# HOW DOES PARENTAL EDUCATION AFFECT INFANT HEALTH?

**Dean Lillard, Kosali Simon, Maki Ueyama**

Estimating the causal effect of education on one's own or one's children's health is complicated by the fact that unobserved variation across individuals could cause both health and education. There is very little investigation of the effect of parent's education on child health in a developed country context. A recent exception is Currie and Moretti (QJE, 2003) who show that exogenous increases in college education induced by college openings has a beneficial impact on an infant's health. No study to date has examined how educational attainment in the form of parental high school completion affects infant health in nationally representative data. Our paper provides answers to this question, and investigates mechanisms that may be responsible for this effect. We exploit variation in state compulsory education policies that lead to exogenous differences in high school educational attainment. We conduct our study with two data sets; the National Longitudinal Survey of Youth, 1979 (NLSY79) cohort. The advantage of using NLSY79 Child data is the extensive list of child health indicators available over the different ages of childhood.

Unlike prior research that has used instrumental variables to estimate the effects of education using variation in policies assuming a person always lived in a particular state of residence (eg Angrist and Kruger 1991), we infer a person's state of residence in each year using current state of residence, state of birth, state of residence at age 14, and date and state of last several residential moves. We then assign compulsory education policies to each parent using his/her state of residence when the compulsory education began.

We first study the impact of educational attainment on infant's health outcomes using state educational policies as instruments. Second, we investigate the mechanisms by which health may be affected- for example- whether it is through the use of health care services, or through health related behaviors.

## Conceptual Framework

An important potential non-market benefit of greater education is better health for oneself as well as one's children. Education could improve ones health through a number of ways-for example, it could make one a better consumer of information or give one greater incentive to live longer to earn more (Grossman..), or it could lead to better health through greater resource availability as more education could improve ones own wages and spouses wages through assortative mating. In the context of infant health, maternal education may also lead to changes in age of child bearing and other fertility decisions that impact infant health. Theories of the role played by education in household and health production functions explained by Becker and Grossman have lead to several empirical attempts to quantify its magnitude, but a recent literature on the effect of maternal education on infant health does not agree on the answer. While the NLSY79 does not allow us to test all possible mechanisms, we will look at some mother's behavioral mechanisms such as mother's tobacco use.

## Identification strategy

The effect of mother's education on infant health is estimated using an IV approach. In the first stage, we estimate mother's education using state's education policies as instruments:

$$E = \alpha_0 + \alpha_1 IV + \varepsilon \tag{1}$$

where E is mother's education (highest grade completed and a dummy variable that equals 0 if the mother is a high school dropout and 1 if she is a high school graduate or more) and IV is a vector of state policy variables that capture the state's compulsory education environment including: the minimum age of permitted entry, the minimum age of compulsory entry, an indicator variable for states that do not have laws on the minimum age of permitted entry, and the minimum age of permitted school leaving. We merge our data with the instruments using mother's state of residence and relevant year. The source of identification is variation in the high school education environment across states over time, driven by (we should say something here about whether its political whim or popular sentiment uncorrelated with health etc..).

In the second stage, we estimate the effect of mother's education on infant health using her predicted education from (11) and other exogenous variables:

$$H = \beta_0 + \beta_1 \hat{E} + \beta_1 C + \beta_2 S + \varepsilon \tag{2}$$

where H is one of the several infant health outcomes including: height adjusted weight, the date of last routine health check-up, and the number of illnesses requiring medial attention or treatment. $\hat{E}$ is mother's predicted education. C is a vector of infant and mother's characteristics including: infant's sex, mother's race, age, and marital status, MSA size of city of residence. One must keep in mind that some of these control variables such as mother's marital status could themselves be causally affected by education and therefore part of the positive effect of maternal education could be acting through the coefficients of these variables. (we should expand this). S is a vector of state characteristics including: the ratio of employed men to women; the seasonally-adjusted unemployment rate; real median wages; the maximum value of the federal and state EITC for a single mother with two children; the income eligibility limit for Medicaid eligibility for pregnant women; the annual employment growth rate; and the amount of federal housing money spent per 1,000 residents in the state. Finally, ε is the error term that captures the remaining unobservables that are not captured in the equation.

Equations (1) and (2) are estimated using a linear probability model for dummy dependent variables and a simple linear model for continuous dependent variables on weighted data. Since the error term is not normally distributed for categorical dependent variables, the use of linear probability model will produce inefficient coefficient estimates. However, this is not a major problem because estimates are generally similar to those produced by nonlinear models when evaluated at the sample means (Greene 1993). Also, Angrist (2001) suggests that the use of linear probability models in the first stage do not produce fundamentally different results as the use of nonlinear probability models in the first stage.

Data

The data for this study comes from the 1979-2002 National Longitudinal Survey of Youth 1979 (NLSY79) and 1986-2002 NLSY79 Child survey. The NLSY79 is a nationally representative dataset that consists of 12,686 individuals who were ages 14-21 as of December 31, 1978. They have been interviewed annually from 1979 to 1994 and biennially since 1994. In 1986, a survey of children born to female NLSY79 respondents began and has been conducted biannually since then. The NLSY79 Child survey includes various information on child's current health condition and health history obtained mostly from mother's report. We use several sample restrictions for the study. We exclude infants with the following mother's characteristics: a foreign residence when the compulsory education most likely began, missing mother's education, and less than age 14 when the infant was born.

Preliminary Results

Our preliminary results using 1989 U.S. Natality Detail Data indicate that maternal high school completion has a small beneficial impact on almost every measure of infant health studied.