	−0.002	−0.018	0.040	−0.005	0.056	−0.097	0.054	0.085
	(0.044)	(0.075)	(0.038)	(0.053)	(0.036)	(0.023)	(0.070)	(0.063)	(0.051)	(0.070)
Biological parent(s)[a]	0.018	0.069	−0.030	−0.043	0.034	−0.008	−0.009	0.012	0.012	−0.034
	(0.027)	(0.058)	(0.031)	(0.049)	(0.031)	(0.022)	(0.054)	(0.046)	(0.036)	(0.052)
Observations	958	958	958	958	958	958	951	953	949	949
R-square	0.02	0.02	0.03	0.02	0.02	0.01	0.02	0.01	0.02	0.03

Notes: Robust standard errors are reported in parentheses.
[a] Household characteristics are measured as the change between the two survey years.
* Statistical level of significance: $p < 0.1$.
** Statistical level of significance: $p < 0.05$.

Table 4A
The impact of school attendance on health knowledge: OLS regressions with parent health knowledge (college-bound students).

Variables	(1) Drinking leads to heart disease	(2) Drinking leads to liver damage	(3) Drinking DOES NOT lead to arthritis	(4) Drinking leads to alcohol addiction	(5) Drinking harms the unborn child
Attendance	−0.004	0.000	0.006***	−0.000	0.000
	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)
Parent health knowledge in first survey	−0.011	−0.023	−0.085**	0.014	−0.113
	(0.053)	(0.064)	(0.040)	(0.046)	(0.089)
Hours worked	−0.009	−0.006	0.004	0.004	−0.004
	(0.013)	(0.008)	(0.011)	(0.007)	(0.003)
Household income[a]	−0.000	0.000	−0.000	−0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Household size[a]	−0.028*	0.010	−0.007	0.003	−0.005
	(0.016)	(0.009)	(0.014)	(0.008)	(0.006)
Urban[a]	−0.078	0.000	−0.024	0.042	−0.005
	(0.075)	(0.038)	(0.052)	(0.035)	(0.023)
Biological parent(s)[a]	0.075	−0.032	−0.044	0.034	−0.005
	(0.058)	(0.031)	(0.049)	(0.030)	(0.022)
Observations	958	958	958	958	958
R-square	0.02	0.03	0.03	0.02	0.01

Notes: Robust standard errors are reported in parentheses.
[a] Household characteristics are measured as the change between the two survey years.
* Statistical level of significance: $p < 0.1$.
** Statistical level of significance: $p < 0.05$.
*** Statistical level of significance: $p < 0.01$.

Table 4B
The impact of highest grade completed on health knowledge: OLS regressions with parent health knowledge (college-bound students).

Variables	(1) Drinking leads to heart disease	(2) Drinking leads to liver damage	(3) Drinking DOES NOT lead to arthritis	(4) Drinking leads to alcohol addiction	(5) Drinking harms the unborn child
Highest grade completed	−0.019	−0.006	0.036	−0.008	0.007
	(0.030)	(0.018)	(0.023)	(0.014)	(0.008)
Parent health knowledge in first survey	−0.010	−0.021	−0.086**	0.015	−0.113
	(0.053)	(0.063)	(0.040)	(0.046)	(0.089)
Aftersemester	0.028	−0.015	0.022	0.012	0.018
	(0.055)	(0.029)	(0.045)	(0.027)	(0.017)
Hours worked	−0.008	−0.006	0.002	0.004	−0.004
	(0.013)	(0.008)	(0.011)	(0.007)	(0.003)
Household income[a]	−0.000	0.000	−0.000	−0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Household size[a]	−0.028*	0.010	−0.007	0.003	−0.004
	(0.016)	(0.009)	(0.014)	(0.008)	(0.006)
Urban[a]	−0.083	0.000	−0.015	0.042	−0.005
	(0.074)	(0.038)	(0.053)	(0.036)	(0.023)
Biological parent(s)[a]	0.069	−0.030	−0.036	0.034	−0.006
	(0.058)	(0.031)	(0.049)	(0.031)	(0.022)
Observations	958	958	958	958	958
R-square	0.02	0.03	0.02	0.02	0.01

Notes: Robust standard errors are reported in parentheses.
[a] Household characteristics are measured as the change between the two survey years.
* Statistical level of significance: $p < 0.1$.
** Statistical level of significance: $p < 0.05$.

ents. Therefore, we are able to estimate only five of the nine models in this analysis.[13] Table 4A uses *Attendance* as the measure of schooling while Table 4B employs *Highest grade completed*.

The extent of parents' health knowledge has no systematic impact on the health knowledge of the child. Furthermore, as expected, controlling for parent health knowledge does not change the estimated impact of schooling on own health knowledge. For example, the coefficient of *Attendance* is 0.006 in the model that explains the probability of a correct answer for the link between drinking and arthritis in Table 4A (column 3); and it is also 0.006 in Table 3A (column 4), which is the

[13] Tables 4A and 4B employ the difference in household attributes. Using the average values of these variables over the years that span the survey years produced almost identical results.

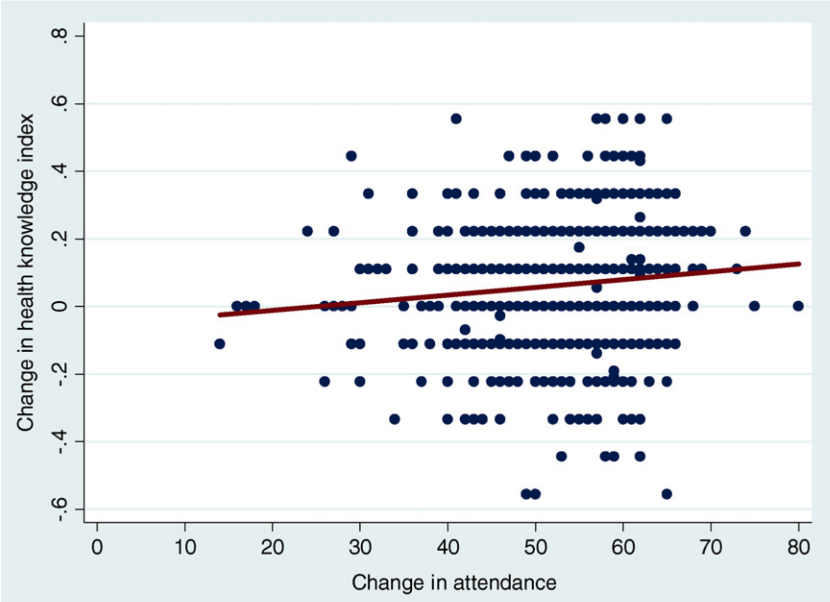


Fig. 3. The change in the health knowledge index versus change in attendance between the two survey waves (college-bound students).

same model with the exception of parent health knowledge variable. The same equivalence is seen between the coefficients of *Highest grade completed* reported in Tables 4B and 3B.

To present the results in a more concise fashion, we created an index where the proportion of correct answers to all health questions is calculated for each individual. The mean (standard deviation) of this index is 0.684 (0.133) in the first wave, and 0.749 (0.131) in the second wave, indicating that overall, there is an increase in the accuracy of the answers between the two survey years. The distributions are displayed in Fig. 2A and B. The change of the index between the survey years is plotted against the change in school attendance for college bound students in Fig. 3.[14] We also created two additional health index variables: one that consist of the five alcohol-related health questions that were also posed to the parents (Index of Health Knowledge – Alcohol), and another one that consists of non-alcohol related health questions (Index of Health Knowledge – Non-alcohol). Descriptive statistics of these indexes are displayed in Table 1.

We run the models with these three health knowledge indexes as dependent variables. The results are presented in Table 5. Column (1) displays the results where the overall health index is the dependent variable. Columns (2) and (3) display the results where the index based on alcohol related questions is the dependent variable. Because these particular questions were also posed to the parents, we created the same index for the parent answers and included it as an explanatory variable in the model reported in column (3). The results presented in column (4) pertain to the health index with non-alcohol questions. In columns (1) to (4) schooling is measured by *Highest grade completed*, and it is measured by *Attendance* in columns (5) to (8). Schooling has no impact on the index based on alcohol-related health questions, and the inclusion or exclusion of parent health information does not alter this result. On the other hand, the index based on non-alcohol questions is influenced by schooling, and the same is true for the overall index. An additional year of schooling increases the health index by about 1.6 percentage points, or about 2.3 percent.

We perform the same analyses using the high school graduate sample. As described earlier, this sample consists of individuals who graduated from high school but have not attended college until they are 24 years of age. Table 6A reports the results with *Attendance* as the measure of schooling. This table is the counterpart to Table 3A. Table 6B is the counterpart to 3B where highest grade completed is the measure of schooling. In Table 6A no coefficient is statistically significant with the exception of the smoking-heart disease question and the condom question, and the latter has the wrong sign. In Table 6B only one education coefficient is significant.

An interesting aspect of the results obtained from the high school graduates sample is that there is some evidence that the hours worked in the labor market has a positive impact on the propensity to answer some health questions correctly. This result makes sense to the extent that working for pay creates exposure to others and provides increased opportunities for gathering health knowledge from other individuals. The impact, however, is modest. An additional 1000 h worked between the two survey years increases the propensity for a correct answer by 2-4 per-

[14] The estimated slope is 0.002 (SE = 0.0007), suggesting that school attendance is positively related to an increase in the health knowledge index.

Table 5
The impact of highest grade completed and school attendance on health knowledge index: OLS regressions (college-bound students).

Variables	(1) Index Of Health Knowledge	(2) Index Of Health Knowledge – Alcohol	(3) Index Of Health Knowledge – Alcohol	(4) Index Of Health Knowledge – Non-alcohol	(5) Index Of Health Knowledge	(6) Index Of Health Knowledge – Alcohol	(7) Index Of Health Knowledge – Alcohol	(8) Index Of Health Knowledge – Non-alcohol
Highest grade completed	0.016**	0.001	0.001	0.035***				
	(0.007)	(0.009)	(0.009)	(0.011)				
Attendance					0.002***	0.000	0.000	0.005***
					(0.001)	(0.001)	(0.001)	(0.001)
Parent health knowledge in first survey			0.008				0.008	
			(0.047)				(0.047)	
Aftersemester	0.020	0.010	0.013	0.032				
	(0.014)	(0.017)	(0.017)	(0.022)				
Hours worked	−0.003	−0.003	−0.002	−0.003	−0.002	−0.002	−0.002	−0.001
	(0.003)	(0.004)	(0.004)	(0.005)	(0.003)	(0.004)	(0.004)	(0.005)
Household income[a]	−0.000	−0.000	−0.000	0.000	−0.000	−0.000	−0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Household size[a]	−0.004	−0.005	−0.005	−0.002	−0.004	−0.005	−0.005	−0.002
	(0.004)	(0.006)	(0.006)	(0.006)	(0.004)	(0.006)	(0.006)	(0.006)
Urban[a]	−0.003	−0.014	−0.013	0.011	−0.006	−0.014	−0.014	0.004
	(0.020)	(0.023)	(0.023)	(0.031)	(0.020)	(0.023)	(0.023)	(0.030)
Biological parent(s)[a]	0.005	0.004	0.004	0.007	0.003	0.004	0.004	0.003
	(0.015)	(0.020)	(0.020)	(0.022)	(0.015)	(0.020)	(0.020)	(0.022)
Observations	958	958	958	958	958	958	958	958
R-square	0.02	0.02	0.02	0.03	0.02	0.02	0.02	0.03

Notes: Robust standard errors are reported in parentheses.
[a] Household characteristics are measured as the change between the two survey years.
** Statistical level of significance: $p<0.05$.
*** Statistical level of significance: $p<0.01$.

centage points. This would roughly translate into working for about 25 h per week for two months each summer over the last five summers.

Tables 6A and 6B show that education is not related to health knowledge in the sample of high school graduates. In models that use *Attendance*, which is a more accurate measure of education, half of the estimated education coefficients have negative signs, and with the exception of smoking and condom use (where the latter has the wrong sign), they are all statistically insignificant. It should be pointed out that the sample size is small in these regressions, about 280 observations.

Tables 7A and 7B present the results from the models that include parent health knowledge. As men-

Table 6A
The impact of school attendance on health knowledge: OLS regressions (eventual high school graduates).

Variables	(1) Smoking leads to heart disease	(2) Drinking leads to heart disease	(3) Drinking leads to liver damage	(4) Drinking DOES NOT lead to arthritis	(5) Drinking leads to alcohol addiction	(6) Drinking harms the unborn child	(7) Pill prevents pregnancy	(8) Condom prevents STDs	(9) Pregnancy most likely	(10) Pregnancy most likely-2
Attendance	0.007**	0.001	−0.000	−0.003	−0.005	0.000	0.008	−0.011***	−0.003	−0.002
	(0.003)	(0.005)	(0.004)	(0.005)	(0.003)	(0.003)	(0.005)	(0.004)	(0.004)	(0.005)
Hours worked	0.024**	0.038*	0.006	−0.051***	−0.004	0.003	0.026	−0.017	0.023*	0.022
	(0.010)	(0.022)	(0.013)	(0.019)	(0.013)	(0.010)	(0.020)	(0.015)	(0.013)	(0.018)
Household income[a]	−0.000	0.000	0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Household size[a]	−0.015	0.015	−0.014	0.009	−0.006	0.002	0.001	0.017	0.015	−0.004
	(0.014)	(0.028)	(0.015)	(0.022)	(0.016)	(0.015)	(0.027)	(0.019)	(0.018)	(0.023)
Urban[a]	0.032	−0.111	−0.076	−0.003	−0.033	−0.021	−0.016	0.011	−0.005	−0.027
	(0.074)	(0.128)	(0.067)	(0.081)	(0.064)	(0.052)	(0.115)	(0.063)	(0.088)	(0.106)
Biological parent(s)[a]	0.055	−0.018	−0.006	0.131*	−0.078	0.035	−0.069	−0.111	0.003	−0.128
	(0.060)	(0.091)	(0.053)	(0.075)	(0.057)	(0.054)	(0.095)	(0.069)	(0.066)	(0.081)
Observations	281	281	281	281	281	281	279	279	280	280
R-square	0.11	0.04	0.06	0.08	0.08	0.01	0.03	0.08	0.03	0.03

Notes: Robust standard errors are reported in parentheses.
[a] Household characteristics are measured as the change between the two survey years.
* Statistical level of significance: $p<0.1$.
** Statistical level of significance: $p<0.05$.
*** Statistical level of significance: $p<0.01$.

Table 6B
The impact of highest grade completed on health knowledge: OLS regressions (eventual high school graduates).

Variables	(1) Smoking leads to heart disease	(2) Drinking leads to heart disease	(3) Drinking leads to liver damage	(4) Drinking DOES NOT lead to arthritis	(5) Drinking leads to alcohol addiction	(6) Drinking harms the unborn child	(7) Pill prevents pregnancy	(8) Condom prevents STDs	(9) Pregnancy most likely	(10) Pregnancy most likely-2
Highest grade completed	0.001	0.062	−0.008	0.016	0.008	0.025	0.016	−0.016	0.027	0.121**
	(0.043)	(0.061)	(0.043)	(0.054)	(0.036)	(0.035)	(0.063)	(0.043)	(0.047)	(0.054)
Aftersemester	−0.008	−0.330***	−0.024	0.091	−0.038	−0.041	−0.210**	−0.084	−0.043	−0.177*
	(0.069)	(0.104)	(0.073)	(0.082)	(0.061)	(0.063)	(0.106)	(0.073)	(0.082)	(0.102)
Hours worked	0.015	0.042**	0.006	−0.047***	0.003	0.004	0.019	−0.005	0.028**	0.032*
	(0.010)	(0.021)	(0.013)	(0.018)	(0.013)	(0.009)	(0.019)	(0.015)	(0.013)	(0.017)
Household income[a]	−0.000	0.000	0.000	−0.000	−0.000	−0.000	−0.000*	−0.000	−0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Household size[a]	−0.012	0.017	−0.015	0.010	−0.008	0.003	0.005	0.011	0.015	0.002
	(0.015)	(0.027)	(0.015)	(0.022)	(0.016)	(0.016)	(0.027)	(0.021)	(0.018)	(0.023)
Urban[a]	0.029	−0.107	−0.077	−0.000	−0.030	−0.019	−0.017	0.012	−0.002	−0.016
	(0.077)	(0.128)	(0.068)	(0.082)	(0.064)	(0.052)	(0.113)	(0.069)	(0.087)	(0.101)
Biological parent(s)[a]	0.056	−0.002	−0.004	0.123*	−0.077	0.036	−0.057	−0.104	0.003	−0.130
	(0.062)	(0.089)	(0.054)	(0.074)	(0.058)	(0.055)	(0.096)	(0.071)	(0.067)	(0.082)
Observations	281	281	281	281	281	281	279	279	280	280
R-square	0.09	0.07	0.06	0.08	0.07	0.01	0.04	0.06	0.03	0.05

Notes: Robust standard errors are reported in parentheses.
[a] Household characteristics are measured as the change between the two survey years.
* Statistical level of significance: $p < 0.1$.
** Statistical level of significance: $p < 0.05$.
*** Statistical level of significance: $p < 0.01$.

tioned earlier, only the five questions listed in the table were posed to the parents. As was the case with the sample of college-bound students, controlling for parent health knowledge does not alter the estimated coefficients.

The frequency distribution of the health index for the eventual high school graduates sample in the two surveys is plotted in Fig. 4. Fig. 5 displays the change in the index versus the change in attendance between the two surveys. Table 8 displays the results obtained from the models that use the health index as the dependent variable. Consistent with previous results, the estimated coefficient of education is statistically not different from zero. This is true for the overall health index in column (1) as well as the indexes

Table 7A
The impact of school attendance on health knowledge: OLS regressions with parent health knowledge (eventual high school graduates).

Variables	(1) Drinking leads to heart disease	(2) Drinking leads to liver damage	(3) Drinking DOES NOT lead to arthritis	(4) Drinking leads to alcohol addiction	(5) Drinking harms the unborn child
Attendance	0.001	−0.000	−0.003	−0.005	0.000
	(0.005)	(0.004)	(0.005)	(0.003)	(0.004)
Parent health knowledge in first survey	−0.041	−0.335***	−0.179**	0.085	0.090
	(0.106)	(0.112)	(0.083)	(0.107)	(0.133)
Hours worked	0.039*	0.004	−0.048**	−0.004	0.004
	(0.022)	(0.013)	(0.019)	(0.014)	(0.010)
Household income[a]	0.000	0.000	−0.000	−0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Household size[a]	0.015	−0.012	0.016	−0.008	0.001
	(0.029)	(0.015)	(0.023)	(0.015)	(0.015)
Urban[a]	−0.112	−0.056	−0.011	−0.028	−0.023
	(0.130)	(0.067)	(0.080)	(0.063)	(0.052)
Biological parent(s)[a]	−0.018	−0.038	0.131*	−0.073	0.036
	(0.093)	(0.054)	(0.076)	(0.057)	(0.054)
Observations	281	281	281	281	281
R-square	0.04	0.10	0.10	0.08	0.01

Notes: Robust standard errors are reported in parentheses.
[a] Household characteristics are measured as the change between the two survey years.
* Statistical level of significance: $p < 0.1$.
** Statistical level of significance: $p < 0.05$.
*** Statistical level of significance: $p < 0.01$.

Table 7B
The impact of highest grade completed on health knowledge: OLS regressions with parent health knowledge (eventual high school graduates).

Variables	(1) Drinking leads to heart disease	(2) Drinking leads to liver damage	(3) Drinking DOES NOT lead to arthritis	(4) Drinking leads to alcohol addiction	(5) Drinking harms the unborn child
Highest grade completed	0.060	−0.010	0.014	0.010	0.025
	(0.061)	(0.044)	(0.055)	(0.036)	(0.035)
Parent health knowledge in first survey	−0.036	−0.341***	−0.191**	0.091	0.087
	(0.104)	(0.111)	(0.083)	(0.108)	(0.133)
Aftersemester	−0.328***	−0.037	0.106	−0.038	−0.037
	(0.105)	(0.073)	(0.085)	(0.062)	(0.063)
Hours worked	0.042**	0.004	−0.045**	0.003	0.005
	(0.021)	(0.013)	(0.018)	(0.013)	(0.009)
Household income[a]	0.000	0.000	−0.000	−0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Household size[a]	0.018	−0.013	0.016	−0.010	0.003
	(0.027)	(0.015)	(0.022)	(0.015)	(0.016)
Urban[a]	−0.108	−0.056	−0.009	−0.026	−0.021
	(0.130)	(0.068)	(0.080)	(0.064)	(0.052)
Biological parent(s)[a]	−0.003	−0.035	0.123	−0.072	0.036
	(0.090)	(0.055)	(0.075)	(0.057)	(0.055)
Observations	281	281	281	281	281
R-square	0.07	0.10	0.11	0.07	0.02

Notes: Robust standard errors are reported in parentheses.
[a] Household characteristics are measured as the change between the two survey years.
** Statistical level of significance: $p < 0.05$.
*** Statistical level of significance: $p < 0.01$.

that consist of alcohol-related questions and non-alcohol-related questions.

To investigate the extent to which education acquired in high school has a different impact on health knowledge than education obtained in college, we used the sample of college-bound students and employed *Attendance – High School* and *Attendance – College* as two separate variables. These variables decompose *Attendance* into two components. More specifically, in Tables 9A and 9B *Attendance-High School* stands for the number of months the students attended high school between the first and second interview waves, and *Attendance-College* stands for the number of months of college attendance. As Table 9A demonstrates, college education has an impact on the arthritis question, on the question about the pill and on the question about the timing of pregnancy. High school attendance has a positive impact on the probability of answering correctly the smoking-heart disease question, and the question on the pill and the question on condom use. The results displayed in Table 9B show that both college attendance and high school attendance have statistically significant impacts on the index of non-alcohol related health knowledge and on the index of overall health knowledge. The coefficient of college attendance is larger than the coefficient of high school attendance, although the difference is not statistically significant in most cases.

4.1. Instrumental variables regressions

Although aging may be correlated with an increase in health knowledge, among the young adults who make up the data set, aging is not related to knowledge. When we ran the models displayed in Tables 3A–9B with the inclusion of age, the coefficient of age was never significant and the other coefficients remained intact. Some of these results are reported in Altindag et al. (2010). We regressed each health knowledge measure used in the paper on highest grade completed, age and the other control variables separately for the 1997 and 2002 samples. In nine of the 10 regressions in each sample, the coefficient of age was insignificant. In the case of smoking-heart attack question the coefficient was significant at the 10-percent level, but it was negative.

As a third exercise, we regressed an indicator for whether the student's grades were mostly As and Bs in the 9th grade on the age of the student in months as of the end of the academic year when he/she finished 9th grade (May). Age was not significant for either college-bound individuals or eventual high school graduates. When we used a dependent variable which takes the value of one if the student's grades were mostly As, the results did not change. Finally, an indicator variable for whether the student's grades were mostly As or Bs in the 8th grade is included as an additional control variable in the regressions to explain whether the student's grades were mostly As or Bs in the 9th grade. The coefficients of age remained insignificant for both college bound students and eventual high school graduates.

As a final check, we constructed a dichotomous variable that takes the value of one if the individual gave the correct answer to the following question when asked in 2007: "*Suppose you had $100 in a savings account and the interest rate was 2 percent per year. After 5 years, how much do you think you would have in the account if you left the money to grow: more than $102, exactly $102, or*

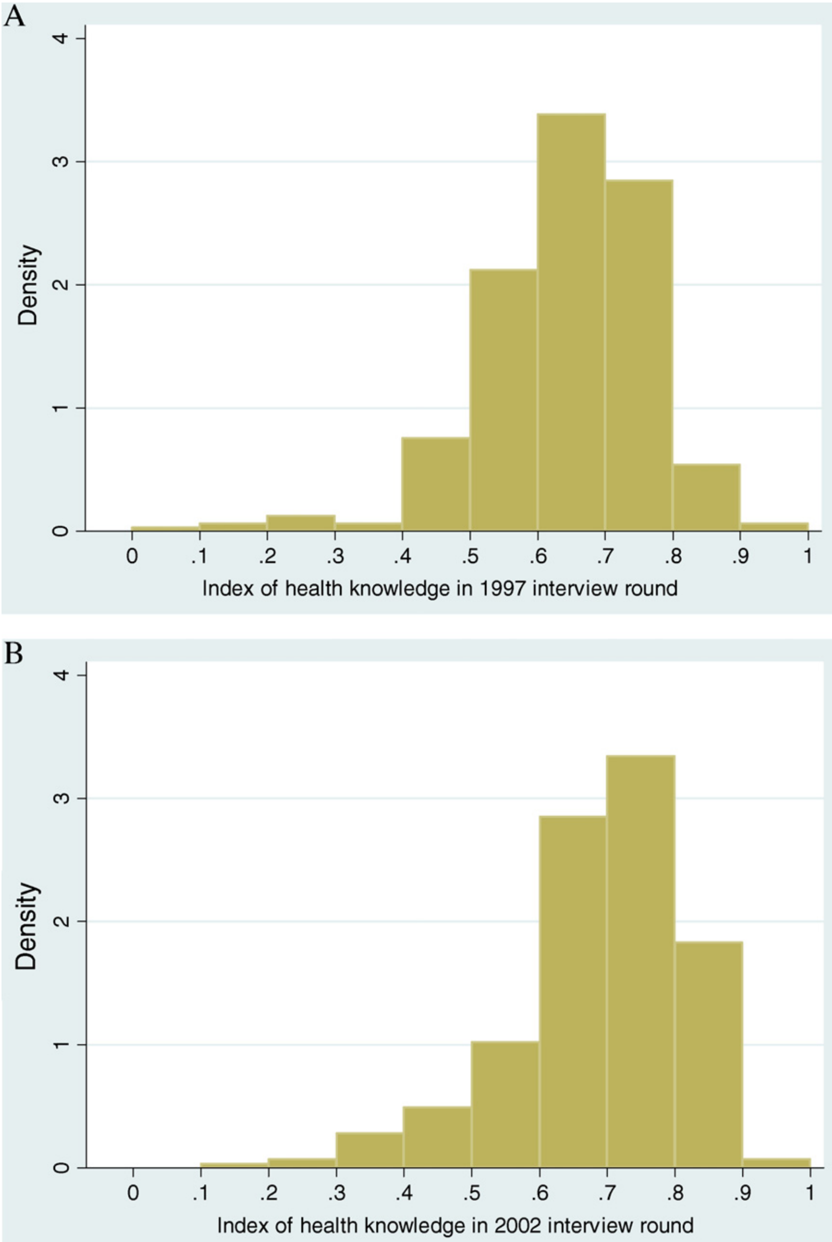


Fig. 4. (A) Distribution of the index of health knowledge during 1997 round of interviews (eventual high school graduates). (B) Distribution of the index of health knowledge during 2002 round of interviews (eventual high school graduates).

less than $*102?*" In a regression of this indicator variable on highest grade completed and age of the individual, the coefficient of age was insignificant. Excluding highest grade completed from this regression increased the size of the coefficient of age, but did not alter its insignificance. This body of evidence suggests that age has no direct impact on health knowledge, school grades, or basic math knowledge.

In this section we report instrumental variables regression results where *Attendance* is instrumented by the distance between the two interviews in months. As demonstrated earlier in the paper, the distance between two interviews is exogenous. Because *Attendance* is the number of months the student attended school between the two interview waves, *Attendance* depends on the distance between the two waves. Distance is also correlated with the age in the second wave, but given the evidence that aging is not related to the change in knowledge, distance is a candidate for a good instrument for *Attendance*.

Tables 10 and 11 present the results for college-bound students where attendance is instrumented by distance

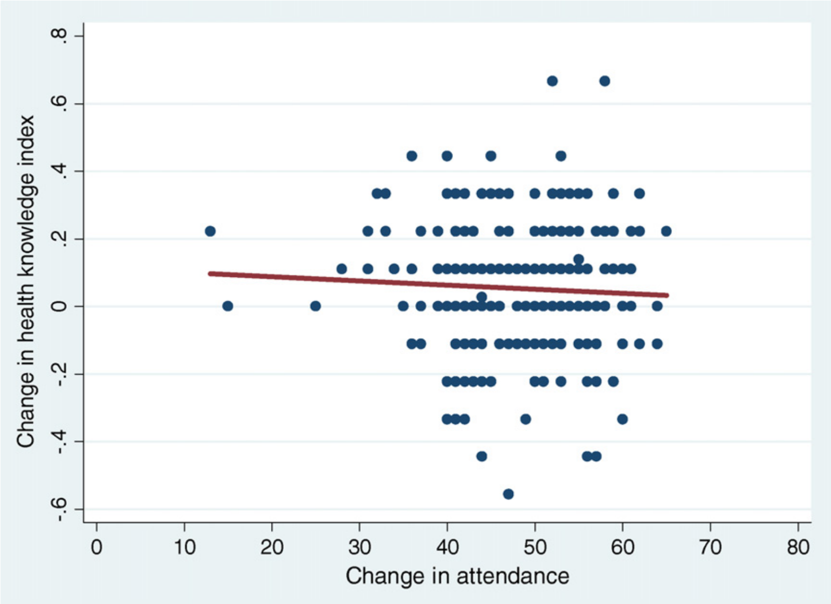


Fig. 5. The change in the health knowledge index versus change in attendance between the two survey waves (eventual high school graduates).

between interviews. As expected the first-stage has high explanatory power with *F*-statistics around 82. Of the 10 health knowledge regressions reported in Table 10, attendance is statistically significant in only one case. Table 11 shows that attendance does not impact alcohol-related health index, but it has a positive effect on the overall health index, which seems to be driven by non-alcohol-related health index.

Tables 12 and 13 present the results obtained from the instrumental variables regressions using the sam-

Table 8
The impact of highest grade completed and school attendance on health knowledge index: OLS regressions (eventual high school graduates).

Variables	(1) Index Of Health Knowledge	(2) Index Of Health Knowledge – alcohol	(3) Index Of Health Knowledge – alcohol	(4) Index Of Health Knowledge – non-alcohol	(5) Index Of Health Knowledge	(6) Index Of Health Knowledge – alcohol	(7) Index Of Health Knowledge – alcohol	(8) Index Of Health Knowledge – non-alcohol
Highest grade completed	0.013	0.021	0.020	0.007				
	(0.016)	(0.024)	(0.024)	(0.023)				
Attendance					−0.001	−0.001	−0.001	0.000
					(0.002)	(0.002)	(0.002)	(0.002)
Parent health knowledge in first survey			−0.089				−0.090	
			(0.095)				(0.094)	
Aftersemester	−0.075**	−0.068	−0.069*	−0.087**				
	(0.031)	(0.042)	(0.042)	(0.041)				
Hours worked	0.007	0.002	0.001	0.015**	0.005	−0.002	−0.002	0.014*
	(0.006)	(0.008)	(0.008)	(0.007)	(0.006)	(0.008)	(0.009)	(0.008)
Household income[a]	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Household size[a]	0.003	0.001	0.002	0.004	0.003	0.001	0.002	0.004
	(0.007)	(0.010)	(0.010)	(0.010)	(0.007)	(0.010)	(0.010)	(0.009)
Urban[a]	−0.024	−0.047	−0.045	0.003	−0.025	−0.049	−0.047	0.003
	(0.035)	(0.048)	(0.048)	(0.041)	(0.036)	(0.048)	(0.048)	(0.042)
Biological parent(s)[a]	−0.002	0.015	0.012	−0.024	−0.006	0.013	0.009	−0.029
	(0.026)	(0.035)	(0.035)	(0.034)	(0.026)	(0.035)	(0.035)	(0.034)
Observations	281	281	281	281	281	281	281	281
R-square	0.07	0.04	0.05	0.06	0.05	0.03	0.04	0.04

Notes: Robust standard errors are reported in parentheses.
[a] Household characteristics are measured as the change between the two survey years.
[*] Statistical level of significance: $p < 0.1$.
[**] Statistical level of significance: $p < 0.05$.

Table 9A
The impact of differential school attendance on health knowledge: OLS regressions (college bound students).

Variables	(1) Smoking leads to heart disease	(2) Drinking leads to heart disease	(3) Drinking leads to liver damage	(4) Drinking DOES NOT lead to arthritis	(5) Drinking leads to alcohol addiction	(6) Drinking harms the unborn child	(7) Pill prevents pregnancy	(8) Condom prevents STDs	(9) Pregnancy most likely	(10) Pregnancy most likely -2
Attendance – High school	0.003**	−0.004	−0.000	0.004	−0.001	−0.001	0.007**	0.006**	−0.000	0.003
	(0.001)	(0.003)	(0.002)	(0.003)	(0.002)	(0.001)	(0.003)	(0.003)	(0.002)	(0.003)
Attendance – College	0.001	−0.004	0.000	0.008***	0.000	0.001	0.011***	0.004	0.006***	0.009***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.001)	(0.001)	(0.003)	(0.003)	(0.002)	(0.003)
Hours worked	0.001	−0.009	−0.006	0.003	0.003	−0.005	−0.000	−0.003	−0.005	−0.021*
	(0.006)	(0.013)	(0.008)	(0.011)	(0.007)	(0.003)	(0.011)	(0.010)	(0.009)	(0.011)
Household income[a]	−0.000	−0.000	0.000	−0.000	−0.000	−0.000	0.000	0.000	0.000	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Household size[a]	0.007	−0.028*	0.010	−0.007	0.003	−0.004	−0.013	−0.012	0.010	0.014
	(0.007)	(0.016)	(0.009)	(0.014)	(0.008)	(0.006)	(0.016)	(0.013)	(0.011)	(0.014)
Urban[a]	0.029	−0.078	−0.002	−0.029	0.040	−0.007	0.041	−0.102	0.045	0.074
	(0.043)	(0.075)	(0.038)	(0.053)	(0.035)	(0.023)	(0.069)	(0.063)	(0.052)	(0.069)
Biological parent(s)[a]	0.016	0.076	−0.032	−0.050	0.034	−0.005	−0.016	0.007	0.014	−0.030
	(0.027)	(0.058)	(0.031)	(0.049)	(0.030)	(0.022)	(0.053)	(0.045)	(0.036)	(0.053)
High school = College (*p*-value)	0.169	0.929	0.849	0.209	0.435	0.057	0.266	0.423	0.009	0.116
Observations	958	958	958	958	958	958	951	953	949	949
R-square	0.02	0.02	0.03	0.02	0.02	0.01	0.03	0.01	0.02	0.03

Notes: Robust standard errors are reported in parentheses.
[a] Household characteristics are measured as the change between the two survey years.
* Statistical level of significance: $p < 0.1$.
** Statistical level of significance: $p < 0.05$.
*** Statistical level of significance: $p < 0.01$.

ple of eventual high school graduates. Attendance is not statistically significant in any of the health questions, and it is not significant in any of the health index regressions.

4.2. Extensions

Taking the difference in household attributes between the survey years may generate noise if their values have

Table 9B
The impact of differential school attendance on health knowledge index: OLS regressions (college-bound students).

Variables	(1) Index Of Health Knowledge	(2) Index Of Health Knowledge – alcohol	(3) Index Of Health Knowledge – alcohol	(4) Index Of Health Knowledge – non-alcohol
Attendance – High school	0.001*	−0.001	−0.001	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)
Attendance – College	0.003***	0.001	0.001	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Parent health knowledge in first survey			0.009	
			(0.047)	
Hours worked	−0.002	−0.003	−0.003	−0.002
	(0.003)	(0.004)	(0.004)	(0.005)
Household income[a]	−0.000	−0.000	−0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Household size[a]	−0.004	−0.005	−0.005	−0.002
	(0.004)	(0.006)	(0.006)	(0.006)
Urban[a]	−0.007	−0.015	−0.015	0.003
	(0.020)	(0.023)	(0.023)	(0.030)
Biological parent(s)[a]	0.004	0.005	0.005	0.004
	(0.015)	(0.020)	(0.020)	(0.022)
High school = College (*p*-value)	0.095	0.170	0.163	0.356
Observations	958	958	958	958
R-square	0.02	0.02	0.02	0.03

Notes: Robust standard errors are reported in parentheses.
[a] Household characteristics are measured as the change between the two survey years.
* Statistical level of significance: $p < 0.1$.
*** Statistical level of significance: $p < 0.01$.

Table 10
The impact of school attendance on health knowledge: Instrumental variables regressions (college bound students).

Variables	(1) Smoking leads to heart disease	(2) Drinking leads to heart disease	(3) Drinking leads to liver damage	(4) Drinking DOES NOT lead to arthritis	(5) Drinking leads to alcohol addiction	(6) Drinking harms the unborn child	(7) Pill prevents pregnancy	(8) Condom prevents STDs	(9) Pregnancy most likely	(10) Pregnancy most likely-2
Attendance	−0.004	0.002	0.000	0.012	−0.001	0.005	0.024***	0.010	0.008	0.008
	(0.004)	(0.009)	(0.004)	(0.008)	(0.004)	(0.003)	(0.009)	(0.007)	(0.006)	(0.008)
Hours worked	−0.003	−0.006	−0.006	0.007	0.004	−0.001	0.009	−0.001	−0.000	−0.018
	(0.007)	(0.014)	(0.008)	(0.012)	(0.007)	(0.004)	(0.012)	(0.011)	(0.009)	(0.012)
Household income[a]	−0.000	−0.000	0.000	−0.000	−0.000	−0.000	0.000	0.000	0.000	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Household size[a]	0.006	−0.027*	0.010	−0.006	0.003	−0.003	−0.010	−0.011	0.011	0.014
	(0.007)	(0.016)	(0.009)	(0.015)	(0.008)	(0.006)	(0.016)	(0.013)	(0.011)	(0.015)
Urban[a]	0.037	−0.087	−0.002	−0.037	0.042	−0.012	0.020	−0.112*	0.042	0.074
	(0.045)	(0.075)	(0.038)	(0.054)	(0.034)	(0.022)	(0.072)	(0.063)	(0.053)	(0.069)
Biological parent(s)[a]	0.035	0.059	−0.033	−0.069	0.036	−0.019	−0.061	−0.006	−0.003	−0.040
	(0.028)	(0.063)	(0.033)	(0.053)	(0.032)	(0.027)	(0.056)	(0.049)	(0.038)	(0.055)
First stage (*F*-statistic)	81.899	81.899	81.899	81.899	81.899	81.899	80.799	82.032	81.393	81.393
Observations	958	958	958	958	958	958	951	953	949	949

Notes: Robust standard errors are reported in parentheses. Instrument: Time between interviews between 1997 and 2002 interview waves.
[a] Household characteristics are measured as the change between the two survey years.
* Statistical level of significance: $p < 0.1$.
** Statistical level of significance: $p < 0.05$.
*** Statistical level of significance: $p < 0.01$.

been temporarily altered in 1997 and/or 2002 from the values that would have been observed in the absence of these temporary shocks. As an alternative measure we used the average values of these variables between the years 1997 and 2002. The results did not change. When the models that employed attendance are re-estimated using linear and quadratic terms of *Attendance*, no clear pattern emerged. Most coefficients were insignificant as before. In some cases, the linear coefficients turned negative.

We also estimated the specification depicted by Eq. (3) using OLS. Selected regressions are displayed in Altindag et al. (2010). The results are very similar to those obtained from Eq. (2), presented in the paper. Although the proportion of correct answers is high in some cases, estimating these models with probit generated marginal effects which were very similar to those obtained from linear probability models. The same result is obtained when we estimated the alternative specification displayed in Eq. (3) by instrumental variables (see Altindag et al., 2010).

To entertain the possibility that the impact of schooling on health knowledge differs between males and females, we re-estimated the models by adding an interaction

Table 11
The impact of school attendance on health knowledge index: instrumental variables regressions (college bound students).

Variables	(1) Index Of Health Knowledge	(2) Index Of Health Knowledge – alcohol	(3) Index Of Health Knowledge – alcohol	(4) Index Of Health Knowledge – non-alcohol
Attendance	0.006**	0.004	0.004	0.010**
	(0.003)	(0.003)	(0.003)	(0.004)
Parent health knowledge in first survey			0.002	
			(0.046)	
Hours worked	0.000	−0.000	0.000	0.001
	(0.003)	(0.005)	(0.005)	(0.005)
Household income[a]	−0.000	−0.000	−0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Household size[a]	−0.003	−0.005	−0.005	−0.001
	(0.004)	(0.006)	(0.006)	(0.006)
Urban[a]	−0.012	−0.019	−0.020	−0.004
	(0.020)	(0.023)	(0.023)	(0.031)
Biological parent(s)[a]	−0.008	−0.005	−0.007	−0.011
	(0.016)	(0.022)	(0.022)	(0.024)
First stage (*F*-statistic)	81.899	81.899	82.996	81.899
Observations	958	958	958	958

Notes: Robust standard errors are reported in parentheses. Instrument: Time between interviews between 1997 and 2002 interview waves.
[a] Household characteristics are measured as the change between the two survey years.
** Statistical level of significance: $p < 0.05$.

Table 12
The impact of school attendance on health knowledge: instrumental variables regressions (eventual high school graduates).

Variables	(1) Smoking leads to heart disease	(2) Drinking leads to heart disease	(3) Drinking leads to liver damage	(4) Drinking DOES NOT lead to arthritis	(5) Drinking leads to alcohol addiction	(6) Drinking harms the unborn child	(7) Pill prevents pregnancy	(8) Condom prevents STDs	(9) Pregnancy most likely	(10) Pregnancy most likely-2
Attendance	−0.000	−0.023	−0.006	0.028	0.002	0.004	0.006	−0.007	0.001	0.002
	(0.013)	(0.023)	(0.016)	(0.023)	(0.011)	(0.012)	(0.030)	(0.018)	(0.017)	(0.024)
Hours worked	0.014	0.008	−0.001	−0.015	0.005	0.007	0.025	−0.012	0.028	0.028
	(0.018)	(0.036)	(0.022)	(0.034)	(0.021)	(0.017)	(0.041)	(0.027)	(0.023)	(0.032)
Household income[a]	0.000	0.000	0.000	−0.000	0.000	−0.000	−0.000	−0.000	−0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Household size[a]	−0.018	0.024	−0.010	−0.005	−0.013	0.008	0.006	0.019	0.008	−0.010
	(0.014)	(0.031)	(0.016)	(0.024)	(0.016)	(0.014)	(0.028)	(0.020)	(0.018)	(0.025)
Urban[a]	−0.001	−0.133	−0.070	0.029	−0.032	0.009	−0.012	0.020	−0.002	−0.021
	(0.072)	(0.134)	(0.071)	(0.094)	(0.063)	(0.045)	(0.119)	(0.067)	(0.092)	(0.111)
Biological parent(s)[a]	0.034	−0.068	−0.042	0.149[*]	−0.094[*]	0.011	−0.074	−0.121[*]	0.034	−0.096
	(0.058)	(0.096)	(0.053)	(0.080)	(0.052)	(0.051)	(0.095)	(0.071)	(0.067)	(0.084)
First stage (*F*-statistic)	13.678	13.678	13.678	13.678	13.678	13.678	13.952	14.136	13.893	13.893
Observations	266	266	266	266	266	266	264	264	265	265

Notes: Robust standard errors are reported in parentheses. Instrument: Time between interviews between 1997 and 2002 interview waves.
[a] Household characteristics are measured as the change between the two survey years.
[*] Statistical level of significance: $p < 0.1$.

Table 13
The impact of school attendance on health knowledge index: instrumental variable regressions (eventual high school graduates).

Variables	(1) Index Of Health Knowledge	(2) Index Of Health Knowledge – alcohol	(3) Index Of Health Knowledge – alcohol	(4) Index Of Health Knowledge – non-alcohol
Attendance	0.001	0.001	0.001	0.001
	(0.006)	(0.008)	(0.008)	(0.010)
Parent health knowledge in first survey			−0.088	
			(0.092)	
Hours worked	0.007	0.001	0.000	0.015
	(0.010)	(0.013)	(0.014)	(0.014)
Household income[a]	−0.000	0.000	0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Household size[a]	0.002	0.001	0.002	0.003
	(0.008)	(0.010)	(0.010)	(0.010)
Urban[a]	−0.021	−0.039	−0.039	−0.000
	(0.036)	(0.048)	(0.048)	(0.044)
Biological parent(s)[a]	−0.018	−0.009	−0.012	−0.031
	(0.024)	(0.032)	(0.032)	(0.034)
First stage (*F*-statistic)	13.678	13.678	13.110	13.678
Observations	266	266	266	266

Notes: Robust standard errors are reported in parentheses. Instrument: Time between interviews between 1997 and 2002 interview waves.
[a] Household characteristics are measured as the change between the two survey years.

variable between attendance and a dichotomous variable to indicate if the person is female. The results, presented in Altindag et al. (2010), demonstrate that the impact of schooling is greater for females, and in some of the alcohol-related questions, such as alcohol-addiction and alcohol-liver damage, education has an impact on the health knowledge for females but not for males.

5. Conclusion

There exists a large body of empirical evidence pointing to a causal effect of education on health (Chou et al., 2010; Conti et al., 2010; Lleras-Muney, 2005). However, the exact mechanism behind the impact of education on health is less clear. Standard theoretical models of health capital, which trace back to Grossman (1972a) predict productive efficiency of education where an increase in schooling improves the efficiency of health production by raising the marginal products of inputs. More schooling could also enable people to understand better the exact nature of a health production function. Thus, in this "allocative efficiency" approach, education improves health knowledge, which in turn translates into a better choice of health inputs.

In this paper we perform a direct test of the allocative efficiency hypothesis where we analyze the impact of education on health knowledge. Health knowledge is measured by the probability of providing the correct answer to nine questions, ranging from the link between drinking and heart disease to the most effective method to prevent a pregnancy. We use data from the 1997 and 2002 waves of the NLSY97, and take advantage of the fact that the respondents were asked the same health questions in both waves. Because the time difference between the two survey years differs between participants, who were in the age range of 13 to 15 in the 1997 wave, participants differ in the amount of schooling they accumulated between the two survey years.

In the sample of individuals who eventually went to college, instrumental variables regressions reveal that education has a statistically significant impact on only one of the nine health knowledge questions. Education has a positive impact on the index of health knowledge, which consists of the proportion of health questions that are answered correctly. It has a positive impact on non-alcohol related health questions, but not on the index that consists of alcohol-related health questions. These effects are weaker for males.

The relationship between education and health knowledge is even weaker in a sample that consists of individuals whose completed schooling is 12 years (those who graduated from high school but have not attended college), where neither OLS nor IV regressions reveal a significant impact of education on health knowledge. Thus, consistent with Kenkel (1991), the results imply that allocative efficiency is not likely to be the main reason for why education improves health.

Appendix A. Correct answers to NLSY 97 health knowledge questions to respondent

1. Does smoking one or more packs of cigarettes per day, INCREASE THE RISK (chance) of getting heart disease?
   Correct answer: Yes
   Sources:
   - 1990 Surgeon General Report[15]
   - American Heart Association, http://www.americanheart.org/presenter.jhtml?identifier=4545 (accessed December 30, 2009)

2. Does having 5 or more drinks of alcohol once or twice each week, INCREASE THE RISK (chance) of damaging the liver?
   Correct Answer: Yes
   Sources:
   - 1988 Surgeon General Report.[16] "Excessive use of alcohol is also associated with liver disease."
   - American Liver Foundation, http://www.liverfoundation.org/education/info/alcohol/ (accessed December 30, 2009)

3. Does having 5 or more drinks of alcohol once or twice each week, INCREASE THE RISK (chance) of getting heart disease?
   Correct answer: Yes
   Sources:
   - 1988 Surgeon General Report[16]
   - American Heart Association, http://www.americanheart.org/presenter.jhtml?identifier=4488 (accessed December 30, 2009); Cardiovascular Institute of the South, http://www.medhelp.org/general/alcohol.HTM (accessed December 30, 2009)

4. Does having 5 or more drinks of alcohol once or twice each week, INCREASE THE RISK (chance) of getting arthritis?
   Correct answer: No
   Sources:
   - Voight et al. (1994) find that "Post menopausal women who averaged more than 14 alcoholic drinks per week had a reduced risk of rheumatoid arthritis." (p. 525) [17]
   - Science Daily, http://www.sciencedaily.com/releases/2007/06/070615110218.htm (accessed December 30, 2009)

5. Does having 5 or more drinks of alcohol once or twice each week, INCREASE THE RISK (chance) of becoming addicted to alcohol?
   Correct answer: Yes
   Sources:
   - 1988 Surgeon General Report[16]
   - American Heart Association, http://www.americanheart.org/presenter.jhtml?identifier=4488, http://www.americanheart.org/presenter.jhtml?identifier=4422 (accessed December 30, 2009)

6. Does having 5 or more drinks of alcohol once or twice each week, INCREASE THE RISK (chance) of harming an unborn child?
   Correct answer: Yes
   Sources:
   - 1988 Surgeon General Report[16]
   - American Heart Association, http://www.americanheart.org/presenter.jhtml?identifier=3017032 (accessed December 30, 2009)

[15] See US Department of Health and Human Services (1990).
[16] See US Department of Health and Human Services (1988).
[17] See Lynda, Koepsell, Lee Nelson, Dugowson, and Daling (1994).

7. Which of these same three methods is the most effective for preventing sexually transmitted diseases (STD) like AIDS or gonorrhea:
   a. Withdrawal
   b. Condom
   c. Birth control pill

   Correct answer: Condom
   Source: American Pregnancy Association, http://www.americanpregnancy.org/preventingpregnancy/malecondom.html (accessed December 30, 2009)

8. Which of these three is the most effective for preventing pregnancy?
   a. Withdrawal
   b. Condom
   c. Birth control pill

   Correct answer: Birth control pill
   Source: American Pregnancy Association, http://www.americanpregnancy.org/preventingpregnancy/birthcontrolpills.html (accessed December 30, 2009)

9. When during the female monthly cycle of menstrual periods is pregnancy most likely to occur?
   a. Right before the period begins
   b. During the period
   c. About a week after the period begins
   d. About two weeks after the period begins
   e. Anytime during the month, makes no difference
   f. Don't know

   Correct answer: About a week after the period begins
   ". . .your fertility period would be from the 8th day of your cycle to the 21st day of your cycle."
   Source: American Pregnancy Association, http://www.americanpregnancy.org/preventingpregnancy/fertilityawarenessNFP.html (accessed December 30, 2009)

References

Aizer, A., & Stroud, L. (2010). Education, knowledge and the evolution of disparities in health. *NBER Working Paper 15840*. http://papers.nber.org/papers/w15840 Accessed 10.05.2010.

Altindag, D., Cannonier, C., & Mocan, N. (2010). The impact of education on health knowledge. *NBER Working Paper No: 16422*.

Berger, M. C., & Leigh, P. J. (1989). Schooling, self-selection, and health. *Journal of Human Resources*, *24*(3), 433–455.

Chou, S.-Y., Liu, J.-T., Grossman, M., & Joyce, T. (2010). Parental education and child health: Evidence from a natural experiment in Taiwan. *American Economic Journal: Applied Economics*, *2*(1), 33–61.

Coates, R. A., Farewell, V. T., Raboud, J., Read, S. E., Macfadden, D. K., Calzavara, L. M., et al. (1990). Cofactors of progression to acquired immunodeficiency syndrome in a chart of male sexual contacts of men with human immunodeficiency virus disease. *American Journal of Epidemiology*, *132*(4), 717–722.

Conti, G., Heckman, J., & Urzua, S. (2010). The education-health gradient. *American Economic Review*, *100*(2), 234–238.

Currie, J., & Moretti, E. (2003). Mother's education and the intergenerational transmission of human capital: Evidence from college openings. *Quarterly Journal of Economics*, *118*, 1495–1532.

Farrell, P., & Fuchs, V. R. (1982). Schooling and health: The cigarette connection. *Journal of Health Economics*, *1*, 217–230.

Grossman, M. (1972a). On the concept of health capital and the demand for health. *Journal of Political Economy*, *80*(2), 223–255.

Grossman, M. (1972b). *The demand for health: A theoretical and empirical investigation*. New York: Columbia University Press for the National Bureau of Economic Research.

Grossman, M. (2008). The relationship between health and schooling. *Eastern Economic Journal*, *34*, 281–292.

Grossman, M. (2006). Education and non-market outcomes. In H. Eric, & W. Finis (Eds.), *Handbook of economics of education* (pp. 577–633). Amsterdam: Elsevier.

Grossman, M. (2000). A. J. Culyer, & J. P. Newhouse (Eds.), *The human capital model in handbook of health economics* (pp. 347–408). Amsterdam: North-Holland.

Kenkel, D. (1991). Health behavior, health knowledge, and schooling. *Journal of Political Economy*, *99*(2), 287–305.

Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States. *Review of Economic Studies*, *72*(1), 189–221.

Mukamal, K. J., Conigrave, K. J., Mittleman, M. A., Camargo, C. A., Stampfer, M. J., Willett, W. C., et al. (2003). Roles of drinking pattern and type of alcohol consumed in coronary heart disease in men. *The New England Journal of Medicine*, *348*, 109–118.

Nieman, R. B., Fleming, J., Harris, J. R., & Mitchell, D. M. (1993). The effect of cigarette smoking on the development of AIDS in HIV-1-seropositive individuals. *AIDS*, *7*(5), 705–710.

Price, J., & Simon, K. (2009). Patient education and the impact of new medical research. *Journal of Health Economics*, *28*(6), 1166–1174.

Rimm, E. B., Williams, P., Fosher, K., Criqui, M, & Stampfer, M. J. (1999). Moderate alcohol intake and lower risk of coronary heart disease: Meta-analysis of effects on lipids and haemostatic factors. *British Medical Journal*, *319*(1), 1523–1528.

Rosenzweig, M. R., & Schultz, P. T. (1982). In V. R. Fuchs (Ed.), *The behavior of mothers as inputs to child health: The determinants of birth weight, gestation and rate of fetal growth in economic aspects of health* (pp. 53–92). Chicago: University of Chicago Press for the National Bureau of Economic Research.

US Department of Health and Human Services. (1990). *The health benefits of smoking cessation: A report of the surgeon general*. Rockville, MD: U.S. Department of Health and Human Services, Public Health Service, Center for Disease Control, Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. Accessed 19.05.2010. http://profiles.nlm.nih.gov/NN/B/B/C/T/_/nnbbct.pdf

US Department of Health and Human Services. (1988). *The surgeon general's report on nutrition and health*. Washington, DC: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control. Accessed 19.05.2010. http://profiles.nlm.nih.gov/NN/B/C/Q/G/_/nnbcqg.pdf

Voight, L. F., Koepsell, T. D., Lee Nelson, J., Dugowson, C. E., & Daling, J. R. (1994). Smoking, Obesity, Alcohol Consumption, and the Risk of Rheumatoid Arthritis. *Epidemiology*, *88*, 525–532.