Annals of Epidemiology 26 (2016) 71–76

Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

Original article

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

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

[a] Department of Epidemiology & Biostatistics, University of California, San Francisco, San Francisco
[b] Department of Biostatistics, Harvard School of Public Health, Boston, MA
[c] Department of Epidemiology, Harvard School of Public Health, Boston, MA
[d] Department of Social and Behavioral Sciences, Harvard School of Public Health, Boston, MA
[e] Health Behavior Branch, Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Rockville, MD
[f] Harvard Center for Population and Development Studies, Harvard School of Public Health, Cambridge, MA
[g] Department of Neurology, Sergievsky Center and the Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Medical Center, New York, NY

ARTICLE INFO

Article history:
Received 21 October 2014
Accepted 18 October 2015
Available online 30 October 2015

Keywords:
Causal inference
Dementia
Education
Instrumental variables
Unmeasured confounding

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; $n = 7981$); and (2) compulsory schooling laws (CSLs) and state school characteristics (term length, student teacher ratios, and expenditures; $n = 10,955$).

Results: Using the genetic risk score as an IV, there was a 1.1% reduction in dementia risk per year of schooling (95% confidence interval, −2.4 to 0.02). Leveraging compulsory schooling laws and state school characteristics as IVs, there was a substantially larger protective effect (−9.5%; 95% confidence interval, −14.8 to −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.

Conclusions: IV analyses suggest education is protective against risk of dementia in older adulthood.

The correlation between educational attainment and later life cognitive function and dementia is well documented [1–4]. Using 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% confidence interval [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 present 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

None of the authors declare any conflicts of interest.

* Corresponding author. University of California, San Francisco, 550 16th Street, 2nd floor, San Francisco, CA 94143-0560. Tel.: +1-617-963-9595; fax: +1-415-514-8150.

E-mail address: Thu.Nguyen@ucsf.edu (T.T. Nguyen).

http://dx.doi.org/10.1016/j.annepidem.2015.10.006
1047-2797/

addressing the causal effects of putative exposures (e.g., in so-called Mendelian randomization studies [8,9]). Recent research identified three 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 present 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 used 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), 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 and/or 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 born outside the U.S., 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 born outside the U.S., or had an unknown place of birth in our analyses using genetic data. Of the respondents who met the above exclusion criteria, 9911 were genotyped. The analytic sample included individuals with 12 or more 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 identifying these SNPs included only Caucasians [15] and the education genetic risk score was not related to education among nonwhites in HRS. The final analytic sample for the genetic IV study included 7981 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 and the Informant Questionnaire for Cognitive Decline 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 nondisease to disease state. For this reason, we considered 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 to 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 years old and minimum drop-out age (CSL) or minimum work permit age (CSL-w) when the respondents were 14 years old. 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 and/or she was born.

State characteristics

Percentage black, urban, and foreign born when the respondents were 6 years of age and manufacturing jobs per capita and manufacturing wages per manufacturing job when the respondents were 14 years of age 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].