Received: 13 September 2023 | Revised: 22 November 2023 | Accepted: 28 November 2023

DOI: 10.1002/hec.4787

Health Economics WILEY

RESEARCH ARTICLE

The effects of education on mortality: Evidence using college expansions

Jason Fletcher[1] | **Hamid Noghanibehambari**[2]

[1]La Follette School of Public Affairs, University of Wisconsin-Madison, Madison, Wisconsin, USA

[2]College of Business, Austin Peay State University, Clarksville, Tennessee, USA

Correspondence
Hamid Noghanibehambari.
Email: noghanih@apsu.edu

Funding information
National Institute on Aging, Grant/Award Numbers: R01AG060109, P30 AG17266

Abstract

This paper explores the long-run health benefits of education for longevity. Using mortality data from the Social Security Administration (1988–2005) linked to geographic locations in the 1940-census data, we exploit changes in college availability across cohorts in local areas. Our treatment on the treated calculations suggest increases in longevity between 1.3 and 2.7 years. Some further analyses suggest the results are not driven by pre-tends, endogenous migration, and other time-varying local confounders. This paper adds to the literature on the health and social benefits of education.

KEYWORDS
college education, education, health, mortality

JEL CLASSIFICATION
I23, I26, H51, H75, D62

1 | INTRODUCTION

It is well documented that the returns to education go beyond labor market outcomes. Education not only has spillover effects for peers, colleagues, and other family members (Fruehwirth, 2014; Jaffe et al., 2006; Martins & Jin, 2008) but also shapes long-run outcomes such as health at older ages (Albouy & Lequien, 2009; Mazzonna, 2014).[1] A recent review by Galama et al. (2018) classified the literature that establishes a causal path between education and health, health behavior, and mortality into three general categories: (1) studies that employ randomized controlled trials such as the Perry Preschool Program and Abecedarian which primarily provides early childhood education intervention (Campbell et al., 2014; Conti et al., 2016; J. Heckman et al., 2013); (2) within-twin variation in years of schooling as the primary education shock (Behrman et al., 2011; Lundborg et al., 2016; Madsen et al., 2010); (3) quasi-experimental settings with primary focus on school reforms which increase minimum school leaving age and compulsory schooling (Black et al., 2015; Clark & Royer, 2013; Fischer et al., 2013; Fletcher, 2015; Gathmann et al., 2015; Lager & Torssander, 2012; Lleras-Muney, 2005; Mazumder, 2008; Meghir et al., 2018; van Kippersluis et al., 2011).

With the exception of twin/sibling studies, the literature concentrates on interventions and improvements at early ages or during K-12 education focusing on years of schooling as the primary explanatory variable and has offered mixed evidence. Some studies point to large positive effects for life expectancy and gains in mortality while other studies do not find any suggestive evidence and in some cases even opposite-signed coefficients (Albouy & Lequien, 2009; Behrman et al., 2011; Campbell et al., 2014; Conti et al., 2016; Gathmann et al., 2015; Lundborg et al., 2016; Madsen et al., 2010; Meghir et al., 2018). A small strand of the research explores the health impacts of college education through exogenous incentives such as college openings (Buckles et al., 2016; Cowan & Tefft, 2020; Currie & Moretti, 2003; Hong et al., 2020; Kamhöfer et al., 2019; Savelyev, 2020). For instance, Cowan and Tefft (2020) use state-level variation in college accessibility to examine the effects of per capita colleges on education and health outcomes. They find increases in education and self-reported health but do not find a positive effect on old-age health outcomes. This paper aims to extend this literature and provide new insight into the long-term effects of education by quantifying the benefits of higher education on old-age mortality in the US.

We investigate whether changes in the number of available colleges in the local area when individuals are 17 years old could influence their age at death. In so doing, we employ death records of the Social Security Administration (observed between the years 1988 and 2005) linked with the 1940 full count census. The linked data allows us to observe county of residence during adolescence years, an important identifier in our setting. We find robust evidence that the availability of 4-year colleges at the age of high school completion can influence their age at death. The reduced form effects are economically meaningful. For example, our most parametrized specification suggests that an additional college opening in own or neighboring counties raises the age of death by 0.16 months. This is equivalent to roughly 5% of white-nonwhite differences in old age mortality. Because college openings have modest direct effects on educational attainments, the treatment on the treated calculations suggest large effects, between 1.3 and 2.7 years of added life for older age individuals.

We implement a wide range of robustness exercises to control for family unobserved heterogeneity, assess for alternative standard error adjustments, consider alternative specifications, functional forms, and various measures of college access. In addition, we explore the possibility of endogeneity of college access, as people with specific demographic characteristics may migrate to areas that have experienced college expansions. In addition, we consider placebo tests that assign college expansions to the year individuals turn age 25, 30, and 35, well after the age that college availability could have an effect on individual's education attainment.

College expansions change the landscape of counties in several ways. For instance, it shifts the occupational compositions or expands new job prospects for individuals and their parents. Alternatively, it can affect education of not individuals per se but that of their parents or peers. All these channels may have long-lasting consequences for individuals' later-life outcomes. Although we provide the battery of tests to rule out concerns over endogeneity and alternative channels, we should note early on that increases in college education is among the many alternative candidates through which college openings may affect later-life longevity.

Our paper contributes to the literature on the social and health benefits of education in two ways. First, this is the first study to link the construction of new colleges on education and later-life mortality. Second, while similar papers have used longitudinal with limited observations, our new longitudinal dataset provides millions of observations which significantly adds power to our statistical tests. In addition, the increased sample size enables a wide range of heterogeneity analysis by cohort, place, and demographic.[2]

The rest of the paper is organized as follows. Section 2 gives a review of related literature. Section 3 introduces the data sources and sample construction. In Section 4, we discuss the econometric method and potential endogeneity concerns. Section 5 offers the main results and discusses additional analysis. Concluding remarks are provided Section 6.

2 | LITERATURE REVIEW

Education has spillover effects across a wide range of outcomes which can operate as a set of channels for health at older age, including longevity (Oreopoulos & Salvanes, 2011). While there is a relatively large literature on education and mortality, most studies focus on years of education in general and exploit compulsory schooling laws or similar reforms as the shock to education and find mixed evidence (Albouy & Lequien, 2009; Buckles et al., 2016; Conti et al., 2010; Everett et al., 2013; Jemal et al., 2008; Kalediene & Petrauskiene, 2005; van Kippersluis et al., 2011; Kravdal, 2008; Lager & Torssander, 2012; Lynch, 2003; Ross et al., 2012; Zajacova, 2006). In her seminal study, Lleras-Muney (2005) takes advantage of changes in compulsory schooling (child labor and compulsory attendance laws) as the instrument for education and builds a synthetic cohort from decennial censuses to measure 10-year mortality rates. She finds that an additional year of education leads to a 6.3% point reduction in 10-year death rates. Mazumder (2008) shows that these findings are not robust to adding state trends. Black et al. (2015) use Vital Statistics death records to investigate the effects of compulsory schooling laws on mortality and find that the gains in mortality can be explained by cohort and state fixed effects. Fletcher (2015) employs survey data (AARP Diet and Health Study) and instruments education with compulsory schooling laws and finds positive health effects and large gains in mortality. His calculations provide similar effects to the findings of Lleras-Muney (2005) but they are statistically insignificant.

Lleras-Muney et al. (2020) examines the association between education and old age longevity for cohorts born between 1906 and 1915 in the US. They link social security death records with 1940 full-count census and find that an additional year of education is associated with 0.4 higher age at death. In a similar study, Halpern-Manners et al. (2020) apply linking techniques to merge 1920 and 1940 full-count census covering males born between 1910 and 1920 with social security administrative mortality data and implement a twin-fixed-effect strategy to study the effect of education on mortality. They find that an additional year of schooling raises age at death by 0.3 years. These effects are slightly smaller than OLS estimates which suggests that common endowments such as genetic factors only modestly influence the education-mortality association in their sample. Savelyev et al. (2022) employ data from the Minnesota Twin Registry and explore the effect of education on longevity and other health outcomes. They find that an additional year of education reduces the probability of death within 20 years of the survey by about 2.3% points. Malamud et al. (2021) explore the impact of a school construction program in Romania on schooling and health outcomes. They find that school construction significantly increased educational attainments but had no impact on mortality, self-reported health, and hospitalization.

A growing literature has used quasi-experimental and sibling methods to examine the effects of college on older age health. Buckles et al. (2016) employ the deferment of the Vietnam War draft for college students as the source of identification to explore the effect of college education on cumulative mortality. They find that a 1% point increase in college completion rate among men reduces their mortality rate by 0.95 fewer death per 1000. Their results suggest that raises in income and health insurance are plausible mechanisms of impact. Savelyev (2020) employs the Terman Life-Cycle Study of Children with High Ability data that contains a sample of individuals with high-IQ and follows them throughout their life cycle. He finds that among high-IQ individuals having a college education significantly increases longevity. Meara et al. (2008) provide descriptive evidence on educational gaps in mortality and life expectancy in the US. They document that in 2000, the life expectancy for a person at age 25 with some college education is 7 years more than a person with a high school diploma.

Cowan and Tefft (2020) use Census and American Community Survey data to explore the effect of college accessibility (number of colleges per capita) on adult outcomes. They find positive effects on education, employment, and self-reported health. However, they fail to detect a significant improvement for old-age health outcomes and mental health outcomes.[3] Kamhöfer et al. (2019) explore education and health effects of new college openings in Germany. They also find positive and significant local average effects on education, wages, and cognitive ability. However, they argue that these findings are partly driven by selection based on unobservables and that the effects on those with the lowest desire to attend college is zero.

Fletcher and Frisvold (2014) compare siblings in the Wisconsin Longitudinal Study to show college attendance is related to old age preventive care decisions and college quality is related to better weight outcomes in old age, respectively.[4] Bautista et al. (2020) use the discontinuity in college attendance as a result of the 1973 military coup in Chile, which significantly reduced access to college for those reaching college age, and show that the reduction in access and enrollment was associated with higher age-adjusted mortality later in life.

Attending college can affect later-life health and old-age mortality through several channels. The primary channel is labor market outcomes as the large and old literature of returns to education has documented (Acemoglu & Angrist, 2000; Angrist & Keueger, 1991; Card, 1999, 2001; Cellini & Chaudhary, 2014; Dickson & Harmon, 2011; Leigh & Gill, 1997; Long, 2010; Psacharopoulos, 1985). For instance, Cellini and Chaudhary (2014) show that individuals who attend a for-profit college compared to their counterfactual case have 10% higher earnings. Income-induced raises in welfare improves other mediatory outcomes such as a healthier living environment, better nutrition, better health insurance, and improved health behavior which in turn have strong effects on mortality and life expectancy (Backlund et al., 1999; Gonzalez & Quast, 2013; Jamison et al., 2007; Lefèbvre et al., 2018; Lindahl, 2005; Muller, 2002; Snyder & Evans, 2006). Lacroix et al. (2019)investigate the health impacts of college education in Canada. They find that college-educated individuals live 4.1 additional years (conditional on being alive at age 51), have 27.3% lower lifetime hospital stays, they have lower rates of diabetes and stroke, and they have higher survival rates conditional on having a health condition.

College-educated individuals utilize healthier behaviors which leave them in trajectories that increase their longevity. For instance, Walque (2007) shows that college-educated people are less likely to smoke, and among those who do smoke, they are more likely to quit. An old and established literature documents the mortality consequences of smoking (Doll & Hill, 1956; Fenelon & Preston, 2012; Preston et al., 2010). In addition, college education facilitates individuals with more white-collar jobs that are comparably safer than jobs in manufacturing, mining, and agriculture. Several studies show that mortality rates vary across individuals in different occupations (Johnson et al., 1999; Luy et al., 2011). Higher educated individuals are less likely to have adverse mental health conditions, too. There is a small literature that documents the negative association between education and mental health issues such as excess anxiety and depression (Chevalier et al., 2004; Cornaglia et al., 2015). Poor mental health, in turn, leads to higher mortality and lower longevity (Cuijpers & Smit, 2002; Schulz et al., 2002; Wulsin et al., 1999). Another channel is social spillovers in education. The benefits of education roll down to health and specifically mortality outcomes of spouse (Jaffe et al., 2006; Spoerri et al., 2014), other family members (Kravdal, 2008), and colleagues in the workplace (Martins & Jin, 2008).

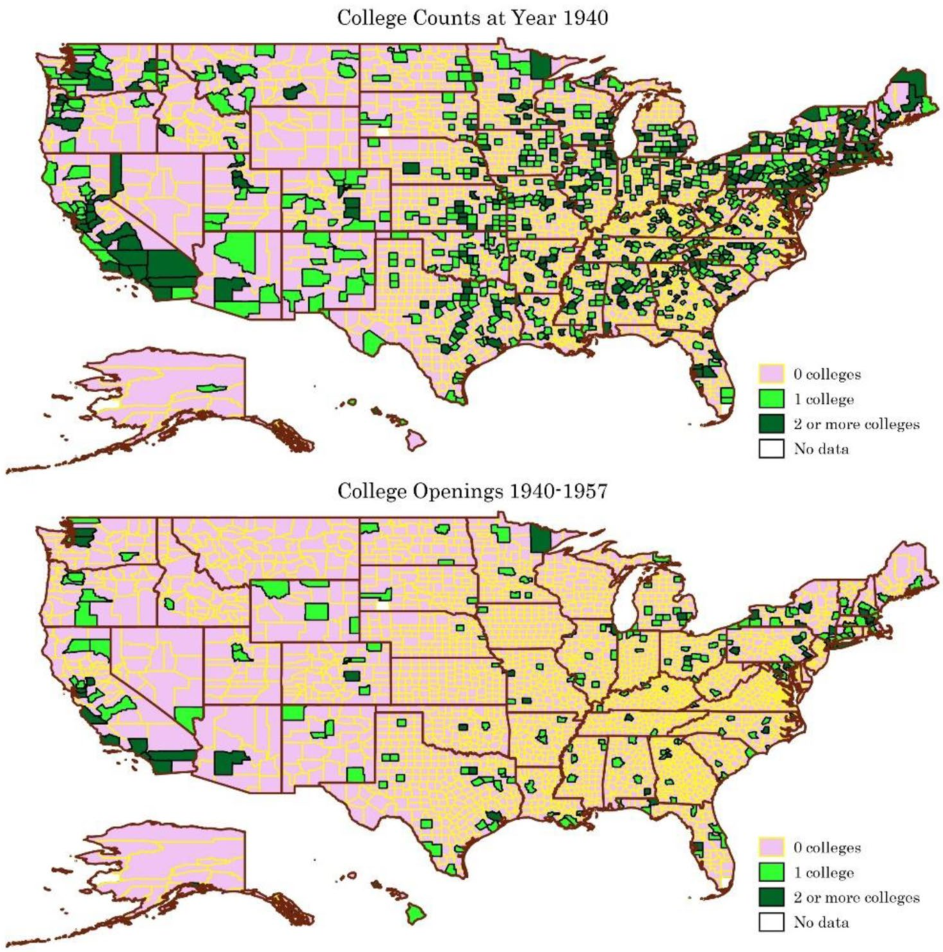


FIGURE 1 Geographic distribution of college inventory at 1940 and college expansion over the years 1940–1957. [Colour figure can be viewed at wileyonlinelibrary.com]

3 | DATA SOURCE

The primary source of data comes from the Censoc project outlined in Goldstein et al. (2021).[5] It uses death records from social security administration for individuals who die in old age and implement data-linkage techniques to link with full-count 1940 census records. We use the Censoc-Numident (hereafter Numident) dataset from this project which has linked 7.9 million individuals to the 1940 census who died between the years 1988–2005. We merge this data with the 1940 census extracted from Ruggles et al. (2020). There are three primary advantages of this linkage that help our identification strategy. First, the resulting sample size consists of millions of individuals, which is considerably larger than available longitudinal datasets and enables much more powerful statistical tests. Second, the county identifier in the 1940 census provides detailed granularity that can be used to match with the county-level college dataset, unlike state-of-birth indicators available in the Health and Retirement Study, Decennial Census, or American Community Survey data, among others. Third, we can observe other family member's socioeconomic characteristics such as parental education that we exploit in our robustness and heterogeneity analyses. Despite these advantages, the primary disadvantage of the data is its selection of death window (restricted to 1988–2005). Moreover, since the data does not report the universe of deaths that occurred between 1988 and 2005, we are unable to calculate mortality rates. Thus, we rely on age-at-death as a proxy for longevity. In sections 4.1, 5.2, and Appendix B, We implement several robustness checks and show that the results are not driven by the nonrandom selection of death window.

The data on county-level college counts are obtained from Currie and Moretti (2003). It reports the total number (as well as some disaggregated categories) of 4-year and 2-year colleges at each county from 1940-onward. Figure 1 shows the geographic distribution of college inventory at 1940 (top panel) and changes in college counts (bottom panel). About 45 and 33 states

TABLE 1 Summary statistics.

Variable	Mean	Std. Dev.	Min	Max
Death age month	832.10672	74.53261	565	995
Death year	1998.5511	4.67835	1988	2005
Birth year	1929.2172	4.75205	1923	1940
Female	0.42651	0.49457	0	1
White	0.90706	0.29035	0	1
Black	0.08903	0.28479	0	1
Other	0.0039	0.06236	0	1
Hispanic	0.01787	0.13246	0	1
First generation immigrant	0.00314	0.05591	0	1
Second generation immigrant	0.14309	0.35017	0	1
4-Years college (own + neighboring counties)	11.79555	14.83874	0	63
2-Years college (own + neighboring counties)	3.42586	5.96345	0	44
4-Years college (own county)	3.10742	5.07035	0	26
2-Years college (own county)	0.97299	2.34944	0	18
Father socioeconomic index missing	0.12383	0.32939	0	1
Father socioeconomic index 1st quartile	0.23414	0.42346	0	1
Father socioeconomic index 2nd quartile	0.21154	0.4084	0	1
Father socioeconomic index 3rd quartile	0.20883	0.40647	0	1
Father socioeconomic index 4th quartile	0.22167	0.41537	0	1
Mother education < high school	0.5729	0.49466	0	1
Mother education = high school	0.29032	0.45391	0	1
Mother education > high school	0.05503	0.22805	0	1
Mother education missing	0.08175	0.27399	0	1
County covariates				
%Blacks	0.09693	0.15217	−0.00032	0.88543
%Whites	0.89839	0.15197	0.11346	1
%Female	0.49331	0.01656	0.1977	0.54919
Number of children <5	0.38842	0.1238	0.11719	1.05243
Occupational income score	23.68711	4.04529	11.78472	30.89858
Observations	3,967,966			

had at least one county with a 4-year and 2-year college opening in our sample.[6] About 80% of counties experienced only one expansion. The counties with the higher number of openings include Los Angeles, CA with five new 4-year colleges and six new 2-year colleges, Nassau, NY with four new 4-year colleges, and Allegheny, PA with four new 2-year colleges. We merge this with our Numident dataset based on the county of residence and the year individuals turn age 17. This leaves us with 3,967,966 observations from cohorts who were born between the years 1923–1940 and died at ages 47–82 between the years 1988–2005.

Summary statistics of the final sample are reported in Table 1. To have a better picture of the demographic composition of this sample, we compare with that of the full-count 1940 census. The sample underrepresents females (0.43 vs. 0.49), underrepresents first-generation immigrants (0.003 vs. 0.089), and overrepresents second-generation immigrants (0.14 vs. 0.09) while the share of whites, blacks, and Hispanics are quite similar to the 1940 census. The average age-at-death is 69.3 years or equivalently 832.1 months. Since the months of longevity provides a more accurate measure, we use age-at-death in months as the primary outcome. Figure 2 shows the geographic distribution of age-at-death by county of residence in 1940.

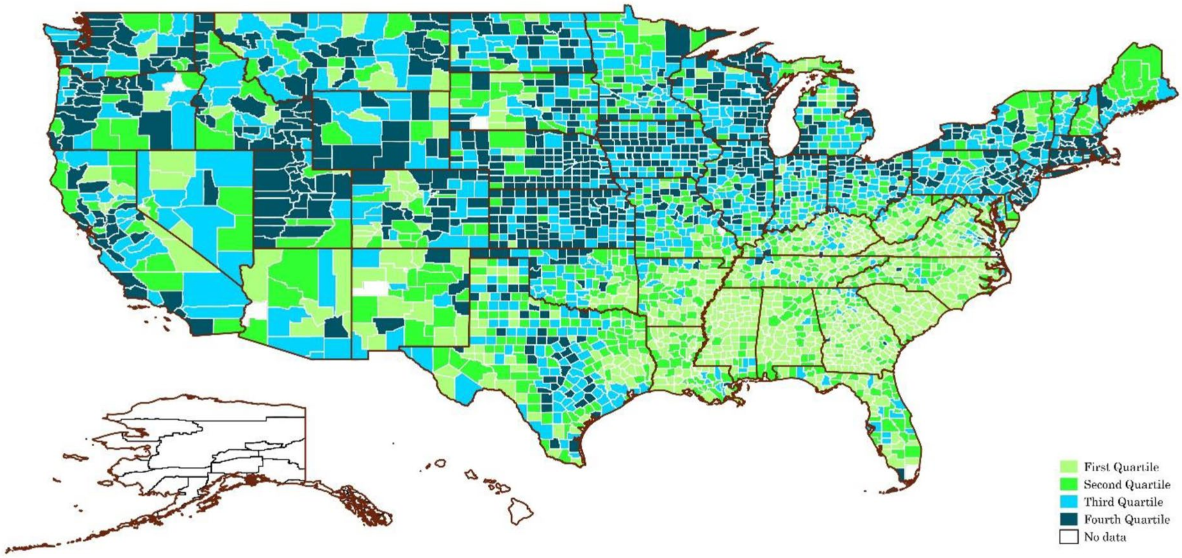


FIGURE 2 Geographic distribution of age-at-death by place of residence during childhood and adolescence at 1940 for cohorts born between 1923 and 1940 and died between 1988 and 2005. [Colour figure can be viewed at wileyonlinelibrary.com]

4 | ECONOMETRIC FRAMEWORK

The identification strategy compares the age of death of individuals who, at the age of 17, resided in counties that experienced a college expansion to those who resided in counties with no college openings, after a college opening compared to before the expansion. We operationalize this difference-in-difference model using the following regression:

$$DA_{icb} = \alpha_0 + \alpha_1 \mathrm{Coll}^{4\mathrm{year}}_{c,b+17} + \alpha_2 \mathrm{Coll}^{2\mathrm{year}}_{c,b+17} + \alpha_3 X_i + \alpha_4 Z_{cb} + \zeta_c + \gamma_b + \varepsilon_{icb} \tag{1}$$

Where the outcome (DA) is the death age of individual i from birth cohort b who, at the age of 17, resided in county c. Since college openings could have spillover effects for college attendance decisions of not only the residents of the county but also the residents of neighboring counties, we also create a second measure that aggregates our measure of college expansion to the college counts of own and neighboring counties, conditional on being in the same state. Therefore, the parameter Coll represents the total number of 4-year and 2-year colleges at the own and within-state neighboring counties where the individual resided at age 17 ($b + 17$).[7] In the difference-in-difference results, we focus on inventory of colleges as the primary independent variable. However, in the event study results we focus on new college openings as the primary independent variable.[8] The main reason is that the county may be treated several times in the event study focuses on only one event while in difference-in-difference we may look at the combined effects of all treatments.

In matrix X, we include some individual and family covariates including indicators for race/ethnicity, gender, first-generation immigrant, second-generation immigrant, dummies for maternal education, and dummies for paternal socioeconomic index. In matrix Z, we include a series of county-by-birth-year covariates. These variables are extracted from decennial censuses 1920–1940 and linearly interpolated for inter-decennial years. The covariates include share of whites, share of blacks, share of females, share of children less than 5, share of white-collar occupation employees, and average occupational income score. The county fixed effects and birth cohort fixed effects are represented by parameters ζ and γ, respectively. ε is a disturbance term. We cluster the standard errors at the county level.

4.1 | Concerns over endogeneity

The parallel trend assumption behind the identification strategy -that the outcomes of those residing in the vicinity of a college opening would have followed the same path and been influenced by the same factors as the outcomes of those not exposed to a new college opening-may be violated for three primary reasons. First, people may migrate to a county that had a college opening either for attending college or for other reasons, such as improved economic conditions, which resulted in college expansion in the first place. If individuals who chose to migrate have characteristics that are correlated with their health in the

TABLE 2 Balancing tests: College expansion and endogenous changes in demographic and socioeconomic characteristics of individuals.

	Outcomes									
	Female	**White**	**Black**	**Father socioeconomic index < median**	**Father socioeconomic index > median**	**Father socioeconomic index missing**	**Mother education < high school**	**Mother education = high school**	**Mother education > high school**	**Mother education missing**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
4-Year college	0.00041	0.00003	0.00001	−0.00488***	0.00481***	0.00007	−0.00742***	0.00498***	0.00094***	0.0015***
	(0.00035)	(0.0005)	(0.00048)	(0.00085)	(0.00052)	(0.00059)	(0.00199)	(0.00162)	(0.00019)	(0.00044)
2-Year college	0.00094***	0.00113***	−0.00076***	−0.00164	0.00228***	−0.00063	−0.0045**	0.00342**	0.0004	0.00068
	(0.00029)	(0.00031)	(0.00026)	(0.00102)	(0.00052)	(0.00076)	(0.00214)	(0.00154)	(0.00026)	(0.00053)
Observations	3,967,966	3,967,966	3,967,966	3,967,966	3,967,966	3,967,966	3,967,966	3,967,966	3,967,966	3,967,966
R-squared	0.00386	0.23174	0.24104	0.13929	0.1329	0.02565	0.07126	0.06775	0.01584	0.07301
Mean DV	0.43	0.91	0.09	0.45	0.43	0.12	0.57	0.29	0.06	0.08

Note: Each column represents a separate regression. Regressions include county and birth year fixed effects. Standard errors, clustered at the county level, are in parentheses.

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

TABLE 3 College openings and endogenous changes in contemporaneous health outcomes.

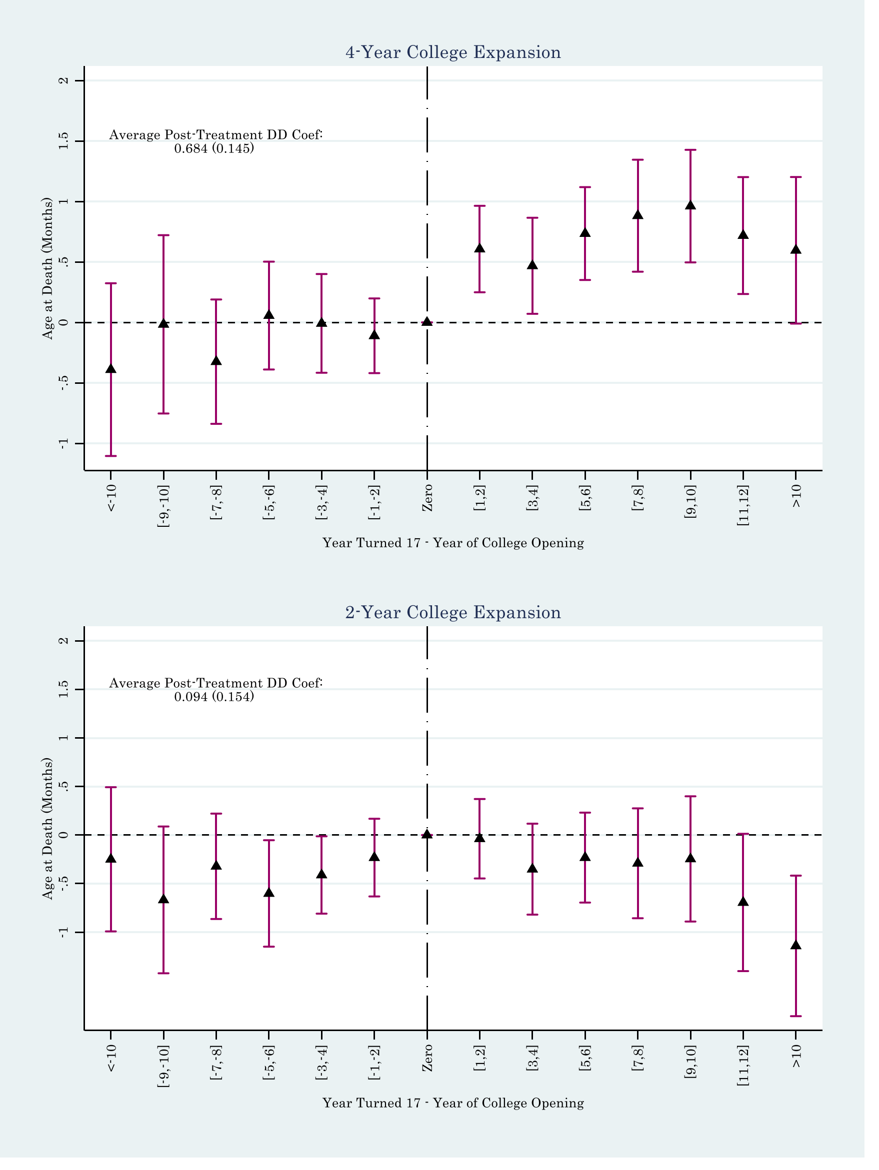
	Infant mortality per 1000 live births (1)	Total mortality per 1000 population (2)	Fertility rate (births per 1000 women) (3)	Share of births born in hospital (4)
4-Year college	0.47224*	0.03853	−0.75252***	−0.00087
	(0.24112)	(0.02831)	(0.22046)	(0.00136)
2-Year college	0.01817	−0.0317**	−0.89424***	−0.00175
	(0.11018)	(0.01588)	(0.24283)	(0.00148)
Observations	27,151	27,151	27,142	15,082
R-squared	0.67239	0.84206	0.88984	0.98479
Mean DV	35.065	9.978	97.888	0.849

Note: Each column represents a separate regression. Regressions include county fixed effects and year fixed effects. The data covers the years 1940–1958. Standard errors, clustered at the county level, are in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

long run, the OLS results of Equation (1) are biased as a result of this self-selection. For example, if whites migrate more than other races, because of their ability and affordability to move or their willingness to be more educated, there will be a sample selection problem and the results overestimate the true effects as whites have higher longevity and better health endowment for other reasons that cannot be fully captured by race dummies. To explore this potential source of endogeneity, we use a series of observable characteristics as the outcome of Equation (1). The results are reported in Table 2 for a specification with county and birth cohort fixed effects. The results offer a mixed and inconsistent pattern of migration. We observe some association between 2-year college expansion and increases in the share of whites and Hispanics and increases in high socioeconomic status fathers. There is also an increase in individuals with higher parental education (columns 7–9) as a response to an increase in 4-year colleges. However, these estimates are not very concerning as their implied effects are minuscule. For instance, the Association between two-year colleges and white is 11 basis points, off a mean of 0.9 suggesting an increase of about 0.1%. In our main results, the coefficient of white in longevity equations is roughly 3.8. Therefore, the potential influence of the observed changes in the share of whites could only induce 0.003 months change in longevity. This is also true when we look at other outcomes., The coefficient of high socioeconomic father with 4-year college implies a 1% change. Considering the magnitudes, we conclude that they endogenous changes in the sample composition based on sociodemographic and socioeconomic characteristics is not likely to induce endogeneity in our regressions. Moreover, in Appendix H, we directly test for potential influence of migration in our regressions as a result of college openings. In so doing, we use information on county and state of residence in 1935 as reported in the 1940 census to calculate within-state cross-county migration and cross-state migration between 1935 and 1940. We find insignificant and small coefficients of two-year and four-year colleges on the probability of migration. As a further test, we link individuals from 1940 to the records in 1930 census.[9] We create a dummy variable to indicate migration from 1930 to 1940 and use it as the outcome in the main regressions. The results are reported and discussed in Appendix H. We find small and insignificant coefficients, ruling out the concern regarding endogenous migration.

Second, changes in the number of colleges in an area could be correlated with changes in local and state-level regulations and legislations that impact, through education or other channels, the longevity of individuals. Moreover, college expansion could reflect other contemporaneous changes in the local economic and non-economic environment that, in turn, has long-term effects for residents. For example, city expansions could lead to additional college openings as well as improving health care access, better job opportunities, and new hospital construction, with plausibly positive long-term effects, or degrading environmental quality by raising pollution, with potentially negative cumulative effects for longevity. These omitted variables result in upward-biased and downward-biased OLS coefficients for the former case and latter case, respectively. The problem of searching for such confounding factors is the scarcity of county-level data over the period of the study (1940–1957). However, there are limited vital statistics information at the county-level which we could exploit as a proxy for general heath trends and health care access. In so doing, we use county-level birth and death data from Bailey et al. (2016) and implement regressions that include county fixed effects and year fixed effects. We regress four outcomes of interest on our measure of 4-year and 2-year college. These outcomes include infant mortality rate, total mortality rate, fertility rate, and share of births attended in the hospital. The results are reported in Table 3. We observe small and positive correlation between 4-year colleges and infant mortality rate. The implied change with respect to the mean of the outcome is about 1%. We also observed negative correlation between 2-year colleges and total mortality rate. The implied percentage change suggests a 0.3% reduction. We also observed small but significant correlations with fertility rate for both 4-year

FIGURE 3 Event-Study Results of College Expansion on Age-at-death. Point estimates and 90 percent confidence intervals are reported. Standard errors are clustered at the county level. Regressions include county and cohort fixed effects. Controls include individual, family, and county-by-cohort controls, as follows. Individual controls include dummies for female, white, black, Hispanic, first-generation immigrant, and second-generation immigrant. Family controls include dummies for maternal education and paternal socioeconomic index. County controls include share of whites, share of blacks, share of females, share of children less than 5, share of white-collar occupation employees, and average occupational income score. [Colour figure can be viewed at wileyonlinelibrary.com]



and 2-year colleges. However, we do not observe any significant and meaningful association with college openings and share of births in hospital as a proxy for general health care access. There are two points to consider interpreting the results of this table. First, the correlations are very small in magnitude and do not point to meaningful changes in health status and healthcare access in counties with college openings. Second, the coefficients do not reveal a consistent pattern. For instance, we observe positive correlation between mortality rate and 4-year college and negative correlation with 2-year college.

Third, the results could be driven by pre-trend changes in health levels that are revealed in age-at-death and cannot be absorbed by birth cohort and county fixed effects. To examine this pre-trend problem, we implement an event study analysis in which the event time is the opening of a new college. This strategy compares the outcomes of individuals aged 17 in different years relative to a college opening in their own and neighboring counties, conditional on fixed effects and covariates.[10] The results, shown in two panels of Figure 3 for 4-year and 2-year colleges, are not consistent with important pre-trends. Compared to unexposed cohorts in counties with no expansion (event time = zero), cohorts who turn age 18-above at the time of college opening (both panels) reveal no differences in their age-at-death. The event-time coefficients are small in magnitude and not statistically different from zero. For 4-year colleges, the coefficients of exposed cohorts start to rise in magnitude and become statistically significant for cohorts who turn age 16-below at the time of college expansion. The overall difference-in-difference estimate that compares post-treatment to pre-treatment cohorts is 0.68 (se = 0.15). However, we observe no post-treatment difference in the case of exposure to 2-year colleges. Important to our empirical strategy is the fact that we do not detect any pre-trend in outcomes for unaffected cohorts for both sets of college expansions.

TABLE 4 Placebo tests: Assigning colleges at ages later than 17.

	Outcome: Age-at-death (Months)					
	Colleges assigned at age 25		Colleges assigned at age 30		Colleges assigned at age 35	
	(1)	(2)	(3)	(4)	(5)	(6)
4-Year college	0.03721	0.02002	0.05411	0.03695	0.04657	0.03002
	(0.04553)	(0.04654)	(0.04547)	(0.0455)	(0.04175)	(0.04325)
2-Year college	0.07825**	0.05367	0.03855	0.02928	0.04043	0.03225
	(0.03296)	(0.03624)	(0.02657)	(0.02791)	(0.02565)	(0.02677)
Observations	3,967,966	3,967,939	3,967,966	3,967,939	3,967,966	3,967,939
R-squared	0.44164	0.44403	0.44164	0.44403	0.44164	0.44403
Mean DV	832.1	832.1	832.1	832.1	832.1	832.1
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes

Note: Each column represents a separate regression. Standard errors, clustered at the county level, are in parentheses. Controls include individual, family, and county-by-cohort controls, as follows. Individual controls include dummies for female, white, black, Hispanic, first-generation immigrant, and second-generation immigrant. Family controls include dummies for maternal education and paternal socioeconomic index. County controls include share of whites, share of blacks, share of females, share of children less than 5, share of white-collar occupation employees, and average occupational income score.

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

TABLE 5 The association between college openings and successful merging of numident with the 1940-census data.

	Outcome: Successful merging between numident and 1940 census		
	(1)	(2)	(3)
4-Year college	0.0002***	0.0005***	−0.0005***
	(0.0001)	(0.0001)	(0.0002)
2-Year college	−0.0008***	−0.0006***	−0.001***
	(0.0001)	(0.0001)	(0.0002)
Observations	40,213,767	35,448,656	4,765,049
R-squared	0.0145	0.015	0.0086
Mean DV	0.1	0.1	0.1
County FE	Yes	Yes	Yes
Birth cohort FE	Yes	Yes	Yes
Individual controls	No	Yes	Yes
Family controls	No	Yes	Yes
County controls	No	No	Yes

Note: Each column represents a separate regression. Standard errors, clustered at the county level, are in parentheses. Controls include individual, family, and county-by-cohort controls, as follows. Individual controls include dummies for female, white, black, Hispanic, first-generation immigrant, and second-generation immigrant. Family controls include dummies for maternal education and paternal socioeconomic index. County controls include share of whites, share of blacks, share of females, share of children less than 5, share of white-collar occupation employees, and average occupational income score.

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

As an additional step we implement a series of placebo tests by assigning the number of colleges to individuals at ages later than 17, specifically, ages 25, 30, and 35. If healthier individuals, who would have otherwise higher longevity, move to counties with college expansions for reasons related to overall trends of the county (e.g., better job opportunities) when they age 25, then the college availability in the new environment should be strongly associated with their health outcomes later in life. The results (shown in Table 4) do not provide evidence for this issue. The effects are quite small (relative to the main results) and in most cases statistically insignificant.

Another concern is sorting individuals based on the probability of linkage between Numident death records and the 1940 census. This becomes problematic for our identification strategy if the linkage is correlated with higher/lower number of colleges in local area of residence. To explore this source of selection, we start with the original population of individuals born between the years 1923–1940 and observed in the full-count 1940 census (roughly 40.2 million observations). We link this with the Numident death records in order to create a new dummy variable that indicates successful merging between the two datasets. We then merge this with county-level college data and regress the successful merging dummy variable on measures of 4-year and 2-year college opening using the same empirical method as discussed in Equation (1). The results are reported in Table 5. Across all models, we observe statistically significant correlations between number of colleges and the probability of being in the final linked sample. However, the point estimates are economically small implying minuscule changes in the outcome. For instance, based on the full specification of column 3, exposure to one additional 4-year college is associated with 0.5 basis points decreases in the likelihood of being in the final sample. This number is equivalent to a reduction of about 0.5% change with respect to the mean of the outcome. We posit that the unobserved differences in longevity of the unmerged cohorts in the original sample and merged cohorts in the final sample should be extremely large in order for this sample selection to induce a meaningful endogeneity into our results.

One may also argue that the window of observation in social security administration death record is narrow and does not include those who die earlier or later. Looking at the Vital Statistics cause-specific death record data from 1959 to 2017, roughly 39% of deaths to birth cohorts of 1923–1940 occur between 1988 and 2005 (Numident window). Comparing to deaths outside of this window (1959–1987, 2006–2017), Numident records are 1.3% points less likely to be white, 4.6% points less likely to be female, and 3.4% points more likely to be black. We will show that, in our sample, the effects are stronger for 4-year colleges and that the effects are not statistically different among different races and ethnicity but more pronounced among males. Therefore, one possible concern in extending the results in this paper to the whole population is the overrepresentation of males for which the effects are stronger.

In another attempt to explore possible issues with the window selection of Numident, we replicate the main results using Censoc-DMF data which links death records of males who die between 1975 and 2005 with the 1940 census. This allows us to explore the effects of deaths that occurred up to 12 years before the start of Numident data. We apply the same sample selections and implement the same econometric method as in the main analysis in the text. These results are reported in Appendix B. The effects of the death window of 1988–2005 (similar to Numident) are virtually the same as the main results of section 5.1. However, the effects are smaller in magnitude when we look at all years covered in DMF. This may suggest that the effects of education on mortality are better detected at older ages. Overall, we should be cautious in interpreting the main results considering left and right truncation of this data. Even though the results of Table 5 do not point to a difference in the probability of being in the death data from the original cohorts, it does not distinguish between too early deaths and too late deaths, that is, it may be that nontreated cohorts died earlier than 1988 and treated cohorts died after 2005. In this scenario, there is no difference in being in the final sample as the effects of too early and too late deaths offset each other. If we had access to post 2005 death, under this assumption, we would have observed larger impacts, as the comparison between Numident and DMF results of Appendix B also suggest.

5 | RESULTS

5.1 | Main results

Figure 4 shows the density distribution of age-at-death for counties with no colleges (never treated, in red) and counties with at least one college opening (treated, in green). For observations below the median of age at death, we observe a higher density for those in no college opening counties. Likewise, for those in counties with at least an opening, the longevity density is higher in higher values of the outcome. Specifically, cohorts exposed to college expansions have 3.6 months higher age-at-death (the raw difference in mortality in counties with at least one opening compared with those in counties with no college opening).

To account for confounding factors, we turn our focus from this visual difference to the difference-in-difference strategy of Equation (1). The main results are reported in Table 6 for specifications that include county and birth year fixed effects (column 1), individual covariates (column 2), family controls (column 3), and county covariates (column 4). The marginal effects of 4-year colleges are positive, significant, and robust across specifications. It implies that an additional increase in the number of 4-year colleges is associated with roughly 0.16 months higher longevity. While this result is a small effect, we can put it into perspective by comparing it with the marginal effects of other covariates, specifically, females, blacks, and other races. This effect is equivalent to 6.2% of the difference in age-at-death between people of other races and whites, 4.2% difference in black-white gap in age-at-death, and 2.5% of the difference in age-at-death of females versus males.

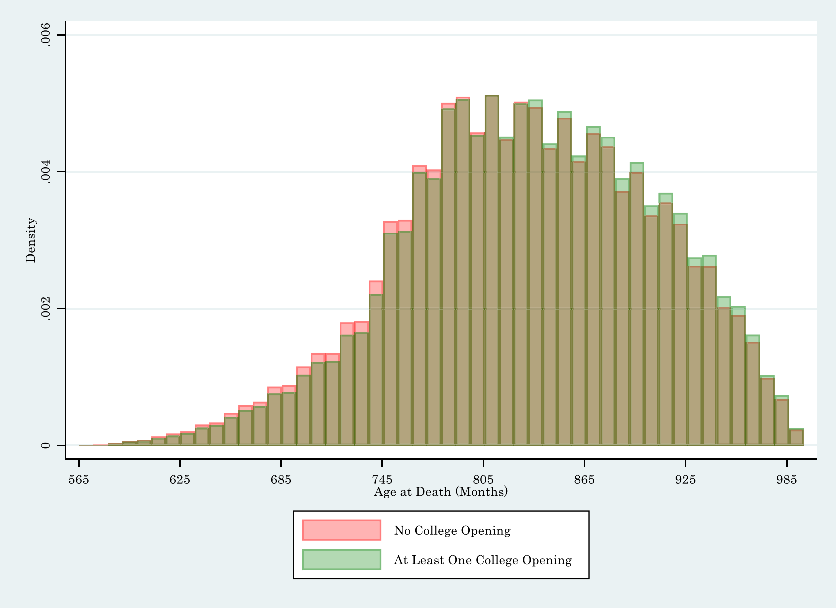


FIGURE 4 Density Distribution of Age-at-death for Counties without/with College Expansion. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 6 Main results: College expansion and age-at-death.

	Outcome: Death age (Months)			
	(1)	(2)	(3)	(4)
4-Year college	0.18033***	0.18671***	0.16775***	0.1635***
	(0.04468)	(0.0462)	(0.04518)	(0.04962)
2-Year college	−0.01423	−0.02115	−0.0282	−0.03029
	(0.05929)	(0.06319)	(0.06163)	(0.06101)
Observations	3,967,966	3,967,966	3,967,966	3,967,939
R-squared	0.44164	0.4438	0.44403	0.44403
Mean DV	832.1	832.1	832.1	832.1
Elasticity of 4-year college	0.00256	0.00265	0.00238	0.00232
Elasticity of 2-year college	−5.9e-05	−8.7e-05	−0.00012	−0.00012
County FE	Yes	Yes	Yes	Yes
Birth cohort FE	Yes	Yes	Yes	Yes
Individual controls	No	Yes	Yes	Yes
Family controls	No	No	Yes	Yes
County controls	No	No	No	Yes

Note: Each column represents a separate regression. Standard errors, clustered at the county level, are in parentheses. Controls include individual, family, and county-by-cohort controls, as follows. Individual controls include dummies for female, white, black, Hispanic, first-generation immigrant, and second-generation immigrant. Family controls include dummies for maternal education and paternal socioeconomic index. County controls include share of whites, share of blacks, share of females, share of children less than 5, share of white-collar occupation employees, and average occupational income score.

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

The average number of 4-year colleges in own county and neighboring counties in our final sample is about 3.4 units. An increase of 3.4 4-year colleges would increase the age-at-death of the population by 0.54 months, which closes the life expectancy gap between the US and other OECD countries by 3.2%.[11] The effect of 2-year college expansion is inconclusive. We observe negative coefficients but the point estimates are very small in magnitude and statistically insignificant. To have a better comparison of the effects of 4-year versus 2-year colleges, we report the elasticities at the end of each column.

The results reported here are intention-to-treat effects as the college expansion leads to college education of only a fraction of the population. Although we explore the direct link between college expansion and college education in section 5.3 and try to convert the marginal effects into treatment-on-treated effects in section 5.4, we should note that these effects provide a minimum benefit of college openings on long-run mortality outcomes.

Moreover, not all the positive effects of college opening operate through increases in own education.[12] For instance, it could improve the education of the spouse which results in gains in mortality for both husband and wife (Jaffe et al., 2006; Spoerri et al., 2014). Therefore, college openings could affect mortality by increases in the education of other current and future family members rather than one's own education. Another aspect is the improvements in education of coworkers which in turn has possible health spillover for own mortality outcomes later in life (Martins & Jin, 2008).

5.2 | Robustness checks

In Table 7, we execute a wide range of robustness checks. In column 2, we add region-of-birth by birth-year fixed effects to account for regional economic shocks that affects cohorts. We observe the comparable coefficient for 4-year college. The coefficient of 2-year college becomes positive but remains small and insignificant. In column 3, we add a county specific linear cohort trend to account for all secular evolution of characteristics across cohorts within the same county. We observe the reduction of about 18% in the coefficient of 4-year college but considerable increase in the coefficient of two-year college. In this model, the coefficient of 2-year college becomes much larger and statistically significant, suggesting an increase of about 0.1 months in longevity.

Several studies suggest that seasonal birth has influences in health at birth and a wide range of later life outcomes, including longevity (Doblhammer & Vaupel, 2001; Vaiserman, 2021). Moreover, there is evidence of the influence of seasonal death on mortality outcomes (Marti-Soler et al., 2014). In columns 4–5, we investigate these sources of endogeneity by adding dummies for month of birth and month of death into regressions, respectively. We observe virtually similar coefficients as the main results.

In column 6, we allow for fixed effects of counties to have differential effects across individuals from different socioeconomic backgrounds. In so doing, we interact county fixed effects with maternal education dummies and paternal socioeconomic status dummies. We observe very similar effects compared with those of column 1.

In column 7, we add a serious of additional family covariates. Specifically, we add fathers' wage income as reported in 1940, father years of schooling, and house value. Since for many observations these covariates are missing, we end up with a substantially smaller sample size. We observe larger effects for 4-year college and small and insignificant effect for 2-year college. To complement this analysis, we add mother fixed effects into regressions to compare sibling outcomes. The sibling strategy reported in column 8 -which compares the age-at-death of different siblings to the same mother who aged 17 in different years relative to a college expansion-suggests that after accounting for unobserved time-invariant family characteristics the effects could be even larger in magnitude than the main results (0.26 vs. 0.16).

One concern is that college openings may occur in areas with higher population and potentially a higher college going subpopulation. Therefore, county population may play a role in the association between college openings and college education, hence college openings and longevity. In column 9, we add the polynomial function of county population. We observe comparable coefficients to the main results.

Additional tests are related to functional form checks. In column 10, to capture the nonlinearities in the effects, we replace the outcome with binary variables that equal one if age at death is greater than 60 years. We observe a positive and significant effect of 4-year college opening on the probability of aging beyond 60. The point estimate suggests 7.9 basis points increases in longevity beyond 60 years for exposure to an additional 4-year college opening, off a mean of 0.9. In column 11, we replace the outcome with the log of age at death. The coefficient of 4-year college suggests an increase of about 0.02%, which is almost identical to the percentage change with respect to the meaning of the outcome implied by column 1.

While in the main results we cluster standard errors at the county level, in column 12–13 we show the robustness to alternative correction methods of standard errors. Specifically, we use Huber White standard errors in column 12 and cluster standard errors at the county-cohort level in column 13. We observe comparable standard errors to the main results.

Another concern is related to endogenous merging that we discussed in section 4.1. As an additional robustness check, we also employ estimation strategy proposed by Heckman (1979) that accounts for endogeneity of sample selection. This method employs a two-step strategy. In the first step, the method estimates the probability of being in the linked sample from the original cohorts of 1940 as a function of observable factors, such as race, ethnicity, gender, and parental characteristics. It then calculates an Inverse Miller Ratio (IMR) that captures potential bias from sample selection and adds it to the second stage model (longevity equation) as a control variable. The results are reported in column 14. We observe very similar coefficients as those reported in the main results.

TABLE 7 Robustness checks of college expansion and age-at-death.

	Column 3 table 6	Adding region by year FE	Adding county linear trend	Adding birth month by birth year FE	Adding month of death FE	Adding interaction of county FE by parental control dummies	Adding father wage, father education, house value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
4-Year college	0.1635***	0.1482***	0.13138**	0.16453***	0.16188***	0.16604***	0.20797***
	(0.04962)	(0.04928)	(0.06341)	(0.04952)	(0.04948)	(0.04925)	(0.06789)
2-Year college	−0.03029	0.00939	0.0903*	−0.03199	−0.02966	−0.03211	−0.04375
	(0.06101)	(0.04344)	(0.04758)	(0.05994)	(0.06166)	(0.05975)	(0.06917)
Observations	3,967,939	3,967,939	3,967,939	3,967,939	3,967,939	3,967,933	1,378,076
R-squared	0.44403	0.44406	0.44452	0.44571	0.445	0.44536	0.42267
Mean DV	832.1	832.1	832.1	832.1	832.1	832.1	839.2

	Adding mother FE	Adding a polynomial function of county population	Outcome: Age at Death>60	Outcome in logarithm (Semi-log specification)	Huber white se	Clustering se at county and birth year	Heckman (1979) estimate
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
4-Year college	0.26736***	0.15137***	0.00079***	0.00021***	0.1635***	0.1635***	0.15813***
	(0.09035)	(0.04412)	(0.00023)	(0.00006)	(0.03867)	(0.05246)	(0.03864)
2-Year college	−0.04285	−0.00778	0.00035**	−0.00003	−0.03029	−0.03029	−0.04652
	(0.09961)	(0.04537)	(0.00017)	(0.00008)	(0.03529)	(0.0371)	(0.03555)
Observations	1,005,445	3,967,939	3,967,939	3,967,939	3,967,939	3,967,939	40,273,164
R-squared	0.68516	0.44404	0.18622	0.44554	0.44403	0.44403	------
Mean DV	831.4	832.1	0.9	6.7	832.1	832.1	832.1

	Propensity score matching method	State in 1940 = birth state	Age in 1940 ≤ 15	Drop colleges if not in IPEDS list	Drop colleges if not granting at least an associate degree	Public colleges	Private colleges
	(15)	(16)	(17)	(18)	(19)	(20)	(21)
4-Year college	0.15565	0.14897***	0.16901***	0.12248*	0.17425***	0.29358***	0.16402**
	(0.16190)	(0.05654)	(0.05394)	(0.06391)	(0.04547)	(0.08383)	(0.0655)
2-Year college	0.03502	−0.03134	−0.01642	0.10276*	−0.06514	−0.07278	0.1886**
	(0.18196)	(0.07077)	(0.05981)	(0.05851)	(0.05465)	(0.04861)	(0.0904)
Observations	3,967,939	3,583,743	3,459,329	3,967,939	3,642,544	3,580,580	3,967,933

TABLE 7 (Continued)

	Propensity score matching method	State in 1940 = birth state	Age in 1940 ≤ 15	Drop colleges if not in IPEDS list	Drop colleges if not granting at least an associate degree	Public colleges	Private colleges
	(15)	**(16)**	**(17)**	**(18)**	**(19)**	**(20)**	**(21)**
R-squared	----	0.4491	0.41858	0.44452	0.45913	0.42427	0.44536
Mean DV	832.1	830.5	823.0	832.1	829.2	825.1	832.1

Note: Each column within each panel represents a separate regression. Standard errors, clustered at the county level, are in parentheses. Regressions include county and cohort fixed effects. Controls include individual, family, and county-by-cohort controls, as follows. Individual controls include dummies for female, white, black, Hispanic, first-generation immigrant, and second-generation immigrant. Family controls include dummies for maternal education and paternal socioeconomic index. County controls include share of whites, share of blacks, share of females, share of children less than 5, share of white-collar occupation employees, and average occupational income score.

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

To further examine the robustness to potential endogenous confounders, we employ a Propensity Score Matching method which creates a control group of counties with no colleges nearby but similar in county characteristics to counties that had a college nearby. We report the results in column 15. We observe very similar coefficients as those reported in the main results.

While we extensively discuss and address the potential migration issues in section 4.1, we add two additional tests to further explore this concern. First, we exclude from the sample observations whose state of birth is different than state of residence in 1940 under the assumption that between state migration is motivated, among other factors, availability and opening of colleges. However, the marginal effect of column 16 is quite comparable to the main effects and are consistent with our assumption regarding selective between-state migrations. Second, since individuals may migrate for college education when they are 17–18 years old, including these cohorts could introduce bias in our estimations. To explore this, we restrict the sample to observations whose age in 1940 is less than 15. The marginal effects for this subsample are reported in column 17 and is almost identical to those in the main results.

As mentioned in Currie and Moretti (2003), while the Integrated Postsecondary Education Data System (IPEDS) is one primary source of college data, some colleges are not reported in this database. Column 18 shows that eliminating those colleges that are not in the IPEDS listing leads to slightly smaller coefficients (0.12 vs. 0.16). Moreover, not all colleges in our sample grant a degree. Column 19 suggests that focusing only on colleges that grant at least an associate degree produces similar coefficients. Decomposing the colleges by public-private status reveals somewhat heterogeneous effects. The 4-year college coefficient increases by 80% when we look at public colleges and remain similar to the main results for private colleges (column 20 and 21).

5.3 | First stage effects

The results so far suggest a consistent and robust reduced-form effect of college openings on later-life mortality. The next step is to explore whether or not college openings increased educational outcomes and if so, to quantify the first stage effects of college openings on education. The Numident data does not report the education or other labor market outcomes of the deceased. One possible approach to estimate the first stage of college openings on educational attainment is to turn to 1960 5% Decennial Census, as it has detailed information on completed education as well as income, most of the 1923–1940 cohorts have completed their education by 1960. The main disadvantage of this data is that the county identifier is not available in the public use version and instead, it reports Public Use Microdata Area (PUMA). In addition, IPUMS de-identifies counties based on other geographic variables (including PUMA) and details about location characteristics. In the 1960 census, there are 1344 PUMAs, and IPUMS de-identifies 435 counties. A PUMA is primarily defined based on the population of an area. In low populated areas (usually rural areas), a PUMA contains several counties while in high populated areas (usually urban areas) a county contains several PUMAs. Therefore, we aggregate college counts at areas where PUMAs cover several counties and use county-level college counts for areas where several PUMAs are included in a large county, conditioning on the fact that the county is de-identified by IPUMS.[13] Therefore, the variation in college expansions across areas is based on a combination of PUMA and county. Using this method, we can merge college data with the 1960 sample and have a match rate of about 97%. We drop respondents below age 22 as the primary outcome of interest is a college education. To mitigate migration issues, we implement two additional sample selection criteria. First, we use the information of state of residence 5 years ago (available in the 1960 census) to exclude those individuals whose current state of residence is different from their residence 5 years earlier. Second, we restrict the sample to respondents below age 30 in the year 1960. This leaves us with 341,834 observations. We implement regressions introduced in Equation (1) while adding a PUMA-county fixed effect instead of county fixed effects. The results are reported in Table 8 for different outcomes in columns. An additional 4-year college raises the probability of having any college education by over 1% point (column 1), equivalent to an increase of roughly 5.3% from the mean of the outcome. The effect of 2-year college is smaller in magnitude and imprecisely estimated. To compare these two coefficients, we focus on the elasticities (reported at the end of columns). A 10% increase in the number of 4-year colleges raises any college education by 2.2%, respectively. Similar to the main results, we observe negative and insignificant coefficients for 2-year college opening.

We also observe increases in years of schooling (column 5), wage income (column 6), and total income (column 7) as a result of 4-year college opening. For instance, an additional 4-year college is associated with approximately 3.6% rise in total income (significant at 5% level) while the 2-year college effects are quite small and statistically insignificant. One possible reason for small and imprecize estimations of 2-year college is that we are observing these individuals at younger ages than the age range over which the labor market returns of college education would appear.[14]

These estimated first-stage effects are partly in line with the results of Cowan and Tefft (2020) and Currie and Moretti (2003). Cowan and Tefft (2020) employ state-level measures of college access over the years 1980–2015 and find that college access

TABLE 8 College expansion and education-income at 1960 census.

	First stage outcomes:					Reduced form outcomes:	
	Education: At least one year of college	Education: At least two years of college	Education: At least three years of college	Education: At least four years of college	Years of schooling	Log of wage income	Log of total income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
4-Year college	0.0106***	0.0088***	0.0069***	0.0048***	0.0921***	0.0552**	0.0362**
	(0.0031)	(0.0028)	(0.0023)	(0.0017)	(0.0263)	(0.0226)	(0.0175)
2-Year college	−0.0018	−0.0033	−0.005**	−0.0023	0.0159	0.0135	0.0006
	(0.0032)	(0.0026)	(0.0022)	(0.0017)	(0.0299)	(0.0263)	(0.0182)
Observations	341,834	341,834	341,834	341,834	341,834	341,834	341,687
R-squared	0.0527	0.0455	0.0369	0.0321	0.1252	0.3146	0.3786
Mean DV	0.204	0.156	0.113	0.086	11.252	5.145	5.404
Elasticity of 4-year college	0.2239	0.2439	0.2636	0.2379	0.03522	0.02881	0.04613
Elasticity of 2-year college	−0.01491	−0.03552	−0.0735	−0.04506	0.00234	0.00019	0.00435
PUMA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Birth cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Each column represents a separate regression. Individual controls include dummies for female, white, black, Hispanic, first-generation immigrant, and second-generation immigrant. Standard errors, clustered at the county-PUMA level, are in parentheses.

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

(per capita) reduces high school graduation but increases college attendance. However, their first-stage results are primarily driven by 2-year public colleges. The first stage results of Currie and Moretti (2003) suggest that both 2-year and 4-year college expansions increase years of schooling with largest effects for 4-year college opening. Their data source is Natality birth records over the years 1970–1999. There are two reasons that the first stage results of both papers are somewhat different than ours. First, both papers employ different years of data and for different cohorts. Second, while the focus of both papers are in concentration and accessibility of colleges (measured by per capita values), our paper focuses on college openings.

While the results of Table 8 point to improvements in education and income of exposed cohorts as likely pathways, another potential mechanism is changes in the composition of occupations in the county following college openings. For instance, if a 4-year college opening is accompanied by the establishment of new firms with better-paid jobs and occupations, then part of the effects on longevity could operate through occupational composition changes. In Appendix G, we explore this channel using decennial census data over the years 1940–1960. Our results fail to provide robust and consistent evidence that there are discernible changes in the composition of occupations following college expansions. Moreover, we use historical County Business Pattern (CBP) database for the years 1946–1958 and investigate the effects of college openings on industry-specific employment per capita. We report and discuss these results in Appendix G. The results do not provide consistent evidence of changes in employment across different industries following a new college opening.

5.4 | A discussion on the magnitude of the main results

The results so far suggest that college expansions have long-run longevity gains and that these gains are primarily driven by 4-year college expansions. These point estimates and larger effects of 4-year college expansions in Table 6 are also in line with the first stage results of Table 8. They suggest that college expansions (and specifically 4-year college openings) increase college education and possibly total income, and through these channels, they positively affect the longevity of individuals. If this story is true, we can combine the first stage results (column 1, Table 8) with the main results (column 3, Table 6) and convert the Intent-to-Treat effects into Treatment-on-Treated effects. A back-of-the-envelope calculation suggests that an additional 4-year college raises the age-at-death of those who would have otherwise not attended college by about 15.1 months.[15] This is smaller than the effects of college education on life expectancy reported by Lacroix et al. (2019) who found that college education raises longevity at age 51 by about 49 months. This effect is comparable to the findings of Halpern-Manners et al. (2020) who implement a twin-strategy and show that each additional year of schooling is associated with roughly 4 months higher age at death.[16]

To add to the TOT effect calculation, we also implement a two-sample two-stage least-square regression by using the information of completed education from the 1960 census and the information on mortality from 1940-census-Numident sample. We aggregate the college counts at the PUMA-county level following the same method as in the first stage results (section 5.3). We apply a two-sample SLS method that includes birth cohort fixed effects, PUMA-county fixed effects, and individual controls. The results are reported in Appendix C. In the fully specified model, having at least a college degree is associated with roughly 33 months higher age-at-death. We also note that measurement error in the first stage analysis as well as the two-sample 2SLS framework (which uses PUMA-county rather than county linkages) would likely attenuate our coefficients and therefore would suggest smaller treatment on the treated effects.

5.5 | Heterogeneity of the results: Heterogeneity by gender-race/ethnicity

The similar patterns in the main results (Table 6) and the findings of the first stage effects (Table 8) suggest that improvements in education and income are likely channels of impact. To better connect the first stage effects and the reduced-form results, we explore how the effects vary across observable characteristics. If our assumption regarding the mechanism channel is correct, then we would observe larger mortality reductions in subpopulations that experience larger improvements in education-income outcomes. Therefore, we explore (and compare) the heterogeneity by gender, race, and ethnicity for both mortality and education-income outcomes.

In Table 9, we explore how the effects vary over demographic characteristics by interacting with college counts a dummy for male, other races, black, and low socioeconomic status father (Duncan socioeconomic index of father being below median) (columns 1–4, in order). The mortality effect of 4-year and 2-year colleges is 0.03 and 0.07 months larger for males than females. The marginal effects of 4-year and 2-year colleges among people of other races/ethnicities (Native Americans, Chinese, Japanese, and 'other' Asian-Pacific Islanders vs. whites and blacks) do not reveal is statistically significant difference with other racial groups. Among blacks, the mortality gains from 2-year college opening are positive and significant, suggesting an additional 0.08 months of longevity.

Heterogeneity results based on father's socioeconomic status document a mixed pattern. We observe smaller effects of 4-year college opening among low socioeconomic status fathers (compared with high socioeconomic status fathers) while we observe increases in longevity of low socioeconomic status fathers individuals as a result of exposure to a 2-year college opening.

We observe similar heterogeneity patterns in the first stage results when we interact the gender/race dummies with measures of college expansion and replace the outcome with various measures of college education using the 1960 census data. These results are reported in Appendix D. The effects of 4-year college expansion are more pronounced among males for all measures of college education. The marginal effects of both 4-year and 2-year college openings are slightly (and insignificantly) larger among other races in comparison with blacks and whites. This consistent pattern between the first stage effects and reduced-form effects also holds when we look at differential effects among blacks versus non-blacks. The 4-year college effects are larger among non-blacks while the effects of 2-year colleges are larger among blacks (compare with the results in Table 9).[17]

5.5.1 | Heterogeneity in 2 × 2 difference-in-difference (DD) estimate

The empirical strategy of Equation (1) operates as a difference-in-difference (DD) model that compares the outcome of cohorts with higher versus lower exposure to college expansions (treatment). However, the OLS estimation of the DD estimator in a two-way fixed effect framework, that the treated group receives the treatment at different points of time (in comparison to a pre-post and treatment-control DD estimation), compares the outcome of all combinations of two-by-two treatment-control/pre-post pairs. The least-square coefficient is finally a weighted average of all these comparisons with the weights in the proportion of how long the pair had received the treatment and also the variance of the treatment (Goodman-Bacon, 2021). For instance, the OLS compares the outcomes of those cohorts who lived in counties that experienced an expansion to those in counties with no colleges at all (treatment vs. never treated), it compares the outcomes of those who experienced a college expansion to those that had an expansion earlier and no new expansion after that (later treatment vs. earlier control), and those who had an expansion earlier versus those who had later (earlier treatment vs. later control). To explore this pairwise heterogeneity of the OLS-produced DD estimation, we implement bacon-decomposition (Goodman-Bacon, 2021), which shows the coefficients of the pairwise DD estimates for 4-year colleges on age-at-death against their respective weights. In so doing, we collapse the data at the county by birth cohort level and assigned a dummy for treatment based on exposure to a college opening.[18] The

TABLE 9 Heterogeneity of the main results by gender, race, and ethnicity.

	Outcome: Death age (Months)			
	(1)	(2)	(3)	(4)
Male × 4-Year college	0.02811***			
	(0.00989)			
Male × 2-Year college	0.07016***			
	(0.02017)			
Male	−7.31946***			
	(0.09259)			
Other × 4-Year college		0.06678		
		(0.07021)		
Other × 2-Year college		0.03932		
		(0.12013)		
Other		−3.05326***		
		(0.6428)		
Black × 4-Year college			0.00559	
			(0.01348)	
Black × 2-Year college			0.08103**	
			(0.03206)	
Black			−3.49192***	
			(0.14402)	
Father SEI low × 4-Year college				−0.01647**
				(0.00772)
Father SEI low × 2-Year college				0.03367*
				(0.01985)
Father SEI low				−0.30176**
				(0.11918)
4-Year college	0.13892***	0.16325***	0.16134***	0.16953***
	(0.05176)	(0.04961)	(0.04929)	(0.04933)
2-Year college	−0.07655	−0.0302	−0.03368	−0.0416
	(0.05845)	(0.06085)	(0.06009)	(0.05718)
Observations	3,967,939	3,967,939	3,967,939	3,967,939
R-squared	0.44406	0.44404	0.44404	0.44404
Birth cohort FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Note: Each column represents a separate regression. Standard errors, clustered at the county level, are in parentheses. Regressions include county and cohort fixed effects. Controls include individual, family, and county-by-cohort controls, as follows. Individual controls include dummies for female, white, black, Hispanic, first-generation immigrant, and second-generation immigrant. Family controls include dummies for maternal education and paternal socioeconomic index. County controls include share of whites, share of blacks, share of females, share of children less than 5, share of white-collar occupation employees, and average occupational income score.

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

results are reported in Figure 5. While there is heterogeneity in the pairwise DD coefficients, all three types of comparisons reveal positive effects. The largest weight is assigned to the comparisons between those with a college expansion versus those with no colleges in their county (weight = 0.92, DD-coefficient = 0.26).[19] Comparing earlier expansions as the treatment to later expansions as the control group (weight = 0.04, DD-coefficient = 0.41) or comparing later expansions as treatment and earlier expansions as control (weight = 0.04, DD-coefficient = 0.26) reveal average DD coefficients that are quite close to the treated-vs-never-treated comparison as well as the overall DD estimation (=0.266). Therefore, though the OLS effects of DD

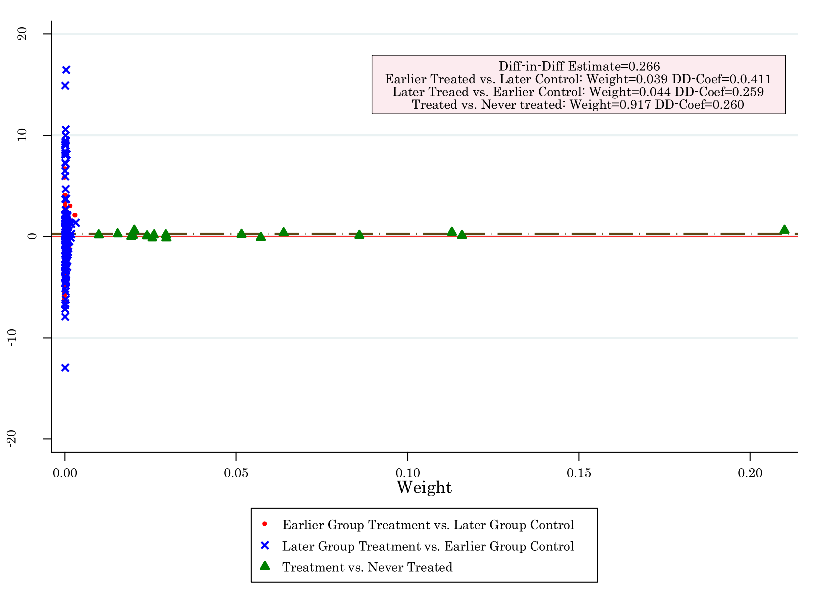


FIGURE 5 Bacon decomposition of difference-in-difference estimates. [Colour figure can be viewed at wileyonlinelibrary.com]

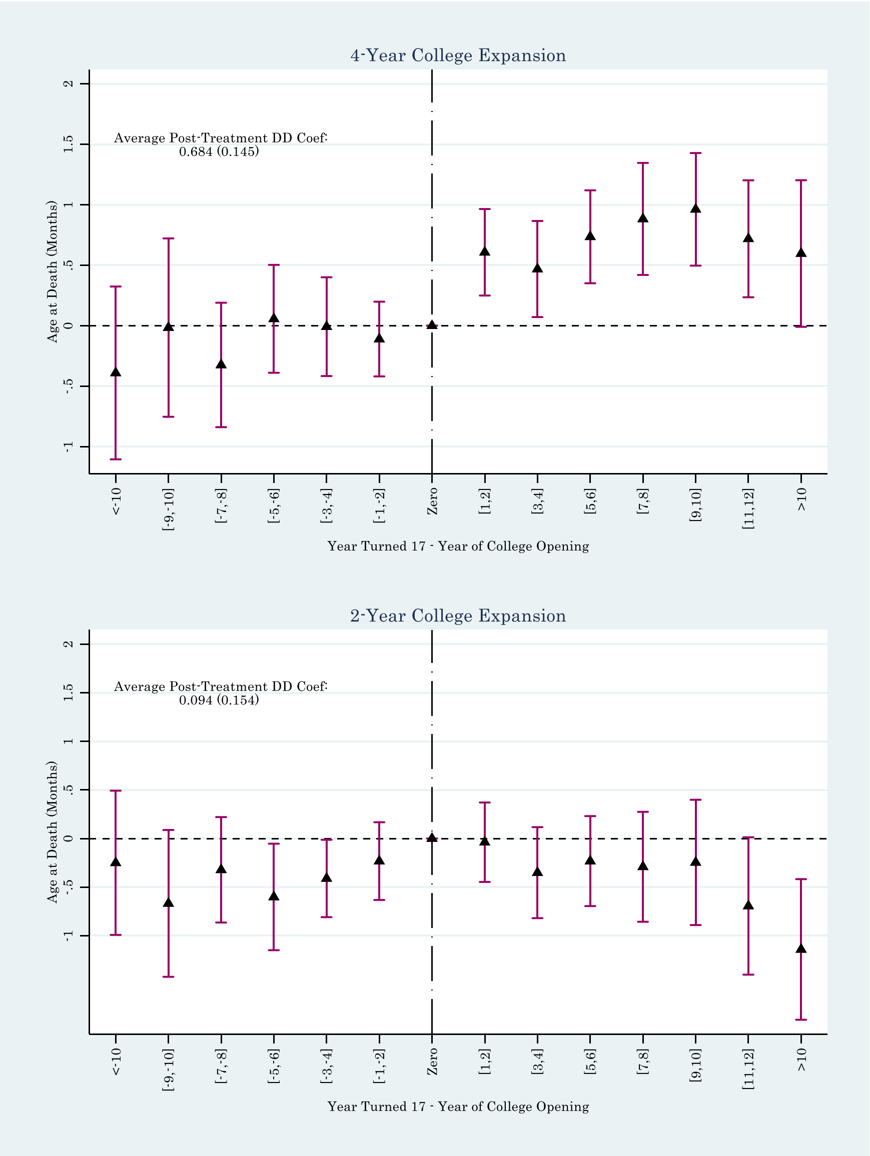
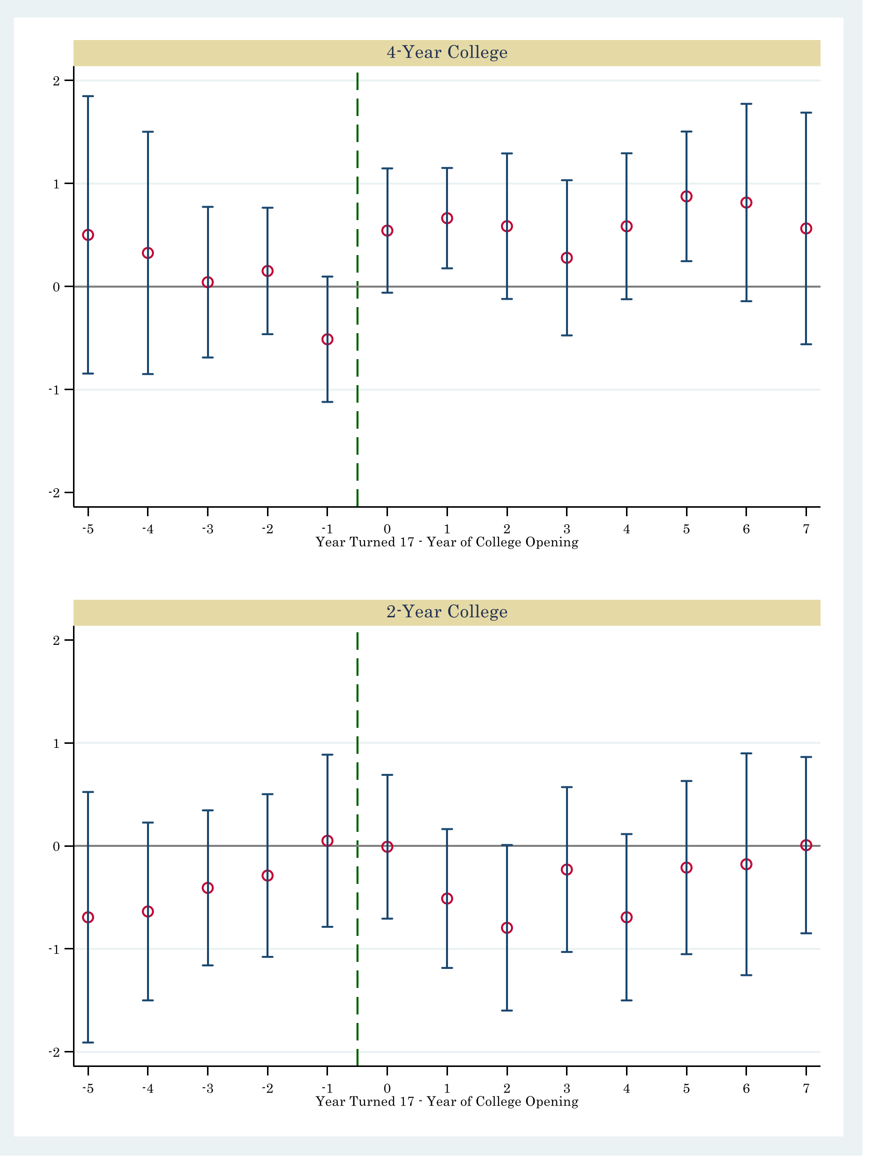


FIGURE 6 Event Study of Sun and Abraham (2021) Dynamic Treatment Difference-in-Difference Estimates. Point estimates and 90 percent confidence intervals are reported. Standard errors are clustered at the county level. Regressions include county and cohort fixed effects. Controls include individual, family, and county-by-cohort controls, as follows. Individual controls include dummies for female, white, black, Hispanic, first-generation immigrant, and second-generation immigrant. Family controls include dummies for maternal education and paternal socioeconomic index. County controls include share of whites, share of blacks, share of females, share of children less than 5, share of white-collar occupation employees, and average occupational income score. [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 7 Event Study of De Chaisemartin and D'Haultfoeuille Dynamic Treatment Difference-in-Difference Estimates. Point estimates and 90 percent confidence intervals are reported. Standard errors are clustered at the county level. Regressions include county and cohort fixed effects. Controls include individual, family, and county-by-cohort controls, as follows. Individual controls include dummies for female, white, black, Hispanic, first-generation immigrant, and second-generation immigrant. Family controls include dummies for maternal education and paternal socioeconomic index. County controls include share of whites, share of blacks, share of females, share of children less than 5, share of white-collar occupation employees, and average occupational income score. [Colour figure can be viewed at wileyonlinelibrary.com]



estimation reveal various effects that are sometimes negative, there is no evidence to suggest that the main results are driven by one specific type of comparison.

As an additional check for the robustness of the effects to other DD estimations, we implement an event study analysis based on two-way fixed effect estimation proposed by Sun and Abraham (2021). The results are depicted in Figure 6. We observe very similar coefficients to the main event study results. As a further check, we also implement the difference-in-difference event study estimation method introduced by de Chaisemartin & D'Haultfœuille (2020). They show that the overall DD estimation produced by OLS could be biased downward due to negative weights assigned to some pairs in group-time DD coefficients. They propose an alternative method that is free of negative weight contamination. We replicate the event-study analysis using their proposed method.[20] The results are reported in Figure 7. We observe a similar pattern as the event-study results produced by OLS. Interestingly, the post-treatment coefficients of 4-year college opening (top panel) reveal larger effects than those reported in. This suggest that some negative weights assigned by OLS (and observed in the bacon-decomposition of Figure 5) bias the overall effects downward.

6 | CONCLUSION

The benefits of education, as shown by a large body of literature, can go beyond its labor market returns. This study investigated the potential long-term effects of college expansion on mortality and longevity. Our results shed new light on the long-run health benefits of education and add to the literature on social returns of education and, specifically, college education. We found that a new 4-year college construction in the local area increases any college education by about 5.2% from the mean and adds to the average schooling by 0.09 additional years. It also contributes to age-at-death by about 0.16 months. A back-of-an-envelope calculation suggests that the effects could be as large as 15 months for those who do attend a 4-year college after a college opening and would have not attended if the college had not been constructed.

An event-study analysis showed that for unaffected cohorts college expansions had no effects, reducing concerns over pre-trends in the outcome, while the effects start to rise for the affected cohorts, those younger than 17 at the time of college expansion. We also explored whether the county-expansion-induced migration of possibly college-educated cohorts to college-expanding counties has driven our results by regressing the observable characteristics of individuals on college counts. The evidence is not consistent and strong enough to point to this selective migration. Moreover, a series of placebo tests, in which we assign the measure of colleges to individuals in ages later than 17, supports our empirical method.

To illustrate the contribution of college expansion to longevity of the US population, we extrapolate our findings to the second half of twentieth century college openings and mortality trends. Between the years 1940–1990, there has been 489 new 4-year college construction across US counties. The share of local area (own and neighboring county) that were affected by the expansion accounts for 0.4% of the US population. Between the years 1970–2018, the average age at death, conditional on survival up to age 47, increased from 72.19 to 77.09 years[21] Considering the fact that college educated individuals accounted for 18.5% of deaths and implementing the results of Table 6 and first stage results of Table 8, one can calculate that the 1940–1990 college expansions increased average age at death for the whole US population by 5.35 months. This is equivalent to 7.5% of the observed increase in the longevity trend between the years 1970–2018.[22]

ACKNOWLEDGMENTS

We thank Enrico Moretti for generously providing the data on college openings. We also thank Joshua Goldstein, Adriana Lleras-Muney, participants of Health-Aging-Place working group at the University of Wisconsin-Madison, the CenSoc Working Group at the University of California-Berkeley, seminar participants of the Midwest Economic Association conference, participants of the American Society of Health Economists Conference, and seminar participants of the Data-Intensive Research Conference for their comments and suggestions. The authors would like to acknowledge financial support from NIA grant R01AG060109 and the Center for Demography of Health and Aging (CDHA) at the University of Wisconsin-Madison under NIA core grant P30 AG17266.

CONFLICT OF INTEREST STATEMENT

The authors claim no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Hamid Noghanibehambari https://orcid.org/0000-0001-7868-2900

ENDNOTES

[1] Many earlier studies within this literature examined cross-sectional associations of education and mortality without fully addressing the common endowments' influence in driving both education and health (Kravdal, 2008; Meara et al., 2008; Ross et al., 2012; Zajacova, 2006).

[2] Our paper also adds to the growing literature on the effects of early life, childhood, and early adulthood exposures and experiences on later-life old-age mortality outcomes. This literature evaluates the relevance of various conditions, environmental factors, and policy exposures for later-life mortality and longevity (Aizer et al., 2016; Almond et al., 2018; Fletcher, 2009, 2012; Hayward & Gorman, 2004; Noghanibehambari & Noghani, 2023; Zhang et al., 2020). For instance, Noghanibehambari and Fletcher (2023a) examine the effects of birth registration policies interacted with compulsory schooling laws and child labor laws for individuals later life mortality. They show that those individuals born in states that has established a birth registration law and are exposed to stricter child labor laws during adolescent years lived longer lives. Atherwood (2022) examining early adulthood exposure to the Dust Bowl on longevity and finds insignificant effects. Noghanibehambari and Fletcher (2023b) focus on childhood exposure to the Dust Bowl and find negative and significant results.

[3] The authors use state-level variation in college access rather than our focus on county level variation in college openings.

[4] Fletcher and Frisvold (2011) compare siblings in Add Health to show that college selectivity is associated with lower likelihood of tobacco and marijuana use in young adulthood.

[5] The name is extracted from the beginnings of *cen*sus and *soc*ial security administration.

[6] The states with no 4-year college expansions include: Delaware, DC, Montana, Utah, Wisconsin, and Wyoming. The states with no 2-year college expansions include: Alaska, Arizona, Arkansas, Delaware, DC, Idaho, Iowa, Montana, Nebraska, Nevada, New Jersey, Rode Island, South Dakota, Tennessee, Vermont, Virginia, West Virginia, and Wisconsin.

[7] Appendix A shows that the results are quite robust and similar when we replace our measure of county by own county number of colleges or with own county and all neighboring counties regardless of the fact that they are within the same state or not.

[8] Here, since the comparison is derived from counties with an incremental inventory of colleges with those of no college counties, the interpretation is similar to having a college opening.

[9] We should notice that the cross census linking induces three important sample restrictions. First, the cross census linking rules are extracted from the census linking project. The links are successful for about 30% of observations. Second, since females change their names, linking rules are only available for male individuals. Third, we are able to link those people born between 1923 and 1930 as the point of observation will be 1930 census.

[10] The event-study regressions are similar to the full specification of Equation (1). Specifically, we use a regression of the following form: $DA_{icb} = \alpha + \sum_{k=-1}^{\underline{T}} \xi_k 1\left(t_{c,b+17} - t^*_{cb} = k\right) + \sum_{k=+1}^{\overline{T}} \eta_k 1\left(t_{c,b+17} - t^*_{cb} = k\right) + \beta X_i + \lambda Z_{cb} + \zeta_c + \gamma_b + \varepsilon_{icb}$. In this formulation, the coefficients ξ represent the average difference in the outcome between cohorts who turned 17 k periods before a new college opening and cohorts who turned 17 at the time of a new opening (pre-treatment coefficients). Likewise, the coefficients η represent the average difference in the outcome between cohorts who turned 17 k periods after a new college opening and cohorts who turned 17 at the time of a new opening (post-treatment coefficients). We include county fixed effects, birth year fixed effects (λ), individual covariates, family controls, and county covariates (as listed in section 4). We cluster standard errors at the county level. As we mentioned in section 4, the comparison of the event study results is based on new college opening in the county, hence the treatment relates to a new college construction between periods t and $t-1$, while in the main difference-in-difference results we focus on inventory of colleges in a county and its neighboring counties, that is, total number of colleges.

[11] Based on OECD health status reports, life expectancy of USA in 2008 was 77.9 and the average OECD countries was 79.3.

[12] This is the primary reason that we prefer a reduced-form analysis and avoid applying 2SLS-IV tests as our main analysis method here. Since college opening could potentially operate through non-own-education channel to improve mortality outcomes, the exclusion restriction assumption will be violated.

[13] For instance, Los Angeles County consists of 45 PUMAs in 1960 census.

[14] While the effects found here are consistent with our main findings, we are aware that the results could be distorted by migration issues as there are evidence of the effects of college education on migration (Malamud & Wozniak, 2012).

[15] This is based on the outcome in column 1 of Table 8: attending at least 1 year of college. This does not include only those who have a college degree but all those who have ever attended any college. This is a better measure for our interpretation as college opening could operate through other channels except education-income to improve old-age health such as peer effects and access to better health-related information and improved critical thinking skills (Cutler & Lleras-Muney, 2006).

[16] We should note that our results measure Local Average Treatment Effects (LATE) and establish the college-opening-induced marginal effects which by construction is different than what Halpern-Manners et al. (2020) estimate.

[17] In Appendix E, we implement additional tests on subsamples based on region of residence, birth cohort, and county characteristics.

[18] To implement this analysis, we assigned a dummy for treatment that takes a value of one if the county experienced a new opening from year t t to year t-1 and zero otherwise. Please refer to footnote 13 for further details of the estimation equation.

[19] In Appendix F, we replicate the event-study analysis for a subsample that includes treated versus never-treated observations as the bacon-decomposition puts a larger weight on this subset. The results are quite similar to the pattern observed in Figure 3.

[20] In footnote 13, we show the regressions that we estimate for the event study analysis.

[21] We look at this time period rather than 1940–1990 (period of college expansion calculation) since we have assigned college expansion at age 17 and we restrict our attention to those aged at least 47, per Numident death coverage. This means that the earlier cohort that we observe must have been 17 years old in 1940 and have reached age 47, which points to a mortality window starting at 1970.

[22] Since the effects of Table 6 are local average treatment effects, which shows the changes in outcome for those in local areas rather than the whole population, we need to multiply the effects by the share of locally affected population to the whole US population. The final number is calculated as the product of total new college construction (489), the inverse value of first stage effect (1/0.0108), the main effect of 4-year college on longevity (0.164), the share of college educated people in death records (0.185), and the average share of local population to the whole US population over the same period (0.004).

REFERENCES

Acemoglu, D., & Angrist, J. (2000). How large are human-capital externalities? Evidence from compulsory schooling laws. *NBER Macroeconomics Annual*, *15*, 9–59. https://doi.org/10.1086/654403

Aizer, A., Eli, S., Ferrie, J., & Muney, A. L. (2016). The long-run impact of cash transfers to poor families. *The American Economic Review*, *106*(4), 935–971. https://doi.org/10.1257/AER.20140529

Albouy, V., & Lequien, L. (2009). Does compulsory education lower mortality? *Journal of Health Economics*, *28*(1), 155–168. https://doi.org/10.1016/J.JHEALECO.2008.09.003

Almond, D., Currie, J., & Duque, V. (2018). Childhood circumstances and adult outcomes: Act II. *Journal of Economic Literature*, *56*(4), 1360–1446. https://doi.org/10.1257/jel.20171164

Angrist, J. D., & Keueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics*, *106*(4), 979–1014. https://doi.org/10.2307/2937954

Atherwood, S. (2022). Does a prolonged hardship reduce life span? Examining the longevity of young men who lived through the 1930s great plains drought. *Population and Environment*, *43*(4), 1–23. https://doi.org/10.1007/S11111-022-00398-W

Backlund, E., Sorlie, P. D., & Johnson, N. J. (1999). A comparison of the relationships of education and income with mortality: The national longitudinal mortality study. *Social Science & Medicine*, *49*(10), 1373–1384. https://doi.org/10.1016/S0277-9536(99)00209-9

Bailey, M., Clay, K., Fishback, P., Haines, M., Kantor, S., Severnini, E., & Wentz, A. (2016). *U.S. County-level natality and mortality data, 1915–2007*. Inter-University Consortium for Political and Social Research. https://doi.org/10.3886/E100229V4

Bautista, M. A., González, F., Martinez, L. R., Munoz, P., & Prem, M. (2020). Does higher education reduce mortality? Evidence from a natural experiment. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.3717849

Behrman, J. R., Kohler, H.-P., Jensen, V. M., Pedersen, D., Petersen, I., Bingley, P., & Christensen, K. (2011). Does more schooling reduce hospitalization and delay mortality? New evidence based on Danish twins. *Demography*, *48*(4), 1347–1375. https://doi.org/10.1007/S13524-011-0052-1

Black, D. A., Hsu, Y. C., & Taylor, L. J. (2015). The effect of early-life education on later-life mortality. *Journal of Health Economics*, *44*, 1–9. https://doi.org/10.1016/j.jhealeco.2015.07.007

Buckles, K., Hagemann, A., Malamud, O., Morrill, M., & Wozniak, A. (2016). The effect of college education on mortality. *Journal of Health Economics*, *50*, 99–114. https://doi.org/10.1016/J.JHEALECO.2016.08.002

Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., & Pan, Y. (2014). Early childhood investments substantially boost adult health. *Science*, *343*(6178), 1478–1485. https://doi.org/10.1126/SCIENCE.1248429/SUPPL_FILE/CAMPBELL_SM.PDF

Card, D. (1999). The causal effect of education on earnings. *Handbook of Labor Economics*, *3*(1), 1801–1863. https://doi.org/10.1016/S1573-4463(99)03011-4

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, *69*(5), 1127–1160. https://doi.org/10.1111/1468-0262.00237

Cellini, S. R., & Chaudhary, L. (2014). The labor market returns to a for-profit college education. *Economics of Education Review*, *43*, 125–140. https://doi.org/10.1016/J.ECONEDUREV.2014.10.001

Chevalier, A., Harmon, C., Walker, I., & Zhu, Y. (2004). Does education raise productivity, or just reflect it? *The Economic Journal*, *114*(499), F499–F517. https://doi.org/10.1111/j.1468-0297.2004.00256.x

Clark, D., & Royer, H. (2013). The effect of education on adult mortality and health: Evidence from Britain. *The American Economic Review*, *103*(6), 2087–2120. https://doi.org/10.1257/AER.103.6.2087

Conti, G., Heckman, J., & Urzua, S. (2010). The education-health gradient. *The American Economic Review*, *100*(2), 234–238. https://doi.org/10.1257/AER.100.2.234

Conti, G., Heckman, J. J., & Pinto, R. (2016). The effects of two influential early childhood interventions on health and healthy behaviour. *The Economic Journal*, *126*(596), F28–F65. https://doi.org/10.1111/ECOJ.12420

Cornaglia, F., Crivellaro, E., & McNally, S. (2015). Mental health and education decisions. *Labour Economics*, *33*, 1–12. https://doi.org/10.1016/J.LABECO.2015.01.005

Cowan, B. W., & Tefft, N. (2020). College access and adult health. https://doi.org/10.3386/W26685

Cuijpers, P., & Smit, F. (2002). Excess mortality in depression: A meta-analysis of community studies. *Journal of Affective Disorders*, *72*(3), 227–236. https://doi.org/10.1016/S0165-0327(01)00413-X

Currie, J., & Moretti, E. (2003). Mother's education and the intergenerational transmission of human capital: Evidence from college openings. *Quarterly Journal of Economics*, *118*(4), 1495–1532. https://doi.org/10.1162/003355303322552856

Cutler, D. M., & Lleras-Muney, A. (2006). *Education and health: Evaluating theories and evidence* (Vol. 37). National Bureau of Economic Research. https://doi.org/10.3386/W12352

de Chaisemartin, C., & D'haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *The American Economic Review*, *110*(9), 2964–2996. https://doi.org/10.1257/AER.20181169

de Walque, D. (2007). Does education affect smoking behaviors? Evidence using the Vietnam draft as an instrument for college education. *Journal of Health Economics*, *26*(5), 877–895. https://doi.org/10.1016/j.jhealeco.2006.12.005

Dickson, M., & Harmon, C. (2011). Economic returns to education: What we know, what we don't know, and where we are going—Some brief pointers. *Economics of Education Review*, *30*(6), 1118–1122. https://doi.org/10.1016/J.ECONEDUREV.2011.08.003

Doblhammer, G., & Vaupel, J. W. (2001). Lifespan depends on month of birth. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(5), 2934–2939. https://doi.org/10.1073/PNAS.041431898/SUPPL_FILE/4318FIG5.PDF

Doll, R., & Hill, A. B. (1956). Lung cancer and other causes of death in relation to smoking. *British Medical Journal*, *2*(5001), 1071–1081. https://doi.org/10.1136/BMJ.2.5001.1071

Eckert, F., Lam, K., Mian, A. R., Müller, K., Schwalb, R., & Sufi, A. (2022). The early county business pattern files: 1946–1974. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.4259255

Everett, B. G., Rehkopf, D. H., & Rogers, R. G. (2013). The nonlinear relationship between education and mortality: An examination of cohort, race/ethnic, and gender differences. *Population Research and Policy Review*, *32*(6), 893–917. https://doi.org/10.1007/S11113-013-9299-0

Fenelon, A., & Preston, S. H. (2012). Estimating smoking-attributable mortality in the United States. *Demography*, *49*(3), 797–818. https://doi.org/10.1007/S13524-012-0108-X

Fischer, M., Karlsson, M., & Nilsson, T. (2013). Effects of compulsory schooling on mortality: Evidence from Sweden. *International Journal of Environmental Research and Public Health*, *10*(8), 3596–3618. https://doi.org/10.3390/IJERPH10083596

Fletcher, J. M. (2009). Childhood mistreatment and adolescent and young adult depression. *Social Science & Medicine*, *68*(5), 799–806. https://doi.org/10.1016/J.SOCSCIMED.2008.12.005

Fletcher, J. M. (2012). The effects of first occupation on long term health status: Evidence from the Wisconsin longitudinal study. *Journal of Labor Research*, *33*(1), 49–75. https://doi.org/10.1007/S12122-011-9121-X/TABLES/13

Fletcher, J. M. (2015). New evidence of the effects of education on health in the US: Compulsory schooling laws revisited. *Social Science & Medicine*, *127*, 101–107. https://doi.org/10.1016/J.SOCSCIMED.2014.09.052

Fletcher, J. M., & Frisvold, D. E. (2011). College selectivity and young adult health behaviors. *Economics of Education Review*, *30*(5), 826–837. https://doi.org/10.1016/J.ECONEDUREV.2011.04.005

Fletcher, J. M., & Frisvold, D. E. (2014). The long run health returns to college quality. *Review of Economics of the Household*, *12*(2), 295–325. https://doi.org/10.1007/S11150-012-9150-0

Fruehwirth, J. C. (2014). Can achievement peer effect estimates inform policy? A view from inside the black box. *The Review of Economics and Statistics*, *96*(3), 514–523. https://doi.org/10.1162/REST_A_00385

Galama, T. J., Lleras-Muney, A., & van Kippersluis, H. (2018). The effect of education on health and mortality: A review of experimental and quasi-experimental evidence. https://doi.org/10.3386/W24225

Gathmann, C., Jürges, H., & Reinhold, S. (2015). Compulsory schooling reforms, education and mortality in twentieth century Europe. *Social Science & Medicine*, *127*, 74–82. https://doi.org/10.1016/J.SOCSCIMED.2014.01.037

Goldstein, J. R., Alexander, M., Breen, C., Miranda González, A., Menares, F., Osborne, M., Snyder, M., & Yildirim, U. (2021). Censoc project. In *CenSoc mortality file*. University of California. Retrieved from https://censoc.berkeley.edu/data/.Version 2.0.

Gonzalez, F., & Quast, T. (2013). Non-linearities in the relationship between aggregate income and mortality rates. *Eastern Economic Journal*, *41*(1), 51–69. https://doi.org/10.1057/EEJ.2013.36

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277. https://doi.org/10.1016/J.JECONOM.2021.03.014

Halpern-Manners, A., Helgertz, J., Warren, J. R., & Roberts, E. (2020). The effects of education on mortality: Evidence from linked U.S. Census and administrative mortality data. *Demography*, *57*(4), 1513–1541. https://doi.org/10.1007/S13524-020-00892-6

Hayward, M. D., & Gorman, B. K. (2004). The long arm of childhood: The influence of early-life social conditions on men's mortality. *Demography*, *41*(1), 87–107. https://doi.org/10.1353/DEM.2004.0005

Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *The American Economic Review*, *103*(6), 2052–2086. https://doi.org/10.1257/AER.103.6.2052

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–161. https://doi.org/10.2307/1912352

Hong, K., Savelyev, P. A., & Tan, K. T. K. (2020). Understanding the mechanisms linking college education with longevity. *Journal of Human Capital*, *14*(3), 371–400. https://doi.org/10.1086/710221

Jaffe, D. H., Eisenbach, Z., Neumark, Y. D., & Manor, O. (2006). Effects of husbands' and wives' education on each other's mortality. *Social Science & Medicine*, *62*(8), 2014–2023. https://doi.org/10.1016/J.SOCSCIMED.2005.08.030

Jamison, E. A., Jamison, D. T., & Hanushek, E. A. (2007). The effects of education quality on income growth and mortality decline. *Economics of Education Review*, *26*(6), 771–788. https://doi.org/10.1016/J.ECONEDUREV.2007.07.001

Jemal, A., Thun, M. J., Ward, E. E., Henley, S. J., Cokkinides, V. E., & Murray, T. E. (2008). Mortality from leading causes by education and race in the United States, 2001. *American Journal of Preventive Medicine*, *34*(1), 1–8.e7. https://doi.org/10.1016/J.AMEPRE.2007.09.017

Johnson, N. J., Sorlie, P. D., & Backlund, E. (1999). The impact of specific occupation on mortality in the U.S. National Longitudinal Mortality Study. *Demography*, *36*(3), 355–367. https://doi.org/10.2307/2648058

Kalediene, R., & Petrauskiene, J. (2005). Inequalities in mortality by education and socio-economic transition in Lithuania: Equal opportunities? *Public Health*, *119*(9), 808–815. https://doi.org/10.1016/J.PUHE.2004.11.004

Kamhöfer, D. A., Schmitz, H., & Westphal, M. (2019). Heterogeneity in marginal non-monetary returns to higher education. *Journal of the European Economic Association*, *17*(1), 205–244. https://doi.org/10.1093/JEEA/JVX058

Kravdal, Ø. (2008). A broader perspective on education and mortality: Are we influenced by other people's education? *Social Science & Medicine*, *66*(3), 620–636. https://doi.org/10.1016/J.SOCSCIMED.2007.10.009

Lacroix, G., Laliberté-Auger, F., Michaud, P.-C., & Parent, D. (2019). The effect of college education on health and mortality: Evidence from Canada. *Health Economics*, *30*(S1), 105–118. https://doi.org/10.1002/HEC.3975

Lager, A. C. J., & Torssander, J. (2012). Causal effect of education on mortality in a quasi-experiment on 1.2 million Swedes. *Proceedings of the National Academy of Sciences*, *109*(22), 8461–8466. https://doi.org/10.1073/PNAS.1105839109

Lefèbvre, M., Pestieau, P., & Ponthiere, G. (2018). Premature mortality and poverty measurement in an OLG economy. *Journal of Population Economics*, *32*(2), 621–664. https://doi.org/10.1007/S00148-018-0688-X

Leigh, D. E., & Gill, A. M. (1997). Labor market returns to community colleges: Evidence for returning adults. *Journal of Human Resources*, *32*(2), 334–353. https://doi.org/10.2307/146218

Lindahl, M. (2005). Estimating the effect of income on health and mortality using lottery prizes as an exogenous source of variation in income. *Journal of Human Resources*, *XL*(1), 144–168. https://doi.org/10.3368/JHR.XL.1.144

Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States. *The Review of Economic Studies*, *72*(1), 189–221. https://doi.org/10.1111/0034-6527.00329

Lleras-Muney, A., Price, J., & Yue, D. (2020). The association between educational attainment and longevity using individual level data from the 1940 census. https://doi.org/10.3386/W27514

Long, M. C. (2010). Changes in the returns to education and college quality. *Economics of Education Review*, *29*(3), 338–347. https://doi.org/10.1016/J.ECONEDUREV.2009.10.005

Lundborg, P., Lyttkens, C. H., & Nystedt, P. (2016). The effect of schooling on mortality: New evidence from 50,000 Swedish twins. *Demography*, *53*(4), 1135–1168. https://doi.org/10.1007/S13524-016-0489-3

Luy, M., di Giulio, P., & Caselli, G. (2011). Differences in life expectancy by education and occupation in Italy, 1980–1994: Indirect estimates from maternal and paternal orphanhood. *Population Studies*, *65*(2), 137–155. https://doi.org/10.1080/00324728.2011.568192

Lynch, S. M. (2003). Cohort and life-course patterns in the relationship between education and health: A hierarchical approach. *Demography*, *40*(2), 309–331. https://doi.org/10.1353/DEM.2003.0016

Madsen, M., Andersen, A.-M. N., Christensen, K., Andersen, P. K., & Osler, M. (2010). Does educational status impact adult mortality in Denmark? A twin approach. *American Journal of Epidemiology*, *172*(2), 225–234. https://doi.org/10.1093/AJE/KWQ072

Malamud, O., Mitrut, A., & Pop-Eleches, C. (2021). The effect of education on mortality and health: Evidence from a schooling expansion in Romania. *Journal of Human Resources*, *56*(2), 1118–9863R2. https://doi.org/10.3368/JHR.58.4.1118-9863R2

Malamud, O., & Wozniak, A. (2012). The impact of college on migration. *Journal of Human Resources*, *47*(4), 913–950. https://doi.org/10.3368/JHR.47.4.913

Martins, P. S., & Jin, J. Y. (2008). Firm-level social returns to education. *Journal of Population Economics*, *23*(2), 539–558. https://doi.org/10.1007/S00148-008-0204-9

Marti-Soler, H., Gonseth, S., Gubelmann, C., Stringhini, S., Bovet, P., Chen, P. C., Wojtyniak, B., Paccaud, F., Tsai, D. H., Zdrojewski, T., & Marques-Vidal, P. (2014). Seasonal variation of overall and cardiovascular mortality: A study in 19 countries from different geographic locations. *PLoS One*, *9*(11), e113500. https://doi.org/10.1371/JOURNAL.PONE.0113500

Mazumder, B. (2008). Does education improve health? A reexamination of the evidence from compulsory schooling laws. *Economic Perspectives*, *32*(Q II), 2–16.

Mazzonna, F. (2014). The long lasting effects of education on old age health: Evidence of gender differences. *Social Science & Medicine*, *101*, 129–138. https://doi.org/10.1016/J.SOCSCIMED.2013.10.042

Meara, E. R., Richards, S., & Cutler, D. M. (2008). The gap gets bigger: Changes in mortality and life expectancy, by education, 1981–2000. *Health Affairs*, *27*(2), 350–360. https://doi.org/10.1377/HLTHAFF.27.2.350

Meghir, C., Palme, M., & Simeonova, E. (2018). Education and mortality: Evidence from a social experiment. *American Economic Journal: Applied Economics*, *10*(2), 234–256. https://doi.org/10.1257/APP.20150365

Muller, A. (2002). Education, income inequality, and mortality: A multiple regression analysis. *BMJ*, *324*(7328), 23. https://doi.org/10.1136/BMJ.324.7328.23

Noghanibehambari, H., & Fletcher, J. (2023a). Childhood exposure to birth registration laws and old-age mortality. *Health Economics*, *32*(3), 735–743. https://doi.org/10.1002/HEC.4643

Noghanibehambari, H., & Fletcher, J. M. (2023b). Dust to feed, Dust to grey: The effect of in-utero exposure to the Dust Bowl on old-age longevity. *Demography*. https://doi.org/10.3386/W30531

Noghanibehambari, H., & Noghani, F. (2023). Long-run intergenerational health benefits of women empowerment: Evidence from suffrage movements in the US. *Health Economics*, *32*(11), 2583–2631. https://doi.org/10.1002/HEC.4744

Oreopoulos, P., & Salvanes, K. G. (2011). Priceless: The nonpecuniary benefits of schooling. *The Journal of Economic Perspectives*, *25*(1), 159–184. https://doi.org/10.1257/JEP.25.1.159

Preston, S. H., Glei, D. A., & Wilmoth, J. R. (2010). A new method for estimating smoking-attributable mortality in high-income countries. *International Journal of Epidemiology*, *39*(2), 430–438. https://doi.org/10.1093/IJE/DYP360

Psacharopoulos, G. (1985). Returns to education: A further international update and implications. *Journal of Human Resources*, *20*(4), 583. https://doi.org/10.2307/145686

Ross, C. E., Masters, R. K., & Hummer, R. A. (2012). Education and the gender gaps in health and mortality. *Demography*, *49*(4), 1157–1183. https://doi.org/10.1007/S13524-012-0130-Z

Ruggles, S., Flood, S., Goeken, R., Grover, J., & Meyer, E. (2020). IPUMS USA: Version 10.0. [dataset]. IPUMS. https://doi.org/10.18128/D010.V10.0

Savelyev, P. A. (2020). Conscientiousness, extraversion, college education, and longevity of high-ability individuals. *Journal of Human Resources*, *58*(1), 0918–9720R2. https://doi.org/10.3368/JHR.58.1.0918-9720R2

Savelyev, P. A., Ward, B. C., Krueger, R. F., & McGue, M. (2022). Health endowments, schooling allocation in the family, and longevity: Evidence from US twins. *Journal of Health Economics*, *81*, 102554. https://doi.org/10.1016/J.JHEALECO.2021.102554

Schulz, R., Drayer, R. A., & Rollman, B. L. (2002). Depression as a risk factor for non-suicide mortality in the elderly. *Biological Psychiatry*, *52*(3), 205–225. https://doi.org/10.1016/S0006-3223(02)01423-3

Snyder, S. E., & Evans, W. N. (2006). The effect of income on mortality: Evidence from the social security Notch. *The Review of Economics and Statistics*, *88*(3), 482–495. https://doi.org/10.1162/rest.88.3.482

Spoerri, A., Schmidlin, K., Richter, M., Egger, M., & Clough-Gorr, K. M. (2014). Individual and spousal education, mortality and life expectancy in Switzerland: A national cohort study. *Journal of Epidemiology & Community Health*, *68*(9), 804–810. https://doi.org/10.1136/JECH-2013-203714

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175–199. https://doi.org/10.1016/J.JECONOM.2020.09.006

Vaiserman, A. (2021). Season-of-birth phenomenon in health and longevity: Epidemiologic evidence and mechanistic considerations. *Journal of Developmental Origins of Health and Disease*, *12*(6), 849–858. https://doi.org/10.1017/S2040174420001221

van Kippersluis, H., O'Donnell, O., & van Doorslaer, E. (2011). Long-run returns to education: Does schooling lead to an extended old age? *Journal of Human Resources*, *46*(4), 695–721. https://doi.org/10.3368/JHR.46.4.695

Wulsin, L. R., Vaillant, G. E., & Wells, V. E. (1999). A systematic review of the mortality of depression. *Psychosomatic Medicine*, *61*(1), 6–17. https://doi.org/10.1097/00006842-199901000-00003

Zajacova, A. (2006). Education, gender, and mortality: Does schooling have the same effect on mortality for men and women in the US? *Social Science & Medicine*, *63*(8), 2176–2190. https://doi.org/10.1016/J.SOCSCIMED.2006.04.031

Zhang, Z., Liu, H., & won Choi, S. (2020). Early-life socioeconomic status, adolescent cognitive ability, and cognition in late midlife: Evidence from the Wisconsin Longitudinal Study. *Social Science & Medicine*, *244*, 112575. https://doi.org/10.1016/J.SOCSCIMED.2019.112575

How to cite this article: Fletcher, J., & Noghanibehambari, H. (2024). The effects of education on mortality: Evidence using college expansions. *Health Economics*, *33*(3), 541–575. https://doi.org/10.1002/hec.4787

APPENDIX A

In the main analysis of the text, we aggregate the number of colleges to those of own county colleges and within-state neighboring counties. This appendix shows the results for two alternative measures of colleges: number of own county colleges and number of own county colleges in addition to all neighboring county colleges regardless of being within the same state or not. The effects, reported in Appendix Table A1, are comparable to the main findings of Table 6.

TABLE A1 Alternative measures of college aggregation.

	Outcome: Death age (Months)/Explanatory variable					
	Only own county colleges			Own County + Neighboring county colleges (no within state restriction)		
	(1)	(2)	(3)	(4)	(5)	(6)
4-Year college	0.2917***	0.2537**	0.2287**	0.2114***	0.194***	0.1986***
	(0.1129)	(0.1097)	(0.1163)	(0.0446)	(0.0439)	(0.0494)
2-Year college	0.0074	−0.0085	−0.0192	−0.0321	−0.0401	−0.0395
	(0.1554)	(0.1512)	(0.1467)	(0.0649)	(0.0634)	(0.064)
Observations	3,973,914	3,973,914	3,973,887	3,967,966	3,967,966	3,967,939
R-squared	0.4438	0.444	0.444	0.4438	0.444	0.444
Mean DV	832.1	832.1	832.1	832.1	832.1	832.1
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual/Family controls	No	Yes	Yes	No	Yes	Yes
County controls	No	No	Yes	No	No	Yes

Note: Each column represents a separate regression. Standard errors, clustered at the county level, are in parentheses. Regressions include county and cohort fixed effects. Controls include individual, family, and county-by-cohort controls, as follows. Individual controls include dummies for female, white, black, Hispanic, first-generation immigrant, and second-generation immigrant. Family controls include dummies for maternal education and paternal socioeconomic index. County controls include share of whites, share of blacks, share of females, share of children less than 5, share of white-collar occupation employees, and average occupational income score.

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

APPENDIX B

This Appendix Table B1 replicates the main results of Table 6 for the DMF dataset which covers male's death between 1975 and 2005 that are linked to the 1940 census. While the marginal effects of the period that matches the death in Numident data are very close to the main results, the earlier death data reveals relatively smaller coefficients.

TABLE B1 Replicating the main results using censoc DMF data 1975–2005.

	DMF, males 1975–2005	DMF, males 1988–2005	Numident, males 1988–2005	Numident, females 1988–2005
	(1)	(2)	(3)	(4)
4-Year college	0.02318	0.13564**	0.18589***	0.13079*
	(0.14373)	(0.06359)	(0.06889)	(0.06919)
2-Year college	−0.12132	0.10066	−0.07402	0.01429
	(0.12498)	(0.06332)	(0.07109)	(0.06778)
Observations	2,081,307	1,591,474	2,275,577	1,692,362
R-squared	0.18784	0.42947	0.43573	0.44969
Mean DV	780.7	826.0	827.3	838.5
Elasticity of 4-year college	0.00037	0.00202	0.00265	0.00184
Elasticity of 2-year college	−0.00056	0.00044	−0.00031	5.8e-05
County FE	Yes	Yes	Yes	Yes
Birth cohort FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Note: Each column represents a separate regression. Standard errors, clustered at the county level, are in parentheses. Regressions include county and cohort fixed effects. Controls include individual, family, and county-by-cohort controls, as follows. Individual controls include dummies for female, white, black, Hispanic, first-generation immigrant, and second-generation immigrant. Family controls include dummies for maternal education and paternal socioeconomic index. County controls include share of whites, share of blacks, share of females, share of children less than 5, share of white-collar occupation employees, and average occupational income score.

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

APPENDIX C

Appendix Table C1 reports the results of two-sample two-stage least square regressions in which the college education (having at least 1 year of college education) is instrumented with the presence of 4-year and 2-year colleges at the local area of residence. The local area in this part is either a county or a PUMA whichever is larger. In areas where a county constitutes several PUMAs, we consider county as the local area. In areas where a PUMA represents several counties, we consider PUMA as the local area. In all other areas, PUMA (which represents a county) is the aggregated measure of local area.

TABLE C1 College education and age-at-death: Two-sample 2SLS results.

	Outcome: Age-at-death (Months)	
	(1)	(2)
Education: Some college or more	34.76539*** (11.7855)	33.77122*** (11.69747)
Observations	1,805,933	1,144,557
Mean DV	786.6	786.6
Birth year FE	Yes	Yes
PUMA-county FE	Yes	Yes
Individual controls	No	Yes

Note: Each column represents a separate regression. The excluded instruments are 4-year and 2-year colleges at the PUMA level. Individual controls include dummies for female, white, black, Hispanic, first-generation immigrant, and second-generation immigrant. Standard errors, clustered at the PUMA level, are in parentheses.

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

In the census 1960, we observe the education of individuals and in the census 1940 we observe individuals' county of residence which is used to capture the number of colleges when they turn 17. Since these two samples are not linked, we apply a two-sample 2SLS method in which the primary independent variable (educational attainment) comes from the 1960 census and the outcome (mortality) comes from the 1940 census that is linked to Numident. To eliminate cohorts who may have not yet completed education, we focus on individuals aged 22 and above. To mitigate the migration issues, we implement two additional restrictions. We keep individuals whose state of residence 5 years ago is different than their current state of residence. We restrict the sample to those who are less than 30 at 1960. The results are comparable without these exclusions.

APPENDIX D

This appendix shows the heterogeneity of the first stage results (Table 8) by race/gender/ethnicity. Each panel in Appendix Table D1 shows the results of first stage where our measures of college expansion is interacted with dummies for gender, race, and ethnicity.

TABLE D1 Heterogeneity in the first stage effects by demographic characteristics.

	Outcomes:			
	Education: At least one year of college	**Education: At least two years of college**	**Education: At least three years of college**	**Education: At least four years of college**
	(1)	**(2)**	**(3)**	**(4)**
Panel A				
4-Year college × Male	0.0018	0.0016	0.0018*	0.0017**
	(0.0012)	(0.001)	(0.001)	(0.0008)
2-Year college × Male	0.0008	0.0004	−0.0009	−0.0009
	(0.0015)	(0.0013)	(0.0012)	(0.001)
4-Year college	0.0098***	0.0081***	0.0061***	0.004**
	(0.003)	(0.0027)	(0.0022)	(0.0016)
2-Year college	−0.0021	−0.0034	−0.0045**	−0.0018
	(0.0032)	(0.0026)	(0.0022)	(0.0017)
Male	0.0533***	0.0516***	0.0428***	0.038***
	(0.004)	(0.0034)	(0.0031)	(0.0024)
Panel B				
4-Year college × other	0.0018	0.003	0.002	0.0038
	(0.0049)	(0.0049)	(0.0048)	(0.0045)
2-Year college × other	0.0106	0.0051	0.0038	−0.0014
	(0.0072)	(0.0072)	(0.0068)	(0.0065)
4-Year college	0.0106***	0.0088***	0.0069***	0.0047***
	(0.0031)	(0.0028)	(0.0023)	(0.0017)
2-Year college	−0.0019	−0.0034	−0.005**	−0.0023
	(0.0032)	(0.0026)	(0.0022)	(0.0017)
Other	−0.0758***	−0.0495***	−0.0454***	−0.0334***
	(0.0167)	(0.0154)	(0.0131)	(0.0123)

(Continues)

TABLE D1 (Continued)

	Outcomes:			
	Education: At least one year of college	**Education: At least two years of college**	**Education: At least three years of college**	**Education: At least four years of college**
	(1)	**(2)**	**(3)**	**(4)**
Panel C				
4-Year college × black	−0.0036	−0.0043**	−0.0042**	−0.004**
	(0.0023)	(0.0021)	(0.0019)	(0.0016)
2-Year college × black	0.0038	0.0039*	0.0035*	0.0036**
	(0.0026)	(0.0024)	(0.0019)	(0.0016)
4-Year college	0.0108***	0.0091***	0.0072***	0.005***
	(0.0032)	(0.0029)	(0.0024)	(0.0017)
2-Year college	−0.0022	−0.0037	−0.0054**	−0.0027
	(0.0032)	(0.0027)	(0.0023)	(0.0017)
Black	−0.1094***	−0.0841***	−0.0649***	−0.0521***
	(0.008)	(0.0067)	(0.0056)	(0.0047)
Observations	341,834	341,834	341,834	341,834
Mean DV	0.204	0.156	0.113	0.086

Note: Each column within each panel represents a separate regression. Individual controls include dummies for female, white, black, Hispanic, first-generation immigrant, and second-generation immigrant. Standard errors, clustered at the county-PUMA level, are in parentheses.

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

APPENDIX E

This appendix shows the heterogeneity of the results across census regions, birth cohorts, and based on county characteristics. We investigate the heterogeneity of the results across subsamples based on other observable dimensions. In so doing, we divide the sample by individuals' census region of residence in 1940 (columns 1–4), birth cohort (columns 5–6), father's education (columns 7–8), and county population (columns 9–10). The results are reported in Appendix Table E1.

Compared to the main results (column 3, Table 6), the effect of 4-year college is larger by a factor of 2 in the Midwest region. It drops by roughly 60% for the South and West regions and is zero for the Northeast region. The effects of 2-year college expansions are considerably larger for the Northeast, Midwest, and South regions. In southern states, an additional 2-year college is associated with 0.35 months higher age at death, almost 7.5 times larger than its aggregate effect in the main results. Both 4-year and 2-year college expansions are more effective for later cohorts than earlier cohorts. The effect of 2-year college expansions among cohorts born after 1932 is not only larger than the main results (by a factor of 6.5) but also statistically significant at 10% level. Not surprisingly, the college expansion is more effective for individuals whose father is more educated. Finally, low population counties benefit more from 4-year college expansions although the marginal effects are not precise and significant at 10% level (column 9 vs. column 10).

TABLE E1 Heterogeneity of the main results by subsamples.

	Outcome: Death age (months)/subsample									
	Region: Northeast	Region: Midwest	Region: South	Region: West	Birth year<1932	Birth year ≥1932	Mother education <12	Mother education ≥12	Below median county population	Above median county population
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
4-Year college	0.02076	0.03376	0.03376	0.06406	0.1073**	0.16685	0.1779***	0.13203*	0.08018*	0.21216*
	(0.0702)	(0.09813)	(0.09813)	(0.10424)	(0.05415)	(0.1358)	(0.05926)	(0.07429)	(0.04712)	(0.12475)
2-Year college	0.09378**	−0.04803	−0.04803	−0.04105	−0.02011	0.10957	−0.0096	−0.04688	−0.01848	0.06629
	(0.04717)	(0.16531)	(0.16531)	(0.07056)	(0.07139)	(0.15964)	(0.06546)	(0.0802)	(0.0457)	(0.12683)
Observations	982,780	1,340,677	1,340,677	385,440	2,735,560	1,232,378	2,962,089	1,005,848	1,965,383	2,002,556
R-squared	0.42726	0.44864	0.44864	0.45818	0.17083	0.21016	0.43803	0.46309	0.44097	0.44708
Mean DV	835.0	833.4	828.1	834.2	860.5	769.2	832.0	832.5	832.4	831.8
Birth cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Each column represents a separate regression. Standard errors, clustered at the county level, are in parentheses. Regressions include county and cohort fixed effects. Controls include individual, family, and county-by-cohort controls, as follows. Individual controls include dummies for female, white, black, Hispanic, first-generation immigrant, and second-generation immigrant. Family controls include dummies for maternal education and paternal socioeconomic index. County controls include share of whites, share of blacks, share of females, share of children less than 5, share of white-collar occupation employees, and average occupational income score.

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

APPENDIX F

The Bacon-decomposition (reported in Figure 5) suggest that the largest effects in producing the OLS estimations are driven by comparing treated and never treated counties (i.e., counties with an expansion to counties with no colleges at all). The event-study in the text (reported in Figure 3) shows no preexisting trend for the whole sample. In this appendix, we replicate the event-study for the sample that excludes counties that received the treatment before the start of the sample. Therefore, we focus on treated and never-treated counties. The results are reported in Appendix Figure F1. The estimated coefficients and the observed pattern for both 4-year and 2-year college openings are quite similar to the event-study of the whole sample.

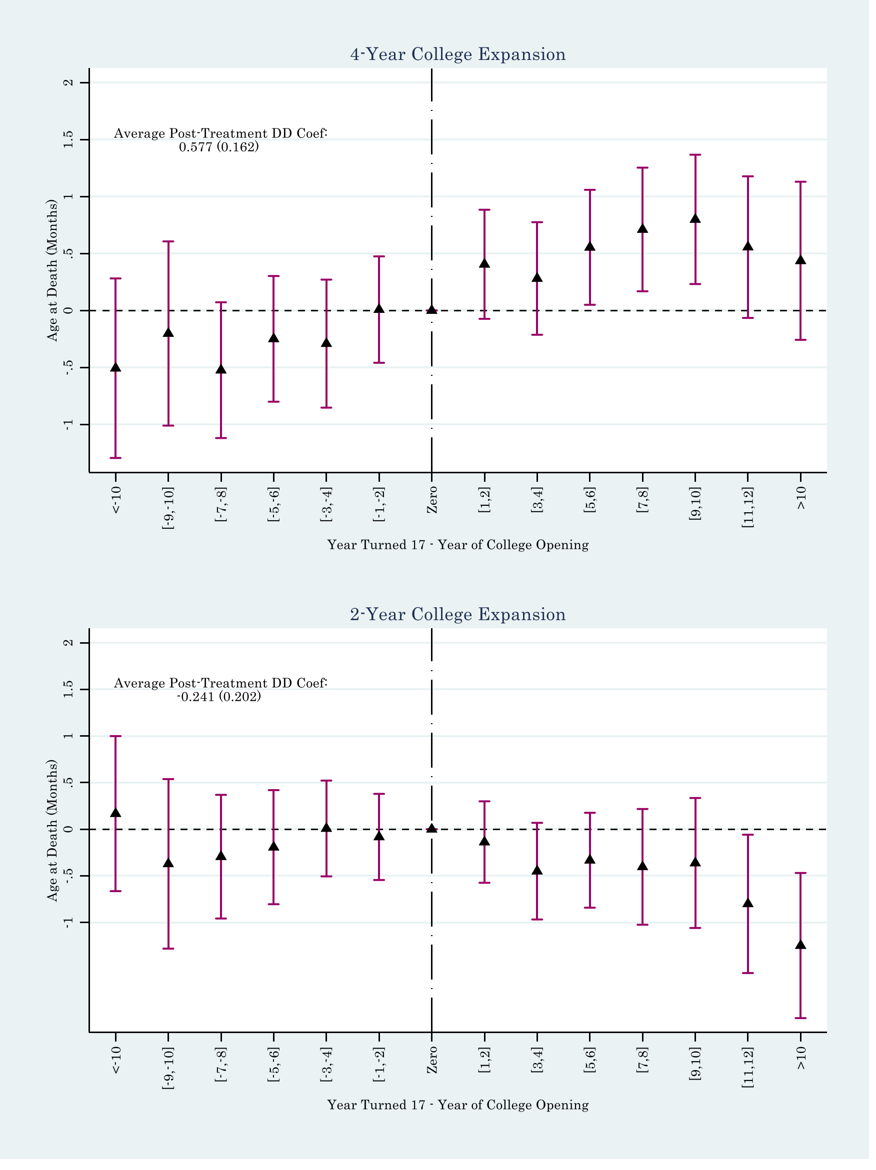


FIGURE F1 Event-Study Results of College Expansion on Age-at-death Using Treated and Never Treated Counties. Point estimates and 90 percent confidence intervals are reported. Standard errors are clustered at the county level. Regressions include county and cohort fixed effects. Controls include individual, family, and county-by-cohort controls, as follows. Individual controls include dummies for female, white, black, Hispanic, first-generation immigrant, and second-generation immigrant. Family controls include dummies for maternal education and paternal socioeconomic index. County controls include share of whites, share of blacks, share of females, share of children less than 5, share of white-collar occupation employees, and average occupational income score. [Colour figure can be viewed at wileyonlinelibrary.com]

APPENDIX G

In the main text, we provided suggestive evidence that education and income profiles improve as a result of new college openings and argue that these improvements are likely mechanism channels. Another possible change in the county characteristics that could partly explain the effects on longevity is compositional changes in occupations and industries. For instance, if a new college opening is accommodated by bringing new firms and establishments into the county and the new firms bring better-paid occupations, then the results could partly be driven by improvements in job quality. However, we do not have county-by-year level data on the share of different occupations over the sample period. One solution is to use decennial census data for the years 1940, 1950, and 1960. The drawback of the census is that for post-1950 years we do not have access to county identifiers. The variable PUMA which we exploited in section 5.3 is also only available for the 1960 census. Therefore, we are left with IPUMS de-identified counties that we do have access to for all the years 1940–1960. We should note that it limits our sample to about 500 counties, usually in metro areas with a large population. We assign the number of colleges to this data based on the year and de-identified county. We then restrict the sample to people aged 20–60 and those who report a specific occupation. We then build several dummy variables indicating the category of occupation. We regress these dummies on college measures, conditional on year fixed effects, birth-year fixed effects, and individual covariates. The results are reported in Appendix Table G1. We do not observe a significant change in the composition of occupations across various categories. However, we find an increase in the share of people employed as managers and officials following a 4-year college opening. The impact

TABLE G1 Changes in occupations following college openings.

	Outcomes: Occupation in the following categories								
	Professionals, instructors, doctors, engineers, other technical	**Farmers and farm managers**	**Managers, officials, and proprietors**	**Sales workers**	**Craftsmen and repairmen**	**Apprentice operatives, mine operatives, other operatives**	**Service workers (private and public)**	**Farm laborers**	**Other laborers**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
4-Year college	0.0039	−0.006*	0.0039	0.0031	0.0034	−0.0026	0.01	−0.0147	0.0008
	(0.0043)	(0.0036)	(0.0028)	(0.0059)	(0.0087)	(0.0066)	(0.0099)	(0.0135)	(0.0028)
2-Year college	0.0001	−0.0002	−0.0015	0.0051**	0.0021	−0.0057**	0.0014	−0.0061	0.0008
	(0.0019)	(0.0016)	(0.0014)	(0.0022)	(0.0033)	(0.0023)	(0.0057)	(0.0085)	(0.0011)
Observations	1,758,358	1,758,358	1,758,358	1,758,358	1,758,358	1,758,358	1,758,358	1,758,358	1,758,358
R-squared	0.031	0.0296	0.0436	0.1595	0.0636	0.1092	0.053	0.0293	0.0448
Mean DV	0.052	0.076	0.087	0.263	0.106	0.070	0.103	0.123	0.013
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year and birth-year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Each column represents a separate regression. Individual controls include dummies for female, white, and black. Standard errors, clustered at the county level, are in parentheses.

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

suggests about a 7% increase from the mean of the outcome. Overall, these results do not offer a discernible pattern of changes in occupation composition as the pathway.

To complement the analysis of this appendix, we also employ data from newly digitized CBP database extracted from Eckert et al. (2022). The CBP data reports employment counts at the county and aggregated industry level. Eckert et al. (2022) impute missing data due to suppressed data entry and provide a county-industry panel covering the years 1946–1974. To make a sample similar to our analysis, we focus on the 1946–1958 period. We calculate employment counts per capita for each county and each aggregated industry. We then merge this data with the college inventory database and implement regressions that include county fixed effects and year fixed effects. The results are report in Appendix Table G2. We do not find any consistent pattern across coefficients and outcomes. The estimated effects are very small and, in most cases, statistically insignificant.

TABLE G2 Changes in county-industry employment following college openings.

	Outcomes: Per capita employment in the following industries										
	All	**Unclassified**	**Agriculture, forestry, fishery**	**Mining**	**Contract construction**	**Manufacturing**	**Transportation and public utilities**	**Wholesale trade**	**Retail trade**	**Finance, insurance and real estate**	**Services**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**
4-Year college	−0.00886	−0.00089	0	0.00129	−0.00095	−0.0002	0.00175	0.00008	−0.0002	0.00051	0.00095
	(0.01265)	(0.00091)	(0.00014)	(0.00089)	(0.00123)	(0.00373)	(0.00176)	(0.00076)	(0.00396)	(0.00072)	(0.0018)
2-Year college	−0.00359	0.00057	−0.00005	−0.00061	0.00081	−0.00037	−0.0002	0.00075	0.0029	0.00054	0.00153
	(0.01148)	(0.0007)	(0.00014)	(0.00187)	(0.00129)	(0.00273)	(0.00142)	(0.00087)	(0.00244)	(0.00079)	(0.00198)
Observations	24,548	17,636	16,102	15,147	18,150	24,412	18,270	18,176	18,414	18,278	18,381
R-squared	0.80026	0.20567	0.37452	0.763	0.61646	0.90305	0.73559	0.732	0.83017	0.71952	0.78009
Mean DV	0.123	0.002	0.001	0.008	0.009	0.055	0.010	0.009	0.034	0.005	0.015
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year and birth-year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Each column represents a separate regression. Standard errors, clustered at the county level, are in parentheses.

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

APPENDIX H

In this appendix, we provide a direct test of migration as a function of new College openings. In so doing, we use the information reported in the 1940 census about place of residence in 1935. Based on this information, we create a variable that captures within state cross county migration from 1935 to 1940. We then use this variable as the outcome in our main regressions. The results are reported in column 1 of Appendix Table H1. We observe very small and statistically insignificant coefficients. Moreover, we define another variable that captures whether the state of residence in 1935 was different than the state of residence in 1940 implying cross state migration. In column 2, report the results for this outcome. We observe small and insignificant coefficients.

To further compliment than analysis of this appendix, we use cross census linking rules to link individuals from 1940 census to 1930 census in order to observe whether they have moved from 1930 or not. Based on their place of residence in 1930 versus 1940, we define a variable that indicates migrant. We use it as the outcome and report the regression results in column 3 of Appendix Table H1. We observe very small coefficients that are statistically significant. The overall picture of this table rules out the concern over endogenous migration.

TABLE H1 Exploring the influence of college opening on migration.

	Outcomes		
	Within-state cross-county migrant (since 1935)	**Cross-state migrant (since 1935)**	**Migrant since 1930**
	(1)	**(2)**	**(3)**
4-Year college	−0.0017	−0.0009	−0.0014
	(0.0019)	(0.0008)	(0.002)
2-Year college	0.0011	−0.0005	0.0014
	(0.0019)	(0.0004)	(0.0012)
Observations	3,967,939	3,967,939	753,991
R-squared	0.1142	0.0471	0.0804
Mean DV	0.442	0.035	0.272
County FE	Yes	Yes	Yes
Birth cohort FE	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes
Family controls	Yes	Yes	Yes
County controls	Yes	Yes	Yes

Note: Each column represents a separate regression. Standard errors, clustered at the county level, are in parentheses. Regressions include county and cohort fixed effects. Controls include individual, family, and county-by-cohort controls, as follows. Individual controls include dummies for female, white, black, Hispanic, first-generation immigrant, and second-generation immigrant. Family controls include dummies for maternal education and paternal socioeconomic index. County controls include share of whites, share of blacks, share of females, share of children less than 5, share of white-collar occupation employees, and average occupational income score.

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.