Does School Education Reduce Childhood Obesity?

Ning Zhang, Cornell University

Backgrounds and Significances

Education is the most important correlate of adults' health (Grossman and Kaestner, 1997). However few researches examine the impact of school education on youth health, particularly in terms of obesity. It is difficult to isolate age from education effects. Moreover, the effects of peer pressure are crucial because adolescents easily influence each other. Finally, young people's understanding of health and obesity issues evolves through schooling. Without properly addressing these problems, one cannot fully determine the impact of education on youth obesity.

This paper employs state-wide first-grade entry policies to identify the impact of school education on youth obesity. A child should be six on or before the school entry dates to attend the first grade. While a child who is born a day after the entry date has to wait a full year to enter school. Therefore children who are born before the school entrance date may have one more year of education than those born after. I exploit this fact and use a child's exact date of birth to construct a dichotomous variable. The dichotomy is the instrument variable for education focusing on the probability of obesity for those born just before and just after the school entrance date. Because few public-use data sets provide the precise date of birth, this research uses restricted-access data from National Longitudinal Survey of Youth 1997 cohort (NLSY97), 1979 cohort, and NLSY79 Children and Young Adults (NLSY79-C). These longitudinal data sets contain measures of education, parental characteristics, and health practices necessary to implement the empirical analysis. The three samples include children and teens aged 5 to 19 since 1964 to the present. This time period is particularly important because a number of exogenous determinants of education and youth health have changed dramatically. For example, more states adopted state-wide entry policies since late 1960s. In addition, many states changed their cutoff dates from the first quarter to the third or forth quarter of a year to encourage children to start school at an older rather than a younger age. During this time frame, youth obesity has captured public attention. State and local governments put efforts to combat the prevalence of obesity, such as publicizing children's nutrition guidelines, and modifying the mandated physical education requirements. The NLSY79, NLSY97 and NLSY79-Children datasets provide the historical comparisons of education effects on youth obesity over time.

This paper examines six important questions. First I demonstrate the instantaneous education effects on obesity. Health returns to education combines the health production and the health process of education. This paper intends to show whether this combination has short-term impact on controlling youth obesity. Second I present whether such short term effects are age- or grade-specific. On average, children in higher grades may have better strategies to maintain healthy weight than children in lower grades. On the other hand, high-grade children may face tougher academic requirements so as to dilute the health gains of education. Results for every age between five and 19, and for grade one through 12, delineate the trends. Next I present a statistical model to explain why fuzzy RD design is correct. In general, the state-wide cutoff date is not the only determinant of school entrance date. Neither early nor late entrance is unusual. Parents may delay sending their children to school if they are relatively small in size, or if they are athletical talents, or if parents believe that late entrance improves better academic performance. Children may attend school earlier than legal age when parents' labor market behaviors are taken into account, or if their ability to succeed in schools is proven. This statistical model will contribute to the literature on application regression discontinuity design. Fourth this paper tests whether peer pressure has significant effects on youth obesity using NLSY97 which collects data on the way that children spend their spare time with friends. Strong peer pressure may weaken education effects, if any. The existing literature on peer impacts focuses on a child's academic achievement (Whitmore, 2005), and our results offer additional evidence on health behavior. Fifth I discuss three channels through which education may affect youth obesity: health knowledge, dietary habits and physical activities. Although no theory favors a particular channel, results suggest their impact vary with age and grade. And last I document school entrance ages and the probability of attending school according to state cutoff dates through examining results from three different datasets. This finding thereby additionally adds to the education literature on sensitivity to education policy.

Identification Strategy

I exploit state-wide first-grade entry policies to identify the impact of school education on youth obesity. The identification strategy thereby is to compare the probability of obesity of students born just before the school entrance dates to those just after it. Under the assumption that all other characteristics affecting body weight and height vary smoothly across entrance date, the difference in probability of obesity may be causally attributed to differences in the number of school years. The regression discontinuity design (RD) plays a central role in this estimation. In particularly the number of the years that a child has been in school at any particular age is a discontinuous function of the birth date. School enrollment begins when the birth date is on or earlier than the State's school entry policy. In subsequent years, the child is always in a different grade at school.[1] In this way, the exact date of birth plays as an instrument variable for identifying the education that a child has obtained and producing credible estimates of the causal impact of education on youth obesity. As in other RD designs, a discontinuous relation between youth health outcome and birth date is identified as this causal impact of education.

This RD design is fuzzy, in a sense that the initial attending school is not just deterministic of birth date. Early entrance is permitted in most states. Parents and teachers can have a mutual agreement to have some children at school at a younger age. Children may go to school early if their mothers are working at their age of five (Gelbach, 2002). Meanwhile, late entrance is practical. In those states where kindergarten attendance is not mandated and where compulsory school ages range from age 6-8, parents can delay a child's entrance into first grade for a year by various reasons. For instance, some parents may send their "big" children to school a year late, to increase the likelihood that the child will perform better on a football

[1] I assume that students do not leave and re-enter school before legal drop-out age and that there is no grade retention.

team. Some others may hold their "small" children back one more year when parents believe that their children are not yet developmentally ready to succeed in school. Taking all these facts into consideration, exact date of birth partially determines the starting age of attending schools among children.

The estimation strategy is a 2SLS. let L_i denote the number of days after the cutoff day that the child is born. That is, a person born on the school entry date is $L_i = 0$, a person born 60 days before the entry date is $L_i = -60$, and a person born 60 days after the entry date is $L_i = 60$. The threshold for applications is $l^*{}_i = 0$, and the participants' status is partly deterministic of L_i , i.e. $d_i = 1\{L_i > 0\}$. Thus children with $L_i \le l_i^*$ enter school one year earlier than children with $L_i \ge l_i^*$ consistently, all throughout the years of school enrollment. This indicator function also satisfies the monotonicity, i.e. $d_i(L_i > l^*) > d_i(L_i < l^*)$. In stage one, participation status d is an instrument for predicting schooling attainment; and in stage two, the predictions of schooling from stage one is used as one regressor. In general, the estimation equations are:

$$H_{it} = \alpha_0 Z_i + d_i \sum_{p=0}^{4} \gamma_{0p} (L_i)^p + (1-d_i) \sum_{p=0}^{4} \gamma'_{0p} (L_i)^p + \delta_0 d_i + \varepsilon_{it}$$

$$S_{it} = \alpha_1 Z_i + d_i \sum_{p=0}^{4} \gamma_{1p} (L_i)^p + (1-d_i) \sum_{p=0}^{4} \gamma'_{1p} (L_i)^p + \delta_1 d_i + \eta_{it}.$$

ε and η are both individually independently distributed, with finite variances. The mean takes the value, $E[\varepsilon_{it} \mid L_i, d_i] = 0$ and $E[\eta_{it} \mid L_i, X_i] = 0$. $E(Z \mid L)$ is continuous at L in both equations. Z includes the vector of a constant and the observed demographic and SES variables, such as gender, race, parents' schooling, parents' marital status, etc. γ_{0p} , γ'_{0p} , γ_{0p} and γ'_{1p} represent the coefficients on the fourth polynomial terms and interaction terms. The parameters of each term of the polynomial and interactions can vary on either side of the cutoff, $L_i = l^*$, meaning the shape of the underlying conditional expectation is different to the left and right of the threshold. If the parameterization in equation (3) is correct, the δ_0 and δ_1 can be constantly estimated via least squares, and their ratio represents the causal effect of schooling on youth obesity, i.e. $\hat{\beta} = \dfrac{\hat{\delta}_0}{\hat{\delta}_1}$. I estimate standard errors clustered on each L in all specifications in this paper.

Finally, I estimate the reduced form as,

$$H_{it} = \alpha_0 Z_i + d_i \sum_{p=0}^{4} \gamma_{0p} (L_i)^p + (1-d_i) \sum_{p=0}^{4} \gamma'_{0p} (L_i)^p + \varphi S_{it} + \varepsilon_{it} ,$$

with the same denotations and assumptions as in equation system (3), except instrumenting S by d, via two-least squares. When specifications are correct, $\hat{\varphi} = \frac{\hat{\delta}_0}{\hat{\delta}_1} = \hat{\beta}$, by econometric theory.

Data Description

The three core datasets in this paper is the restricted-access data (geocode) from NLSY79, NLSY97 and Children and Young Adults of NLSY79 Women. Since the paper focuses on youth obesity, respondents who are older than 19 drop out of the sample. Consequently sample respondents age from 14 to 19 from NLSY79, from 12 to 19 from NLSY97, and from 5 to 19 from NLSY79-Children. I restrict the data to those whose body weight measurements are reasonable. That is, 1) BMI is between 10 and 50; and 2) height is greater than 3 feet and less than 8 feet, and weight is greater than 40 pounds and less than 300 pounds. Next I drop all respondents who have missing birth date information or missing state of residence identifier. These three criteria lead to around 20 percent deletion of the total sample.

Since both body weight and height were self-reported, a technique that corrects the reporting error is necessary. Throughout all three data, I use the correcting method which is introduced by Cawley and Burkhauser (2006), where we run a regression of an actual weight and height measuring in NHANESIII on reporting measures. This correcting step is applied to all body measurements in this study, including children's body measurements and biological mothers' measurements.

Selected References:

Cawley, J.H., and R. Burkhauser. 20006. "Beyond BMI: The Value of More Accurate Measures of Fatness and Obesity in Social Science". NBER Working Paper #12219.

Gelbach, J. (2002) "Public Schooling for Young Children and Maternal Labor Supply." The American Economic Review, vol. 92, no. 1 (Mar 2002), 307-322.