**Challenging the Education Gradient of Health: the Case of the GED Recipients**

Hyun Sik Kim
hskim@ssc.wisc.edu

Postdoctoral Researcher

Wisconsin Scholars Longitudinal Study

Wisconsin Center for Education Research

University of Wisconsin-Madison

Direct all correspondence to Hyun Sik Kim, Wisconsin Scholars Longitudinal Study, University of Wisconsin-Madison, 1025 West Johnson Street, Suite 575, Madison, WI (hskim@ssc.wisc.edu)

Abstract

Large body of health literature has extensively reported the linearly aligned education gradient supporting thick theoretical building on the education gradient that has become one of few established facts. In this proposal, I attempt to describe apparently contradictory empirical observations to the gradient argument: the recipients of the General Educational Development (GED) certificate. The NHIS data show that GED recipients enjoyed no advantage in the global health status measured by self-reported health over high school dropouts, which do not appear support the widely accepted education gradient. Exploratory analysis using the Add Health also indicated that moderate benefits GED recipients reported in comparison with high school dropouts may be entirely attributable to pre-existing differences in school years. I suggested several competing hypotheses for these observations and plans for more rigorous analyses are outlined.

Challenging the Education Gradient of Health: the Case of the GED Recipients

**RESEARCH INTEREST**

The socioeconomic status (SES) gradient in health status is one of few established facts in population health research (Adler et al. 1994). This patterned observation states that those of higher SES as opposed to lower SES not only enjoy better health status but also the advantages are distributed in a linear fashion that each ascending step of SES ladder adds a similar amount of benefits. the SES gradient has been reported for extensive domains of adolescent health such as self-reported health and depression (Goodman 1999), chronic diseases (Goldman and Smith 2011), health behaviors such as smoking and drinking (Cutler and Lleras-Muney 2010), psychological symptoms such as depression (Lorant et al. 2003), functional limitations in older ages (Minkler, Fuller-Thomson, and Guralnik 2006), and infant mortality (Singh and Kogan 2007) as well as adult mortality (Elo and Preston 1996; Kitagawa and Hauser 1968) even though controversies remain to be solved for some outcomes such as body mass index (Sánchez-Vaznaugh et al. 2009).

Together with income, educational attainment has been the primary variable to proxy SES measure because the variable does not change over time above a certain age and income and occupation have been regarded as outcomes of educational attainment. Completed years of education has become such a pivotal variable in literature on the SES gradient that the education gradient has acquired its own independent meaning indicating linearly aligned differentials of health outcomes by education (Kimbro et al. 2008). It is notable that the recent stream of research on the education gradient goes beyond just reporting the patterned constellation of health differentials by assessing various theoretical hypotheses on which factors are responsible

for it (Baum II and Ruhm 2009; Conti and Heckman 2010). For instance, Cutler and Lleras-Muney (2010) examined numerous data sources to find that material advantages, knowledge and cognitive ability, and social networks may account for about 30, 30, and 10 percent of the education gradient, respectively.

Despite of this rich body of theoretical formulations and empirical evidence, recent development of education literature suggests that one group of educational attainment begs further investigation on the time-honored observation of the education gradient: the recipients of the General Educational Development (GED) certificate (Heckman and Rubinstein 2001). Under the theoretical prediction of the education gradient, on the one hand, GED recipients are supposed to be placed, if not higher than high school graduates, between high school graduates and high school dropouts, in their health status particularly because the GED is designed to test equivalence of cognitive skills to high school graduates (Heckman, Humphries, and Mader 2010). On the other, however, GED recipients would not benefit much for their efforts and skills after accounting for a sufficient set of confounding variables such as cognitive skills especially given that their ambiguous positions in labor market (Heckman and Rubinstein 2001).

I attempts to explore diverse theoretical formulations regarding this obscure dimension of the education gradient and to document empirical evidence accessible through several nationally representative data sets. To attain these research goals, this proposal will be developed into three main directions. First, I will report health disparities among high school dropouts, GED recipients, and high school graduates using self-reported global health measure collected from 1997 to 2003 in the National Health Interview Survey (NHIS: National Center for Health Statistics [NCHS] 2000). This step is designed to understand basic pictures of the education gradient and special attention will be given to heterogeneous constellation by age particularly

because some empirical evidence pointed toward lessening of the gradient in older ages (e.g. Cutler and Lleras-Muney 2006). In addition, I also note that further investigation by subgroups such as gender and race/ethnicity and other health measures is warranted for more promising research.

It is widely accepted that cross-sectional data are gravely limited in disentangling causes and consequences of the observed association without strongly theoretically-driven instrumental variable(s) (Rosenbaum 2002; Wooldridge 2002). Under our current context, for instance, we are not certain whether it is because severe health problems would prevent GED recipients from completing their high school courses or because their health would be deteriorated after GED test taking if we find lowered health status of GED recipients as opposed to high school graduates. The second part of this proposal will coarsely describe changes of health status among high school dropouts, GED recipients, and high school graduates before and after GED reception using the National Longitudinal Study of Adolescent Health (Add Health: Harris and Udry 2011). Based on these two broad observations, I develop analytical strategies that serve as building blocks of later analyses in the last part.

**GLOBAL HEALTH COMPARISON USING THE NHIS**

To get general sense of the education gradient of self-reported health among high school dropouts, GED recipients, and high school graduates, I provide Figure 1 which uses the pooled data of the NHIS from 1997 and 2003 (NCHS 2000). The global health measure asked "would you say health in general is very good, good, fair, or poor?" and I will treat the choice set as a continuous variable that included exclusively and exhaustively "excellent (=1)", "very good (=2)", "good (=3)", "fair (=4)", "poor (=5)". Self-reported global health has been repeatedly

reported to be strongly associated with future mortality and has become an essential measure to gauge a person's health status (Goldman and Smith 2011; Idler and Benyamini 1997). It is also worth mentioning that GED recipients included recipients of not only GED but also other equivalent certificates (question wording was "GED or equivalent"). To enhance data reliability and representativeness, I select only those respondents who reported the global health status on their own and all observations are weighted using the "wtfa" variable.

[Figure 1 about here]

Several observations stand out from Figure 1. 1) In general, self-reported global health seems to have deteriorated as people aged until they reached 50s and then leveled off, 2) this trend by age does not fit to high school graduates who appears to have experienced continuous declines over time. Of course, we can not determine whether these patterns are due to age effects or cohort effects from this figure specifically and from my analyses using the NHIS generally (Preston, Heuveline, and Guillot 2001). More relevant to my current research objectives, 3) there was no advantage gained from receiving the GED credential in early years such 20s and 30s, 4) however, some advantages emerged from 40s so that linearly aligned health disparities by education is evident in their 70s and over. It is tempting to argue that, related to the first point above, the trajectory of the self-reported health of GED recipients follow an inverted U-shape with a peak at 50s but comparison of confidence intervals across ages seems to indicate that this argument is too strong to be accepted.

To further explore the education gradient to more detailed extent, I disaggregate the high school dropouts by their specific years of schooling using the same data and methods. Figure 2

5

summarizes graphical patterns of point estimates and confidence intervals. The smallest number of observations in cells of Figure 2 is 483 so that problems of small numbers of observations may not be an issue in the two graphs, which is exactly why I pooled 7 years of the NHIS data in exchange for confounding among age, period, and cohort effects.

[Figure 2 about here]

The most striking finding from Figure 2 is that the overall health of GED recipients lagged behind that of respondents who attended 12th grade but not graduated high schools except the age interval of 70s and over for which I observed a well-slated education gradient. More specifically, it deems reasonable to state that the global health of GED recipients remained in the level of those respondents with less than 12 years of schooling and traced the similar trajectories with them across ages. To summarize my findings up to now, 1) GED recipients did not appear to enjoy health advantages that should have derived from their cognitive skills as opposed to high school dropouts except very limited age intervals and 2) they seems to have experienced deteriorated, rather than improved, health status compared to 12th grade dropouts. 3) However, the global health of GED recipients was likely to have been enhanced over the life course compared to dropouts before 12th grade.

How can we explain these observations? Several hypotheses would be come up with. Particularly regarding the second observation, measurement problem may have been involved in the completed years of schooling in a way that many high school graduates would have misreported to be 12th grade dropouts. When it comes to comparisons with high school graduates, it is plausible to hypothesize that GED recipients would be thought to have lower

levels of noncognitive traits (Heckman et al. 2010; Heckman and Rubinstein 2001) so that they were treated equal to high school dropouts in labor market as well as in society in general, which may have induced psychological distress leading to deterioration of the global health status. Or selection argument would maintain that compared to high school graduates, GED recipients were in worse health status before they took the GED test. As to a similar level of the global health with dropouts in early ages but emerging disparities over time, a causal argument would state that the global health of GED recipients improved after receiving the credentials to the level of high school dropouts. By contrast, selection perspective would argue that those dropouts who felt better would have taken the GED test for various reasons. In a word, it is very difficult to choose any one among competing hypotheses using Figure 1 and 2. To get more concrete evidence for or against the competing hypotheses, I describe some observations from the Add Health in the next section.

**GLOBAL HEALTH COMPARISONS USING THE ADD HEALTH**

Figure 3 displays point estimates and their confidence intervals among three groups of educational attainment, recovered from the Add Health (Harris and Udry 2011). It should be clear that all respondents in the sample were in schools when they responded in the initial survey. Thus, the education group of the initial survey in Figure 3 was that observed in the third follow-up. This figure is designed to get a descriptive picture on constellation of the global health before GED recipients took the test. It is also informative to mention that there were 298, 194, and 724 observations for high school dropouts, GED recipients, and high school graduates respectively. To obtain population estimates, I weighted the sample data using the "gswgt4_2" variable.

[Figure 3 about here]

The main findings from Figure 3 are three-fold: 1) GED recipients showed a similar level of the global health with high school graduates rather than high school dropouts, which is very different from the NHIS data in comparable age intervals, 2) GED recipients were at a lower level than high school graduates but a higher level than high school dropouts even before they took the test and 3) the initial health disparities did not change after they succeed obtaining the credential. Namely, Figure 3 appears to support the selection hypothesis rather than the causal argument, which is quite contradictory to the education gradient argument while being consistent with Conti and Heckman's work (2010) showing that dominant portion of the education gradient is attributable to early childhood characteristics. However, these observations are all descriptive and quite limited because of existence of confounding variables not accounted for in these figures. To reach more solid conclusions, we need to undertake statistical tests under some restrictive assumptions as discussed in the following section.

**FUTURE ANALYSIS PLANS**

It should be clear that my research objectives to employ statistical models are to see whether there are distinguishable selection effects of health status on GED test-taking and to determine whether there are causal effects of the GED credential on health status after taking the selection effects into account. To attain those goals, I will concentrate on comparisons between GED recipients and high school dropouts using the Add Health data to unravel selection and causal effects of the GED credentials. I will experiment four statistical models widely used in the

modern statistics literature: Ordinary Least Squares, a matching estimator (Abadie and Imbens 2006; Rosenbaum and Rubin 1983), a doubly robust estimator (Bang and Robins 2005; Lunceford and Davidian 2004) and Heckman selection model (Heckman 1979; Heckman and Navarro-Lozano 2004).

Each model has its own advantages and disadvantages so that comparing results across several different statistical methods will shed more illuminating light on the data generating process than just using one method. For instance, propensity scores matching method is widely used in recent causal inference literature but it has limited utility for my data set because the sample sizes not only of the treated but also of the controlled are small. In this respect, Heckman selection model (or control function approach) look more promising for my data but its distributional assumption is quite burdensome. Doubly robust estimator has its own advantages by modeling both the treatment variable (the GED credential) and outcome variable (the global health) but it is known that this method is also subject to strong distributional assumptions (for a debate on the doubly robust estimator, for instance, see *Statistical Science* 2007 [Volume 22 Number 4]). Therefore, multiple analyses will provide more reliable and robust inference for the underlying data generating processes.

Reference

Abadie, Alberto and Guido W. Imbens. 2006. "Large Sample Properties of Matching Estimators for Average Treatment Effects." *Econometrica* 74(1):235-267.

Adler, Nancy E., Thomas Boyce, Margaret A. Chesney, Sheldon Cohen, Susan Folkman, Robert L. Kahn, and S. Leonard Syme. 1994. "Socioeconomic Status and Health: The Challenge of the Gradient." *American Psychologist* 49: 15-24.

Bang, Heejung and James M. Robins. 2005. "Doubly Robust Estimation in Missing Data and Causal Inference Models." *Biometrics* 61: 962-72.

Baum II, Charles L. and Christopher J. Ruhm. 2009. "Age, Socioeconomic Status and Obesity Growth." *Journal of Health Economics* 28: 635-648.

Conti, Gabriella and James J. Heckman. 2010. "Understanding the Early Origins of the Education-Health Gradient: A Framework That Can Also Be Applied to Analyze Gene-Environment Interactions." *Perspectives on Psychological Science* 5: 585-605.

Cutler, David M. and Adriana Lleras-Muney. 2010. "Understanding Differences in Health Behaviors by Education." *Journal of Health Economics* 29: 1-28.

Cutler, David M. and Adriana Lleras-Muney. 2006. "Education and Health: Evaluating Theories and Evidence." NBER Working Paper Series 12352. National Bureau of Economic Research. http://www.nber.org/papers/w12352

Elo, Irma T. and Samuel H. Preston. 1996. "Educational Differentials in Mortality: United States, 1979-85." *Social Science and Medicine* 42: 47-57.

Goldman, Dana and James P. Smith. "The Increasing Value of Education to Health." *Social Science and Medicine* 72: 1728-37.

Goodman, Elizabeth. 1999. "The Role of Socioeconomic Status Gradients in Explaining Differences in US Adolescents' Health." *American Journal of Public Health* 89: 1522-28.

Harris, Kathleen M. and J. Richard Udry. 2011. *National Longitudinal Study of Adolescent Health (Add Health), 1994-2008 [Computer file]*. ICPSR21600-v4. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2011-03-18. doi:10.3886/ICPSR21600.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47: 153-61.

Heckman James J., John E. Humphries, and Nicholas S. Mader. 2010. "The GED." NBER Working Paper Series 16064. National Bureau of Economic Research. http://www.nber.org/papers/w16064

Heckman, James J. and Salvador Navarro-Lozano. 2004. "Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models." Review of Economics and Statistics 86(1):30-57.

Heckman James J. and Yona Rubinstein. 2001. "The Importance of Noncognitive Skills: Lessons from the GED Testing Program." *American Economic Review* 91: 145-149.

Idler, Ellen L. and Yael Benyamini. 1997. "Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies." *Journal of Health and Social Behavior* 38: 21-37.

Kimbro, Rachel T., Sharon Bzostek, Noreen Goldman, and Germán Rodrĭguez. 2008. "Race, Ethnicity, and the Education Gradient in Health." *Health Affairs* 27: 361-372.

Kitagawa, Evelyn M. and Philip M. Hauser. 1968. "Education Differentials in Mortality by Cause of Death: United States, 1960." *Demography* 5: 318-53.

Lorant,Vincent, D. Deliège, W. Eaton, A. Robert, P. Philippot, and M. Ansseau. 2003.

 "Socioeconomic Inequalities in Depression: A Meta-Analysis." *American Journal of*

 *Epidemiology* 157: 98-112.

Lunceford, Jared K. and Marie Davidian. 2004. "Stratification and Weighting via the Propensity

 Score in Estimation of Causal Treatment Effects: A Comparative Study." *Statistics in*

 *Medicine* 23: 2937-60.

National Center for Health Statistics. 2000. *Data File Documentation, National Health Interview*

 *Survey, 1997 (machine readable data file and documentation).* Hyattsville, MD: National

 Center for Health Statistics.

Minkler, Meredith, Esme Fuller-Thomson, and Jack M. Guralnik. 2006. "Gradient of Disability

 across the Socioeconomic Spectrum in the United States." *New England Journal of*

 *Medicine* 355: 695-703.

Preston, Samuel H., Patrick Heuveline, and Michel Guillot. 2001. *Demography: Measuring and*

 *Modeling Population Processes.* Malden, MA: Blackwell Publishers.

Rosenbaum, Paul R. 2002. *Observational Studies*, 2nd ed. New York: Springer-Verlag.

Rosenbaum, Paul R. and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in

 Observational Studies for Causal Effects." Biometrika 70(1):41-55.

Sánchez-Vaznaugh, Emma V., Ichiro Kawachi, S.V. Subramanian, Brisa N. Sánchez, and

 Dolores Acevedo-Garcia. 2009. "Do Socioeconomic Gradients in Body Mass Index Vary

 by Race/Ethnicity, Gender, and Birthplace?" *American Journal of Epidemiology* 169:

 1102-12.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data.*

 Cambridge, MA: MIT press.

Figure 1. Self-reported Health by Usual Education Category and Age
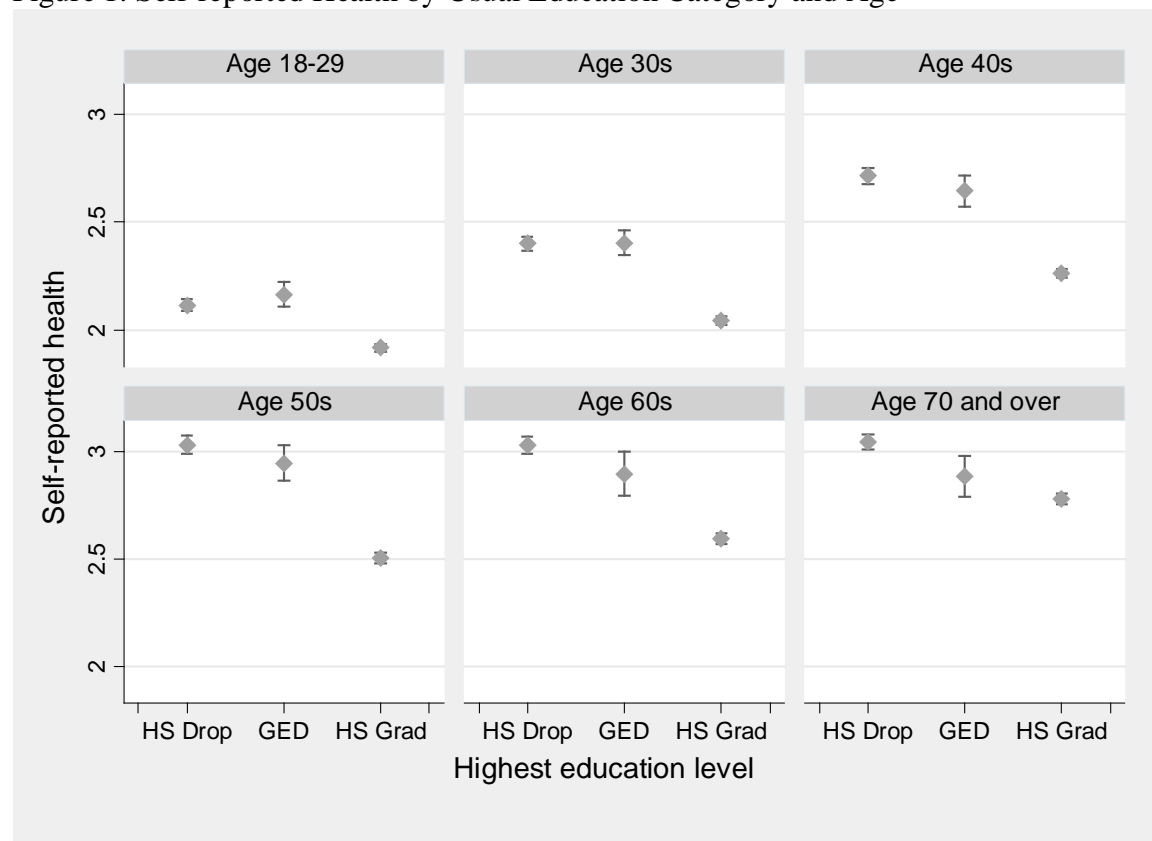


Notes. Pooled data of the National Health Interview Survey from 1997 to 2003. Point estimates and confidence intervals are presented. Value 1 means "excellent" and value 5 denotes "poor." See the text for more information.

Figure 2. Self-reported Health by One-Year Education Category and Age
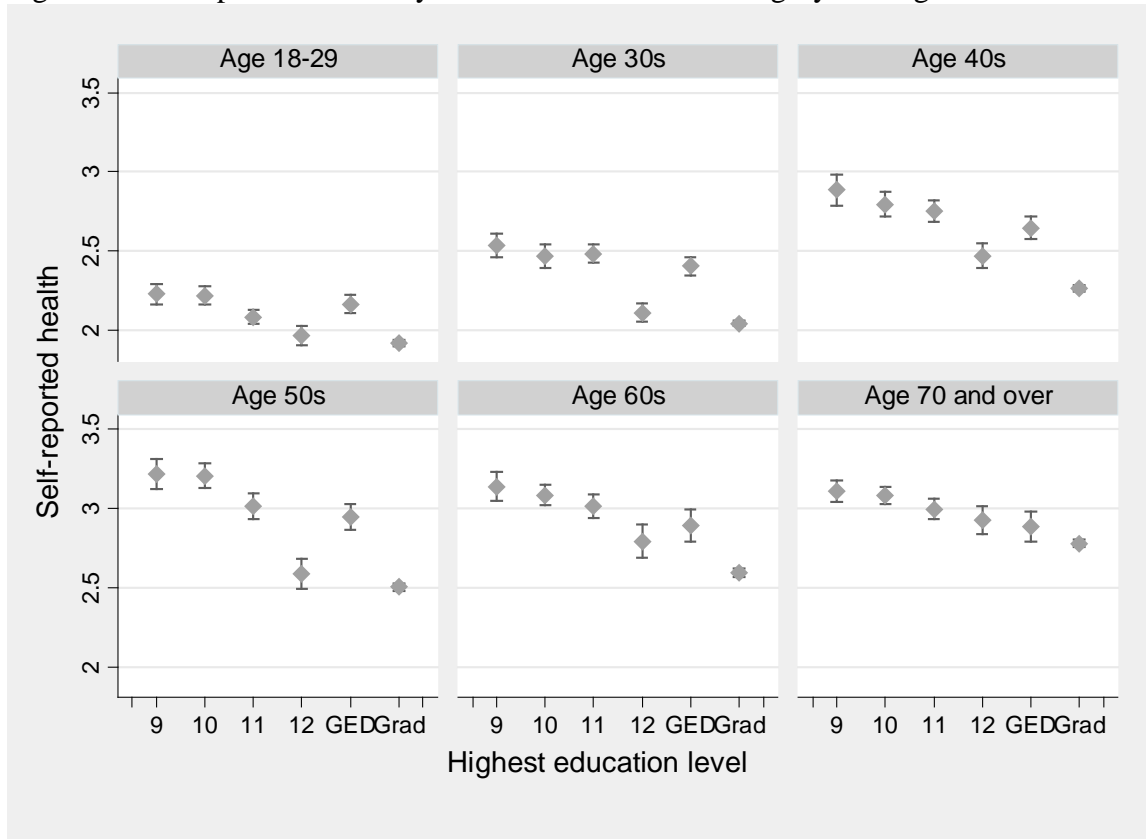


Notes. Same data are used as in Figure 1.

Figure 3. Self-reported Health by Education and Wave from the Add Health
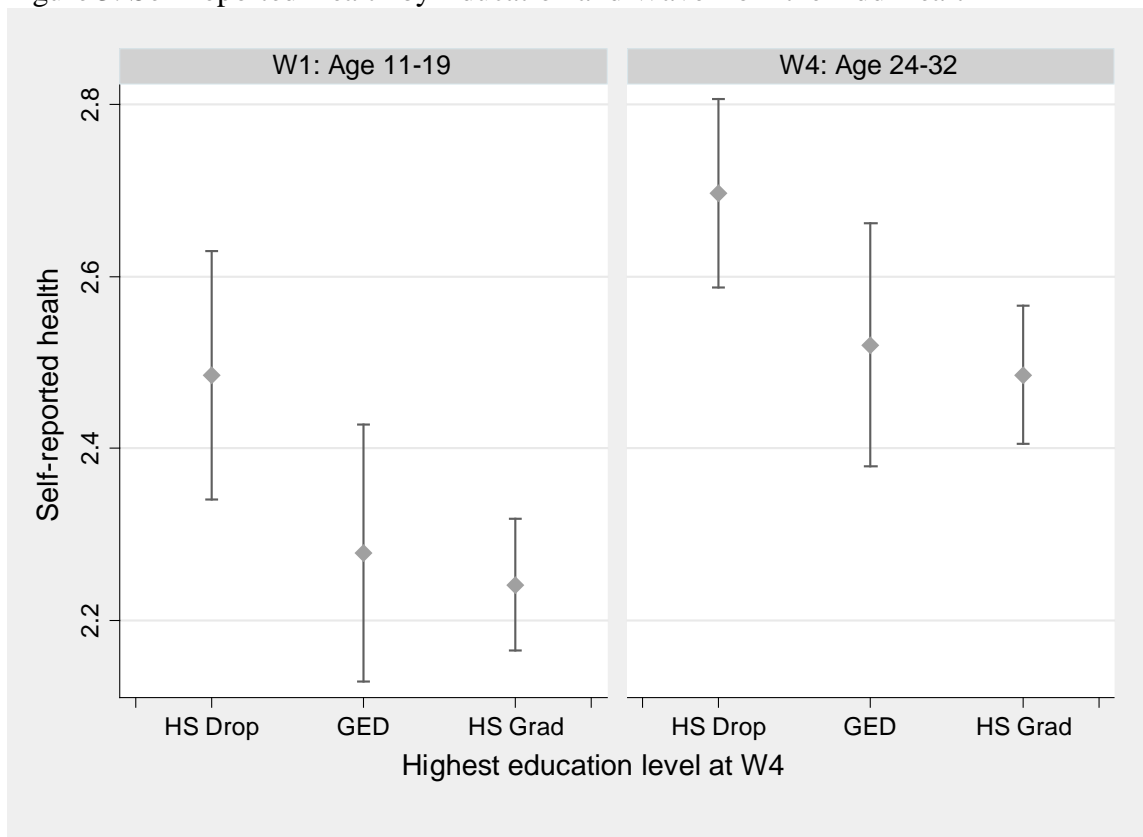


Notes. W1 means the initial survey and W4 denotes the third follow-up survey.