

Schooling, Self-Selection, and Obesity

Schooling and health have long been known to be positively correlated. There is, however, considerable debate surrounding explanations for this empirical association. In this paper, the effect of schooling on the odds that an individual is obese is estimated. Particular attention is given to the issue of self-selection bias and the omission of "third" variables measuring risk/time preferences. The results in this paper indicate that the direct effect of schooling on obesity remains even with the correction of self-selection bias and risk/time preferences in the model. However, the results suggest that the magnitude of the schooling effect is sensitive to the adjustments made for self-selection bias.

Rodolfo M. Nayga, Jr., Texas A&M University¹

Introduction

A large number of health economics studies have established a strong relation between schooling and better health outcomes. This relation has been called "one of the strongest generalizations to emerge from empirical research on health in the United States" (Farrell & Fuchs, 1982). These studies have been so persuasive in their demonstration of the importance of schooling that medical researchers are increasingly emphasizing lack of schooling as a "risk factor" in a variety of diseases (Leigh, 1990). Yet there is little agreement among economists concerning the mechanisms through which schooling enhances health. Some have argued for a direct effect whereby schooling increases the allocative efficiency of household health production and, therefore, due to schooling people may have better knowledge about the relation between certain health production inputs and health outputs (Grossman, 1975; Grossman & Joyce, 1987; Rosen & Taubman, 1982; Berger & Leigh, 1989). Others assert that the estimated relation between schooling and health outcomes may be due to the relation between health outcomes and schooling with the same unobservables (Fuchs, 1982; Farrell & Fuchs, 1982). So instead of schooling causing better health outcomes, some unobservable factors or "third" variables may be related to both schooling and better health outcomes.

Several economic studies have tested the theories presented above. Examples of such studies relate to the consumption of alcohol, cigarettes, and exercise (Kenkel, 1991), smoking (Sander, 1995), disability and functional limitations as well as systolic and diastolic blood pressures (Berger & Leigh, 1989), and seat belt use (Leigh, 1990). No other known published study, however, has focused on the relationship between schooling and a very important public health issue, obesity. Obesity is now one of the most important health problems confronting Americans (Centers for Disease Control, 1997).

Empirical Framework and Data

The objective of this empirical work is to revisit the relationship between schooling and health by focusing on a relatively neglected but very important health issue, obesity. The purpose of this study also is to determine whether the statistical association between schooling and obesity is affected when selectivity bias and time/risk preferences among individuals are taken into account in the model.

Self-selection bias arises because it is only possible to observe individuals making optimal choices involving, in this case, level of schooling and obesity. This process truncates the underlying error terms of the equations so the sample of individuals who make each choice is nonrandom (Berger & Leigh, 1989). Heckman (1983), Lee (1978), and Willis and Rosen (1979) have developed self-selection models applied to choices for two outcomes. Garen (1984) developed and applied the theory to continuous choice variables (i.e., years of schooling). Garen's methodology is used in this paper.

The obesity model can be expressed as:

$$O_i = b_1S + b_2X + b_3Z + \epsilon_i \quad (1)$$

where O_i is obesity variable given a value of 1 if individual is obese and zero otherwise, S is years of schooling, X is a vector of variables which affect both schooling and the likelihood of being obese, and Z represents risk/time

preference variables which affect likelihood of being obese. The ϵ is an error term reflecting the effects of unobservables on the likelihood of being obese.

The variables in X include ethnicity, income, age, household size, employment status, gender, regional variables, and urbanization. The Z variables include factors that may reflect health related risk or time preferences of individuals. These variables are:

Pref1 -- whether an individual strongly believed that some people are born to be fat or thin and that there is not much one can do to change it (=1, 0 otherwise)

Pref2 -- whether it is very important to the individual to maintain a healthy weight (=1, 0 otherwise),

Pref3 -- whether the individual has heard of any health problems caused by being overweight, (=1, 0 otherwise), and

Pref4 -- whether the individual often uses the nutrition panel on food labels that tells the amount of calories, protein, fat, and such in a serving of food (=1, 0 otherwise).

Findings from previous studies on the effect of schooling on health imply that b_1 is negative and significant. It is unclear if this finding represents a direct influence of schooling on health. Equation (1), however, do not correct for selectivity bias. In order to allow for these effects, a self-selection model (Garen, 1984) is adapted in this paper using a maximum likelihood heteroskedasticity probit model. The explanatory variables included in the equation are fairly standard in health economics studies (Feldstein, 1983; Kenkel, 1990).

The data set used in this study is the 1994 Diet and Health Knowledge Survey (DHKS) from the U.S. Department of Agriculture. The target individuals in this survey were randomly selected from among eligible 1994 Continuing Survey of Food Intakes by Individuals (CSFII) sample persons 20 years of age and older. The exclusion of younger individuals in the sample is not deemed inappropriate for the present study because they are less likely to have completed their schooling (Sander, 1995). Data in this survey were collected by computer assisted telephone interviews (in-person interviews for those without telephones). A total of 1,879 individuals participated in the survey. Due to incomplete data in some of the variables, 1,546 observations were used in the analysis.

Basic Results

The probit results of the principal variables explaining the odds of being obese are exhibited in Table 1. In the interest of brevity, the discussion of the results will be limited to the principal variables that are the focus of the paper: schooling, schooling residuals, and risk/time preferences. Column (1) corresponds to the first probit regression in which no adjustments are made for self-selection bias and risk/time preferences. The estimated coefficient on schooling is -0.055 and is statistically significant. This result suggests that the level of schooling is inversely related to the odds of being obese. Column (2) reports the probit regression which removes the self-selection bias. The estimated coefficient on schooling increases in magnitude to -0.457. This value represents a roughly 730 percent increase in the magnitude of the schooling coefficient.

Column (3) exhibits the probit regression which includes variables depicting risk/time preferences. Interestingly, the schooling coefficient is almost identical in magnitude to the schooling coefficient in column (1) without correction for self-selection and risk/time preferences. This finding implies that schooling effect is not sensitive to the inclusion of risk/time preference variables used in this study.

Column (4) represents the full model combining the technique for removing self-selection bias together with risk/time preferences. The magnitude of the schooling coefficient is significantly larger than those in columns (1) and (3) but is only slightly larger than that of column (2). These results indicate that the schooling coefficient is less sensitive to the addition of the risk/time preference variables. The significant coefficient of the residuals from the schooling equation suggests that unobservables in schooling equation have an influence on the odds of being obese. These unobservables may represent factors other than the risk/time preference variables used in this paper (e.g., genetic and other personality traits). The results of the risk/time preference variables in column (4) are similar to those in column (3) discussed above.

The last column of Table 1 reports the results of the probit regression with the risk/time preference variables using the instrumental variable approach (estimated using the two-stage estimation method described by Maddala, 1983). As mentioned earlier, this approach does not remove the self-selection bias in the model. Interestingly, the difference in the magnitude of the schooling coefficients in columns (4) and (5) is only 0.05. The magnitude of the coefficients of the risk/time preference variables are also comparable except for the statistical insignificance of "pref3" variable in column (5).

Table 1
 Probit Estimates of Principal Variables in Obesity Equation

| Variables | Coefficients (std errors) | | | | |
|---|------------------------------------|--------------------------------------|--|--|---|
| | (1) Actual Schooling Only | (2) Schooling and Residuals | (3) Schooling and Preferences | (4) Schooling, Residuals, & Preferences | (5) Predicted Schooling w/ Preferences |
| <i>Schooling</i> | -0.055* (0.012) | -0.457* (0.180) | -0.056* (0.013) | -0.460* (0.181) | |
| Actual School | | | | | -0.455* (0.180) |
| Predicted School | | 0.403* (0.181) | | 0.405* (0.181) | |
| Residuals from Schooling Equation | | | | | |
| <i>Risk/Time Preference</i> | | | | | |
| Pref1 | | | 0.055 (0.097) | 0.046 (0.097) | 0.105 (0.096) |
| Pref2 | | | -0.138* (0.078) | -0.137* (0.078) | -0.142* (0.078) |
| Pref3 | | | -0.268* (0.139) | -0.273* (0.139) | -0.226 (0.138) |
| Pref4 | | | -0.012 (0.075) | -0.018 (0.075) | -0.035 (0.074) |
| -2Ln(Likelihood) | 1863.74* | 1858.80* | 1857.11* | 1852.14* | 1869.43* |

Note: asterisk denotes statistically significance at 0.05 level. Obesity in this paper is defined as having a body mass index of 27.8 for men and 27.3 for women (as defined in the health/medical literature). Parameter estimates of other explanatory variables (i.e., black, other, age, income, hhsz, male, unemployed, northeast, midwest, west, city, and nonmetro) are available from the author.

Concluding Remarks

No other known published study has focused on the effect of schooling on obesity. In this paper, particular attention is given to the issue of self-selection bias and the omission of "third" variables measuring risk/time preferences. The results indicate that the magnitude of the schooling coefficient is much more sensitive to the omission of the variable that accounts for self-selection bias than the risk/time preference variables used in this study. Once the self-selection bias is accounted for, the magnitude of the effect of schooling on the odds of being obese increases by roughly 730 percent. Overall, the statistical significance of schooling is consistently one of the largest among the variables in the regressions. The schooling coefficient also consistently has a negative sign which implies that schooling has a negative effect on the odds of being obese and hence, a positive effect on reducing the incidence of obesity, independent of risk/time preference and general unobserved heterogeneity. This finding has important public policy implications because it suggests that education programs can be used to improve public health by reducing the incidence of obesity in the population.

Appendix

For question about this paper you may contact Rudy Nayga by telephone, fax or electronic mail.
 Telephone: (979) 845-8376; FAX: (979) 862-8679; E-Mail: rnayga@tamu.edu

References

- Berger, M. C. & Leigh, J. P. (1989). Schooling, Self-Selection, and Health, Journal of Human Resources, 24, 433-455.
- Centers for Disease Control (1997). More Americans of all Ages are Overweight, U.S. Department of Health and Human Services, HHS News, March.
- Farrell, P. and Fuchs, V.R.(1982). Schooling and Health: The Cigarette Connection, Journal of Health Economics, 1, 217-230.
- Fuchs, V. R. (1982). Time Preference and Health: An Exploratory Study, In Economic Aspects of Health, ed. V.R. Fuchs, University of Chicago Press, Chicago, pp.93-120.
- Feldstein, P. (1983). Health Economics, New York, John Wiley and Sons.
- Garen, J. (1984). The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable, Econometrica, 52(5), 1199-1218.
- Grossman, M. (1975). The Correlation between Health and Schooling, In Household Production and Consumption, edited by Nestor E. Terleckyj. New York: Columbia Univ. Press (for NBER).
- Grossman, M. & Joyce, T. J. (1987). Socioeconomic Status and Health: A Personal Research Perspective, Paper presented at the Kaiser Foundation Conference on Socioeconomic Status and Health, March 26-27.
- Heckman, J. J. (1983), Sample Selection Bias as a Specification Error, Econometrica, 47(1), 153-161.
- Kenkel, D. (1990), Consumer Health Information and the Demand for Medical Care, Review of Economics and Statistics, 587-595.
- Kenkel, D. (1991), Health Behavior, Health Knowledge, and Schooling, Journal of Political Economy, 99(2) 287-305.
- Lee, L. (1978). Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables, International Review, 19(2), 415-433.
- Leigh, J. P. (1990), Schooling and Seat Belt Use, Southern Economic Journal, 57(1), 195-207.
- Maddala, G. S. (1983). Limited Dependent and Qualitative Variables in Econometrics. New York: Cambridge University Press.
- Rosen, S. & Taubman, P. (1982). Some Socioeconomic Determinants of Mortality, In Economics of Health Care, eds. J. van der Gagg, W. B. Neeman, and T. Tsukahara, Jr., Praeger Publishers, New York, pp. 255-271.
- Sander, W. (1995), Schooling and Quitting Smoking, Review of Economics and Statistics, 77(1), 191-199.
- Variyam, J., Blaylock, J. & Smallwood, D. (1998). Informational Effects of Nutrient Intake Determinants on Cholesterol Consumption, Journal of Agricultural and Resource Economics, 23 (1), 110-125.
- Willis, R. J. & Rosen, S. (1979). Education and Self-Selection, Journal of Political Economy, 87(5), S7-S36.

Endnotes

- ¹ Associate Professor, Department of Agricultural Economics