Anthony Fontana

Econ 201

Final Project

Literature Review

Economists have been concerned with parental education and its relation to child outcomes for quite some time now. This is primarily because in other studies economists have found that poor child health leads to lower educational attainment, lower adult life health, and other negative socio-economic outcomes. It is clearly important to find out what causes poor child health. Some of the major works have found clear indication that parental education is an important in determining a variety of socio-economic outcomes for children. Other research has found no relation between the two indicating that the relation may be spurious. Although there is correlational evidence tying higher parental education to favorable child outcomes many studies find this relation to be non-causal. Other studies find that this relation is causal. When surveying the literature, results are not robust to model specification meaning that some authors who choose a different identification strategy come to different results. Also, when adjusting the main research hypothesis even slightly, results appear to change. Authors do seem to be converging on an agreed upon method of identification. Many people are using Joshua Angrist's strategies to identify causal relations. Perhaps the literature is trending into the correct direction.

A variety of causal mechanisms have been proposed that link parental education to the health outcome of a child. Most of these claim that more educated parents tend to have higher incomes so they can use this money to tend to their children. Other studies relating family income to socioeconomic outcomes of their children find a positive relation so it doesn't seem unreasonable that this mechanism might be plausible. Other direct causal paths suggest that more education parents make smarter fertility choices. Some even suggest a multiplier effect meaning that women with higher education tend to choose partners of higher education which further increases the household income.

Historically, in the literature we find that specific health outcomes of children tend to be measured by nutritional status, morbidity, and mortality. These are seen in early studies before more sophisticated measures of health were created or data was unavailable. Older studies tend to use different evidence and data when coming to research related questions. For example, (Cochrane, Leslie, O'Hara 1982) use evidence from field studies in developing countries to identify the effect of education on child health outcomes. Typically, the older studies tend to have a more difficult time acquiring reliable data on nutritional status or even defining a coherent measure of such a variable. Most of the data was also survey evidence surveying parents directly. This can lead to severe measurement errors if parents are not willing to be truthful about their own educational status and the health of their children. In addition, much of the data is collected at an aggregate level which makes it more difficult to control for correlated errors.

More recently, anthropometric measures such as height-for-age Z-scores (HAZ), low birth weight measured by body mass index, acute or chronic mental conditions etc. and some of the

data is longitudinal. These more sophisticated measures are now capable of being constructed because we have better access to data. Some of the data is now collected via government bodies or through documents created by physicians performing the physicals to identify key health indicators. Some studies still use subjective measures of health but the data does seem to come from more reliable sources (Doyle, Harmon, Walker 2005).

As identification strategies have evolved over time, so have the final results of the studies. Older studies tend to only use simple bivariate or multivariate regressions and rely heavily on R-squared as a criterion for explanatory power. Many of these studies also do simple t-tests to compare populations on the relevant outcome variable. There are no matching strategies and very few attempts to identify an exogenous shock to determine variation. It seems the main concern is sample size and range of data rather than more sophisticated methods for determining causal relations. These studies do include controls to help alleviate potential omitted variable bias but they generally suffer from problems faced by all cross-sectional studies used to identify causal relations. Some of these downsides include reverse causality and omitted variables. They also suffer from selection issues which may be a reason why they find positive results because their sample participants are more likely to be positively related to the outcome variable. Quasi-experimental designs typically were not seen.

More recent literature still adopts this methodology but many of the issues are avoided because of better measurement constructs and data. In addition, some of these studies include sensitivity analyses as a way of determining the extent to which this methodology might be flawed. For example, in (Chen and Li, 2009) they use modern measurement constructs as controls but test for potential non-linearities or interactions between the variables. Although their research is attempting to identify the effect of non-biological parental education on adopted children outcome, the researchers are converging to using the same measurements. These researchers find a positive relation between the variables of interest although suffering from selection issues such as children more likely to be adopted (parents might adopt a certain type of child more likely to be healthy). Identification strategy seems to be important in determining the outcome but results also seem sensitive to data and the construction of measures.

Some studies including (Doyle, Harmon, Walker 2005) use an ordered probit model to identify the effect of the explanatory variables on subjective health measures (low to high). Their approach is an Instrumental Variable approach to account for endogeneity issues. Many authors are trending to the newer identification strategies such as Regression Discontinuity, Differences-in-Differences, and even Regression Kink. Some are combining Regression Discontinuity designs with Instrument Variables to exploit exogenous shocks which can help investigate causal claims. These approaches are superior to older methods and seem to all arrive at similar conclusions about the effect of parental health and income on socioeconomic outcomes of their children. These methods are superior at eliminating endogeneity issues such as education being related to family background which could be related to school through non-education mechanisms that we don't know of (McCray, Royer 2006). However, they are not without their flaws and they do tend to lack external validity and will always be subject to whether the instrument truly meets the exclusion criteria. The results can also only be interpreted as local

average treatment effects. Opportunities to determine causal mechanisms might be slim because they normally depend on exogenous school policy shocks which presuppose that the policies can’t be predicted by the people being treated.

In general, older methods tend to find positive relationships between the variables in question. Newer methods do not find such a relation and many of them tend to be converging to similar answers. The discrepancies in the literature might be due to question framing.

Proposal

This topic seems to be heading in the right direction. Authors have been able to replicate studies conducted by other authors with different datasets. There also seems to be consensus on what methods we should use to answer questions of causality and to deal with endogeneity. Many people are also using the same measures of health so construct validity seems to not be an issue. To expand upon this research topic, I have a few suggestions.

Researchers in this area should try and expand external validity and develop more of a thorough understanding of how different levels of education might lead to health effects. Perhaps we can't detect an effect of parental education on health because the range of additional education is too narrow. Instead, find out the effect of an additional year to five years and estimate this effect. Also, it is difficult to determine the quality of education received by a student and their likeliness to participate in the learning process. Maybe these students who received the additional seven months did not retain any material or thorough understanding and hence were unaffected. This is like the arguments I made earlier for determining construct validity of the outcome variable health. Additional education might not reflect or measure what really affects child health (perhaps increased aptitude for example). This has implications for policy such as directing resources towards improving the quality of education rather than just increasing or decreasing the required amount of schooling (in terms of years). Instead of using education in this way, perhaps we should direct our attention to see how the quality of education received will impact the health of a child. External validity seems to be an issue here as well since no study can see if these results hold for other populations of interest in different geographical/social contexts and at different times.

My main proposal is to expand this project to find different exogenous shocks in different settings to see if these results are consistent elsewhere. Also, I suggest finding a better measure of education (or consider the implications of using an alternative measure). It could be that these dropouts are more likely to not retain any information from the additional schooling they acquire so the additional education might appear not effective for them but it could be effective for students who are more likely to participate and retain knowledge (heterogeneity). This appears to be a selection issue. Therefore, if the researchers in this area can find a way to use propensity score matching or some other matching procedure we might see results change. Lastly, I suggest using quantile regression to estimate the impact of these shocks at the tails of the education distribution. If researchers do not take this approach, then the best alternative will be to include non-linear terms in the regression discontinuity design to see if results change.

References

Lindeboom M, Llena-Nozal A, Klaauw B (2009). "Parental Education and Child Health: Evidence from a Schooling Reform". *Journal of Health Economics.* (103-131).

Cochrane, S, Leslie, H, O'Hara, D (1983) "Parental Education And Child Health: Intracountry Evidence" *Health Policy and Education.* (213-250)

Li, H & Chen, Y (2009). "Mother's education and child health: Is there a nurturing effect?" *Journal of Health Economics.* (413-426)

Doyle, O & Harmon C, & Walker, I (2005) "The Impact of Parental Income and Education on the Health of their Children". IZA DP No. 1832.

McCrary, J & Royer, H (2006) "The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth". *Journal of Economic Literature.*