Intergenerational Effect of Education on Physical and Mental Health

By: Timothy-Paul Chevalier and Parmis Mehdiyar

tp.chevalier@mail.utoronto.ca, parmis.mehdiyar@mail.utoronto.ca

1005803784, 1005781307

ECO375: Applied Econometrics

University of Toronto

Department of Economics

November 2022

Abstract

To determine whether education has an intergenerational effect on physical and mental health of the next generation, we separately model the relationship between the average years of parental education and self-report scores of their children's physical and mental health as measured through the SF12 questionnaire. We find that there is no significant relationship between one's SF12 mental or physical health scores and the average years of parental education, especially after holding confounding variables like the individual's own education, race, family income and access to a healthcare provider constant. We do however see repeated evidence of an individual's own education being correlated with their health scores, which emphasizes the importance of providing adequate access to education to every generation for optimal societal health outcomes, especially since these effects are not intergenerationally persistent. Finally, the paper discusses the potential limitations of our model in terms of external validity and internal validity - namely omitted variable bias, imprecise variables and potentially misspecified functional form, and suggests modifications for future replications of the study.

Introduction

While the socioeconomic effects of education have been extensively studied in economic research - for example measuring the effects of increased schooling on income, crime, mortality, health etc. - a relatively smaller and newer field focuses on the intergenerational transmission of these effects. This is a critical area of study for determining the longer-term effects of schooling policies that cross generations. Inspired by Chalfin et Al's research on the effects of parental education on children's use of alcohol, cannabis and other substances (Chalfin, 2018), in this paper we aim to build on this newer field of research while studying a different array of costly social problems: physical and mental wellbeing of the next generation. Considering previous research has shown that education has positive returns to self-reported physical and mental wellbeing (Oreopoulos, 2007), we especially hope to discover if these positive effects transfer intergenerationally as well, and if so, to what extent. The focus of this paper is therefore to robustly estimate the effect of the years of parental education on the physical and mental health of their children, as measured through SF-12 physical and mental health self-assessment scores.

Broadly, we find that after controlling for an individuals' own education and other lurking variables like family income, race and access to healthcare providers the intergenerational effects of education on physical and mental health of the next generation are negligible. While disappointing, this lack of intergenerational transmission emphasizes the importance of providing every generation with adequate education opportunities, irrespective of parental education levels, for optimal societal health. However, we make no definitive claims regarding the magnitude of our estimates, as our measures of physical and mental health are self-reported and potentially linked with many omitted factors that our data does not track.

The Context and Data

The data used in this study were drawn from the National Longitudinal Survey of Youth (NLSY97): a panel survey conducted from 1979 to 2018 on a nationally representative random sample of 12,686 adolescents aged 14 to 21 on December 31, 1978. A unit of observation in this study is therefore a respondent. The main explanatory variable in the study is the average years of education of a respondent's father and mother (Table 1), which ranges anywhere between 0 (parents have no education at all) to 20 years (both parents have graduate degrees). This relatively large range of an explanatory variable allows for the variations in the independent variables to be observed precisely. Average parental education seems to be distributed almost normally, and on average a respondent's parents have 10.9 years of education (Table 1).

The primary dependent variables include proxies of physical and mental health as measured by the SF-12 scores of the respondents recorded when they turned 40 years of age. The SF-12 is a 12-question self-assessment measuring general health through the individual's everyday life. It has a physical component and a mental component, each assessed out of 100, and we look at each separately as our dependent variables. Table 1 provides summary statistics for these measures. The average physical score amongst our sample is 46.53 percentage points and has a standard deviation of 11.41 percentage points. The lowest score recorded was 11.31 and the highest score recorded was 65.45, thus a relatively small range of variation within this dependent variable. Furthermore, in the case of the scores for the mental component, the mean average is 52.86 percentage points, with an 8.72 percentage point standard deviation. This variable also has a relatively limited range of 58 percentage points, suggesting no extreme observations in terms of physical or mental health capacities.

Regression Analysis

For estimating the relationship between the average parental education and physical and mental health scores of an individual, we first started with a simple linear regression framework to estimate the linear relationship between parental education and physical and mental SF-12 scores. Table 2.1 (1) reflects the model for physical health, as specified by:

$$SF12\ Physical\ Component\ Score_i = \beta_{10} + \beta_{11}(Average\ Years\ of\ Parental\ Education)_i + e_i$$

and Table 2.2 (1) reflects the model for mental health, as specified by:

$$SF12\ Mental\ Component\ Score_i = \beta_{20} + \beta_{21}(Average\ Years\ of\ Parental\ Education)_i + e_i$$

where i is a person. We use robust standard errors to account for heteroscedasticity. According to the first model, a 1-year increase in average parental education is associated with a 0.467 percentage point increase in the SF-12 physical component scores of their children on average. With the relatively large t-statistic of 5.28, we can reject the null (H_0: $\beta_{11} = 0$) that there is no relationship between parental education and physical health scores. While statistically significant at the 1% level, this result is not economically significant. To put this into context, completing a 4-year college degree by both parents would on average only increase their child's physical health score by less than 2 percentage points, which is negligible given the trade-off for 4 years of college education.

Similarly, according to the second model, a 1-year increase in average parental education is associated with a 0.087 percentage point increase in the SF-12 mental component scores of their children on average. With the small t-statistic of 1.22 however, we fail to reject the null (H_0: $\beta_{21} = 0$) that there is no relationship between parental education and mental health scores. These results are also not economically significant, as they estimate the difference between mental health scores of a respondent whose parents have had no education even at the primary level, and the mental health scores of someone whose parents have both studied for 20 years to be about 1.75 percentage points.

Extension: Quadratic Functional Form

The mentioned specifications could be biased if the relationship between parental education and health scores are in fact non-linear, as hinted by our preliminary visualizations of SF-12 scores against parental education (Figure 1 and 2). This could be the case for example if higher levels of parental education had adverse effects on the next generations' health for a certain threshold of education, and additional years of parental education only had a positive effect on the next generations' awareness of, and care for mental and physical stimulus after parents completed that minimum threshold of education. Table 2.1 (2) and Table 2.2 (2) explore the potential for such non-linearity, with the following specifications:

Table 2.1 (2)

$$SF12\ Physical\ Score_i = \beta_{20} + \beta_{21}(Parental\ Education)_i + \beta_{22}(Parental\ Education)_i^{\,2} + e_i$$

Table 2.2 (2)

$$SF12\ Mental\ Score_i = \beta_{20} + \beta_{21}(Parental\ Education)_i + \beta_{22}(Parental\ Education)_i^{\,2} + e_i$$

The fitted regression estimates suggest that in terms of physical health of the next generation, returns to parental education are negative when parents have less than 6 years of education, and if parents completed high school (12 years of education), their children would score on average about 3 percentage points higher on the SF-12 Physical Health Questionnaire than if they only had 6 years of education. Furthermore, if parents completed 8 years of post-secondary education, their children would score on average about 10 percentage points higher on the SF-12 Physical Health Questionnaire than if they only graduated high school. These results are not only statistically significant, but also economically significant, particularly for higher education policy.

Similarly, in terms of mental health, the quadratic model estimates that returns to parental education become positive when parents have closer to 7 years of education. However, after that, completing even up to 8 years of post-secondary education results only in about a 2.5 percentage point increase in mental health, which is negligible compared to the additional education and thus economically insignificant.

The mental health estimates are also statistically insignificant with very small t-statistics. Note that although the quadratic extension to the model improves the adjusted R-squared, thus reflecting that it better fits the underlying data, the adjusted R-squared for the models in Table 2.1 (2) and Table 2.2 (2) remain low in absolute terms.

Extension - Control Multiple Linear Regression

The aforementioned specifications fail to account for certain characteristics that could likely be correlated with both parental education and physical and mental health scores of their children. This could lead to the simple OLS estimates obtained above to suffer from omitted variable bias as the OLS assumption 1 could be violated. One of these variables is the person's own education, as suggested by existing research on the positive effects of education on health (Oreopoulos, 2007). Given that an individual's education is likely positively correlated with those of their parents, not including this term in our model would bias the estimates positively. Similarly, in our model for the physical health scores, we control for family income in 1979, ethnicity (white or non-white) and whether the respondent has a healthcare provider in 1979, as they could be correlated both with measures of health and education of the parents in the household. By allowing the models to attribute variations in SF-12 scores to the mentioned control variables, we can isolate the variation in SF-12 that is due only to the variations in average parental education (Table 2.1 (3) and 2.2 (3)) and the quadratic form of average parental education (Table 2.1 (4) and Table 2.2 (4)) for unbiased estimates:

Table 2.1 (4)

$$\begin{aligned} SF12\ Physical\ Score_i &= \beta_{40} + \beta_{41}(Parental\ Education)_i + \beta_{42}(Parental\ Education)_i^{2} \\ &+ \beta_{43}\ (Years\ of Education)_i + \beta_{44}\ (White)_i\ +\ \beta_{45}\ (Healthcare\ Provider)_i \\ &+ \beta_{46}\ (Net\ Family\ Income)_i + e_i \end{aligned}$$

Table 2.2 (4)

$$\begin{aligned} SF12\ Mental\ Score_i &= \beta_{40} + \beta_{41}(Parental\ Education)_i + \beta_{42}(Parental\ Education)_i^{2} \\ &+ \beta_{43}\ (Years\ of Education)_i + \beta_{44}\ (White)_i\ +\ \beta_{45}\ (Healthcare\ Provider)_i \\ &+ \beta_{46}\ (Net\ Family\ Income)_i + e_i \end{aligned}$$

The addition of control variables improves the model for both SF12 Physical Score and SF12 Mental Score, increasing the adjusted R-squared from 0.030 (Table 2.1 (2)) to 0.043 (Table 2.1 (4)) and from 0.001 (Table 2.2 (2)) to 0.017 (Table 2.2 (4)). Furthermore, the results from the controlled quadratic specifications confirm the presence of omitted variable bias in the base quadratic specifications; given that there are significant reductions in the OLS point estimators for parental education and parental education squared. Meaning, a considerable part of the variation of SF-12 scores that we believed was attributed to changes in parental education was actually attributed to these other factors that were previously omitted.

The results from Table 2.1 (4) suggest that for a couple of parents with a high school diploma on average, completing a 4-year post-secondary degree increases the SF12 physical scores of their child by less than 2 percentage points on average. Similarly, completing an 8-year post-secondary degree increases the SF12 physical scores of their child by about 5 percentage points on average - a much smaller effect than that of Table 2.1 (2), which suffered from omitted variable bias. The results display no statistically significant relationship between the predictor of interest and SF12 Physical Score, given the small T-statistics. However, the results show a highly statistically significant relationship between two of the control variables and SF-12 Physical Score; Years of Education and Family Income. Particularly, holding all else constant, an additional year of respondent education resulted in almost a 0.5 percentage point increase in the respondents physical SF-12 score, which is highly consistent with Oreopoulos' study's results (Oreopoulos, 2007).

Similarly, the results from Table 2.2 (4) display no statistically significant relationship between the predictors of interest and SF-12 Mental Score, even at the 10% level. Furthermore, they suggest that for a couple of parents with a high school diploma on average, completing an 8-year post-secondary degree increases the SF-12 mental scores of their child by about 1 percentage points on average - an extremely negligible increase, economically speaking. Once again, the small T-statistics show a high level of statistical significance for the relationship between the respondent's own education and their mental health scores, which is consistent with existing literature on the relationship between education and health (Oreopoulos, 2007).

Limitations of results

The specifications outlined above have considerable limitations. They continue to suffer from omitted variable bias, suggesting major threats to internal validity, and there are doubts regarding the interpretability and replicability of the results, threatening external validity. Overall, given the very low fit of the specifications across the board, as shown by the universally low adjusted R-squared, it is evident that the results of the model are neither strong estimates for causal inference, nor do they hold rigorous descriptive power.

Even after introducing controls in Table 2.1 (4) and Table 2.2 (4), it is likely that the models are subject to threats of internal validity - meaning, they deviate from the true intergenerational effect of education on health. Firstly, it is very likely that there are other factors that are correlated with both measures of health and parental education that the lack of data prevented this study from controlling, such as cultural factors, lifestyle and geographical situation, which could explain the variations in physical and mental health scores. Future replications of this study could certainly benefit from observing and controlling for such factors. Future studies could also look into the intergenerational effects on health outcomes using measures of parental education other than the mean average, such as the maximum years of parental education amongst the two parents, or a measure that gives more importance to one parent's education. For simplicity, this paper gave equal weights to maternal and paternal years of education when producing our measure of education levels in the prior generation, but it could be the case that one parent's education drives physical and mental wellbeing more strongly.

Next, the survey-based nature of our data could have implications for the precision of the observed variables used in the study. For example, it could be the case that due to societal expectations, low levels of parental education or family income were not always truthfully reported. Moreover, SF12 scores may also be biased measures of health, given that they are calculated using only 12 questions based on self-reported implications of health on daily life, and were collected when the respondent turned 40. It is possible that at that age, a person's health would be very minimally related to their upbringing setting and experiences as an adolescent, including parental education. Future modifications of this study could use more precise measures of health that are not subject to personal biases, and recorded during adolescence years. Alternatively, future research could control for individual or year fixed effects by including intercepts in the regression model, thus removing any bias caused by aging and the passing of time to better isolate the effect of parental education on health scores.

In terms of estimating the true functional form of the relationship of interest, we are doubtful of the quadratic relationship holding for the mental health estimates (given the fall in adjusted R-squared between Table 2.2 (3) and Table 2.2 (4). Meaning, it is possible that the model for mental health scores is, unlike that of the physical health scores, not a quadratic function of average parental education. It is therefore advised that future replications of this study re-evaluate other functional forms that this model can take, especially if other lurking variables are introduced or precision of current measures are improved. Moreover, while respondents were originally randomly selected in 1979, there is likely a relationship between education or wellbeing, and choosing to participate in the follow up at the age of 40, where SF-12 surveys were answered. This means that our sample may not be fully reflective of the true intergenerational impacts of parental education on health. In the random selection setting of the NLSY97 data, collecting measures of health at the same occasion as the explanatory variables could potentially address this issue as it would ensure people randomly selected for the survey do not drop out of the study in a systematic manner.

Finally, the main threats to external validity in our research are the constraints in populations and settings. This paper's data were collected as part of an American Longitudinal Survey spanning from 1979 to 2018. Given that the results here are heavily dependent on state and national education and healthcare policies, the results could vary drastically over time and space. For example, it is uncertain whether we can assume that the outcomes from the present research could also be applied to the Canadian demographic due to the vast difference in healthcare policies between Canada and the US. Even if we hold country or state constant, it is likely that the change in local health policies in the US and US states since when the original data were collected would lead to different results if repeated today.

Conclusion

In conclusion, our analysis does not provide sufficient evidence to reject the claim that parental education has no intergenerational effect on children's mental and physical health. It cannot be concluded however, that such a relationship does not exist in reality, given that there were many lurking variables unaccounted for by our model and unaccountable with the data at hand. Moreover, improving the precision of the control variables used in this study, and using more precise measures of physical and mental health could help bring future models close to the real intergenerational effects of education on the next generation's health.

Within the bounds of this research project and the limits of the underlying data, we were not able to include enough control variables to significantly increase the model fit and to mitigate the omitted variable bias that could affect our main predictor. However, we could confirm findings of previous studies that demonstrated a significant relationship between education and measures of health within a generation (for the same person). This suggests the importance of providing adequate education to every generation for optimal societal health outcomes, as intergenerational effects of education do not seem to be significantly persistent.

References

"12-Item Short Form Survey (SF-12)." Physiopedia. Accessed November 23, 2022. https://www.physio-pedia.com/12-Item_Short_Form_Survey_(SF-12).

Chalfin, Aaron, and Monica Deza. "The Effect of Parental Education on Children's Drug and Alcohol Use." *AEA Papers and Proceedings* 108 (2018): 373–78. https://doi.org/10.1257/pandp.20181105.

"National Longitudinal Surveys." National Longitudinal Surveys | A Program of the U.S. Bureau of Labor Statistics. Accessed November 20, 2022. https://www.nlsinfo.org/.

Oreopoulos, Philip. "Do Dropouts Drop out Too Soon? Wealth, Health and Happiness from Compulsory Schooling." *Journal of Public Economics* 91, no. 11-12 (2007): 2213–29. https://doi.org/10.1016/j.jpubeco.2007.02.002.

Tables and Figures

Table 1: Summary Statistics

Variable	(1)	(2)	(3)	(4)	(5)
	count	**mean**	**SD**	**min**	**max**
SF-12 Physical Component Score	1438	46.53581	11.40869	11.31	65.45
SF-12 Mental Component Score	1438	52.85978	8.728457	11.4	69.4
Net Family Income (1979)	1134	15512.25	12720.18	26	75001
Years of Education	1438	13.53199	2.725945	0	20
Average Years of Parental Education	1187	10.93387	3.404448	0	20
Has a Healthcare Provider (1979)	1366	.8814056	.3234291	0	1
White	1438	.5055633	.500143	0	1

Note: Unit of observation is a respondent. Average years of parental education is the mean average of respondent's mother and father's years of education, obtained in 1979. The respondents were surveyed again at the age of 40, when their own education information and SF-12 responses were recorded.

Table 2.1: Regression Analysis of SF-12 Physical Health Score

SF-12 Physical Component Score	(1)	(2)	(3)	(4)
Average Years of Parental Education (1979)	0.467***	-0.750*	0.197	-0.522
	(5.28)	(-2.40)	(1.59)	(-1.42)
Average Years of Parental Education - Squared		0.0623***		0.0375*
		(4.34)		(2.24)
Years of Education (Respondent at age 40)			0.538***	0.490***
			(3.75)	(3.38)
White			-0.789	-0.744
			(-0.97)	(-0.91)
Has a Healthcare Provider (1979)			0.548	0.732
			(0.45)	(0.60)
Net Family Income (1979)			0.0000924***	0.0000860**
			(3.31)	(3.08)
Constant	41.99***	47.12***	36.18***	39.72***
	(39.44)	(27.84)	(17.74)	(14.66)
Observations	1187	1187	896	896
Adjusted R-squared	0.019	0.030	0.040	0.043

Note: Unit of observation is a respondent. T-statistics are in parenthesis and based on robust standard errors. Average years of parental education is the mean average of respondent's mother and father's years of education, obtained in 1979. The respondents were surveyed again at the age of 40, when their own education information and SF-12 responses were recorded.

* p<0.05, ** p<0.01, *** p<0.001

Table 2.2: Regression Analysis of SF-12 Mental Health Score

SF-12 Mental Component Score	(1)	(2)	(3)	(4)
Average Years of Parental Education (1979)	0.0869	-0.203	0.0322	-0.0838
	(1.22)	(-0.78)	(0.35)	(-0.30)
Average Years of Parental Education - Squared		0.0148		0.00604
		(1.12)		(0.43)
Years of Education (Respondent at age 40)			0.446***	0.438***
			(4.46)	(4.35)
White			-0.577	-0.570
			(-0.90)	(-0.89)
Has a Healthcare Provider (1979)			0.741	0.771
			(0.79)	(0.82)
Net Family Income (1979)			0.0000140	0.0000129
			(0.52)	(0.49)
Constant	52.14***	53.37***	45.82***	46.39***
	(63.23)	(41.69)	(29.09)	(23.30)
Observations	1187	1187	896	896
Adjusted R-squared	0.000	0.001	0.018	0.017

Note: Unit of observation is a respondent. T-statistics are in parenthesis and based on robust standard errors. Average years of parental education is the mean average of respondent's mother and father's years of education, obtained in 1979. The respondents were surveyed again at the age of 40, when their own education information and SF-12 responses were recorded.
* p<0.05, ** p<0.01, *** p<0.001

Figure 1: Scatter Plot for SF12 Physical Scores Against Average Years of Parental Education

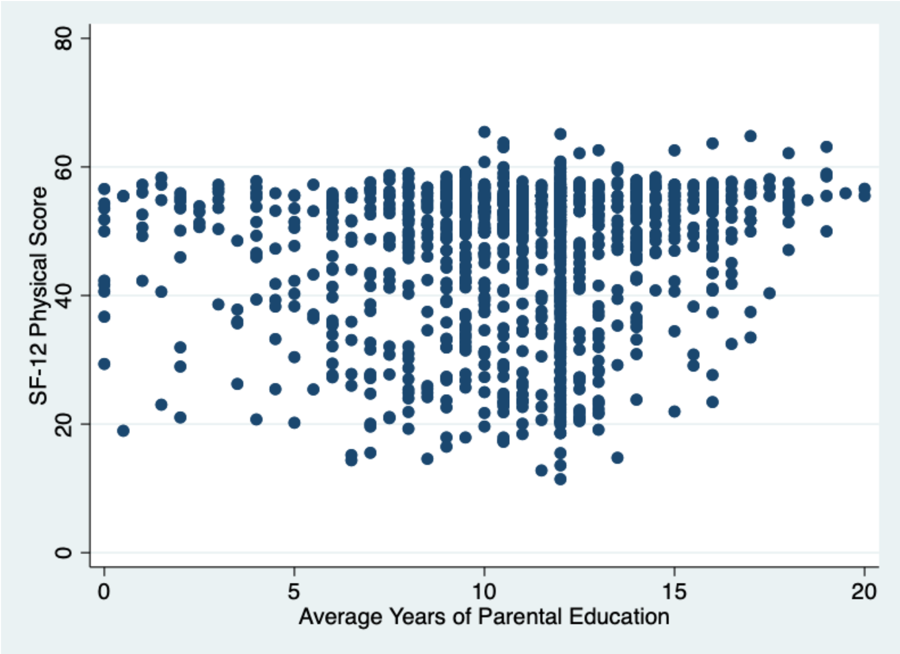

Figure 2: Scatter Plot for SF12 Mental Scores Against Average Years of Parental Education