Discussion Papers
Department of Economics
University of Copenhagen

No. 15-01

Life Expectancy and Education: Evidence from the Cardiovascular Revolution

Casper Worm Hansen and Holger Strulik

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

Life Expectancy and Education: Evidence from the Cardiovascular Revolution*

Casper Worm Hansen† and Holger Strulik**

January 2015.

Abstract. In this study we investigate the causal impact of increasing adult longevity on higher education. We exploit the fourth stage of the epidemiological transition, i.e. the unexpected decline of deaths from heart attack and stroke in the 1970s as a large positive health shock that affected predominantly old age mortality. Using a differences-in-differences estimation strategy we find across U.S. states that the cardiovascular revolution led to an increase in adult life expectancy by about 2 years, which caused higher education enrollment to increase by 7 percentage points, i.e. 30 percent of the observed increase from 1970 to 2000. Our findings are robust to the inclusion of state-specific health trends and a host of confounding variables. They suggest large effects of improving longevity on higher education enrollment.

Keywords: adult life expectancy; higher education; cardiovascular diseases; 2SLS strategy; differences-in-differences first-stage.
JEL: I15; J24; N30; O10; O40.

* We would like to thank Uwe Sunde for discussion and helpful comments.
† Department of Economics, University of Copenhagen, Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen, Denmark; email: casper.worm.hansen@econ.ku.dk.
** University of Goettingen, Department of Economics, Platz der Goettinger Sieben 3, 37073 Goettingen, Germany; email: holger.strulik@wiwi.uni-goettingen.de

1. Introduction

One of the most controversial issues in the economics of long-run development is whether the increasing life expectancy, observable across the world during the 20th century, exerted a positive effect on the evolution of per capita income (e.g., Acemoglu and Johnson, 2007; Lorentzen et al., 2008; Cervellati and Sunde, 2011). The most obvious channel for a positive income effect is based on the higher incentive to invest in education induced by the expectation of a longer life. In theory, this mechanism was established more than 40 years ago (Ben-Porath, 1967). In practice, the relevance of the Ben-Porath mechanism has been recently challenged by Hazan's (2009) study, which then triggered a series of refinements of the original theory (Hansen and Lønstrup, 2012; Strulik and Werner, 2012; Cervellati and Sunde, 2013). In this paper we attempt to shed new light on these issues with an empirical strategy that captures the ideas of the original Ben-Porath model more closely than the existing literature by focussing on the impact of adult longevity on higher education.

The existence of a strong positive association between life expectancy and education is a well documented fact. According to one popular study, US Americans aged 25 in the year 2000 lived 7 years longer with any college education compared to those with only high school education or less (Meara et al., 2008). In many other countries a similar association between education and health has been observed. One of the major problems, however, hampering identification of the impact of longevity on education is reverse causality. Several mechanisms are conceivable generating a positive impact of education on longevity (e.g., Grossman, 2006; Strulik, 2013) and there exists a rich microeconometric literature suggesting that indeed a part of the observable positive correlation is explained by a causal impact of education on health and longevity (see Cutler et al., 2011 for a survey of the literature).

In order to uncover the causal effect of life expectancy on education and GDP per capita, Acemoglu and Johnson (2006; 2007) suggested to exploit the third stage of the epidemiological transition, i.e. the diffusion of antibiotics and new vaccines in the 1940, as an exogenous health shock. Their seminal approach has triggered a series of follow-up studies and extensions, mainly with the focus on GDP or GDP growth, which either challenged (Aghion et al., 2011; Cervellati and Sunde, 2011; Bloom et al., 2014) or confirmed (Acemoglu and Johnson, 2014; Hansen 2014; Hansen and Lønstrup, 2014) the original results. These studies vary in their specific estimation strategies but so far no attempts have been made in order to exploit a different health shock.

The third stage of the epidemiological transition, however, is a less convincing health shock for the identification of the Ben-Porath mechanism, i.e. the causal impact of adult longevity on education. The diffusion of antibiotics and other new drugs in the 1940 had a large impact on death from *communicable* diseases, in particular tuberculosis, pneumonia, and malaria. It thus affected predominantly infant and child mortality (WHO, 2008). Adult mortality, in contrast, is mainly determined by degenerative, *non-communicable* diseases caused by the aging body. In the U.S., for example, heart disease has been the leading cause of death since 1921 until the present day (DHHS, 1999a; American Heart Association, 2014). Worldwide, heart disease is the leading cause of death for adults aged 60 and above and it comes second (after HIV) for those aged 15-59 (WHO, 2008). Naturally, the health shock of the third stage of the epidemiological transition in the 1940s had no impact on deaths from cardiovascular disease, which were continuously on the rise from the beginning of the 20th century until about 1970 (American Heart Association, 2014, Chart 13-4).

The focus on communicable diseases and infant and child mortality, which follows inevitably from using the third stage of the epidemiological transition as a health shock, appears to be less appropriate to uncover the nexus between longevity and education for several reasons. Firstly, it is already conceivable from theory that lower child mortality causes human capital per capita to decline. This may be the case because of the scarring effect of diseases such that cured survivors perform worse at school than non-infected children (Aksan and Chakraborty, 2014, see e.g. Snow et al., 2003, and Bleakley, 2007 for evidence). Secondly, education of children is either compulsory or decided upon by their parents. According to standard economic theory parents would decide jointly on the number and education of their children. In this respect the conventional overlapping generation model predicts that improving child mortality should leave net fertility and thus education unchanged (Galor, 2005, 2011) while the Becker-Barro model as well as extensions of the OLG model predict increasing net fertility (Doepke, 2005; Strulik and Weisdorf, 2014), a response which is confirmed by time series analysis Eckstein et al., 1999; Herzer et al., 2012). Increasing net fertility, in turn, would delay the fertility transition and the take-off of education and economic growth. These considerations are consistent with results from Cervellati and Sunde (2011) showing that increasing life expectancy had a positive effect on income per capita only in those countries, in which net fertility declines as a response to declining child mortality, i.e. in countries where the demographic transition had advanced far

already in 1940. One interpretation of their result could be, in light of the present paper, that life expectancy at birth in these countries was strongly correlated with adult life expectancy because child mortality was already very low.

A paper which, like us, focusses on adult mortality is provided by Lorentzen et al. (2008). This study, however, refrains from exploiting a health shock for identification and uses instead malaria ecology, climatic variables, and geographic features as instruments. Yet again, these instruments capture predominantly exogenous influences on deaths from communicable diseases and are presumably less suitable to capture the influence of natural aging and non-communicable diseases. Lorentzen et al. (2008) find that adult mortality, instrumented in this way, exerts a significant effect of on growth, investment, and fertility but, like Acemoglu and Johnson (2006), they fail to establish a significant effect on average years of schooling in the population. A recent study by Hansen (2013a), in contrast, demonstrates a positive effect of increasing life expectancy on years of schooling using a cohort based measure and the identification strategy of Acemoglu and Johnson (2006, 2007). Inspecting different potential drivers of these results, the study concludes that declines in pneumonia mortality seem to be mainly responsible. Interestingly, lower respiratory infections (including pneumonia) come fourth, after the cardiovascular diseases, as the leading cause of death for adults aged 60+ (WHO, 2008).

The question, however, remains what can be learned from the epidemiological transition of the 1940s or, more generally, from the historical decline of communicable diseases for the future of longevity and higher education in developed countries. In developed countries, primary and secondary education is largely compulsory and tertiary education, i.e. the decision to go to college or university is presumably made by adults on their own behalf, in line with the assumptions of the original Ben-Porath model and its implementation in theories of long-run growth (Boucekinne et al., 2002; Cervellati and Sunde, 2005, 2014). These adults experience little threat of dying prematurely from communicable diseases, which are affecting predominantly the mortality of children in developing countries. Instead they face the risk of dying in old age by non-communicable diseases, most importantly by heart attack and stroke. Human capital from tertiary education, moreover, is presumably the most important input in R&D and thus the ultimate driver of productivity growth (Strulik, 2005; Strulik et al., 2013).

In order to uncover the causal effect of adult life expectancy on higher education we develop a new instrument, namely the strong, permanent, and unexpected decline of death from heart

disease and stroke that began in the 1970, and has been dubbed the cardiovascular revolution (Foege, 1987; Vallin and Mesle, 2009; Thelle, 2011). We focus on education enrollment of 18 to 24 old US-Americans and on life expectancy at 30 (or, alternatively, at 50). This means that there is already by design relatively little concern for reverse causality, i.e. that the education decision of 18-24 year old may affect life expectancy of the contemporaneously 30 or 50 year old. Measurement error and time varying omitted variables, however, are still an issue, which we attempt to take care of with our instrumental variable approach, which follows the same principal logic as the empirical strategies employed in, e.g., Bleakley (2007), Acemoglu and Johnson (2007), and Hansen (2014).

The paper is organized as follows. The next section introduces the cardiovascular revolution. Section 3 and 4 describe the data and estimation strategy. Section 5 establishes the impact of the cardiovascular revolution on adult life expectancy and Section 6 presents our main results on the impact of adult longevity on higher education enrollment. Section 7 concludes.

2. The Cardiovascular Revolution

It may be illuminating to begin with some insights from gerontology in order to contrast our approach with the existing health-shock literature. In gerontology there exists a very strong association between age and the force of mortality, i.e. the conditional probability to die in the next period (e.g., Arking, 2006). This association, known as Gompertz-Makeham law, is formally given by $\mu = A + R\mathrm{e}^{\alpha x}$, in which x is age and μ is the force force of mortality. The parameter A captures "background mortality", i.e. age-unrelated causes of death, whereas the parameters R and μ capture the impact of aging on mortality. In human history, background mortality has come down big time, particularly during the epidemiological transition of the 1940s, and it is now close zero in developed countries (Gavrilov and Gavrilova, 1991). In developed countries most people die from degenerative, aging related diseases whose impact is covered by initial frailty R and the speed of increasing mortality with age α.

It is easy to see that any positive trends of life expectancy would soon come to an end if the trend were solely determined by falling background mortality, including the prevention and treatment of communicable diseases. Life expectancy would converge towards a constant life span determined by the aging parameters R and α. The observation that life expectancy does not (yet) seem to converge, i.e. that we observe "broken limits to life expectancy" (Oeppen

and Vaupel, 2002) indicates that we, unlike all other animals, managed to manipulate the aging process itself. Inspecting the aging parameters we find indeed evidence that since the 1970s not only life expectancy (a population-specific characteristic) but also human life span (a species-characteristic) is on the rise (Strulik and Vollmer, 2013). The most likely driver of this "manufactured life-time" (Carnes and Olshansky, 2007) is medical technological progress that interfered with the natural way of human aging. The observation that human life span seems to be manipulated for the first time in the 1970 inspired our research on the cardiovascular revolution.

"Man is as old as his arteries". This old aphorism, cited according to Dantas et al. (2012), describes succinctly that cardiovascular diseases are degenerative and inevitably related to the "natural" process of human aging.[1] Our arteries are simply not built to last. They dilate and stiffen through natural physical stress and fatigue (Ungvari et al., 2010). The most common type of heart disease and cause of heart attacks is coronary heart disease also known ischemic heart disease. It is caused by plaque building up along the inner walls of the arteries of the heart, which narrows the arteries and reduces the blood flow to the heart.

Given that background mortality is continuously declining such that humans die increasingly of degenerative, aging-related diseases, we would expect that the incidence of death from cardiovascular diseases increases over time. Indeed, extrapolating historical trends for US-Americans, the National Heart, Lung, and Blood Institute expected about 1.6 million deaths of coronary heart disease (i.e. an age-adjusted death rate of 0.55 percent) at the end of the 20th century while actually it were "only" 0.4 million (a death rate of 0.15 percent), see NHLBI (2012, chart 3-24). The beforehand increasing trend was reverted around the year 1970. Figure 1 shows a replication of this finding with the data used in our empirical analysis. The death rate is measured in per 100,000 and displayed in logs. For both men and women the positive trend was reverted in about 1970. A similar though less impressive break in trend is observed for deaths from stroke (see Figure A.1 in the Appendix).

[Figure 1 about here]

[1] In biology aging is understood as the "intrinsic, cumulative, progressive, and deleterious loss of function that eventually culminates in death" (Arking, 2006). For an introduction to the evolutionary foundations of human aging see Kirkwood (1999). For a detailed formal description of the aging process by reliability theory see Gavrilov and Gavrilova (1991).

The trend reversal was preceded by a series of medical innovations, notably the pacemaker (in 1958), the cardiopulmonary resuscitation through external chest compressions (1960), the beta blocker (1962), the artificial heart (1963), the portable defibrillator (1965) and the heart transplant (1967). While there was comparatively little medical breakthroughs concerning cardiovascular diseases before the 1960s, medical progress continued afterwards, notably with the introduction of cholesterol lowering drugs (1987) and the intravascular stent (1988). The decline was also promoted by changing behavior, notably the decline in smoking since the mid 1960s (DHHS, 1999b). These changes in smoking behavior were presumably triggered by the first publications establishing a link between smoking and cancer appearing in the 1950s and United States Surgeon General's Report on Smoking and Health from 1964, which led to the banning of certain advertising, warning labels on tobacco products, and large scale anti-smoking campaigns USPHS, 1964). Similar to the 1940s, the health shock starting in the 1960s could thus be conceptualized as a compound of medical and institutional innovations, improving prevention and treatment. The main difference is that the cardiovascular revolution mainly affected health of the elderly, which makes it an appropriate instrument to investigate the impact of adult longevity on education.

Another perspective to compare our contribution with is the literature on the "epidemiological transition", originally proposed by Omran (1971) as a three-stage theory of historical health transitions, from the age of pestilence to the age of receding pandemics to the age degenerative diseases. The unexpected decline of death from cardiovascular diseases after the 1970s inspired sociologists and gerontologists to introduce a fourth stage, the "age of delayed degenerative diseases" (Olshansky and Ault, 1986), in which declining death rates are predominantly concentrated at advanced ages. Omran (1998) acknowledged this phenomenon and included the fourth stage in his revised theory of health transitions. The related literature on the longevity-education nexus thus focusses on the historical third stage of epidemiological transition while our approach focusses on the still ongoing fourth stage. Taking a within country perspective we expect that those US American states are benefitting the most from the cardiovascular revolution in which the pre-shock mortality from these diseases were highest.

3. Data and Pretreatment Differences

3.1. **Data and Descriptive Statistics.** This section describes the construction of the dataset and reports descriptive statistics. In the main analysis, the units of observation are the 50 US states observed in the year 1960 (pretreatment) and in the year 2000 (posttreatment). However, we also estimate specifications using decadal observations between 1960 and 2000, and some robustness checks require that the period of observation starts in 1940, which restricts the sample to the 48 contiguous states.

The higher education enrollment rate, which is the main outcome of interest, is obtained from Census data (IPUMS; Ruggles et al., 2008). To construct aggregate representative statistics for the number of people from the white population (in the age group 18–24) attending school that leads to a college/graduate degree, we apply the personal weight (PERWT) —mapping a number of individuals in the US population to a representative individual in an IPUMS sample—to the variable "grade level attending" (GRADEATT). The higher education (net) enrollment rate is then obtained by dividing this measure with the number of white people in the age group 18–24, which also is obtained from the IPUMS. As seen from Table 1, we find that the state mean higher education enrollment rate increases from 0.15 in 1960 to 0.36 in 2000. Because GRADEATT is available only in the period 1960–1980, and 2000 in the Censuses, we construct—in the same way—the schooling enrollment rate age 18–24 using the variable SCHOOL as it is available all the years between 1940 and 2000. Not surprisingly, the state mean schooling enrollment rate is higher than the corresponding higher education enrollment rate, however, it follows the same trend, and the correlation between the two measures of enrollment is 0.99.

The main explanatory variable is life expectancy for the white population at various ages, obtained on a decennial basis (1939–1941, 1949–1951, ..., 1999–2001) from the US Decennial Life Tables (NCHS, 2014). Due to data restrictions in the early years, we use the unweighted average of male and female life expectancy. Table 1 reveals that the state mean life expectancy at age 30 has increased 5.25 years over the observation period, while the corresponding increase in life expectancy at age 50 is 4.82 years.

We capture the intensity of the health shock by the number of deaths due to major cardiovascular-renal diseases per 100 white people (age adjusted), measured prior to the breakthroughs in the understanding of cardiovascular diseases (i.e., in 1960). This variable is gathered from Grove and Hetzel (1968) and is referred to as *CVD*. Figure 2 shows that Northeastern states generally

have higher levels *CVD*.

Motivated by the evidence described in Section 2, we code the (medical) interventions, which are argued to have decreased mortality from cardiovascular diseases, as occurring after the 1960s. This means that the time indicator (*Post*) equals one from 1970 onwards. We use the interaction between these two variables (i.e., $CVD \cdot Post$) to estimate how the decline of cardiovascular-disease mortality influenced the subsequent development of life expectancy and education. This health-shock variable constitutes the IV in our 2SLS strategy, and we refer to it to as the CVD shock.

To control for the possible influence of non-cardiovascular mortality on life expectancy, we also construct the variable *Initial Mortality*, which is the age-adjusted mortality rate from the causes of death that are *not* attributable to cardiovascular diseases in 1960. This variable is also obtained from Grove and Hetzel (1968) and interacted with the time indicator.

Finally, we consider the following variables (measured in 1960): average years of schooling in the workforce, log GDP per capita, and log capital per worker. These variables are taken from Turner et al. (2007), and, to account for this variation in our regressions, they are also interacted with the time indicator. In addition, we report results from specifications that control directly for log GDP per capita.

[Table 1 about here]

[Figure 2 about here]

3.2. **Pretreatment Differences in the Outcomes.** Because our empirical strategy exploits a differences-in-differences (DD) application to estimate the effect of the cardiovascular revolution on life expectancy and education (see also the next section), the identifying assumption is identical pretreatment trends in the outcomes between treatment and control states. Due to the continuous nature of the treatment, this is the same as saying that *CVD* should be uncorrelated with changes in the outcomes before treatment. Therefore, as an initial step of testing this assumption, we now explore pretreatment differences in the level and the change of the outcomes for states with different levels of *CVD*. In the level specification, we regress the outcome j in state s (Y_s^j) on the cardiovascular mortality rate in 1960 (CVD_s):

$$Y^{j}_{s1960} = \alpha + \beta_j \ CVD_s + \mathbf{X}'_s\gamma + \varepsilon_s, \tag{1}$$

where $\mathbf{X}'_s$ is vector of controls and ε_s is the error term. The estimated β_j reflects pretreatment (conditional) differences in the outcome j for states with different mortality rates from cardiovascular diseases in 1960. Before turning to our change specification, it is important to note that $\hat{\beta}_j \neq 0$ does necessarily imply that the identifying assumption is violated as the DD model allows us to non-parametrically control for state fixed effects.

In the change specification, we regress the pretreatment changes in the outcome between 1940 and 1960 (ΔY^j_s) on CVD_s and $\mathbf{X}'_s$:

$$\Delta Y^{j}_{s} = \bar{\alpha} + \beta^{\Delta}_{j} \ CVD_s + \mathbf{X}'_s\gamma + \varepsilon_s. \tag{2}$$

The estimate of β^{Δ}_j provides the same information as a falsification exercise testing if the CVD-shock variable ($CVD \cdot Post$) in the DD model has any predictive power for the development of the outcomes in the pretreatment period (i.e., 1940–1960). If $\hat{\beta}^{\Delta}_j = 0$ cannot be rejected, the falsification test would support the identifying assumption of common pretrends.

Table 2 reports weighted least squares estimates of β_j and β^{Δ}_j in equations (1) and (2), respectively. Columns 2 and 4 report the unconditional estimates, whereas columns 3 and 5 report these estimates conditional on two or three state-level controls.[2] Unsurprisingly, the relationship between life expectancy at age 30 (or 50) in 1960 and CVD is negative and statistically significant at the 1 percent level (columns 2 and 3). More importantly, however, we also find that CVD is completely unrelated to changes in life expectancy between 1940 and 1960, which imply that the CVD shock has no any predictive power for life expectancy in the pretreatment period (columns 4 and 5). Figure 3 shows graphically that this lack of relationship is *not* driven by some outlier states.

Furthermore, the estimates reported in Table 2 reveal that states with different rates of cardiovascular-disease mortality were similar in terms of higher education enrollment and schooling enrollment at age 18–24 in 1960, and that these states experienced similar changes between 1940 and 1960 in the schooling enrollment age 18–24. This suggests that the CVD-shock variable

[2]For the life expectancy outcomes, the covariates are: Initial Mortality, Initial Enrollment, and Initial Income, whereas for the schooling outcomes, the covariates are: Initial Mortality and Initial Income.

is not likely to capture preexisting different trends in the education outcomes.

[Table 2 and Figure 3 about here]

4. Empirical Framework

This section explains our main estimation strategy. We use an instrumental-variable approach to identify the effect of life expectancy on education. The second-stage relationship in that framework is given by the following estimation equation:

$$Y_{st} = \theta \ \textit{Life Expectancy}_{st} + (\mathbf{X}'_s \cdot \textit{Post}_t)\eta + \lambda_s + \tau_t + \epsilon_{st}, \tag{3}$$

where Y_{st} is the higher education enrollment rate (or the school enrollment rate age 18-24) for the white population in state s at time t. In our baseline specification, we consider the model using the two years: $t_1 = 1960$ and $t_2 = 2000$. The main variable of interest is life expectancy at age 30 (or 50) for the white population ($\textit{Life Expectancy}_{st}$). We also include a vector of controls ($\mathbf{X}'_s$) containing the age-adjusted mortality rate from non-CVD causes, the initial enrollment rate, initial log GDP per capita. These cross-sectional variations are measured in 1960 and interacted with the indicator ($\textit{Post}_t$), which equals one in the years after 1960. The panel structure of our dataset allows us to non-parametrically control for state (λ_s) and time (τ_t) fixed effects. We cluster the error term (ϵ_{st}) at the state level, ensuring that the standard errors of our estimates are robust to arbitrary serial correlation in the states (e.g., Cameron and Miller, 2013).

The main first-stage specification is:

$$\textit{Life Expectancy}_{st} = \bar{\pi}(\textit{CVD}_s \cdot \textit{Post}_t) \ + (\mathbf{X}'_s \cdot \textit{Post}_t)\bar{\eta} + \bar{\lambda}_s + \bar{\tau}_t + \bar{\epsilon}_{st}, \tag{4}$$

where $\textit{CVD}_s$ is the age-adjusted death rate from major cardiovascular-renal diseases for the white population in 1960. The remaining variables are as defined above. The baseline specification estimates equation (4) using the years 1960 and 2000. In order to test the identifying assumptions of the DD model, we also report results from 10-year panel models over the period 1940–2000.

Because the uptake of the new cardiovascular-disease fighting technologies is endogenous, we exploit the fact that some states stood to benefit more from the shock as they had higher levels of cardiovascular-disease mortality prior to the breakthrough. Following this approach, our estimate of $\bar{\pi}$ should not be biased by the different rates of actual adoption of the new

technologies. In other words, we are measuring the *intention-to-treat* effect with the estimation of $\bar{\pi}$, which is smaller than the average causal effect of on those who were actually treated (Angrist and Pischke, 2008). This means that $\bar{\pi}$ provides a conservative estimate of how the cardiovascular revolution influenced the development of life expectancy.

Our instrument for life expectancy ($CVD_s \cdot Post_t$) computes the DD estimate of the shock to cardiovascular-disease mortality in the 1970s since we are comparing the relative change in life expectancy in the pretreatment period relative to the posttreatment period between states with different treatment intensities (i.e., CVD_s). As also argued in previous section, the main issue in the DD model is to check for pretrends in the outcomes that are correlated with the intensity of treatment. While the results in Tables 2 indicate that this is indeed *not* the case, we also estimate a fully flexible model that takes the following form:

$$Life\ Expectancy_{st} = \sum_{k=1950}^{2000} \pi_k (CVD_s \cdot Post_t^k) \ + (\mathbf{X}_s' \cdot Post_t)\bar{\eta} + \bar{\lambda}_s + \bar{\tau}_t + \bar{\epsilon}_{st}, \tag{5}$$

where the only difference to equation (5) is that CVD_s is interacted with time-period fixed effects for the period 1950–2000, implying that this model is estimated using a 10-year panel between 1940 and 2000, where 1940 is the omitted year of comparison. A test of the DD (pretrends) assumption is $\pi_k = 0 \ \forall k < 1970$, whereas as the effect after treatment could increase, decrease or stay constant. This means, in contrast to the pretreatment $\pi_j's$, that the $\hat{\pi}_j's \ \forall k \geq 1970$ are allowed not be identical.

5. Cardiovascular Diseases and Life expectancy

As a first step in our empirical analysis, this section presents the result from estimating the first-stage relationship. Throughout the analysis, the regressions include state and year fixed effects and are weighted by the white population in 1960, implying that the estimates reflect the average effect for a white individual in the US.[3]

This section is structured to gradually build up from the flexible-estimation approach, which is used to validate our identifying assumption of common pretrends, to the main DD estimate of the CVD shock (i.e., the first stage) used to the identify the effect of life expectancy on enrollment rate in Section 6.

[3]In general, similar conclusions are obtained from using the white population age 18–24 as an alternative weight of from the unweighted results.

5.1. **Flexible Estimates.** Table 3 reports the results from estimating the flexible model given in equation (5). Columns (1)–(3) report estimates for life expectancy at age 30, and columns (4)–(6) report estimates for life expectancy at age 50.

In column 1, we see that $\hat{\pi}_{1950} = -0.29$ (s.e. = 1.08) and $\hat{\pi}_{1960} = 0.04$ (s.e. =1.91), indicating that before treatment life expectancy was not trending any differently between treatment and control states (compared to 1940). By 1970, in contrast, we find a positive and statistical significant relationship between *CVD* and life expectancy. This effect accumulates over time: the estimated coefficient $\hat{\pi}_{1970}$ is equal to 5.51 (s.e. = 2.77), whereas the estimated coefficient $\hat{\pi}_{2000}$ is equal to 14.59 (s.e. = 5.74).

Columns 2 and 3 add controls for the initial mortality environment (measured by the death rate from non-cardiovascular diseases), initial log GDP per capita, and the initial higher education enrollment rate. In order to estimate the contribution of these cross-sectional measures, they are interacted with the time indicator as all the specifications include state-fixed effects. We find that the inclusion of these controls reinforces the observed pattern of the estimated coefficients in column 1. Interestingly, the effect of the initial mortality environment is positive but highly insignificant, indicating that the decline in cardiovascular-disease mortality was the main contributor to the increase in life expectancy during this period. It should also be noted that the same pattern, though slightly more *precisely* estimated, is obtained by controlling flexibly for the cross-sectional measures, that is, interacting them with time-period fixed effects instead of the indicator.

In an unreported specification, we also included state-specific linear time trends to column 3 allowing for differential trends across states. This exercise provides the following pattern in the estimated coefficients: $\hat{\pi}_{1960} = 0.82$ (s.e. = 1.55); $\hat{\pi}_{1970} = 6.51$ (s.e. = 2.74); $\hat{\pi}_{1980} = 9.18$ (s.e. = 3.96); $\hat{\pi}_{1990} = 13.01$ (s.e.= 4.93); $\hat{\pi}_{2000} = 16.27$ (s.e. = 6.56).[4] Thus, our basic conclusion, derived from the evidence reported in Table 3, survives in this much more demanding specification.

Columns 4–6 report the evidence for life expectancy at age 50. We arrive at the same conclusion that there are no trends of the estimated interaction effects prior to the cardiovascular revolution, whereas by 1970 "more treated" states begin to experience greater increases in life expectancy. Overall, the identifying assumption in the DD framework of common pretrends is

[4]Notice, adding (linear) state trends, we lose one observation period (i.e., 1940), so the estimates should be compared to 1950.

strongly supported by the evidence reported in Table 3.

[Table 3 about here]

Before turning to our main first-stage relationship, Table 4 reports the results from estimating equation (5) using the 10-year panel dataset from 1940–2000. Thus, this model provides a DD estimate of the CVD shock on life expectancy using multiple periods.[5] We see that the DD model confirms our conclusion so far. For example, column 3, which includes the baseline control interactions, shows an estimated coefficient on the CVD shock of 9.46 (s.e. = 2.81). This suggest that for states with a higher mortality rate from cardiovascular diseases in 1960, life expectancy at age 30 was growing significantly faster in the period 1970–2000. The coefficient estimate implies that increasing treatment intensity (i.e., the cardiovascular-disease mortality rate in 1960) from the lowest to the highest level is associated with an increase in life expectancy of 1.80 years. In terms of the average decrease in cardiovascular-disease mortality between 1970 and 2000,[6] this estimate suggests that the cardiovascular revolution contributed to an increase in life expectancy with about 1.9 years.

[Table 4 about here]

5.2. **The First-Stage Relationship.** Table 5 documents our main results for the first-stage relationship, which is a DD estimate of the CVD shock on life expectancy. The estimation equation is (4) including one year prior to treatment (1960) and one year after treatment (2000). These two-dates specifications are referred to as a long panel (or DD model). The table structure follows that of the previous regression tables, so that columns 1–3 (4–6) report the results for life expectancy at age 30 (50).

Consistent with the argument that the advancement of cardiovascular-disease technologies in the 1970s contributed significantly to the subsequent development of life expectancy at older ages, the estimates reveal that states with a higher treatment intensity experienced larger increases in life expectancy between 1960 and 2000. In comparison to the point estimates from the 10-year panel model (reported in Table 4), the effects are generally larger in magnitude.

[5]The only difference to the flexible model being that we now estimate the interaction between *CVD* and the indicator variable rather than the time-period fixed effects

[6]Data from the National Heart, Lung, and Blood Institute indicate that the cardiovascular disease mortality rate decreased by about 50 percent from 1970–2000.

For example, columns 3 indicates that a one-standard-deviation increase in *CVD* leads to an increase in life expectancy at age 30 of 0.50 year, whereas the 10-year panel model suggests an increase of about 0.36 year.[7] Previous studies using the same empirical framework also find larger effects in the long panel specifications; see, e.g., Acemoglu and Johnson (2006; 2007) and Hansen (2013; 2014).

Comparing the estimates in columns 1–3, the effect of the shock remains stable in magnitude and statistical significance by adding controls for initial variation in mortality, income, and the higher education enrollment rate. The relationship reported in column 3 is depicted as a partial correlation plot in Figure 4. Panel A indicates that Hawaii could be an influential observation that drives our finding, however, because the regressions are weighted by population size, our baseline estimate is almost unaffected by the exclusion of Hawaii from the sample as shown in Panel B of Figure 4.

Before reporting the evidence from the robustness analysis, it is noteworthy that similar results are obtained for life expectancy age 50 (columns 4–6) and the estimated coefficient on the interaction *Initial Mortality* $\cdot$*Post* is negative but statistically highly insignificant in all the reported specifications.

[Table 5 about here]

[Figure 4 about here]

Table 6 subjects the specifications reported so far to state-specific linear time trends. This test requires at least three data points per state, so the long-panel specification is augmented with one additional pretreatment year (i.e., 1940).[8] This influences our estimate of π only marginally, demonstrating that our baseline estimate of the shock is not likely to capture differential pretrends in life expectancy related to, for example, the decline in infectious diseases from the 1920s to the 1950s.[9]

[7] Notice that these differences are not statistically significant.

[8] This implies that only the 48 contiguous states are included in this particular robustness check.

[9] As Hansen (2013b) demonstrates that states with higher levels of infectious disease mortality in 1940 experienced greater increase in human capital between 1950 and 2000, the CVD-shock variable could be picking this pretrend up. Moreover, in an unreported specification, we find that infectious disease mortality in 1940 does not predict the development of life expectancy at age 30 or 50 but only life expectancy at birth. This finding is consistent with infectious-disease mortality was weighted towards child mortality.

[Table 6 about here]

Table 7 considers various extensions to the baseline specification for life expectancy at age 30.[10] The first four columns add potential confounders. Firstly, columns 1–3 demonstrate that our conclusion is robust to initial variation in average years of schooling in the workforce and physical capital per worker (both interacted with the indicator). For example, subject to both controls in column 3, the estimated coefficient increases in magnitude to 16.28 (s.e. = 2.88). Secondly, while log income per capita is likely to be a bad control variable, columns 4 reveals that the effect of the shock to life expectancy remains reassuringly stable in both magnitude and statistical significance controlling for log income per capita (i.e., *Income*).

The remaining columns report the results from various sample splits. Columns 5–7 make the split by geographical area. This evidence demonstrates that our first-stage relationship seems mainly to be driven by Northern and Southern states as we find positive and statistical significant relationships in these sub samples, whereas the coefficient estimate for the Western states is 7.43 (s.e. = 8.48). Nevertheless, as an alternative check along the same lines, estimating a model for the full sample of US states, which non-parametrically takes into account the variation occurring between these three regions over time by including region-by-year fixed effects, we find that the estimate of the first-stage continues to be positive and statistically significant at the 5 percent level ($\hat{\pi} = 11.90$; s.e.= 5.00). This finding documents that the effect of the shock on life expectancy is not only driven by variation between states in the north and west as the map in Figure 2 of the geographical distribution of *CVD* could indicate.

The final two columns of the table report the results from splitting the sample into rich and poor states as measured by median GDP per capita in 1960. The effect is larger for initially poor states, however, the difference in point estimates is not statistically significant. Nonetheless, for a one-standard-deviation increase in *CVD* among the poor states, life expectancy increases by 0.54 years, while the point estimate for the rich states suggests an increase of 0.43 years.

[Table 7 about here]

[10]The robustness results for life expectancy at age 50 are not reported in order to save space. However, they are available upon request.

6. Life expectancy and Higher Education

6.1. **Main results.** We turn now to our main empirical results in which we attempt to establish a causal relationship between the higher education enrollment rate and life expectancy. Table 8 presents the baseline results for the effects of life expectancy on education enrollment. The estimation equation is (3) using the first-stage relationship in (4). Panel A reports the least-square (LS) estimates, while panel B reports two-stage least square (2SLS) estimates.

We find positive and statistical significant LS estimates. Taken at face value, the baseline LS estimate of θ, reported in column 3, suggests that for one additional year of life expectancy, education enrollment increases by about 2 percentage points. As seen from columns 1–6, this estimate is relatively stable across specifications. However, due the usual endogeneity concerns, such as time varying omitted variable, reverse causality, and measurement error, the LS estimates cannot be interpreted causally. For these reasons, we next exploit the cardiovascular revolution as an exogenous source of within-state variation in life expectancy as documented in the previous section.

The 2SLS estimates, reported in panel B, show that the effect of life expectancy on the higher education enrollment rate is positive and highly significant. The 2SLS estimates are also larger in magnitude compared to the corresponding LS estimates. Continuing our calculation from the previous section, where a 1.9 years increase in life expectancy was ascribed to the cardiovascular revolution, and using the 2SLS estimate in column 3, this translates into a 7 percentage points increase in the enrollment rate, which is about 30 percent of the observed increase between 1960 and 2000.

The instrument quality in terms of the Kleibergen-Paap F-statistic is reported in bottom of the table. Because the F-statistics are above the rule-of-thumb level of 10 in all the specifications, concerns about weak-instrument bias are moderate.

The reduced-form relationship, matching the specification in column 3, is graphed as a partial correlation plot in Figure 5. Panel A shows the relationship for the full sample of 50 states, whereas panel B shows it for the 48 contiguous states. In line with the two-stage evidence, the estimated coefficient for the reduced form is positive and significant. According to these estimates, a one-standard-deviation increase in CVD is associated with 2 percentage points increase in the enrollment rate.

[Table 8 about here]

[Figures 5 about here]

Table 9 presents the same type of evidence for the schooling enrollment rate age 18–24 as the outcome variable. As also mentioned in the data section, this extension is necessary in checking for differential pretrends in enrollment as the US Census does not provide any information on the degree attending before 1960. Nonetheless, the schooling enrollment rate age 18–24 appears to be a reasonable proxy for the higher education enrollment rate. The correlation is 0.99 in the years for which data on both variables are available (i.e., in our case 1960 and 2000). The LS and 2SLS estimates, reported in Table 9, are similar in magnitude and significance to the estimates with the higher education enrollment rate as outcome, and Figure 6 also graphs similar reduced-form relations. Next, we augment these long-panel specifications for schooling enrollment with the pretreatment year 1940 and subject them to state-specific trends. The 2SLS estimates in Table 10 show little change to this robustness test.

[Tables 9 and 10 about here]

[Figures 6 about here]

Finally, we study the robustness of the 2SLS estimates for life expectancy at age 30 with the higher education enrollment rate as the outcome variable. Inspecting the estimated coefficients presented in Table 11, we see that the effect of life expectancy on the education is not driven by some convergence process in education related to the initial level of enrollment, years of schooling in the workforce, or income (columns 1–4). Moreover, motivated by the observation that *CVD* is generally larger in Northern and Southern states, we have also checked whether the effect is explained by the fact that states in these areas historically had more colleges, and this historical persistence could have produced a head start when college education became more widespread in the second half of the 20th century. When we accounted for this possible explanation by including the (state) number of permanent colleges before 1860 interacted with the indicator, the baseline estimate remained positive and significant (not reported). We also see from column

4 that the 2SLS estimate on life expectancy is not influenced by controlling for log income.

The results from the sample splits, reported in columns 5–9 demonstrate that the effect is prevalent in the Northern and Southern states, and in initially rich and poor states, whereas there is no effect considering states in the West only. Although the first-stage F-statistics in columns 5–8 are below 10, our conclusions seem not to be driven by weak first stages as indicated by the p-values from Anderson-Rubin Wald tests reported below the first-stage F-statistics.

[Table 11 about here]

7. Conclusion

This paper documents a positive effect of adult life expectancy on higher education enrollment in the US over the period 1960–2000 using the cardiovascular revolution as plausible source of conditional exogenous variation in life expectancy. While many existing studies exploit the decline in infectious-disease mortality—known as the third stage of the epidemiological transition—to identify the effect of health on economic outcomes, we focus on the cardiovascular revolution as the incidence of cardiovascular diseases is concentrated in adulthood and, thus, more suitable for testing the ideas of the original Ben-Porath model. Nevertheless, our findings remain reduced-form evidence support for this model. Future research could use our empirical strategy to study how the expected length of working life impacts on schooling choices. This would be even a more direct test of the Ben-Porath model.

References

[1] Acemoglu, D. and Johnson, S. (2006). Disease and development: The effect of life expectancy on economic growth, NBER Working Paper 12269.

[2] Acemoglu, D. and Johnson, S. (2007). Disease and development: The effect of life expectancy on economic growth, *Journal of Political Economy* 115, 925-985.

[3] Acemoglu, D. and Johnson, S. (2014) Disease and development: A reply to Bloom, Canning, and Fink, *Journal of Political Economy*, forthcoming.

[4] Aksan, A.M. and Chakraborty, S. (2014). Mortality versus morbidity in the demographic transition, *European Economic Review*, forthcoming.

[5] American Heart Association (2014). Heart disease and stroke statistics–2014 update: A report from the American Heart Association, *Circulation* 129, e28-e292.

[6] Angrist, J.D., and Pischke, J.S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

[7] Arking, R. (2006), *The Biology of Aging: Observations and Principles*, Oxford University Press, Oxford.

[8] Ashraf, Q.H., Lester, A., and Weil, D.N. (2008). When does improving health raise GDP?, NBER Macroeconomics

[9] Aghion, P., Howitt, P, Murtin, F. (2011). The relationship between health and growth: When Lucas meets Nelson-Phelps, *Review of Economics and Institutions* 2(1).

[10] Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings, *Journal of Political Economy* 75, 352–365.

[11] Bleakley, H. (2006). Disease and development: comments on Acemoglu and Johnson (2006), Working Paper.

[12] Bleakley, H. (2007). Disease and development: Evidence from hookworm eradication in the American South, *Quarterly Journal of Economics* 122, 73-117.

[13] Bleakley, H. (2009). Economic effects of childhood exposure to tropical disease. The American Economic Review, 21

[14] Bloom, D., Canning, D., and Fink, G. (2014). Disease and development revisited, *Journal of Political Economy*, forthcoming.

[15] Boucekinne, R, de la Croix, D., and Licandro, O. (2002). Vintage human capital, demographic trends, and endogenous growth, *Journal of Economic Theory* 104, 340.375.

[16] Carey, J.R. (2003). Life span: a conceptual overview, Population and Development Review, 29,1-18.

[17] Carnes, B.A. and Olshansky, S.J. (2007). A realist view of aging, mortality, and future longevity, *Population and Development Review* 33, 367-381.

[18] Cameron, A.C. and Miller, D.L. (2014). A Practitioner's Guide to Cluster-Robust Inference, *Journal of Human Resources*, (forthcoming).

[19] Case, A. and Paxson, C. (2008). Height, health, and cognitive function at older ages. American Economic Review, 98, 463–467.

[20] Cervellati, M. and Sunde, U. (2005). Human capital formation, life expectancy, and the process of development, *American Economic Review* 95, 1653-1672.

[21] Cervellati, M. and Sunde, U. (2011b). Disease and development: the role of life expectancy reconsidered. Economics Letters 113 (3), 269–272.

[22] Cervellati, M. and Sunde, U. (2011). Life expectancy and economic growth: the role of the demographic transition, *Journal of Economic Growth* 16, 99-133.

[23] Cervellati, M. and Sunde, U. (2013). Life Expectancy, schooling, and lifetime labor supply: Theory and evidence revisited, *Econometrica* 81, 2055-2086.

[24] Cervellati, M. and Sunde, U. (2014). The economic and demographic transition, mortality, and comparative Development, *American Economic Journal: Macroeconomics*, forthcoming.

[25] Cutler, D.M., Lleras-Muney, A. and Vogl, T. (2011). Socioeconomic status and health: dimensions and mechanisms, in: Glied, S. and Smith, P.C. (eds), *The Oxford Handbook of Health Economics*, Oxford University Press, Oxford, pp. 124-163.

[26] Dalgaard, C.J. and Strulik, H. (2014). Optimal aging and death, Journal of the European Economic Association, 12, 672–701.

[27] Dantas, A. P., Jimenez-Altayo, F., and Vila, E. (2012). Vascular aging: facts and factors. *Frontiers in Physiology 3*, 1-2.

[28] DHHS (1999a). Achievements in Public Health, 1900–1999: Decline in Deaths from Heart Disease and Stroke—United States, 1900–1999, Department of Health and Human Services, Centers for Disease Control and Prevention, Morbidity and Mortality Weekly Report 48, 649-646.

[29] DHHS (1999b). Achievements in Public Health, 1900–1999: Decline in Deaths from Heart Disease and Stroke—United States, 1900–1999, Department of Health and Human Services, Centers for Disease Control and Prevention, Morbidity and Mortality Weekly Report 48, 986-993.

[30] Doepke, M. (2005). Child mortality and fertility decline: Does the Barro-Becker model fit the facts? *Journal of Population Economics* 18, 337-366.

[31] Eckstein, Z. Mira, P., and Wolpin, K. (1999). A quantitative analysis of Swedisch fertility dynamics. *Review of Economic Dynamics*, 2, 137–165.

[32] Epstein, F.H. (1996). Cardiovascular disease epidemiology: A journey from the past into the future, Circulation, 93, 1755-1764.

[33] Foege, W.H. (1987). Public health: moving from debt to legacy, *American Journal of Public Health* 77, 1276-1278.

[34] Galor, O., 2005, From stagnation to growth: unified growth theory, in: *Handbook of Economic Growth*, Amsterdam: North-Holland.

[35] Galor, O., (2011). Unified Growth Theory. Princeton University Press.

[36] Gavrilov, L.A. and Gavrilova, N.S., (1991), *The Biology of Human Life Span: A Quantitative Approach*, Harwood Academic Publishers, London.

[37] Grossman, M., (2006), Education and nonmarket outcomes, in: *Handbook of the Economics of Education* Vol. 1, 577–633, Elsevier, Amsterdam.

[38] Grove R.D., Hetzel, A.M. (1968). Vital Statistics rates in the United States 1940-1960. *United States Government Printing Office*, Washington, 1968.

[39] Hansen, C.W. (2013a). Life Expectancy and Human Capital: Evidence from the International Epidemiological Transition. Journal of Health Economics, 32(6), 1142–1152.

[40] Hansen, C.W. (2013b). Health and development: a neoclassical perspective. Journal of Human Capital, 7(3), 274–295.

[41] Hansen, C.W. (2014). Cause of Death and Development in the US. Journal of Development Economics, 109, 143–153.

[42] Hansen, C.W. and Lønstrup, L. (2012) Can higher life expectancy induce more schooling and earlier retirement?, *Journal of Population Economics* 25, 1249-1264.

[43] Hazan, M. (2009). Longevity and lifetime labor supply: Evidence and implications, *Econometrica* 77, 1829-1863.

[44] Hazan, M. and Zoabi, H. (2006). Does longevity cause growth? A theoretical critique, Journal of Economic Growth, 11, 363-76.

[45] Herzer, D., Strulik, H., and Vollmer, S. (2013). The long-run determinants of fertility: One century of demographic change 1900-1999, *Journal of Economic Growth* 17, 357-385.

[46] Kirkwood, T.B.L. (1999). *Time of Our Lives: The Science of Human Aging*, Oxford University Press, Oxford.

[47] Lorentzen P., McMillan, J., and Wacziarg, R. (2008). Death and development, *Journal of Economic Growth* 13, 81-124.

[48] Meara, E.R., Richards, S., and Cutler, D.M. (2008), The gap gets bigger: changes in mortality and life expectancy, by education, 1981–2000, *Health Affairs* 27, 350-360.

[49] NCHS (2014), National Center for Health Statistics, (`http://www.cdc.gov/nchs/`.)

[50] NHLBI (2012). Morbidity and Mortality Chart Book, National Heart, Lung, and Blood Institute (`www.nhlbi.nih.gov/research/reports/2012-mortality-chart-book.htm`).

[51] Oeppen, J. and Vaupel, J.W. (2002). Broken limits to life expectancy, *Science* 296, 1029-1031.

[52] Olshansky, S.J., and Ault, A.B. (1986). The fourth stage of the epidemiologic transition: the age of delayed degenerative diseases, *Milbank Quarterly* 64, 355-391.

[53] Olshansky, S.J. and Carnes, B.A. and Cassel, C., (1990). In search of Methuselah: estimating the upper limits to human longevity, Science, 250, 634-640.

[54] Omran, A.R. (1971). The epidemiologic transition: a theory of the epidemiology of population change, *Milbank Quarterly* 49, 509–538.

[55] Omran, A. R. (1998). The epidemiologic transition theory revisited thirty years later, *World Health Statistics Quarterly* 51, 99-119.

[56] Snow, R.W., Craig, M.H., Newton, C.R.J.C., and Steketee, R.W., 2003. The public health burden of Plas-

modiumfalciparummalaria, in Africa: Deriving the numbers, Washington DC: The Disease Control Priorities Project (DCPP) Working Paper Number 11.

[57] Strulik, H. (2005). The role of human capital and population growth in R&D-based models of economic Growth, *Review of International Economics* 13, 129-145.

[58] Strulik, H., (2013). Health and Education: Understanding the Gradient, University of Goettingen Cege Discussion Paper 172.

[59] Strulik, H., Prettner, K., and Prskawetz, A. (2013). The past and future of knowledge-based growth, *Journal of Economic Growth* 18, 411-437.

[60] Strulik, H. and Vollmer, S. (2013). Long-run trends of human aging and longevity, *Journal of Population Economics* 26, 1303-1323.

[61] Strulik, H. and Werner, K. (2013). 50 is the New 30 – Long-run Trends of Schooling and Retirement Explained by Human Aging. University of Goettingen Cege Discussion Paper 152.

[62] Strulik, H. and Weisdorf, J. (2014). How child costs and survival shaped the industrial revolution and the demographic transition, *Macroeconomic Dynamics* 18, 114-144.

[63] Ruggles, S., J. T. Alexander, K. Genadek, R. Goeken, M. B. Schroeder, and M. Sobek (2010). Integrated Public Use Microdata Series: Version 5.0 *[Machine-readable database]. Minneapolis: University of Minnesota.*

[64] Thelle, D.S. (2011). Case fatality of acute myocardial infarction: an emerging gender gap, *European Journal of Epidemiology* 26, 829-831.

[65] Turner, C., Tamura, R., Mulholland, S.E., Baier, S. (2007). Education and income of the states of the United States: 1840-2000, *Journal of Economic Growth*, 12, 101-158.

[66] Ungvari, Z., Kaley, G., de Cabo, R., Sonntag, W. E., and Csiszar, A. (2010). Mechanisms of vascular aging: new perspectives, *Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 65, 1028-1041.

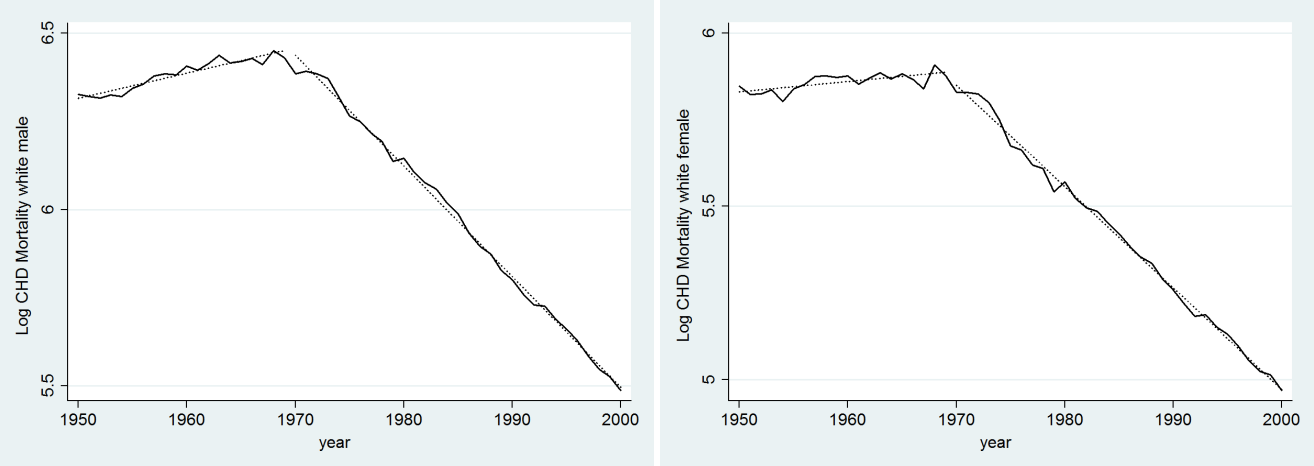
[67] USPHS (1964). Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service, United States, Public Health Service Publication No. 1103.

[68] Vallin, J., and Mesle, F. (2009). The segmented trend line of highest life expectancies, *Population and Development Review* 35(1), 159-187.

[69] WHO (2008). *The Global Burden of Disease: 2004 Update*, World Health Organization, Geneva.

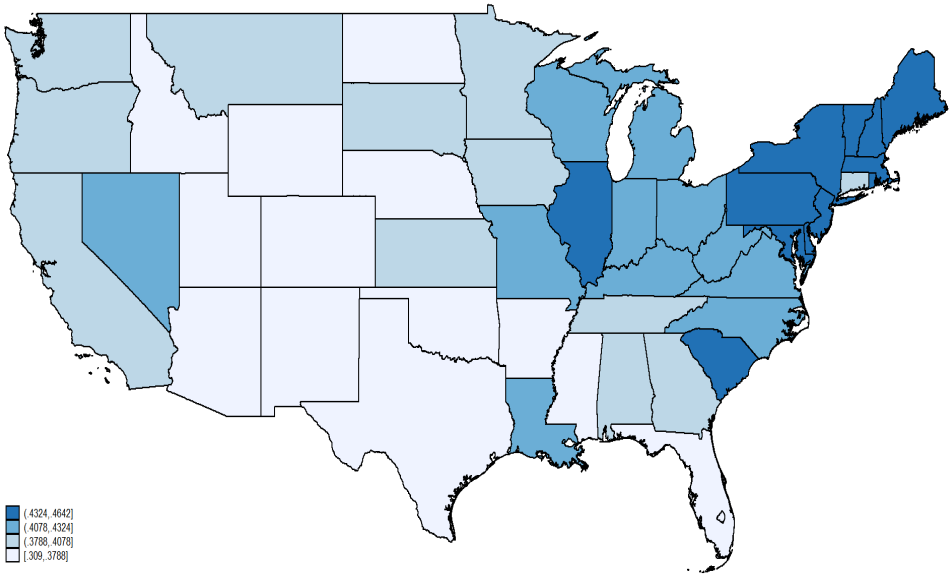
[70] Yusuf, S., Reddy, S., Ounpuu, S., and Anand, S. (2001). Global burden of cardiovascular diseases part I: general considerations, the epidemiologic transition, risk factors, and impact of urbanization, Circulation, 104, 2746-2753.

Figure 1: Annual Development of US CVD-mortality 1950–2000 (Coronary Heart Disease Mortality by Sex)



Notes: The left (right) figure shows the annual development in the log age-adjusted mortality rate from Coronary Heart Disease, CHD, for white males (females) in the US. Data source: National Heart, Lung, and Blood Institute.

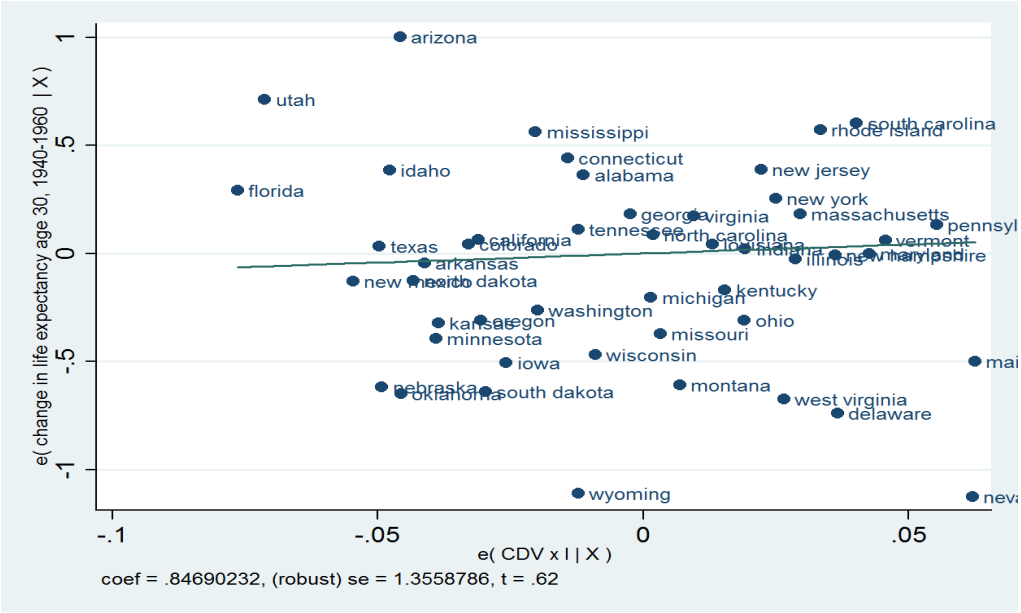
Figure 2: Spatial Distribution of 1960-CVD Mortality



Notes: This map pictures the geographical distribution of *CVD* for the 48 contiguous states.

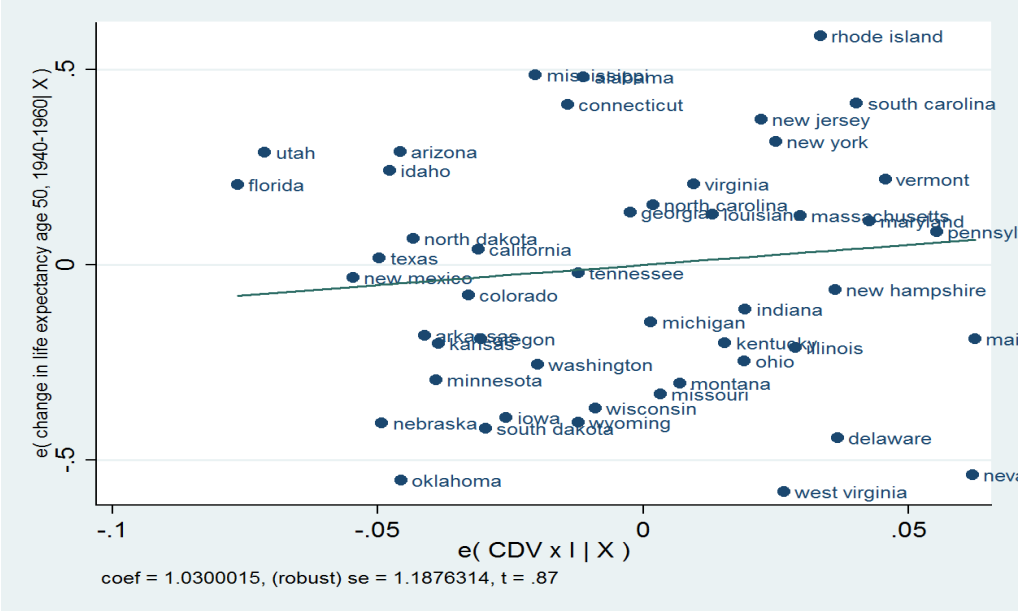
Figure 3: Pretreatment Changes in Life Expectancy and the CVD Shock

Panel A: Life Expectancy at age 30



Notes: This figure shows the relattionship between the change in pretreatment (i.e, 1940-1960) life expectancy at age 30 and the change in the CVD-shock variable 1960-2000 (i.e, CDV_s). Source is column 5 of Table 1.

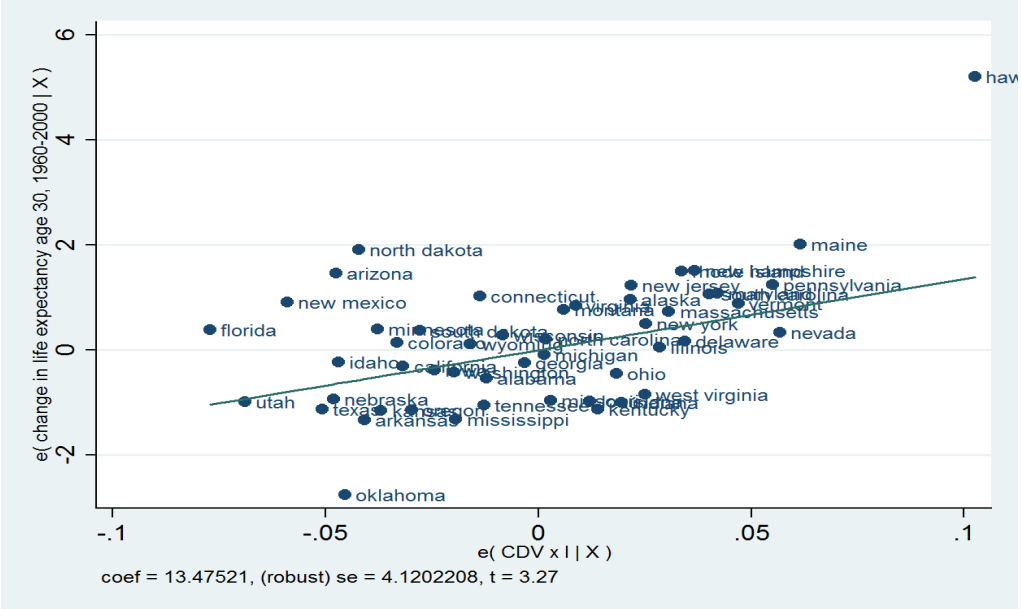
Panel B: Life Expectancy at age 50.



Notes: This figure shows the relattionship between the change in pretreatment (i.e, 1940-1960) life expectancy at age 50 and the change in the CVD-shock variable 1960-2000 (i.e, CDV_s). Source is column 5 of Table 1.

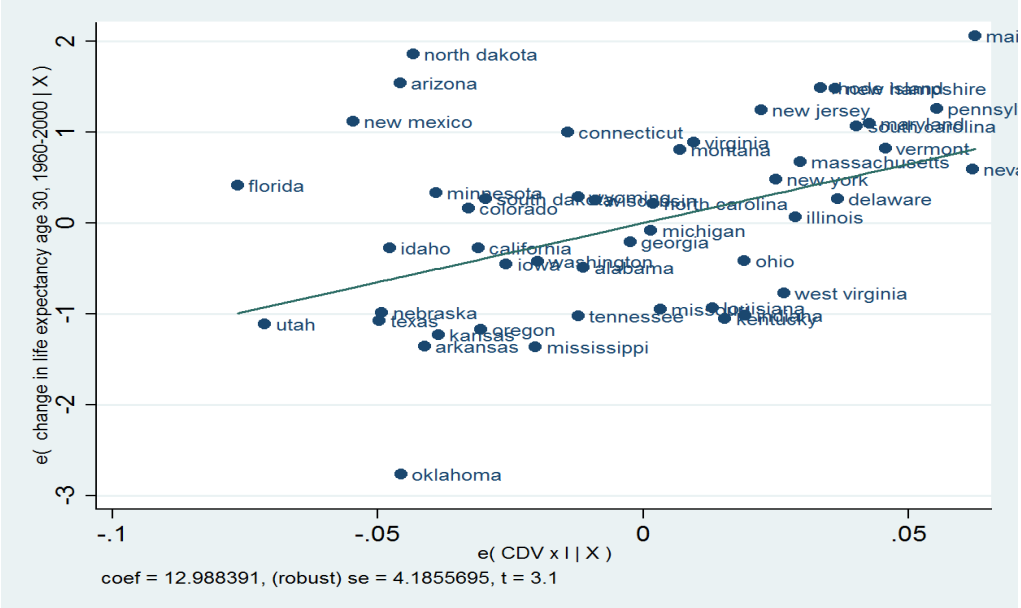
Figure 4: Baseline First-stage Result, 1960-2000

Panel A: Baseline sample with the 50 US states



Notes: This figure shows the partial correlation plot between life expectancy at age 30 and the CVD-shock variable for the period 1960-2000 for the 50 US states. Source is column 3 of Table 5.

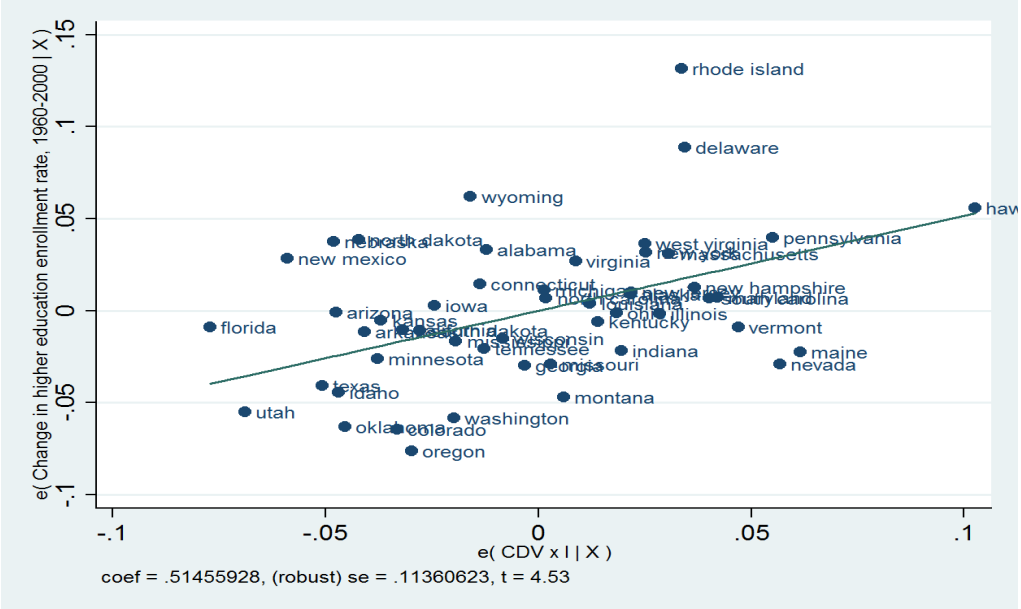
Panel B: Sample with the 48 contiguous US states



Notes: This figure shows the partial correlation plot between life expectancy at age 30 and the CVD-shock variable for the period 1960-2000 for the 48 contiguous US states.

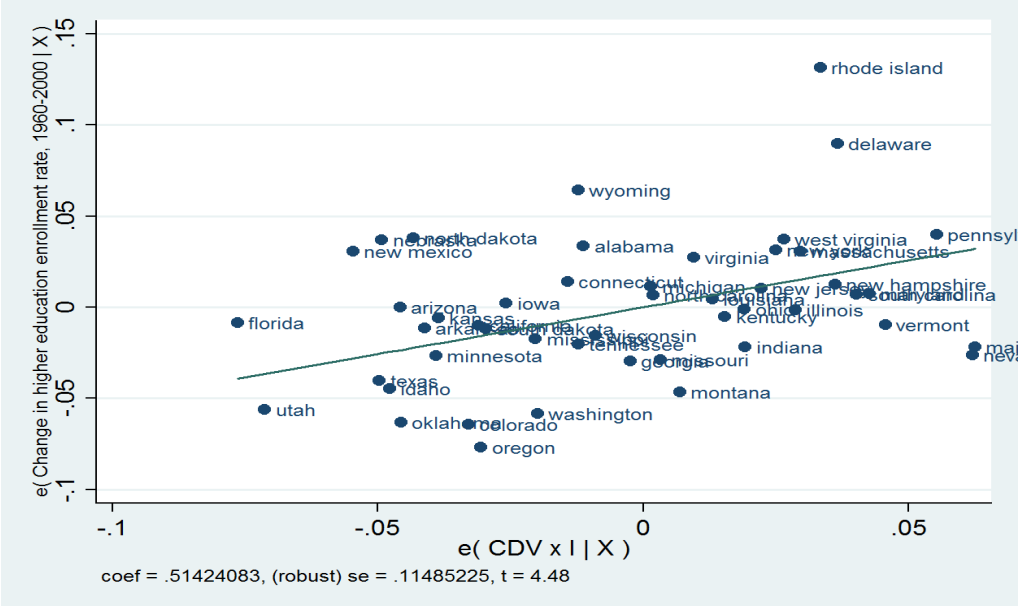
Figure 5: The Reduced Form for Higher Education Enrollment, 1960-2000

Panel A: Baseline sample with the 50 US states



Notes: This figure shows the reduced-form partial correlation plot between the higher education (net) enrollment rate and the CVD-shock variable for the period 1960-2000 for the 50 US states.

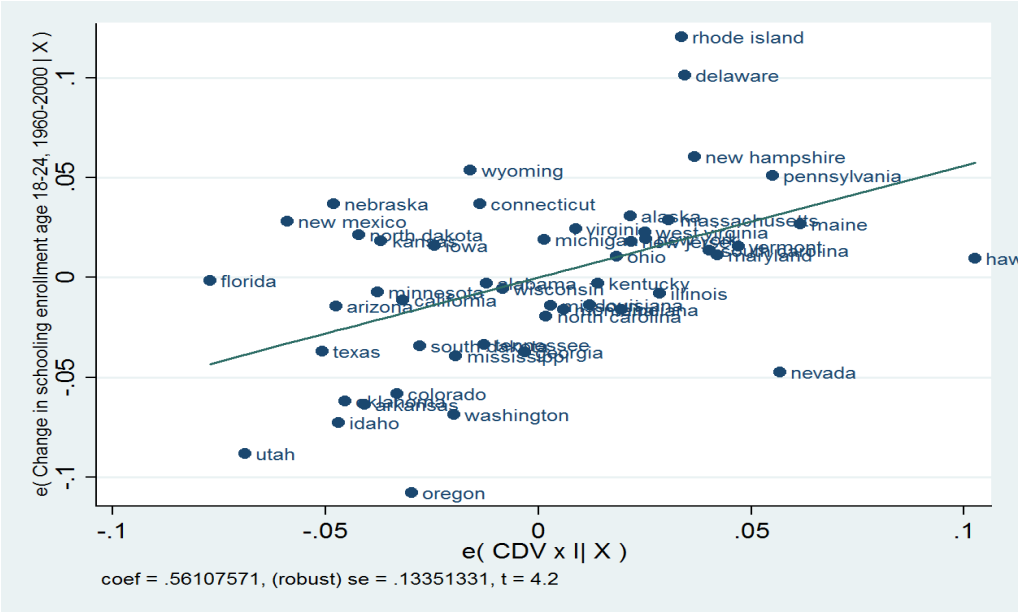
Panel B: Sample with the 48 contiguous US states



Notes: This figure shows the reduced-form partial correlation plot between the higher education (net) enrollment rate and the CVD-shock variable for the period 1960-2000 for the 48 contiguous US states.

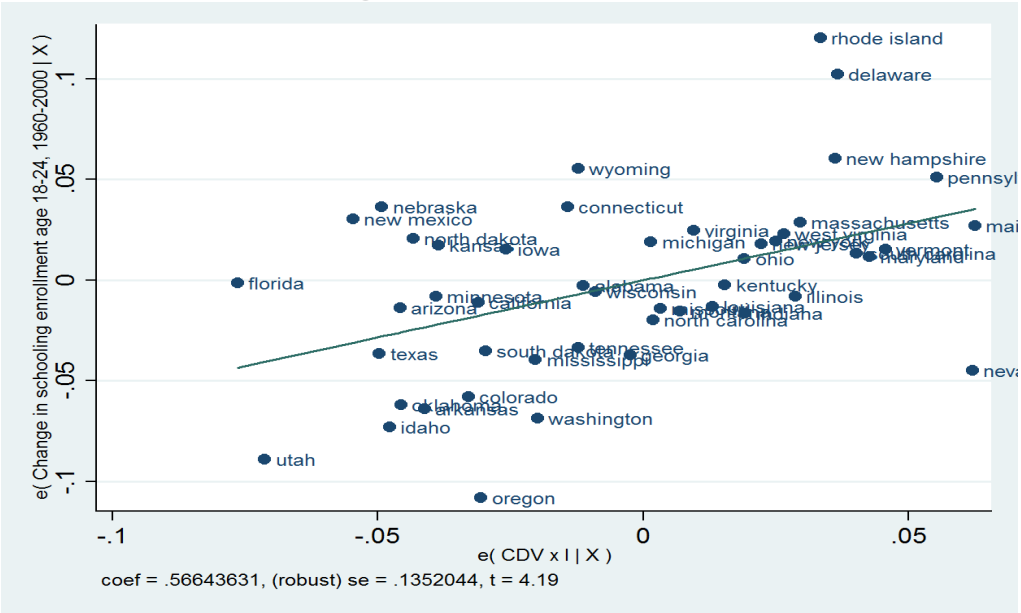
Figure 6: The Reduced Form for Schooling Enrollment (age 18-24), 1960-2000

Panel A: Baseline sample with the 50 US states



Notes: This figure shows the reduced-form partial correlation plot between the schooling enrollment rate age 18-24 and the CVD-shock variable for the period 1960-2000 for the 50 US states.

Panel B: Sample with the 48 contiguous US states



Notes: This figure shows the reduced-form partial correlation plot between the schooling enrollment rate age 18-24 and the CVD-shock variable for the 48 contiguous US states.

Table 1—Descriptive Statistics by Year

	Obs:	Mean	Std. Dev.	Min	Max
Year 1960:					
Life Expectancy at age 30	50	43.98	0.771	42.18	45.35
Life Expectancy at age 50	50	25.85	0.745	23.91	27.00
CVD · Post	50	0.000	0.000	0.000	0.000
Initial Mortality · Post	50	0.000	0.000	0.000	0.000
Higher Education Enrollment	50	0.145	0.042	0.031	0.244
Schooling Enrollment age 18-24	50	0.236	0.476	0.081	0.342
Year 2000:					
Life Expectancy at age 30	50	49.23	1.118	46.99	52.16
Life Expectancy at age 50	50	30.67	0.967	28.71	33.19
CVD · Post	50	0.404	0.038	0.309	0.499
Initial Mortality · Post	50	0.372	0.031	0.340	0.472
Higher Education Enrollment	50	0.360	0.049	0.239	0.494
Schooling Enrollment age 18-24	50	0.462	0.052	0.320	0.570

Table 2—Pretreatment State Characteristics by 1960-CVD Mortality

		Pretreatment levels		Pretreatment changes	
	Pretreatment mean	Without Covariates	With Covariates	Without Covariates	With Covariates
	(1)	(2)	(3)	(4)	(5)
Life Expectancy at age 30	42.41	-17.54***	-18.60***	0.244	0.847
	(1.862)	(1.626)	(1.153)	(1.888)	(1.283)
Life Expectancy at age 50	24.77	-18.86***	-19.40***	0.815	1.030
	(1.383)	(1.369)	(0.915)	(1.516)	(1.124)
Higher Education Enrollment	0.150	0.0465	0.0211	NA	NA
	(0.036)	(0.133)	(0.137)		
Schooling Enrollment age 18-24	0.193	-0.144	-0.140	0.140	0.161
	(0.060)	(0.139)	(0.149)	(0.123)	(0.107)
# States	48	48	48	48	48

Notes: The table reports least squares estimates (weighted by the white population size in 1960) of CVD in equations (**??**) and (**??**). The left-hand-side variables are: life expectancy at age 30, life expectancy at age 50, the higher education (net) enrollment rate for the white population, the schooling enrollment rate age 18-24. Pretreatment level refers to the level of the outcome in 1960, whereas pretreatment changes refers to the change in the outcome in the period 1940–1960. For the life expectancy outcomes the covariates are: the number of deaths from all other causes than CVD per 100 white population in 1960; the higher education enrollment rate for the white population in 1960; log GDP per capita in 1960. For the schooling outcomes the covariates are: the number of deaths from all other causes than CVD per 100 white population in 1960; log GDP per capita in 1960. Constants and controls are not reported. Standard errors are robust clustered at the state level.

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

Table 3—Flexible Model: 10-year Panel (1940–2000)

	Dependent Variable:					
	Life Expectancy at age 30			Life Expectancy at age 50		
	(1)	(2)	(3)	(4)	(5)	(6)
CVD · 1950	-0.288	-0.288	-0.288	0.630	0.630	0.630
	(1.081)	(1.081)	(1.081)	(0.939)	(0.939)	(0.939)
CVD · 1960	0.0395	0.0463	-0.00859	0.615	0.616	0.604
	(1.910)	(1.904)	(1.907)	(1.542)	(1.541)	(1.534)
CVD · 1970	5.506**	5.793**	4.960**	5.847**	5.899**	6.909***
	(2.772)	(2.631)	(2.245)	(2.434)	(2.427)	(2.245)
CVD · 1980	7.951**	8.238**	7.405**	8.222***	8.274***	9.284***
	(3.382)	(3.237)	(2.903)	(3.061)	(3.040)	(2.787)
CVD · 1990	11.56***	11.85***	11.02***	12.17***	12.22***	13.23***
	(3.363)	(3.230)	(3.592)	(3.527)	(3.522)	(3.539)
CVD · 2000	14.59**	14.87***	14.04***	16.48***	16.54***	17.55***
	(5.735)	(5.652)	(4.925)	(5.430)	(5.426)	(5.006)
Controls (· *Post*):						
Initial Mortality		3.325	2.842		0.603	0.353
		(5.970)	(6.103)		(5.510)	(5.626)
Initial Enrollment			7.198***			8.305***
			(2.737)			(2.586)
Initial Income			1.602***			1.542***
			(0.316)			(0.298)
# Observations	346	346	346	346	346	346
# States	50	50	50	50	50	50

Notes: The table reports least squares estimates, weighted by the white population size in 1960. The observation period is 1940–2000 (comparison year is 1940); 10-year panel. All regressions include state and year fixed effects. The left-hand-side variable is life expectancy at the ages 30 and 50 for the white population. CVD is the age-adjusted number of deaths from cardiovascular diseases per 100 white population in 1960. Initial mortality is the number of deaths from all other causes than CVD per 100 white population in 1960. Initial Education is the higher education enrollment rate for the white population. Initial Income is log GDP per capita in 1960. Post is an indicator that equals one after 1960. Constants are not reported. Standard errors are robust clustered at the state level.

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

Table 4—First Stage: 10-year Panel (1940–2000)

	Dependent Variable:					
	Life Expectancy at age 30			Life Expectancy at age 50		
	(1)	(2)	(3)	(4)	(5)	(6)
CVD · Post	9.984***	10.27***	9.455***	9.984***	10.32***	9.492***
	(3.043)	(2.984)	(2.805)	(3.043)	(3.019)	(2.745)
Controls (· Post):						
Initial Mortality		3.324	2.842		0.602	0.352
		(5.971)	(6.104)		(5.510)	(5.626)
Initial Enrollment			7.198***			8.307***
			(2.736)			(2.586)
Initial Income			1.602***			1.542***
			(0.316)			(0.298)
# Observations	346	346	346	346	346	346
# States	50	50	50	50	50	50

Notes: The table reports least squares estimates, weighted by the white population size in 1960. The observation period is 1940–2000; 10-year panel. All regressions include state and year fixed effects. The left-hand-side variable is life expectancy at the ages 30 and 50 for the white population. CVD is the age-adjusted number of deaths from cardiovascular diseases per 100 white population in 1960. Initial mortality is the number of deaths from all other causes than CVD per 100 white population in 1960. Initial enrollment is the higher education enrollment rate for the white population. Initial Income is log GDP per capita in 1960. Post is an indicator that equals one after 1960. Constants are not reported. Standard errors are robust clustered at the state level.

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

Table 5—First Stage: Baseline DD Estimates

	Dependent Variable:					
	Life Expectancy at age 30			Life Expectancy at age 50		
	(1)	(2)	(3)	(4)	(5)	(6)
CVD · Post	14.55***	14.56***	13.48***	15.87***	15.87***	14.86***
	(4.277)	(4.302)	(3.909)	(4.268)	(4.331)	(4.363)
Controls (· Post):						
Initial Mortality		0.159	-0.161		-0.00119	-0.190
		(7.393)	(7.682)		(7.366)	(7.672)
Initial Enrollment			10.99***			10.64***
			(3.831)			(3.556)
Initial Income			2.217***			2.039***
			(0.485)			(0.465)
# Observations	100	100	100	100	100	100
# States	50	50	50	50	50	50

Notes: The table reports least squares estimates, weighted by the white population size in 1960. The observation years are 1960 and 2000. All regressions include state and year fixed effects. The left-hand-side variable is life expectancy at the ages 30 and 50 for the white population. CVD is the age-adjusted number of deaths from cardiovascular diseases per 100 white population in 1960. Initial mortality is the number of deaths from all other causes than CVD per 100 white population in 1960. Initial Enrollment is the higher education enrollment rate for the white population. Initial Income is log GDP per capita in 1960. Post is an indicator that equals one after 1960. Constants are not reported. Standard errors are robust clustered at the state level.

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

Table 6—First Stage: Robustness to Linear State Trends

	Dependent Variable:					
	Life Expectancy at age 30			Life Expectancy at age 50		
	(1)	(2)	(3)	(4)	(5)	(6)
CVD · Post	14.05***	13.06***	12.14***	14.82***	14.26***	13.38***
	(3.358)	(3.459)	(3.335)	(3.490)	(3.622)	(3.999)
Controls (· Post):						
Initial Mortality		-11.40	-10.67		-6.452	-6.290
		(7.298)	(7.158)		(7.226)	(7.318)
Initial Enrollment			14.38***			11.15***
			(3.377)			(3.118)
Initial Income			1.071**			1.283***
			(0.490)			(0.448)
State specific linear time trends	Yes	Yes	Yes	Yes	Yes	Yes
# Observations	144	144	144	144	144	144
# States	48	48	48	48	48	48

Notes: The table reports least squares estimates, weighted by the white population size in 1960. The observation years are 1940, 1960 and 2000. All regressions include state and year fixed effects and linear state trends. The left-hand-side variable is life expectancy at the ages 30 and 50 for the white population. CVD is the age-adjusted number of deaths from cardiovascular diseases per 100 white population in 1960. Initial mortality is the number of deaths from all other causes than CVD per 100 white population in 1960. Initial Enrollment is the higher education enrollment rate for the white population. Initial Income is log GDP per capita in 1960. Post is an indicator that equals one after 1960. Constants are not reported. Standard errors are robust clustered at the state level.

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

Table 7—First Stage: Sensitivity Tests and Sample Splits

	Dependent Variable is Life Expectancy at age 30								
	Addtional Controls:				Sample Splits: Northern States	Southern States	Western States	Rich States	Poor States
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
CVD · Post	12.24***	15.39***	16.28***	15.88***	27.58***	14.43***	7.426	9.964**	16.39***
	(3.338)	(3.739)	(2.886)	(3.059)	(6.106)	(5.306)	(8.476)	(4.846)	(5.406)
Income				1.969**					
				(0.859)					
Controls (· Post):									
Initial Mortality	1.950	-3.050	-4.095	-0.0973	-30.82**	-55.80***	5.079	-8.374	1.584
	(5.953)	(8.720)	(7.153)	(7.810)	(15.70)	(13.36)	(7.271)	(11.52)	(12.18)
Initial Enrollment	0.563	11.20***	-0.846	-2.507	19.10***	2.550	-1.533	3.364	16.11***
	(5.806)	(3.689)	(4.690)	(4.098)	(5.155)	(8.183)	(6.274)	(7.251)	(4.084)
Initial Income	0.843	1.535	-0.920	0.277	-0.226	4.294***	0.477	3.391***	3.850***
	(0.573)	(0.985)	(1.045)	(1.252)	(0.783)	(1.057)	(0.945)	(1.119)	(1.483)
Initial Education	0.821***		0.968***	0.993***					
	(0.242)		(0.225)	(0.202)					
Initial Physical Capital		0.635	1.412*	1.223*					
		(0.774)	(0.774)	(0.731)					
# Observations	100	100	100	100	42	32	26	50	50
# States	50	50	50	50	21	16	13	25	25

Notes: The table reports least squares estimates, weighted by the white population size in 1960. The observation years are 1960 and 2000. All regressions include state and year fixed effects. The left-hand-side variable is life expectancy at the ages 30 for the white population. CVD is the age-adjusted number of deaths from cardiovascular diseases per 100 white population in 1960. Initial mortality is the number of deaths from all other causes than CVD per 100 white population in 1960. Initial Enrollment is the higher education enrollment rate for the white population. Initial Income is log GDP per capita in 1960. Initial Education is average years of schooling in the workforce in 1960. Initial Physical Capital is log capital per worker in 1960. Income is log GDP per capita. Post is an indicator that equals one after 1960. Northern States are states in the US North. Southern States are states in the US South. Western States are states in the US West. Rich states are states with above median income in 1960. Poor states are states with below median income in 1960. Constants are not reported. Standard errors are robust clustered at the state level.

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

Table 8—Second Stage: Life Expectancy and Higher Education

	Dependent Variable is the Higher Education Enrollment Rate					
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: LS estimates					
Life Expectancy at age 30	0.0163***	0.0159***	0.0205***			
	(0.00336)	(0.00336)	(0.00355)			
Life Expectancy at age 50				0.0175***	0.0171***	0.0217***
				(0.00330)	(0.00332)	(0.00317)
	Panel B: 2SLS estimates					
Life Expectancy at age 30	0.0370***	0.0358***	0.0382***			
	(0.00735)	(0.00722)	(0.00761)			
Life Expectancy at age 50				0.0339***	0.0329***	0.0346***
				(0.00582)	(0.00569)	(0.00705)
Controls (· Post):						
Initial Mortality	No	Yes	Yes	No	Yes	Yes
Initial Enrollment	No	No	Yes	No	No	Yes
Initial Income	No	No	Yes	No	No	Yes
First-stage F-stat.	11.22	11.00	11.18	13.41	12.89	10.91
# Observations	100	100	100	100	100	100
# States	50	50	50	50	50	50

Notes: Panel A reports least squares estimates, weighted by the white population size in 1960. Panel B reports two stage least squares estimates, weighted by the white population size in 1960, using the instrument CVD ×Post. The observation years are 1960 and 2000. All regressions include state and year fixed effects. The left-hand-side variable is the (net) higher education enrollment rate for the white population. CVD is the age-adjusted number of deaths from cardiovascular diseases per 100 white population in 1960. Initial mortality is the number of deaths from all other causes than CVD per 100 white population in 1960. Initial Enrollment is the (net) higher education enrollment rate for the white population. Initial Income is log GDP per capita in 1960. Post is an indicator that equals one after 1960. Constants are not reported. Standard errors are robust clustered at the state level.

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

Table 9—Second Stage: Life Expectancy and Schooling Enrollment (age 18-24)

	Dependent Variable is the Schooling Enrollment age 18-24					
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: LS estimates					
Life Expectancy at age 30	0.0218***	0.0212***	0.0237***			
	(0.00265)	(0.00273)	(0.00379)			
Life Expectancy at age 50				0.0233***	0.0226***	0.0252***
				(0.00258)	(0.00273)	(0.00329)
	Panel B: 2SLS estimates					
Life Expectancy at age 30	0.0423***	0.0401***	0.0416***			
	(0.00897)	(0.00842)	(0.00738)			
Life Expectancy at age 50				0.0388***	0.0367***	0.0378***
				(0.00703)	(0.00643)	(0.00622)
Controls (· Post):						
Initial Mortality	No	Yes	Yes	No	Yes	Yes
Initial Enrollment	No	No	Yes	No	No	Yes
Initial Income	No	No	Yes	No	No	Yes
First-stage F-stat.	11.22	11.00	11.18	13.41	12.89	10.91
# Observations	100	100	100	100	100	100
# States	50	50	50	50	50	50

Notes: Panel A reports least squares estimates, weighted by the white population size in 1960. Panel B reports two stage least squares estimates, weighted by the white population size in 1960, using the instrument CVD ×Post. The observation years are 1960 and 2000. All regressions include state and year fixed effects. The left-hand-side variable is the schooling enrollment rate for the white population age 18-24. CVD is the age-adjusted number of deaths from cardiovascular diseases per 100 white population in 1960. Initial mortality is the number of deaths from all other causes than CVD per 100 white population in 1960. Initial Enrollment is the (net) higher education enrollment rate for the white population. Initial Income is log GDP per capita in 1960. Post is an indicator that equals one after 1960. Constants are not reported. Standard errors are robust clustered at the state level.

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

Table 10—Second Stage: Robustness to Linear State Trends

	Dependent Variable is the Schooling Enrollment age 18-24					
	(1)	(2)	(3)	(4)	(5)	(6)
Life Expectancy at age 30	0.0343***	0.0343**	0.0343***			
	(0.0131)	(0.0146)	(0.00996)			
Life Expectancy at age 50				0.0325***	0.0314**	0.0311***
				(0.0122)	(0.0127)	(0.00961)
Controls (· Post):						
Initial Mortality	No	Yes	Yes	No	Yes	Yes
Initial Enrollment	No	No	Yes	No	No	Yes
Initial Income	No	No	Yes	No	No	Yes
State specific linear time trends	Yes	Yes	Yes	Yes	Yes	Yes
First-stage F-stat.	16.96	13.67	12.43	17.48	14.87	10.51
# Observations	144	144	144	144	144	144
# States	48	48	48	48	48	48

Notes: The table reports two stage least squares estimates, weighted by the white population size in 1960, using the instrument CVD ×Post. The observation years are 1940, 1960 and 2000. All regressions include state and year fixed effects and linear state trends. The left-hand-side variable is the schooling enrollment rate for the white population age 18-24. CVD is the age-adjusted number of deaths from cardiovascular diseases per 100 white population in 1960. Initial mortality is the number of deaths from all other causes than CVD per 100 white population in 1960. Initial Enrollment is the higher education enrollment rate for the white population. Initial Income is log GDP per capita in 1960. Post is an indicator that equals one after 1960. Constants are not reported. Standard errors are robust clustered at the state level.
*** p<0.01, ** p<0.05, * p<0.1.

Table 11—Second Stage: Sensitivity Tests and Sample Splits

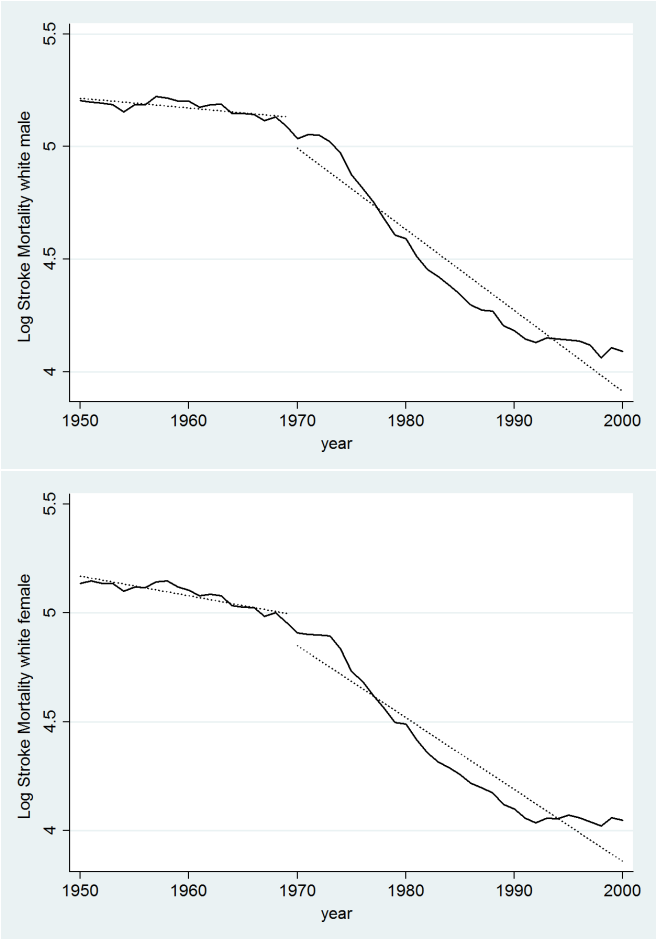
	Dependent Variable is the Higher Education Enrollment Rate								
	Addtional Controls:				Sample Splits: Northern States	Southern States	Western States	Rich States	Poor States
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Life Expectancy at age 30	0.0420***	0.0377***	0.0359***	0.0366***	0.0266***	0.0308***	-0.0874	0.0474***	0.0282**
	(0.00795)	(0.00869)	(0.00606)	(0.00628)	(0.00829)	(0.00996)	(0.131)	(0.0142)	(0.0113)
Income	No	No	No	Yes	No	No	No	No	No
Controls (· Post):									
Initial Mortality	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial Enrollment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial Income	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial Education	Yes	No	Yes	Yes	No	No	No	No	No
Initial Physical Capital	No	Yes	Yes	Yes	No	No	No	No	No
First-stage F-statistics	12.51	15.77	29.28	24.55	17.54	6.035	0.595	3.727	8.108
Anderson-Rubin [p-value]	[0.0000]	[0.0000]	[0.0000]	[0.000]	[0.0000]	[0.0007]	[0.0146]	[0.0014]	[0.0003]
# Observations	100	100	100	100	42	32	26	50	50
# States	50	50	50	50	21	16	13	25	25

Notes: The table reports two stage least squares estimates, weighted by the white population size in 1960, using the instrument CVD ×Post. The observation years are 1960 and 2000. All regressions include state and year fixed effects. The left-hand-side variable is the (net) higher education enrollment rate for the white population. CVD is the age-adjusted number of deaths from cardiovascular diseases per 100 white population in 1960. Initial mortality is the number of deaths from all other causes than CVD per 100 white population in 1960. Initial Enrollment is the net higher education enrollment rate for the white population. Initial Income is log GDP per capita in 1960. Initial Education is average years of schooling in the workforce in 1960. Initial Physical Capital is log capital per worker in 1960. Income is log GDP per capita. Post is an indicator that equals one after 1960. Northern States are states in the US North. Southern States are states in the US South. Western States are states in the US West. Rich states are states with above median income in 1960. Poor states are states with below median income in 1960. Constants are not reported. Standard errors are robust clustered at the state level.

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

Appendix

Figure A1: Annual Development of US CVD-mortality 1950–2000 (Stroke Mortality by Sex)



Notes: The top (lower) figure shows the annual development in the log age-adjusted mortality rate from Strokes for white males (females) in the US. Data source: National Heart, Lung, and Blood Institute.