



ELSEVIER

Contents lists available at ScienceDirect

## Preventive Medicine

journal homepage: [www.elsevier.com/locate/ypmed](http://www.elsevier.com/locate/ypmed)

# Educational attainment and allostatic load in later life: Evidence using genetic markers

Xuejie Ding<sup>a,b,\*</sup>, Nicola Barban<sup>c</sup>, Melinda C. Mills<sup>a,b,d</sup>

<sup>a</sup> Department of Sociology, University of Oxford, United Kingdom of Great Britain and Northern Ireland

<sup>b</sup> Nuffield College, United Kingdom of Great Britain and Northern Ireland

<sup>c</sup> University of Essex, Institute for Social and Economic Research (ISER), United Kingdom of Great Britain and Northern Ireland

<sup>d</sup> Leverhulme Centre for Demographic Science, University of Oxford, United Kingdom of Great Britain and Northern Ireland

## ARTICLE INFO

## Keywords:

Allostatic load  
Polygenic risk score  
Instrumental variable  
Years of education  
Mendelian randomization

## ABSTRACT

Education is strongly correlated with health outcomes in older adulthood. Whether the impact of education expansion improves health remains unclear due to a lack of clarity over the causal relationship. Previous health research within the social sciences has tended to use specific activities of daily living or self-reported health status. This study uses a broader and objective health measure – allostatic load (AL) – to take into consideration the exposures that accumulate throughout the life course. This paper applies a Mendelian Randomization (MR) approach to identify causality in relation to education on health as measured by AL. Using the Health and Retirement Study 2008 ( $N = 3935$ ), we adopt a polygenic score built from genetic variants associated with years of education. To test whether our analyses violate the exclusion assumption, we further run MR Egger regressions to test for bias from pleiotropy. We also explore the potential pathways between education and AL, including smoking, drinking, marital length, health insurance, etc. Using this genetic instrument, we find a 0.3 unit (19% of a standard deviation) reduction in AL per year of schooling. The effect is mainly driven by BMI and Hba1c. Smoking and marital stability are two potential pathways that also causally influenced by education. If our main and sensitivity analyses are valid, the results find support that a higher level of education is causally related to better health in older adulthood.

## 1. Introduction

The empirical relationship between education and health over the life course is well-established. The mechanisms linking educational attainment to health outcomes, however, are not fully understood. One conceptualisation is that low education, as an indicator of socioeconomic disadvantage, is subject to environmental, psychological and behavioral characteristics. Such experiences and exposures accumulate throughout the life course, more often place demands on the biological system (e.g., immune, cardiovascular and metabolic systems), ultimately leading to more significant system dysregulation, and subsequently enhancing the risk for poor health and functioning (Gruenewald et al., 2012). The concept of allostatic load (AL) has been proposed as a more comprehensive, multisystem measure of the cumulative biological dysregulation across major physiological systems resulting from the accumulation of stressful exposures (McEwen and Stellar, 1993). AL assesses risk across a wide array of biomarkers and across multiple systems, to capture the cumulative burden that may

have a considerable impact on future health risks.

There are three ways in which educational attainment may be related to health in the literature (Cutler and Lleras-Muney, 2006; Eide and Showalter, 2011). The first argues that education causes to better health. The second holds that the direction of causality is reversed and runs from poorer health to lower educational attainment. The third suggests that both schooling and health are affected by third omitted factors such as family background, parental investment to children, non-cognitive ability, and time preferences.

A few studies investigate the association between socioeconomic gradient and AL measured health, but the results remain inconclusive. Analysing a nationally representative sample of the US, Seeman et al. (2010) provide evidence of an inverse association between AL and level of education in all age groups. A recent study by Merkin et al. (2014) shows that lower socioeconomic attributes are associated with faster accumulation of AL. Despite the growing body of replications over time and across different socio-political and economic contexts, evidence for a causal effect of education on biomarkers is much more limited. To our

\* Corresponding author at: Department of Sociology, University of Oxford, United Kingdom of Great Britain and Northern Ireland.

E-mail address: [xuejie.ding@sociology.ox.ac.uk](mailto:xuejie.ding@sociology.ox.ac.uk) (X. Ding).

<https://doi.org/10.1016/j.ypmed.2019.105866>

Received 8 April 2019; Received in revised form 9 September 2019; Accepted 11 October 2019

0091-7435/ © 2019 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

knowledge, the only study attempting to explore the association between education and AL using a co-twin control design failed to find causality (Hamdi et al., 2016).

The purpose of this paper is to establish whether additional years of education have a causal impact on AL. This is particularly relevant for older individuals since they have not only have been exposed to stressors imposed by the current environment, but also by their survival and the ever-changing environments they have experienced. One challenge in the estimation of the causal relationship between education and AL is that educational attainment is potentially confounded by a broad range of confounders, including childhood health status, cognitive abilities, and familial socioeconomic status (Kawachi et al., 2010). We use the information on a variety of genes known to be related to educational attainment as an instrumental variable. Such an approach is also referred to as Mendelian Randomization (MR) (Davey Smith and Ebrahim, 2003). Genetic variants are randomly allocated at conception, their effect on the observable exposure of interest, i.e., educational attainment, are randomly assorted in relation to potential confounders (Davey Smith and Ebrahim, 2003). For this reason, genetic information is considered to fulfil the independent assumption. Based on the findings from a recent genome-wide association study (GWAS) of educational attainment (Okbay et al., 2016), we construct a genetic risk score to predict educational attainment with data from the Health and Retirement Study (HRS).

Few studies have examined the causal relationship between education and health indicators using a genetic instrument (Böckerman et al., 2017; Nguyen et al., 2016; Tillmann et al., 2017; Viinikainen et al., 2018). Using genetic score based on 74 genetic variants that have found to be associated with the years of education, Böckerman et al. (2017) found a negative causal effect of education and BMI. Tillmann et al. (2017) reported supportive evidence that low education is a causal risk factor in developing coronary heart disease. No prior studies of education and AL have used an MR design. The paper is one of the first to exploit genetic variants for education in attempt to estimate the causal effect of schooling on AL in later life.

## 2. Methods

### 2.1. Data

The HRS is a national representative study of individuals 50 years of age or older and their spouses in the United States (Ofstedal et al., 2011). The survey contains detailed socio-demographic information in addition to a genetic sample. The first survey wave was collected in 1992, with biennial interviews available through 2010.

In 2006, HRS initiated an enhanced face-to-face (EFTF) interview that collects biological and genetic information from respondents. Between 2006 and 2008, the HRS genotyped 12,507 respondents. Our study uses the 2008 biomarkers. We focus on individuals for whom the genetic data and biomarker data are available after the quality control. 13,643 non-Hispanic white respondents older than 50 years old were interviewed in the 2008 wave. Among these people, 4495 have provided information on biomarkers. 3935 have been genotyped. Our analyses include the sample weights provided by the HRS. The weights are produced to adjust for non-random sampling and selective non-response to participation, so that the biomarker sample can closely match the HRS sample composition by age, gender and race. Since the HRS only collected genomic data for those who lived until 2006, only the subset of birth cohort members who have survived to the time of data collection is sampled. To correct the mortality selection, we include inverse probability weights based on the procedure that Domingue and colleagues used in their study (Domingue et al., 2017).

### 2.2. Measures

#### 2.2.1. Exposure

The main exposure of interest was educational attainment, operationalized as the number of completed years of schooling (ranging from 0 to 17). Since our genetic instrument for educational attainment is extracted from the GWAS of Okbay et al. (2016), the same measures of educational attainment are used as in the GWAS.

#### 2.2.2. Outcomes

The health outcome is measured by an overall summary index of multi-system risk (AL) based on nine biomarkers, to reflect the cumulative effect of physiological dysregulation across multiple systems. The nine biomarkers are subsets of five physiological systems: the inflammation system includes C-reactive protein; HbA1c, cholesterol ratio, BMI, and waist circumference are subsets of metabolic system; diastolic/systolic blood pressure are indicators of cardiovascular system function; cystatin C belongs to hematopoietic system, and handgrip is a measure of muscle stress.

For each of the nine biomarkers, a dichotomous indicator was created, reflecting those with “high risk” values (assigned a score of “1”) and “low risk” of values (assigned a score of “0”) based on cut-off values commonly accepted in clinical practice and the literature (Ding et al., 2017; Gruenewald et al., 2012; Juster et al., 2010). Respondents who reported taking medication for hypertension and diabetes are also categorized as “high risk” for blood pressure and HbA1c. The indicators are then summed to create AL. AL is equal to the sum of “high risk” conditions scaled by the ratio of “number of items in the index” to “number of non-missing values”. Missing values do not pose a problem for this study since only 2% of the respondents are missing more than three biomarkers. Descriptive statistics for biomarkers are reported in Table 1. The average AL value was 2.14 (SD = 1.57, range = 0–9).

#### 2.2.3. Instrumental variables

We construct a polygenic score (PGS, also known as a genetic risk score) derived from a recent GWAS of educational attainment conducted by Okbay et al. (2016), which identified 74 independent SNPs associated with an individual's total years of schooling. PGS is a single quantitative summary of an individual's cumulative genetic predisposition to a specific disease or traits, weighted by effect size on the trait of interest. A PGS for individual  $i$  can be calculated as the sum of the allele counts  $a_{ij}$  (0, 1, or 2) for each SNP  $j = 1, \dots, M$ , weighted by association strength  $p_j$ :

$$PGS = \sum_{j=1}^M p_j a_{ij} \quad (1)$$

**Table 1**

Descriptive statistics and cut-off points for high risk values of individual biomarkers.

Biomarkers	N	M	SD	Cut-off points
BMI (kg/m <sup>2</sup> )	3905	0.30	0.46	≥ 30 or < 18.5
Waist circumferences (cm)	3862	0.64	0.48	Male > 102; female > 88
HbA1c (%)	3899	0.11	0.31	≥ 6.5
Cholesterol ratio	2342	0.09	0.28	Total cholesterol to HDL ≥ 5.92
SBP (mmHg)	3814	0.20	0.40	> 140 in all three measurements
DBP (mmHg)	3814	0.08	0.26	> 90 in all three measurements
High Cystatin C (mg/L)	2596	0.04	0.20	> 1.55
C-reactive protein (μg/mL)	3858	0.18	0.38	≥ 3
Handgrip (kg)	3749	0.32	0.46	Male ≤ 30; female ≤ 20
AL (unstandardized)	3935	2.10	1.52	Range: 0–9

Note: N = sample size; M = mean; SD = standard deviation; HDL = high density lipoprotein.

Since the HRS was a part of the GWAS sample, we obtained the list of association results calculated excluding the HRS from the meta-analysis from the Social Science Genetic Association Consortium. Using these summary statistics, we selected SNPs that have a statistical association of  $p\text{-value} < 5 \times 10^{-8}$ , and excluded the SNPs that have been found to be associated with BMI and cognitive abilities (Okbay et al., 2016). This yields a list of 35 SNPs (see Supplementary Materials). We then constructed a linear PGS weighted for their effect sizes in the meta-analysis using the software PLINK and PRSice (Euesden et al., 2015; Purcell et al., 2007).

### 2.2.4. Covariates

We present the main findings, both with and without adjustment for covariates. All models control for age, age squared, gender and the first ten genetic principal components for each individual using genome-wide principal components that function as ancestry markers (Price et al., 2006). Controlling for population stratification and focusing on the white non-Hispanic population would further ensure that the independence assumption is not violated.

In the analysis that adjusts for covariates, we additionally control for the area of birth, and parental years of education.

## 3. Results

### 3.1. Descriptive statistics

Columns 1, 2 and 3 in Table 2 present the descriptive statistics of the key variables. Column 4–6 show the raw association between these measures, the covariates and the genetic variants, obtained from a regression of years of schooling or each covariate on the polygenic allele score. The top row of these columns presents the relationship between years of education and the instrument, showing a strong positive relationship for educational attainment ( $\beta = 0.221, p < .001$ ). This significant and substantial effect suggests that there is a non-zero effect of the genetic IV on education. The remaining rows show no clear patterns or statistically significant associations in the relationship between the contextual variables and the genetic instrument, except the parental educational attainments. The association with parental education implies potential violation of the exclusion assumption (see Supplementary Materials). We, therefore, include parental education as covariates in the models.

### 3.2. OLS versus IV estimates

Table 3 presents OLS estimates of the regression of individuals' years of education on health in later life. Using 3935 individuals, we found that a higher level of education was strongly correlated with individuals with lower AL, indicating better health. Column (1) shows unadjusted results and column (2) presents the results adjusted for covariates.

**Table 2**

Descriptive statistics of education and the covariates: column 1, 2 and 3 show their mean and standard deviation. Columns 4–6 present the coefficients, standard error and  $p$ -value of the variables shown in the first column regressed on the genetic instrument (polygenic allele score of educational attainment).

	(1)	(2)	(3)	(4)	(5)	(6)
	N	Mean	Std. dev.	Coeff.	Std. err.	$p$ value
Years of education	3935	13.2	2.5	0.221	0.039	< 0.001
Age	3935	70.4	9.9	0.228	0.150	0.128
Male	3935	0.4	0.5	0.008	0.008	0.271
Married = 1	3935	0.6	0.5	0.004	0.007	0.621
Mother years of education	3935	10.3	3.0	0.104	0.048	0.031
Father years of education	3935	9.9	3.5	0.141	0.057	0.017

**Table 3**

OLS estimates of years of education on AL.

	(1) Covariate unadjusted	(2) Covariate adjusted
Years of education	-0.078 [-0.096, -0.059]	-0.057 [-0.080, -0.034]
Observations	3935	3935
R <sup>2</sup>	0.042	0.044

Note: 95% CI is reported in the bracket. All regressions include respondent's age, age squared, gender, and top 10 principal components for their respective population stratification. The covariates adjusted model additionally includes mother's year of education, father's years of education, and area of birth.

**Table 4**

First stage and 2SLS estimates of years of education on allostatic loads.

	(1) IV = PGS		(2) IV = PGS, covariates adjusted	
	First stage	2SLS	First stage	2SLS
Years of education		-0.308 [-0.573, -0.044]		-0.583 [-1.220, -0.019]
Polygenic score	0.184 [0.109, 0.259]		0.097 [0.025, 0.169]	
Observations	3935	3935	3935	3935
F statistics	22.94	-	10.89	-
$p$ -Value DWH test	-	0.0669	-	0.0395

Note: 95% CI is reported in the bracket. All Regressions include respondent's age, age square, gender, and top 10 principal component for their respective population stratification. The covariates adjusted model additionally includes mother's year of education, father's years of education, and area of birth. DWH refers to the Durbin-Wu-Hausman test.

Column (1) shows that years of education is negatively correlated with AL. This estimate indicates that an additional year of schooling reduces AL by 0.08. After controlling for covariates, the point estimate reduced to -0.057 (95% CI = -0.080, -0.034).

Table 4 presents the instrumental variable analysis. The IV results reveal that our instrument is relevant since the first stage F-statistic in column (1) is 22.94, and the instrument is significant in the first stage. Genetic IV analyses provided evidence that the association between education and AL was partly causal. Covariate adjusted estimates indicate similar results showing that years of schooling has a protective effect on AL. Overall, our results consistently suggest that the actual impact of education on health maybe underestimated. The hypothesis that the OLS and the IV estimates are similar to one another cannot be rejected for the covariate unadjusted model (Durbin Hausman Wu test  $p\text{-value} = .669$ ). However, this may be due to insufficient power. We further conduct sensitivity analyses to test the relationship between highest degree completed and allostatic load. The causal relationship is robust and can be found in the supplementary material (Table D).

Our operationalization of AL contains more metabolic and cardiovascular biomarkers than biomarkers from other physiological system. In Table 5, we present OLS and 2SLS estimated effects of education on each of the nine biomarkers that constitute our overall AL. AL findings are largely driven by the BMI and Hba1c biomarkers. With the exception of diastolic blood pressure, the OLS estimate is negative for each high-risk biomarker, the most for waist circumference, and the least for Cystatin C (indicator for kidney inflammation) and Cholesterol ratio. The IV analyses implied that educational attainment is causally related to BMI and diabetes. Education is also a potential determinant for Hba1c, another indicator for metabolic dysregulation, with each additional year of schooling decreases Hba1c by 0.06.

**Table 5**  
OLS and 2SLS estimates of years of education on single health biomarker.

	(1) BMI		(2) Hba1c	
	(a) OLS	(b) 2SLS, PGS	(a) OLS	(b) 2SLS, PGS
Years of education	−0.013 [−0.019, −0.007]	−0.080 [−0.161, −0.003]	−0.008 [−0.012, −0.004]	−0.064 [−0.120, −0.008]
Observations	3905	3905	3899	3899
R <sup>2</sup>	0.034		0.012	
First stage F statistics		22.22		23.26
p-Value DWH test		0.0879		0.0300
	(3) Systolic blood pressure		(4) Diastolic blood pressure	
	(a) OLS	(b) 2SLS, PGS	(a) OLS	(b) 2SLS, PGS
Years of education	−0.006 [−0.011, −0.001]	−0.023 [−0.091, 0.044]	0.010 [−0.035, 0.056]	0.010 [−0.035, 0.056]
Observations	3814	3814	3814	3814
R <sup>2</sup>	0.034		0.070	
F statistics		21.838		21.838***
p-Value DWH test		0.6113		0.6267
	(5) Cholesterol ratio		(6) Cystatin C	
	(a) OLS	(b) 2SLS, PGS	(a) OLS	(b) 2SLS, PGS
Years of education	−0.057 [−0.078, −0.034]	−0.025 [−0.292, 0.242]	−0.004 [−0.007, −0.001]	−0.024 [−0.075, 0.027]
Observations	2342	2342	2596	2596
R <sup>2</sup>	0.030		0.001	
F statistics		17.28		10.47
p-Value DWH test		0.8174		0.4147
	(7) C-reactive protein		(8) aist circumferences	
	(a) OLS	(b) 2SLS, PGS	(a) OLS	(b) 2SLS, PGS
Years of education	−0.007 [−0.011, −0.002]	−0.025 [−0.067, 0.059]	−0.016 [−0.022, −0.010]	−0.045 [−0.124, 0.035]
Observations	3858	3858	3862	3862
R <sup>2</sup>	0.008		0.024	
F statistics		22.57		23.31
p-Value DWH test		0.9353		0.4707
	(9) Handgrip			
	(a) OLS		(b) 2SLS, PGS	
Years of education	−0.013 [−0.018, −0.007]		−0.036 [−0.109, 0.036]	
Observations	3749		3749	
R <sup>2</sup>	0.252			
F statistics			19.86	
p-Value DWH test			0.5212	

Note: All regressions include respondent's age, age square, gender, and top 10 principal components for their respective population stratification. DWH refers to the Durbin-Wu-Hausman test.

### 3.3. Potential pathways

As discussed previously, the pathways through which education affect health are multi-faceted. With better health knowledge, the well-educated may improve health by adopting a healthier lifestyle. Education also increases income and wealth since additional economic resources improve health by investing in health care. More education also provides individuals with better interpersonal skills and is linked with higher marital satisfaction and lower levels of stress-related diseases. Thus we performed additional analyses to investigate whether the factors suggested in health literature can serve as a pathway for the

effect of education on AL (Table 6). Due to lack of good instrument of the mediators, we did not perform two-step MR to study mediation. Our tests of potential mechanisms should be treated as suggestive. Our IV analyses found that education causally reduces the probability of smoking – each additional year of schooling reduces the probability of smoking by 4%. Our results is in line with recent findings reported by Sanderson et al. (2019) who used UK Biobank data to study the education-smoking association, adjusting for cognitive ability. We also found suggestive evidence that education is causally linked with marital stability and Spouse's educational level. This result is consistent with the past literature among the US population that spousal education

**Table 6**  
OLS and 2SLS estimates of years of education on potential mechanisms.

	(1) Smoking		(2) Drinking	
	(a) OLS	(b) 2SLS, PGS	(a) OLS	(b) 2SLS, PGS
Years of education	−0.018 [−0.022, −0.014]	−0.038 [−0.067, −0.086]	0.085 [0.081, 0.088]	−0.113 [−0.950, 0.724]
Observations	3912	3912	2136	2136
R <sup>2</sup>	0.052		0.045	
F statistics		18.34***		19.74**
p-Value DWH test		0.1746		0.8172
	(3) Exercise		(4) Marital stability	
	(a) OLS	(b) 2SLS, PGS	(a) OLS	(b) 2SLS, PGS
Years of education	−0.095 [−0.099, −0.091]	−0.510 (0.347) [−1.190, 1.170]	0.237 [0.059, 0.415]	0.271 [0.167, 0.375]
Observations	3565	3565	3841	3841
R <sup>2</sup>	0.024		0.256	
F statistics		13.22***		16.19***
p-Value DWH test		0.1762		0.0368
	(5) Spouse's education level		(6) Life satisfaction	
	(a) OLS	(b) 2SLS, PGS	(a) OLS	(b) 2SLS, PGS
Years of education	0.505 [0.472, 0.538]	0.578 [0.153, 1.003]	0.069 [0.065, 0.073]	−0.074 [−0.174, 0.026]
Observations	2554	2554	3629	3629
R <sup>2</sup>	0.266		0.027	
F statistics		13.41***		14.35***
p-Value DWH test		0.0693		0.4251
	(7) Household wealth		(8) Health insurance	
	(a) OLS	(b) 2SLS, PGS	(a) OLS	(b) 2SLS, PGS
Years of education	0.251 [0.210, 0.292]	0.273 [−0.121, 0.667]	0.028 [0.022, 0.034]	0.038 [−0.042, 0.118]
Observations	3841	3841	3841	3841
R <sup>2</sup>	0.046		0.058	
F statistics		13.42***		16.91***
p-Value DWH test		0.1562		0.7648

Significance level: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Note: All regressions include respondent's age, age squared, gender, and top 10 principal components for their respective population stratification. Smoking is measured by whether the respondent smokes at wave 9 (yes = 1,  $M = 0.12$   $SD = 0.33$ ); drinking is measured by number of days per week the respondent drinks alcohol ( $M = 2.44$   $SD = 2.59$ ); exercise is measured by how often the respondent walks for 20 min (daily 1 – not in last month 6,  $M = 3.46$   $SD = 1.94$ ); Marital stability is measure by length of longest marriage in years ( $M = 35.86$ ,  $SD = 16.53$ ); Spousal educational attainment is measure by current spouse's years of education completed ( $M = 13.25$ ,  $SD = 2.56$ ); life satisfaction is measure by "I am satisfied with my life" (strongly disagree 1 = strongly agree 7,  $M = 5.47$ ,  $SD = 1.68$ ); household wealth is measure by natural log transformed household total wealth ( $M = 11.41$ ,  $SD = 3.35$ ); health insurance is measured by the level of health insurance the respondent has ( $M = 0.72$ ,  $SD = 0.58$ ). DWH refers to the Durbin-Wu-Hausman test.

attenuates the association between individual's own education and self-reported health, and highlights the importance of education as it is a shared resource in marriage for producing health (Brown et al., 2014) (Table 6).

#### 4. Discussion

The current study made use of genetic instrumental variables to control for confounders in examining the association between education and a direct, biologically-based measure of AL that captures physiological dysregulation across multiple, major regulatory systems. Our IV analysis with the polygenic score of educational attainment showed a significant, negative relationship between education and AL, meaning that a higher level of education leads to better health.

Overall, our results consistently suggest that the actual degree of higher education on health may be underestimated. The greater

magnitude of 2SLS compared to the OLS estimates is not a rare case among MR studies (Sanderson et al., 2019; Schmitz and Conley, 2017; Willage, 2018). One explanation is that the instrumental variable estimates the local average treatment effect (LATE), while OLS tries to estimate the average treatment effect (ATE) over the entire population. Genetic predisposition of education may shift the behavior of a subgroup of individuals for whom the effect of education on health are larger than average. OLS may underestimate the true effect of education on health due to omitted variables that are positively related to education but negatively related to health, or vice versa. Nevertheless, both IV and OLS estimates are generally small compared to the standard deviation of years of education. Measurement error in the explanatory variable could also result in larger estimates in the IV, as the OLS estimation was likely to be biased toward zero. However, there is little evidence indicating that our results are driven by measurement error. Moreover, since the DWH test shows that the difference between OLS

and IV estimates are insignificant, it is also possible that the IV estimates are imprecise.

Compared to other IV used in traditional sociological studies, such as compulsory schooling laws and quarter of birth, which capture a LATE (i.e. the estimated return to education for whose behavior is affected by the policy), the MR IV works across the whole distribution of education leaving ages. It may be able to recover effects closer to an ATE rather than a LATE. The MR is likely to estimate compliers with “intention to treat” or “genetic endowment of education”, as not everyone will attain increased education given these genetic variants. Our results should be interpreted with caution because our sample only focused on older adults in the US. The genetic predisposition may interact with environmental changes and have different effects on the average US population. In addition, genetic variants may operate through different biological mechanisms and may have different LATEs on individuals (Böckerman et al., 2019).

This study uses a multisystem measure of AL as the outcome since a comprehensive indicator of health is preferred. Our analyses showed that years of education were significantly associated with reduced risk for BMI and HbA1c, which are indicators for metabolic dysregulation (2SLS  $p < .05$ ). This finding aligns with previous research that education has a protective effect on BMI (Brunello et al., 2013; Kemptner et al., 2011). We found no consistent evidence of a causal relation between higher education and a lower value of biomarkers. One possibility is that the effect size of a single physiological measure is too small to detect, making the genetic IV a weak instrument in these analyses. Research with larger sample sizes on separate biomarker may provide a more accurate estimation. Exploring possible pathways suggests that education may affect AL via smoking, marital stability, and spousal educational level.

## 5. Limitation and conclusion

Our paper contributes to both the substantive and methodological literature of health research, but also has several limitations. First, while previous studies employing neurotransmitter genes often suffer from unconvincing exclusion restrictions, the polygenic score employed here satisfy the exclusion assumption based on the current knowledge of the gene's function. However, there is still a possibility that the genetic instrument may be correlated with other factors besides education that may also affect AL. One source of violation is pleiotropy, which means the genetic IV may have direct effects on both educational attainment and health (Ding et al., 2019), we applied MR-Egger, weighted median and weighted mode approaches (Bowden et al., 2016; Hartwig et al., 2017) to test for causality even when the genetic IV is invalid. The sensitivity analyses found no evidence for pleiotropic effect in our models (Supplementary Materials, Table E).

Second, another concern on the validity of this study comes from the dynastic effects or bias due to assortative mating. Tests of such violation require parental genotype, or via sibling designs, which is unfortunately not available in HRS genotypic data. However, Okbay et al. (2016) found little evidence that the effects of the polygenic score for education attenuated after controlling for family structure. After controlling for parental education and childhood health status, our IV results are still robust. These findings all suggest that parental education have little contribution to the causal pathway between participants' education and health in old age.

Third, genetic variants may not be powerful enough to identify causal effects. While the IVs are not weak in a statistical sense, the IV's effect may be too small to impact the biomarkers and on the possible pathways to health. In other words, a 1–2 year increase in education may not lead to a large drop in morbidity. Hence, it is not surprising that we find no significant effect on most of the biomarkers and some potential pathways. Future research needs to use larger samples to test the relationship between education and a specific pathway or disease.

Finally, our study is based on mostly homogeneous groups of non-

Hispanic Caucasian older adults in the U.S. Including other racial and ethnic groups may lead to weak instrument problem (Martin et al., 2017). Past literature has shown that the education-AL association differs by ancestral groups, and the pathway in which biomarkers attribute to the accumulation of AL differs by ancestral origins and by educational level (Howard and Sparks, 2015). The findings may not extend to individuals of other ethnic or cultural backgrounds, or later-born cohorts. Moreover, the HRS genetic and biomarker sample weights did not adjust the bias that education may be correlated with respondents' willingness to be genotyped. If less educated individuals are more likely to opt-out genotyping, this might result in well educated individual's responses being over-emphasized. Future studies with larger samples may be useful to generalize the conclusion to people from different ancestral and socioeconomic groups (Mills and Rahal, 2019).

In summary, our findings support the hypothesis that education reduces the risk of poor health in older adults. This would support potential preventive interventions based on educational attainment. Policymakers should also bear in mind that this recommendation has the assumption that the education induced by genetic variants have the same biological and psychological effects as the education induced by policy intervention, which is not necessarily true. The mechanisms of how genetic variants work remain largely unknown. More years of education induced by genetic variants may work through different biological pathways compared with education resulting from policy change (Tillmann et al., 2017). In addition, due to population stratification, policymakers should be cautious when generalizing our results to different ancestral groups (Mills and Rahal, 2019). Nevertheless, even though our results do not directly lead to clear policy interventions, we are discreetly optimistic that education-promoting policy could lead to better health in older adults. Future research on genetic variants as IVs needs to be used and interpreted with care. This paper is the first to exploit genetic variation in education to examine the causal effects of education on AL. We instrument education with educational polygenic score and find that education has a protective effect on health. This study also provides promising evidence that genetic factors seem to be useful instruments for studying behavioral effects.

## Acknowledgement

Open access funding is provided by University of Oxford. The research leading to this project was funded by the European Research Council Consolidator Grant SOCIOGENOME, awarded to M. Mills (Grant no. 615603). The authors would like to thank Yanchun Bao, Ben Domingue, John Ermisch, Christiaan Monden, Cecilia Potente, Felix Tropsf, Matthias Qian, the PM editors, and anonymous reviewers for their helpful comments and suggestions. N. Barban received funds from the ESRC through the Research Center on Micro-Social Change MiSoC at the University of Essex, grant numbers ES/L009153/1 and ES/S012486/1.

## Ethics approval

This paper uses secondary data. Data usage has been approved by the Health and Retirement Survey. We did not directly work with any of human subjects in our data.

There is no conflict of interest in this study.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ypmed.2019.105866>.

## References

Böckerman, P., Viinikainen, J., Pulkki-Råback, L., Hakulinen, C., Pitkänen, N., Lehtimäki,

- T., Raitakari, O.T., 2017. Does higher education protect against obesity? Evidence using Mendelian randomization. *Prev. Med.* 101, 195–198.
- Böckerman, P., Cawley, J., Viinikainen, J., Lehtimäki, T., Rovio, S., Seppälä, I., Raitakari, O., 2019. The effect of weight on labor market outcomes: an application of genetic instrumental variables. *Health Econ.* 28 (1), 65–77.
- Bowden, J., Davey Smith, G., Haycock, P.C., Burgess, S., 2016. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* 40 (4), 304–314.
- Brown, D.C., Hummer, R.A., Hayward, M.D., 2014. The importance of spousal education for the self-rated health of married adults in the United States. *Popul. Res. Policy Rev.* 33 (1), 127–151.
- Brunello, G., Fabbri, D., Fort, M., 2013. The causal effect of education on body mass: evidence from Europe. *J. Labor Econ.* 31 (1), 195–223.
- Cutler, D.M., Lleras-Muney, A., 2006. Education and Health: Evaluating Theories and Evidence. National Bureau of Economic Research.
- Davey Smith, G., Ebrahim, S., 2003. Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease. *International journal of epidemiology* 32 (1), 1–22.
- Ding, X., Barban, N., Tropf, F.C., Mills, M.C., 2019. The relationship between cognitive decline and a genetic predictor of educational attainment. *Social Science & Medicine* 239, 112549.
- Ding, X., Billari, F.C., Gietel-Basten, S., 2017. Health of midlife and older adults in China: the role of regional economic development, inequality, and institutional setting. *Int. J. Publ. Health* 62 (8), 857–867.
- Domingue, B.W., Belsky, D.W., Harrati, A., Conley, D., Weir, D.R., Boardman, J.D., 2017. Mortality selection in a genetic sample and implications for association studies. *Int. J. Epidemiol.* 46 (4), 1285–1294.
- Eide, E.R., Showalter, M.H., 2011. Estimating the relation between health and education: what do we know and what do we need to know? *Econ. Educ. Rev.* 30 (5), 778–791.
- Euesden, J., Lewis, C.M., O'Reilly, P.F., 2015. PRSice: polygenic risk score software. *Bioinformatics* 31 (9), 1466–1468.
- Gruenewald, T.L., Karlamangla, A.S., Hu, P., Stein-Merkin, S., Crandall, C., Koretz, B., Seeman, T.E., 2012. History of socioeconomic disadvantage and allostatic load in later life. *Soc. Sci. Med.* 74 (1), 75–83.
- Hamdi, N.R., South, S.C., Krueger, R.F., 2016. Does education lower allostatic load? A twin control study. *Brain Behav. Immun.* 56, 221–229.
- Hartwig, F.P., Smith, G.D., Bowden, J., 2017. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* 46 (6), 1–14.
- Howard, J.T., Sparks, P.J., 2015. The role of education in explaining racial/ethnic allostatic load differentials in the United States. *Biodemograp. Soc. Biol.* 61 (1), 18–39.
- Juster, R.P., McEwen, B.S., Lupien, S.J., 2010. Allostatic load biomarkers of chronic stress and impact on health and cognition. *Neurosci. Biobehav. Rev.* 35 (1), 2–16.
- Kawachi, I., Adler, N.E., Dow, W.H., 2010. Money, schooling, and health: mechanisms and causal evidence. *Ann. N. Y. Acad. Sci.* 1186 (1), 56–68.
- Kempton, D., Jürges, H., Reinhold, S., 2011. Changes in compulsory schooling and the causal effect of education on health: evidence from Germany. *J. Health Econ.* 30 (2), 340–354.
- Martin, A.R., Gignoux, C.R., Walters, R.K., Wojcik, G.L., Neale, B.M., Gravel, S., Kenny, E.E., 2017. Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* 100 (4), 635–649.
- McEwen, B.S., Stellar, E., 1993. Stress and the individual: mechanisms leading to disease. *Arch. Intern. Med.* 153 (18), 2093–2101.
- Merkin, S.S., Karlamangla, A., Roux, A.V.D., Shrager, S., Seeman, T.E., 2014. Life course socioeconomic status and longitudinal accumulation of allostatic load in adulthood: multi-ethnic study of atherosclerosis. *Am. J. Public Health* 104 (4).
- Mills, M.C., Rahal, C., 2019. A scientometric review of genome-wide association studies. *Communications biology* 2 (1), 1–11.
- Nguyen, T.T., Tchetgen, E.J.T., Kawachi, I., Gilman, S.E., Walter, S., Liu, S.Y., Glymour, M.M., 2016. Instrumental variable approaches to identifying the causal effect of educational attainment on dementia risk. *Ann. Epidemiol.* 26 (1), 71–76.e3.
- Ofstedal, M., Weir, D., Chen, K., Wagner, J., 2011. HRS Documentation Report: Updates to HRS Sample Weights. University of Michigan, Arbor, MI.
- Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., ... & Oskarsson, S. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, 533(7604), 539.
- Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., Reich, D., 2006a. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38 (8), 904–909.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81 (3), 559–575.
- Sanderson, E., Davey Smith, G., Bowden, J., Munafò, M.R., 2019. Mendelian randomisation analysis of the effect of educational attainment and cognitive ability on smoking behaviour. *Nat. Commun.* 10 (1), 2949.
- Schmitz, L.L., Conley, D., 2017. The effect of Vietnam-era conscription and genetic potential for educational attainment on schooling outcomes. *Econ. Educ. Rev.* 61, 85–97.
- Seeman, T., Epel, E., Gruenewald, T., Karlamangla, A., McEwen, B.S., 2010. Socio-economic differentials in peripheral biology: cumulative allostatic load. *Ann. N. Y. Acad. Sci.* 1186 (1), 223–239.
- Tillmann, T., Vaucher, J., Okbay, A., Pikhart, H., Peasey, A., Kubinova, R., ... Holmes, M. V. (2017). Education and coronary heart disease: Mendelian randomisation study. *BMJ (Online)*, 358.
- Viinikainen, J., Bryson, A., Böckerman, P., Elovainio, M., Pitkänen, N., Pulkki-Råback, L., Pehkonen, J., 2018. Does education protect against depression? Evidence from the Young Finns Study using Mendelian randomization. *Prev. Med.* 115, 134–139.
- Willage, B., 2018. The effect of weight on mental health: new evidence using genetic IVs. *J. Health Econ.* 57, 113–130.