1-2016

Instrumental variable approaches to identifying the causal effect of educational attainment on dementia risk

Thu T. Nguyen

Eric Tchetgen Tchetgen

Ichiro Kawachi

Stephen E. Gilman

Stefan Walter

See next page for additional authors

HHS Public Access

Author manuscript

Ann Epidemiol. Author manuscript; available in PMC 2017 January 01.

Published in final edited form as:

Ann Epidemiol. 2016 January ; 26(1): 71–76.e3. doi:10.1016/j.annepidem.2015.10.006.

Instrumental variable approaches to identifying the causal effect of educational attainment on dementia risk

Thu T. Nguyen[a], **Eric J. Tchetgen Tchetgen**[b,c], **Ichiro Kawachi**[d], **Stephen E. Gilman**[c,d,g], **Stefan Walter**[a], **Sze Y. Liu**[e], **Jennifer Manly**[f], and **M. Maria Glymour**[d]

[a]University of California, San Francisco, Department of Epidemiology & Biostatistics

[b]Harvard School of Public Health, Department of Biostatistics

[c]Harvard School of Public Health, Department of Epidemiology

[d]Harvard School of Public Health, Department of Social and Behavioral Sciences

[e]Harvard School of Public Health, Harvard Center for Population and Development Studies

[f]Columbia University Medical Center, Department of Neurology, Sergievsky Center and the Taub Institute for Research on Alzheimer's Disease and the Aging Brain

[g]*Eunice Kennedy Shriver* National Institute of Child Health and Human Development, Division of Intramural Population Health Research

Abstract

Purpose—Education is an established correlate of cognitive status in older adulthood, but whether expanding educational opportunities would improve cognitive functioning remains unclear given limitations of prior studies for causal inference. Therefore, we conducted instrumental variable (IV) analyses of the association between education and dementia risk, using for the first time in this area, genetic variants as instruments as well as state-level school policies.

Methods—IV analyses in the Health and Retirement Study cohort (1998–2010) used two sets of instruments: 1) a genetic risk score constructed from three single nucleotide polymorphisms (SNPs) (n=8,054); and 2) compulsory schooling laws (CSLs) and state school characteristics (term length, student teacher ratios, and expenditures) (n=13,167).

Results—Employing the genetic risk score as an IV, there was a 1.1% reduction in dementia risk per year of schooling (95% CI: −2.4, 0.02). Leveraging compulsory schooling laws and state school characteristics as IVs, there was a substantially larger protective effect (−9.5%; 95% CI: −14.8, −4.2). Analyses evaluating the plausibility of the IV assumptions indicated estimates derived from analyses relying on CSLs provide the best estimates of the causal effect of education.

Corresponding author: Thu T. Nguyen, University of California, San Francisco, 550 16th Street, 2nd floor, San Francisco, CA 94143-0560, Telephone: 617.963.9595, Thu.Nguyen@ucsf.edu.

Conflict of interest: none declared.

Conclusion—IV analyses suggest education is protective against risk of dementia in older adulthood.

Keywords

causal inference; dementia; education; instrumental variables; unmeasured confounding

The correlation between educational attainment and later life cognitive function and dementia is well documented (1–4). Using a large population based cohort studies, Brayne et al. found a dose effect of education such that more education was associated with reduced dementia risk independently of severity of pathology (5). A recent meta-analysis of 19 observational studies reported a relative risk of 1.33 (95% CI: 1.15, 1.54) comparing all cause dementia among those with low or medium levels of education compared to those with a high level of education (6). Despite the numerous replications, all prior studies share an essential weakness, in that inferences rest on the strong assumption that there are no unmeasured common causes of educational attainment and dementia. Potential confounders include childhood health status, cognitive abilities, and socioeconomic circumstances, all of which influence educational attainment and are likely risk factors for dementia (1, 7).

The current study attempts to mitigate the confounding biases present in existing observational studies of education and dementia using instrumental variables (IV). IVs provide an opportunity for causal inference even in the presence of unmeasured confounders. Genetic variants have proven to be powerful instruments for addressing the causal effects of putative exposures (e.g., in so-called Mendelian randomization studies (8, 9)). Recent research identified 3 single nucleotide polymorphisms (SNPs) that together predict education, thus allowing for the first time the possibility of using genetic variants as instruments for the effects of education. In addition, the current study uses a second set of instruments based on state-level schooling policies. School policies have previously been used as instruments to estimate the effects of education on health, with the most promising results related to cognitive outcomes (10–13). Recognizing that IVs depend on strong assumptions, we used two different sets of instruments to investigate this research question.

METHODS

The Health and Retirement Study (HRS) is a national, longitudinal study of individuals 50 years of age or older and their spouses. The first survey wave was collected in 1992, with biennial interviews (or proxy interviews for decedent participants) available through 2010. New cohorts were added in 1993, 1998, 2004, and 2010. We utilized follow-up data from 1998 – 2010 and includes individuals from all enrollment cohorts except 2010. Survey response rates ranged from 70 to 82%, and retention rates through 2008 ranged from 86% to 91%. HRS was approved by the University of Michigan Health Sciences Human Subjects Committee, and the Harvard School of Public Health Human Subjects Committee determined the current analyses were exempt.

Sample

The two IV analyses based on school policies and genetic information used different analytic samples. Three data sources were used for the analyses involving school policies: HRS; historical federal reports on compulsory schooling laws (CSLs) and school characteristics, and state characteristics; and the 1980 census micrososample (n=2,536,876) (14). The census sample was used to estimate effects of CSLs and school characteristics on education. We restricted the HRS and census sample to match on race/ethnicity, state of birth, nativity, birth year, and education.

Individual health outcome data came from HRS. From an initial total sample of 30,670 members in HRS, we excluded individuals younger than 50 at the beginning of follow-up (defined as 2004 for the 2004 enrollment cohort or 1998 for all others), were foreign-born, with unknown place of birth, with more than 12 years of education, or missing data on education, covariates, or dementia risk. Dementia outcomes were not available for Hispanics, so they were excluded. Participants with greater than 12 years of schooling were excluded because CSLs and school characteristics did not influence years of schooling beyond primary and secondary school in our analyses. The final analytic sample for the school policy IV study included 10,955 participants.

Similarly to the previous sample, for the analyses using genetic data, we excluded participants who were younger than 50, were foreign-born, or had an unknown place of birth in our analyses using genetic data. Of the respondents who met the above exclusion criteria, 9,911 were genotyped. The analytic sample included individuals with 12+ years of education because the single nucleotide polymorphisms (SNPs) were found to predict college completion as well as average years of education. However, racial and ethnic minorities were excluded from these analyses because the genome-wide association study (GWAS) identifying these SNPs included only Caucasians (15) and the education genetic risk score was not related to education among Non-Whites in HRS. The final analytic sample for the genetic IV study included 7,981 respondents.

Measures

Exposures—The main exposure of interest was educational attainment operationalized as self-reported years of schooling.

Outcomes—Immediate and delayed recall of a 10-item word list, the Telephone Interview for Cognitive Status (TICS), and the Informant Questionnaire for Cognitive Decline (IQCODE) were used to construct an overall dementia probability score. The dementia probability score achieved a c-statistic of 94.3% in predicting DSM-IV diagnosed dementia (16) Scores can range from 0 (no chance this individual would meet diagnostic criteria) to 1 (individual certain to meet diagnostic criteria). Our current knowledge of dementia is that it is an insidious disease that can develop over decades (17). The moment of diagnosis is somewhat arbitrary and does not necessarily mark the transition from non-disease to disease state. For this reason, we consider the continuous score to more closely correspond with the underlying disease. To reduce random measurement error in dementia score, we averaged available repeated measurements of dementia probability from 1998–2010.

Compulsory schooling laws and school characteristics—HRS respondents were linked to school policy characteristic that would have affected their schooling, based on the year and state of the respondent's birth. Compulsory schooling laws (CSLs) from 1906 to 1978 were compiled by Lleras-Muney (18), Angrist and Acemoglu (19), and Glymour (12) using federal education reports usually available biennially. Data were collected on mandatory age at school enrollment, youngest age when it was legal to drop out of school, and youngest age when individuals could receive a work permit. For years without data, we carried forward the most recently reported value of the state policy variable. For each respondent, years of compulsory schooling were calculated by taking the difference between enrollment age when respondents were 6 and minimum drop-out age (CSL) or minimum work permit age (CSL-w) when the respondents were 14. Ranges of CSL and CSL-w were 6–12 and 0–10 years, respectively with 0 indicating the state did not have a law specifying work permit age.

State average school term length (1905–1957), student-teacher ratios (1907–1955), and per-pupil expenditures (1907–1943) were compiled by Glymour and Manly (personal communication, November 2, 2012) from biennial state reports. For each respondent, we calculated the average term length, student-teacher ratio, and per pupil expenditure when that respondent was 6–14 years of age in the state where he/she was born.

State characteristics—Percentage black, urban, and foreign born when the respondents were 6 and manufacturing jobs per capita and manufacturing wages per manufacturing job when the respondents were 14 were included as covariates in IV models using compulsory schooling laws and school characteristics as instruments. The state characteristics were compiled by Glymour (12) and Lleras-Muney (20) using Statistical Abstracts of the United States and federal manufacturing employment data. School and state characteristics were linearly interpolated for the years between reports.

Genotyping—DNA samples were collected in 2006 and 2008 from HRS respondents. Details regarding the quality control procedures are available elsewhere (21).

Genetic risk score—Three independent single nucleotide polymorphisms (SNPs) (rs11584700, risk allele: A, rs4851266, T, rs9320913, A) have been previously identified as genome-wide significant ($p<5 \times 10^{-8}$) predictors of educational attainment in a large genome-wide association study (GWAS) meta-analysis (15). A genetic risk score for years of schooling was calculated for each individual by summing the product of the number of risk alleles at each locus (0–2) with its meta-analyzed beta estimate (bolded) from Rietveld et al. (15) (Eq 1).

$$\text{Genetic Risk Score} = (\text{rs11584700\_ A} * -\mathbf{0.095} + \text{rs4851266\_ T} * \mathbf{0.082} + \text{rs9320913\_ A} * \mathbf{0.101}) \quad \text{(Equation 1)}$$

Statistical Analysis

To yield valid results, all instruments must meet three main assumptions. First, the instruments must predict the exposure. Second, the instruments must affect the outcome *only*

through the exposure. Lastly, the instruments must not share unmeasured common causes with the outcome.

We completed two sets of IV analyses, one using CSL and school characteristics and the other using education genetic risk score as the instrument. For both sets of instruments, we used a separate-sample approach to implement IV analyses. In a separate sample approach, external data are used for beta estimates for the effect of the IVs on years of schooling, eliminating "weak instruments bias," which can bias IV estimates toward conventional estimates (22, 23). To understand why, we can imagine using set of instruments that has no causal effect on our exposure of interest. However, when combining a large number of putative instruments, they will slightly predict the exposure of interest. The predicted value of the exposure will be correlated with unmeasured causes of the exposure within that sample, including confounders that influence both the exposure and outcome. When the predicted value is used in the second stage, it will be related to the outcome because it is correlated with the confounders. This results in an estimate that is biased towards the conventional covariate estimate. This phenomenon does not occur when using separate samples for the two stages because the instruments that by chance best predicts the exposure in one sample is unrelated to the confounders of the exposure-outcome association in the second sample (23).

Statistical analysis using CSLs and school characteristics—The first stage was implemented using the 1980 census 5% sample (14), which is nearly 200 times larger than HRS. In the first stage, we fit linear regression models using compulsory schooling laws, school characteristics, and covariates to predict educational attainment (Models 1). Beta coefficients from the first stage were used to obtain predicted years of schooling, which was then employed as an independent predictor of dementia in HRS (Model 2a–b).

$$1^{\text{st}}\ \text{stage:Predicted}_{\text{education}}=\beta_0+\beta_1\text{CSL}+\beta_2\text{CSLw}+\beta_3\text{Term\_ length}+\beta_4\text{Student- teacher\_ ratio}+\beta_5\text{Per- pupil\_ expenditures}+\beta_6\text{Student- teacher\_ ratio}+\beta_7\text{C} \quad (1)$$

$$2^{\text{nd}}\ \text{stage:Dementia risk}=\eta_0+\eta_1\text{Predicted}_{\text{education}}+\eta_2\text{C}+\text{e} \quad (2a)$$

In these equations, C is a vector of covariates included in both equations: sex, self-assessed race/ethnicity (White, Black, Other), birth year indicators, state of birth indicators, and the state characteristics mentioned above.

Statistical analysis using education genetic risk score—We implemented the separate sample IV analyses using the education genetic risk score by applying the meta-analyzed beta estimates for the association between the SNPs and years of schooling reported from Rietveld et al. (15) as described above (23). We then regressed dementia probability and logit transformed dementia scores on the education genetic risk score, adjusting for sex, age at first outcome assessment, age^2, early life socioeconomic status (SES), and eigenvectors to control for population substructure. Early life SES combined into a single scale father's occupational status, birth in southern US, rural residence during childhood, and mother's and father's educational attainment, following previous research in

HRS (24). We formally tested whether IV estimates were significantly different from the conventional estimate where covariates are added to the model to control for confounding with the Wu-Hausman endogeneity test (25). In this paper, we refer to conventional estimates as covariate adjusted (CA) estimates.

As a sensitivity analysis, we logit transformed the dementia scores because the untransformed dementia probability scores clustered at the low end and were not normally distributed. However, the results were qualitatively similar to the untransformed scores. The results are presented in supplemental materials (eTable 1). Analyses were performed using Stata 12.

RESULTS

Table 1 shows the demographics of the analytic sample used for the two IV analyses. Using the 1980 census 5% sample in the first stage, compulsory schooling laws (CSLs), school to work laws (CSL-w), and school characteristics were significantly associated with years of schooling. F statistics, an indication of the strength of the instruments for the excluded instrument, were well above 10 (Table 2). The linear regression coefficients for the effects of CSLs and school characteristics on years of schooling between the census and HRS samples were similar, demonstrating that comparable associations existed in both samples but using the census afforded much greater precision in the first-stage estimation (Table 3).

Results Using CSLs and school characteristics—IV estimates indicated each additional year of schooling predicted lower dementia probability ($\beta = 9.5$ percentage points; 95% CI: −14.8, −4.2; $P < 0.001$). Similarly, results from CA estimates indicated a protective effect of years of schooling on dementia. The probability of dementia decreased by 2.1 percentage points (95% CI: −2.3, −1.9; $P < 0.001$) for each year increase in schooling (first row of Table 4). Wu-Hausman endogeneity tests showed the IV estimate was significantly different than the CA estimates for the effect of years of schooling on dementia probability (Table 4).

Results Using Education Genetic Risk Score—Covariate adjustment estimates revealed years of schooling was inversely related to dementia. In the IV analysis, dementia probability decreased by −1.1 percentage points for each year of schooling (95% CI: −2.4, 0.02; $P = 0.11$). Wu-Hausman endogeneity tests showed the IV estimate was not different than the CA estimate (Table 4).

Sensitivity Analyses

For both sets of instruments, sensitivity analyses were conducted using 2010 dementia scores rather than average 1998–2010 scores and yielded qualitatively similar (results not shown). In addition, we performed additional analyses to examine the IV assumption that the instruments only influence dementia via education, also known as the exclusion restriction (See eFigure 1, eTable 2–3) (26).

DISCUSSION

The IV estimates indicate education reduces dementia probability. These results are consistent with previous IV literature reporting a protective effect of education on cognitive function (10–12). Education may be an indicator of cognitive reserve, with higher educational attainment indicating larger reserve to delay the onset of dementia (27, 28).

In the IV analyses, we used state of birth and year of birth to classify the compulsory schooling required for each participant. There may be misclassification of the CSLs and school characteristics for participants in which the state they were born is different from the state where they received their schooling. Ninety-one percent of children 5–12 years of age in the 1940 Census IPUMS data lived in the stated they were born in (29), suggesting moves out of state during school years may be limited.

This study attempts to address a major threat to validity in observational studies of education and dementia. Compared to the 19 observational studies reported in the Caamano-Isorna et al.'s meta-analysis (6), this analysis is the *only* study that is valid even if there are unmeasured confounders of education and dementia. This strength comes at the expense of a different set of assumptions. We assume the IV effects are fully mediated by education, i.e. no pathway between instruments and dementia that does not involve education. A pathway that we worry about is through intelligence or cognitive function. This is not plausible for CSLs or state school characteristics, which are identified from state and year of birth, thus avoiding correlations with any individual level variables. However, it may be that the education genetic risk score influences cognitive function (not through education) and goes on to influence dementia risk. We found that the GRS predicted dementia (Table 4) as strongly as it predicted education (Table 3) (see Online Supplemental Materials for analyses examining IV assumptions). To our knowledge, this paper represents for the first time a genetic risk score has been used as an instrument to investigate the relationship between education and health. One contribution of this paper is to clearly present the challenges of using genetic data for social science research. However, looking at the results and the analyses, we consider CSLs more plausible IVs and therefore prefer the estimates given by CSLs and school characteristics.

IV estimates using state policies are larger than estimates based on the education genetic risk score. There are a few potential reasons for this difference. It may that the IV estimates apply to different populations since we used different subsamples. IV estimates utilizing compulsory schooling laws and state characteristics are generalizable to the subgroup of individuals who receive more years of schooling because of their state's CSL. IV estimates utilizing the SNPs apply to people whose behavior is influenced by the SNPs. Of note, alternative interpretations of IV estimates with different assumptions are possible (30, 31). Also while a 1% and 9% point estimates may seem dissimilar, both estimates have 95% confidence intervals that are just 2% apart. Another explanation may be that education genetic risk score is not a valid IV, specifically violating assumption 2.

For the IV analyses using CSLs, we restricted the sample to those with 12 or fewer years of education because these laws did not refer to college attendance and did not influence years

of schooling beyond primary and secondary school. Education beyond 12 years may further protect against dementia. For the IV analyses using the education genetic risk score, we did not need to make this exclusion since the instrument predicted education beyond 12 years.

IV analyses are a promising approach to identify determinants of dementia when hypothesized risk factors are potentially confounded. However, the validity of IV models rests on strong assumptions, so presentation of results alongside sensitivity analyses and clarifying assumptions of the assessments is critical. For many substantive important research questions, RCTs are not feasible. In such situations, triangulation of evidence from multiple sources and study designs, each relying on different assumptions, is most compelling. This paper represents such an effort in triangulation.

In summary, our findings support the hypothesis that education reduces the risk of dementia in older adults, and increases in educational attainment may lead to lower risk of dementia in later life. This would support potential preventive interventions based on educational activities and also predicts that global increases in average educational attainment will lead to global reductions in dementia.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding:

E.J.T.T.'s work was funded by the National Institutes of Health (NIH) grants AI113251 and ES020337.

S.E.G.'s work was supported by the Intramural Research Program of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development

References

1. Cobb JL, Wolf PA, Au R, White R, D'Agostino RB. The effect of education on the incidence of dementia and Alzheimer's disease in the Framingham Study. Neurology. 1995 Sep; 45(9):1707–12. [PubMed: 7675231]
2. The Canadian Study of Health and Aging: risk factors for Alzheimer's disease in Canada. Neurology. 1994 Nov; 44(11):2073–80. [PubMed: 7969962]
3. Glatt SL, Hubble JP, Lyons K, Paolo A, Troster AI, Hassanein RE, et al. Risk factors for dementia in Parkinson's disease: effect of education. Neuroepidemiology. 1996 Jan-Feb;15(1):20–5. [PubMed: 8719045]
4. Plassman BL, Welsh KA, Helms M, Brandt J, Page WF, Breitner JC. Intelligence and education as predictors of cognitive state in late life: a 50-year follow-up. Neurology. 1995 Aug; 45(8):1446–50. [PubMed: 7644038]
5. Brayne C, Ince PG, Keage HA, McKeith IG, Matthews FE, Polvikoski T, et al. Education, the brain and dementia: neuroprotection or compensation? Brain. 2010 Aug; 133(Pt 8):2210–6. [PubMed: 20826429]
6. Caamano-Isorna F, Corral M, Montes-Martinez A, Takkouche B. Education and dementia: a meta-analytic study. Neuroepidemiology. 2006; 26(4):226–32. [PubMed: 16707907]
7. Kawachi I, Adler NE, Dow WH. Money, schooling, and health: Mechanisms and causal evidence. Annals of the New York Academy of Sciences. 2010; 1186(1):56–68. [PubMed: 20201868]

8. Asvold BO, Bjorngaard JH, Carslake D, Gabrielsen ME, Skorpen F, Davey Smith G, et al. Causal associations of tobacco smoking with cardiovascular risk factors: a Mendelian randomization analysis of the HUNT Study in Norway. Int J Epidemiol. 2014 May 26.
9. Palmer TM, Sterne JAC, Harbord RM, Lawlor DA, Sheehan NA, Meng S, et al. Instrumental Variable Estimation of Causal Risk Ratios and Causal Odds Ratios in Mendelian Randomization Analyses. American Journal of Epidemiology. 2011; 173(12):1392–403. [PubMed: 21555716]
10. Huang W, Zhou Y. Effects of education on cognition at older ages: evidence from China's Great Famine. Soc Sci Med. 2013 Dec.98:54–62. [PubMed: 24331882]
11. Schneeweis, N.; Skirbekk, V.; Winter-Ebmer, R. Institute for the Study of Labor (IZA) Discussion Paper. 2012. Does schooling improve cognitive functioning at older ages?.
12. Glymour MM, Kawachi I, Jencks CS, Berkman LF. Does childhood schooling affect old age memory or mental status? Using state schooling laws as natural experiments. J Epidemiol Community Health. 2008 Jun; 62(6):532–7. [PubMed: 18477752]
13. Banks J, Fabrizio M. The effect of education on old age cognitive abilities: evidence from a regression discontinuity design. Econ J. 2012; 122:418–47.
14. Ruggles, S.; Sobek, M.; Alexander, T.; Fitch, CA.; Goeken, R. Integrated Public Use Microdata Series: version 3.0 (machine-readable database), 2004. 2014. [cited 2014]; Available from: http://www.ipums.org
15. Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science. 2013 Jun 21; 340(6139):1467–71. [PubMed: 23722424]
16. Wu Q, Tchetgen Tchetgen EJ, Osypuk TL, White K, Mujahid M, Maria Glymour M. Combining direct and proxy assessments to reduce attrition bias in a longitudinal study. Alzheimer Dis Assoc Disord. 2013 Jul-Sep;27(3):207–12. [PubMed: 22992720]
17. Jack CR, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet neurology. 2010; 9(1):119. [PubMed: 20083042]
18. Lleras-Muney A. Were compulsory attendance and child labor laws effective? An analysis from 1915 to 1939. Journal of Law and Economics. 2002; (45):401–35.
19. Acemoglu, D.; Angrist, JD. Research NBoE. NBER Working Paper Series; working paper no 7444. Cambridge, MA: 1999. How large are the social returns to education? Evidence from compulsory schooling laws.
20. Lleras-Muney A. The Relationship between Education and Adult Mortality in the United States. The Review of Economic Studies. 2005; 72:189–221.
21. Weir, M. Quality Control Report for Genotypic Data. University of Washington; 2012.
22. Angrist JD, Krueger AB. Split-Sample Instrumental Variables Estimates of the Return to Schooling. Journal of Business & Economic Statistics 1995. 1995 Apr 01; 13(2):225–35.
23. Tchetgen, Tchetgen EJ.; Walter, S.; Glymour, MM. Commentary: building an evidence base for mendelian randomization studies: assessing the validity and strength of proposed genetic instrumental variables. Int J Epidemiol. 2013 Feb; 42(1):328–31. [PubMed: 23508417]
24. Nandi A, Glymour MM, Kawachi I, VanderWeele TJ. Using marginal structural models to estimate the direct effect of adverse childhood social conditions on onset of heart disease, diabetes, and stroke. Epidemiology. 2012 Mar; 23(2):223–32. [PubMed: 22317806]
25. Hausman JA. Specification Tests in Econometrics. Econometrica. 1978; 46(6):1251–71.
26. Glymour MM, Tchetgen Tchetgen EJ, Robins JM. Credible Mendelian Randomization Studies: Approaches for Evaluating the Instrumental Variable Assumptions. Am J Epidemiol. 2012; 175(4):332–9. [PubMed: 22247045]
27. Barnes DE, Tager IB, Satariano WA, Yaffe K. The Relationship Between Literacy and Cognition in Well-Educated Elders. J Gerontol A Biol Sci Med Sci. 2004; 59(4):390–5. [PubMed: 15071084]
28. Farfel JM, Nitrini R, Suemoto CK, Grinberg LT, Ferretti RE, Leite RE, et al. Very low levels of education and cognitive reserve: a clinicopathologic study. Neurology. 2013 Aug 13; 81(7):650–7. [PubMed: 23873971]

29. Liu SY, Manly JJ, Capistrant BD, Glymour MM. Historical Differences in School Term Length and Measured Blood Pressure: Contributions to Persistent Racial Disparities among US-Born Adults. PLoS One. 2015; 10(6):e0129673. [PubMed: 26076495]

30. Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? Epidemiology. 2006 Jul; 17(4):360–72. [PubMed: 16755261]

31. Swanson SA, Hernan MA. Commentary: how to report instrumental variable analyses (suggestions welcome). Epidemiology. 2013 May; 24(3):370–4. [PubMed: 23549180]

Table 1

Demographic Characteristics of Eligible HRS Sample Members

	CSLs & school characteristics		Genetic risk score sample	
	n	(%)	n	(%)
n	13,167	100.0	8,054	100.0
Male	5,654	42.9	3,556	44.2
Birth year				
<1914	248	1.9	53	0.7
1914–1921	2,524	19.2	509	6.3
1922–1930	3,137	23.8	1,692	21.0
1931–1941	5,545	42.1	3,360	41.7
1942–1947	1,505	11.4	1,365	17.0
1948–1953	208	1.6	1,075	13.4
Years of schooling				
<6	536	4.1	42	0.5
6–8	1,860	14.1	319	4.0
9–11	3,315	25.2	865	10.7
12	7,456	56.6	2,924	36.3
>12			3,904	48.5
Non-Hispanic White	10,460	79.4	8,054	100.0
Non-Hispanic Black	2,508	19.1		
Other	198	1.5		

Table 2

First Stage Linear Regression Models Predicting Years of Schooling with CSLs and School Characteristics as Instruments in 1980 Census 5% Sample

Instrument	N	1st stage estimate[a]	95% CI	P value	F statistic	Partial R^2
	2,536,876				496.95	0.08%
CSL		0.044	(0.039, 0.049)	<0.001		
CSL-w		0.033	(0.029, 0.037)	<0.001		
Term length[b]		0.029	(0.023, 0.034)	<0.001		
Student-teacher ratio		−0.014	(−0.015, −0.013)	<0.001		
Per-pupil expenses		−0.007	(−0.007, −0.006)	<0.001		

[a] Models were adjusted for sex, self-assessed race/ethnicity (Non-Hispanic White, Non-Hispanic Black, Other, birth year and state of birth indicators, and state characteristics (% black, % urban, and % foreign born when the respondents were 6 and manufacturing jobs per capita and manufacturing wages per manufacturing job when the respondents were 14).

[b] Term length in 10 day increments.

Table 3

First Stage Linear Regression Models Predicting Years of Schooling with CSLs, School Characteristics, and Education Genetic Risk Score in HRS Sample

Instrument	N	1st stage estimate	95% CI	P value	F statistic	Partial R^2
Using CSLs and school characteristics[a]	13,167				9.38	0.29%
CSL		0.046	(−0.025, 0.116)	0.21		
CSL-w		0.044	(−0.011, 0.098)	0.11		
Term length[b]		0.019	(−0.052, 0.089)	0.61		
Student-teacher ratio		−0.036	(−0.056, −0.016)	<0.001		
Per-pupil expenses		−0.016	(−0.023, −0.009)	<0.001		
Using genetic data[c]	7,981				2.85	0.11%
Education GRS		0.401	(−0.064, 0.866)	0.09		

[a] Models were adjusted for sex, self-assessed race/ethnicity (Non-Hispanic White, Non-Hispanic Black, Other), birth year and state of birth indicators, and state characteristics (% black, % urban, and % foreign born when the respondents were 6 and manufacturing jobs per capita and manufacturing wages per manufacturing job when the respondents were 14).

[b] Term length in 10 day increments.

[c] Models were adjusted for sex, age at first outcome assessment, age^2, early life SES (mother's and father's educational attainment (<8, 8–11, 12, 12+), father's occupational status (manual/unskilled service, professional/white collar), birth in southern US (yes/no), and rural residence during childhood (yes/no), and eigenvectors to control for population substructure. While most waves of HRS asked respondents to report the highest grade their parents completed, in the early waves between 1993 and 1996, respondents were asked whether their parents completed 8 or more years of schooling. 7.5 years was assigned if less than 8 years was reported and 8.5 if 8 or more years were reported.

Table 4

Covariate Adjustment (CA) and Instrumental Variable (IV) Estimates for the Effect of Years of Schooling on Dementia Using CSLs and School Characteristics and Education Genetic Risk Score (GRS) as Instrumentsf

	Instrumental Variable Estimate				Covariate Adjustment Estimate			
	N	**β_{IV}**	**95% CI**	**P value**	**β_{CA}**	**95% CI**	**P value**	**Wu-Hausman P value[c]**
Years of Schooling[a]	10,955	−0.095	(−0.148, −0.042)	<0.001	−0.021	(−0.023, −0.019)	<0.001	0.006
Edu GRS[b]	7,981	−0.011	(−0.024, 0.002)	0.11	−0.005	(−0.006, −0.004)	<0.001	0.39

[a]Model was adjusted for sex, self-assessed race/ethnicity (Non-Hispanic White, Non-Hispanic Black, Other), birth year and state of birth indicators, and state characteristics (% black, % urban, and % foreign born when the respondents were 6 and manufacturing jobs per capita and manufacturing wages per manufacturing job when the respondents were 14).

[b]Models was adjusted for sex, age at first outcome assessment, age^2, early life SES (mother's and father's educational attainment (<8, 8–11, 12, 12+), father's occupational status (manual/unskilled service, professional/white collar), birth in southern US (yes/no), and rural residence during childhood (yes/no), and eigenvectors to control for population substructure. While most waves of HRS asked respondents to report the highest grade their parents completed, in the early waves between 1993 and 1996, respondents were asked whether their parents completed 8 or more years of schooling. 7.5 years was assigned if less than 8 years was reported and 8.5 if 8 or more years were reported.

[c]Wu-Hausman test for equality between IV and CA estimates.