Children of the Mortality Revolution – Infectious Disease and Long-run Outcomes

Xiaohan Zhang*

JOB MARKET PAPER

Abstract

This paper studies the effect of infectious disease exposure in early childhood on adult labor market outcomes. To do this, I exploit the exogenous variations in public health projects and new drugs during the Mortality Revolution (1901-1955) in the United States, an era with unmatched mortality decline, driven by innovations in disease control technologies. I create an index of early childhood disease exposure that exploits cross-state variation in pre-intervention disease prevalence, and time variation arising from medical innovations during this period. The results indicate that higher disease prevalence in childhood reduces adult education attainment and earnings, and that public health interventions contributed to roughly 10% of the changes in labor market outcomes between the 1901 and 1955 cohorts. The effect per unit of mortality decline is stronger in the second half of this period (1937-1955), when medications such as penicillin and sulfa drugs were introduced. My findings also shed light on the benefit of controlling infectious diseases in the developing world.

*Zhang: University of California Davis, One Shields Avenue, Davis, CA, 95616, xhzhang@ucdavis.edu. I thank the participants of the labor economic brownbag and the economic history workshop in UC Davis; Professors Douglas Miller, Ann Stevens, Alan Omstead, and Colin Cameron for their feedback on the paper throughout the process; Professors Price Fishback and Robert McGuire for the provision of important data and comments. Professors Hilary Hoynes and Aline Butikofer for providing helpful feedbacks. I am especially grateful to my advisor Professor Marianne Page for her teaching and support.

1 Introduction

Health in early childhood has long lasting impacts (Barker, 1995; Almond and Currie, 2011a,b; Elo and Preston, 1992). One of the most prevailing childhood health hazards in the early 20th century U.S. was infectious disease. And it continues to be so in the developing countries today. Great resources have been devoted to control infectious diseases worldwide. The Bill and Melinda Gates Foundation alone has donated over 7 billion dollars to this cause. Do investments in reducing infectious diseases generate large returns? This paper measures a major portion of the return – the adult productivity gains from lower early childhood exposure to common infectious diseases. I examine this question by looking at the impact of the public health interventions in the U.S. in an era later known as the Mortality Revolution and the productivity gains of children growing up during this era.

Between 1900 and 1955, the U.S. witnessed a substantial decline in infectious disease prevalence (Armstrong, Conn, and Pinner, 1999). The Centers for Disease Control and Prevention refers to the control of infectious diseases as one of the top ten achievements of public health in the 20th century U.S. (CDC, 1999). Durand (1960) referred to this period as a "revolution in the technology of disease control". Historians named this period the "Mortality Revolution" (Easterlin, 1995), because the deaths due to infectious causes of death declined by 90% within these years (Cutler and Meara, 2003, also see Figure 1). The Mortality Revolution is divided into two periods according to the type of technologies and public health interventions used to reduce infectious diseases. In the First Mortality Revolution (1900-1936), the interventions are mainly health behavior campaigns, sanitation and hygiene actions by the government, vaccination, and advancements in detection techniques. The innovation in antibiotics and other antimicrobial medicine are often characterized as the Second Morality Revolution (1937-1955) (Cutler and Miller, 2005). This paper is the first to gauge the overall impact of these interventions on the adult labor market gains of children growing up in this era, and the differential benefits from technologies during the First versus the Second Mortality Revolution.

For children born and raised during the first half of the 20th century, these health interventions

operated together to create a series of positive health shocks. Multiple interventions often took place simultaneously and targeted a similar set of diseases[1]. Using an underutilized data set from McGuire and Coelho (2011) and an index I named "predicted childhood disease exposure", this paper is the first to examine the effects of a set of most common infectious diseases on children's future economic outcomes.

A handful of studies have looked at the long run impact of public health interventions during the Mortality Revolution era. Most of these studies focus on local public health interventions that affected small parts of the population (Bleakley, 2007, 2010; Beach, Ferrie, Saavedra, and Troesken, 2014). An exception is Bhalotra and Venkataramani (2011), who estimate the impact of sulfa drugs, the most effective antibiotic prior to the discovering of penicillin. All these interventions together account for less than 16% of the decline in death from infectious diseases during this period[2]. This paper substantially expands the set of public health interventions under examination. In addition, the existing studies found drastically different results. This paper proposes a framework that helps reconcile the differences in the findings from the past literature according to the types of interventions studied.

The biggest identification challenge is to meaningfully capture improvements in health that were caused by public health interventions rather than economic growth or other confounding factors. I overcome this problem by adopting and improving on an identification strategy proposed by Acemoglu and Johnson (2007) and used in Hansen (2014)[3]. The method in Bleakley (2007, 2010) also captures a similar type of variation, but in a slightly different way.

My identification strategy is similar to an identification strategy that has recently been har-

[1] Additionally, some infectious diseases were often misreported and misdiagnosed as others (McGuire and Coelho, 2011; Troesken, 2004). When the morbidity or mortality of one disease is used as a proxy for the effectiveness of public health interventions, measurement errors arise.

[2] Sulfa drugs can reduce mortalities from multiple causes. So I estimated an upper and a lower bound for the impact of sulfa drugs. The upper bound is the change in total mortality between 1937 and 1942. The lower bound is the change in all-age mortality from pneumonia and influenza between 1937 and 1942. The latter is from Bhalotra and Venkataramani (2011). These numbers put the estimated impact of sulfa drugs between 65 and 116 per 100,000 population. Using the upper(lower) bound, past studies can account for 16% (9%) of the decline in infectious deaths.

[3] The early specification used in Acemoglu and Johnson (2007) is a long-difference model that compares observations from two cross sections. This method was criticized by Aghion, Howitt, and Murtin (2011) and Bloom, Canning, and Fink (2014) for lack of initial health controls by country. This problem is resolved in both Hansen (2014) and in the current paper.

nessed to better understand the extent to which cross country and sub-country differences in developing countries' contemporaneous Gross Domestic Product (GDP) relate to public health improvements (Acemoglu and Johnson, 2007; Hansen, 2014). That literature currently overlooks the biggest beneficiaries of the public health interventions–children. Instead, current studies focus on the contemporaneous health environment as it relates to the health of the *adult* population and concurrent GDP growth. Figure 4 shows that the "disease burden[4]" falls heaviest on infants and children. In this paper, I match adults to the disease environment during *early childhood*. In doing so, I estimate the effect of exposure to infectious diseases during the part of the life-cycle when individuals are most vulnerable to them, which provides an important extension of the existing literature.

Based on the method in Acemoglu and Johnson (2007), my index of "predicted childhood exposure" captures the degree to which individuals were exposed to infectious diseases during early childhood. The index is based on the number of deaths in each states due to infectious diseases. Mortality is commonly used as a proxy for morbidity and it is highly correlated with the probability of catching an infectious disease.

This method exploits two types of variations. The first is the variation in disease intensity across states in a period prior to the interventions as a proxy for variation in treatment intensity generated from medical advancement. Figure 2 (a) shows that there was a great deal of geographic variation in the 12 common infectious diseases prior to the interventions[5]. The second is the time variation arising from a series of medical innovations and public health interventions. I test the validity of the identification strategy by testing its exclusion restrictions. The list includes, but is not restricted to, per capita income, education expenditure and other aspects of health. The results are robust to these tests, signaling that "predicted childhood exposure" is uncorrelated with state

[4]Disease Burden is the impact of a health problem as measured by financial cost, mortality, morbidity, or other indicators.

[5]The 12 common infectious diseases I used are smallpox, dysentery, typhoid fever, diphtheria, croup, tuberculosis (all forms), pertussis, measles, scarlet fever, malaria, influenza and pneumonia. According to McGuire and Coelho (2011), the most common infectious diseases also include cholera, typhus, and yellow fever. I did not include these diseases because of data limitation in the national mortality rates from the vital statistics. These three diseases are also less common than the 12 diseases I included.

economic development and other public investment.

I find a large and significant improvement in education attainment and earnings among adults who benefited from the U.S. public health interventions during childhood. *Ceteris paribus*, moving from a state with high initial exposure to disease to a state with low exposure (from the 80th to the 20th percentile of all states) increases high school completion by 4 percentage points, increases overall education attainment by half a year, and increases earnings by 7 percent[6]. The interventions explain roughly 10% of the increase in high school graduation rates between the 1900 and the 1960 cohorts. This is comparable to the large productivity gain that has brought about by the improvement of the U.S. education system during the 20th century (Goldin and Katz, 2010)[7]. These findings suggest that health capital may be an important channel of productivity gain during the 20th century.

I also find that the benefit from a unit of mortality decline was much smaller during the First Mortality Revolution (1900-1936) compared to the Second Mortality Revolution (1937-1955). This pattern may be explained by the different types of innovations that occurred during the two periods. The first Mortality Revolution focused on reducing the contraction of infectious diseases, while the second introduced new drugs that treated and cured diseases (Connolly, Golden, and Schneider, 2012; Bhalotra and Venkataramani, 2011).

Disease intervention has two off-setting effects on the health of the population, commonly referred to as scarring and selection effects. Mortality rates cannot fully capture the health of the survivors due to the existence of both scarring and selection effects. An increase in mortality may imply better health if selection effect dominates. Therefore, the basic specification, which uses a basic version of the predicted childhood exposure, will capture the net effect of selection and scarring effects combined. The selection effect is positive, which means that the basic specification in this study, as well as findings in most papers in this literature (Rawlings, 2012; Almond, 2006; Almond, Currie, and Herrmann, 2011; Bozzoli, Deaton, and Quintana-Domeque, 2009), are biased

[6]I define high school graduation as finished 12th grade or higher. It does not necessarily imply obtaining a high school degree. I cannot directly identify individuals with high school degrees in early censuses.

[7]Through growth accounting, Goldin and Katz (2010) argued that the change in secondary and higher education accounted for 15% of the total growth in the twenties century.

toward zero. In the extensions of the basic results, I disentangled scarring effects from selection effects. The results suggest that selection effect after birth is tiny comparing to the scarring effect. However, the selection effect *in utero* and around birth is stronger than scarring effect.

The rest of the paper proceeds as follows: Section 4 introduces the empirical strategy; Section 5 presents the data used to realize the empirical analysis; Section 6 provides the empirical results with robustness checks and extensions; and Section 7 concludes.

2 Literature Linking Infectious Diseases to Productivity

Recent epidemiological studies suggest that infectious disease exposure in early childhood generates an inflammatory immune response that diverts nutritional resources away from physical and mental development. Severe or repeated infections can lead to long run decline in adult health or cognitive development (Finch and Crimmins, 2004, Crimmins and Finch, 2006, Eppig, Fincher, and Thornhill, 2010). Findings in the economic literature support the importance of initial health on adult productivity[8]. Particularly, emerging literature on public health intervention and the long run outcomes of childhood disease environments support the hypothesis that early exposure to disease can leave permanent scars on human capital (Bleakley, 2007, 2010; Bhalotra and Venkataramani, 2011; Beach, Ferrie, Saavedra, and Troesken, 2014; Almond, Currie, and Herrmann, 2011; Case and Paxson, 2010).

Studies also suggest that better life prospects cause individuals to invest more in human capital (Ben Porath 1967, Kalemli-Ozcan, Ryder and Weil 2000, Soares 2005, Murphy and Topel 2005). Young people with better health make greater human capital investments, simply because the stream of returns from an investment is expected to last longer (Jayachandran and Llera-Muney,2009). If lower disease exposure increases investment in human capital, it should also spur economic growth.

It is also well documented that health in early childhood has long lasting impacts on human

[8]The literature on early life influences can be divided into the in utero influence literature and the childhood influence literature. The former, which is motivated by the "fetal origins hypothesis" (Barker, 1995), was reviewed in detail by Almond and Currie (2011b), and Currie (2011). The latter is reviewed in Elo and Preston (1992) and in Almond and Currie (2011a). Both branches of literature suggests that health in early life has long term consequences on education, earnings, and future health.

capital (Almond and Currie, 2011b,a). And there is a consensus that the decline in the prevalence of infectious diseases in the U.S. was one of the most important health changes experienced by children in the early 20th century (Costa, 2013). Only a small set of papers made causal connections between public health events in this era and children's long term outcomes. The interventions examined include hookworm and malaria eradication in the southern U.S. (Bleakley, 2007, 2010) ; water chlorination in the northeast region of the U.S. (Beach, Ferrie, Saavedra, and Troesken, 2014); and the introduction of sulfa drugs after 1937 (Jayachandran, Lleras-Muney, and Smith, 2010; Bhalotra and Venkataramani, 2011. Another related paper discusses the increased access to these medical innovations among black children after the racial desegregation in the south due to the 1964 Civil Right Act (Almond and Chay, 2006)[9]).

While these studies are empowered by the exogenous events examined, they are simultaneously limited in their focus on specific geographic areas, time spans, and types of diseases. Many public health interventions, such as vaccines and antitoxins, health behavior campaigns, penicillin (1941), streptomycin (1943), erythromycin(1952), etc., may have contributed to even bigger declines in mortality, but these have not been accounted for in the literature (See Figure 1 for a illustration of the timing on some of the interventions). Yet the combination of these interventions may account for more than two-thirds of the mortality decline from infectious causes in the twentieth century (Figure 1). Using an underutilized data set from McGuire and Coelho (2011), this paper is the first to examine and compare the effects of a set of most common infectious diseases on children's future economic outcomes.

In the eyes of children born and raised during the Mortality Revolution, the health interventions operated together to create a series of positive health shocks. Because these interventions took place nearly simultaneously and targeted a similar set of diseases, there is substantive value in understanding their overall effects. Additionally, because some causes of infectious death can be misreported (e.g. typhoid is often misreported as typhus (McGuire and Coelho, 2011), and often misdiagnosed as malaria (Troesken, 2004)). Focusing on one type of disease morbidity or mortal-

[9]These researches on sulfa drugs and desegregation are also inspired by Jayachandran, Lleras-Muney, and Smith (2010); Chay, Guryan, and Mazumder (2009), which study the short run outcomes of the same events.

ity as a proxy for the effectiveness of public health interventions can lead to measurement error. I utilize a measure of the disease environment that captures multiple interventions and diseases during childhood. This overcomes issues of misreporting and misdiagnoses.

A branch of the development literature also studies the connection between public health interventions and productivity, focusing on the *immediate* impact of disease on the working population (see a summary in Weil (2014)).). Among these studies, Hansen (2014) examines the disease-GDP relationship in the U.S. He found that life expectancy in a given state and year is negatively but insignificantly associated with per capita GDP in the same year. The same result is supported by a couple of other seminal research in this field that argues the reduction of disease only leads to population explosion and no growth (Young, 2005; Acemoglu and Johnson, 2007[10]).

The findings in Hansen (2014) does not imply that the elimination of diseases have little impact on productivity in the U.S. because the paper, as well as similar studies on other countries, currently overlooks the biggest beneficiaries of the public health interventions–children.

The public health interventions had huge impact on children's health. The age profile of mortality for most infectious diseases peaks at age 0-5[11]. In 1900, children aged 0-5 accounted for 30.4% of all deaths, however, after the Second Mortality Revolution, this number fell to around 1%. Hoyert, Kochanek, and Murphy (1999) found that the decline in infectious causes of death contributed to a sharp drop in infant and child mortality, and to the 29.2-year increase in life expectancy at birth[12].

[10]Some international studies on disease and country GDP or the income of individuals living outside of U.S. found large effects of exposure to disease (Shastry and Weil, 2003; Lorentzen, McMillan, and Wacziarg, 2008), but more research supports a modest effect of disease on productivity both in the short-run or the long-run (Ashraf, Lester, and Weil, 2009; Weil, 2010; Werker, Ahuja, and Wendell, 2007Ashraf, Lester, and Weil (2009) found the effect of disease elimination on long-term growth is approximately 15% in magnitude. It even decreases per capita GDP if the population size explodes in the short-run (Young, 2005; Acemoglu and Johnson, 2007; Rodriguez and Sachs, 1999; Barro and i Martin, 1995) as a result of the Malthusian effect.

[11]Most infectious diseases have a overwhelmingly high death rate between 0-5 comparing to other ages, with the exception of typhoid fever and tuberculosis of the lungs which mostly kill young adults. It is worth mentioning that mortality of other forms of tuberculosis, unlike tuberculosis with the lungs, peaks at 0-5.

[12]This relationship is mechanical because life expectancy is a function of mortality at different ages, with special emphasis on infant mortality. For example, Cutler and Meara (2003) estimate that in the U.S., 80% of life expectancy improvements between 1900-1940 were due to reductions in death before age 45, 57% before age 15.

3 Background on the Mortality Revolution

This project will investigate the long term impacts of public health interventions associated with the "Mortality Revolution" which occurred between 1900 and 1955. During the Mortality Revolution, the adoption of several new drugs and technologies greatly reduced the prevalence of a number of infectious diseases and the deaths (Durand, 1960; Easterlin, 1995).

The general public started to realize the importance of public health service in the mid-19th century. The report, *The Sanitary Condition of the Labouring Population of New York* (1848), eventually lead to the establishment of the first public agency for health, the New York City Health Department, in 1866 (The Future of Public Health,1988). By 1900, forty states had established health departments. Following this, in the early half of the 20th century, the U.S. experienced a period noted by an unprecedented decline in infectious diseases. Mortality from infectious causes of death declined by 90% (Cutler and Meara, 2003) (Figure 1; Durand, 1960; Easterlin, 1995).

The "Mortality Revolution" is often divided into two parts. The First Mortality Revolution was between 1900 and 1936, and is associated with an almost constant decrease in mortality of 1% per year[13]. During this period, the roles of state and local public health departments expanded greatly, as did identification and treatment of individual causes of diseases (Committee for the Study of the Future of Public Health; Division of Health Care Services, 1988). Antitoxins and vaccines achieved remarkable success in lowering death from diseases such as smallpox, cholera, rabies, plague, typhoid, diphtheria, and tuberculosis[14] (CDC, 1999). Water filtration and chlorination also reduced water-borne diseases such as typhoid and cholera (Troesken, 2004; Cutler and Miller, 2005). More emphasis was put on health education and the promotion of healthy habits (Ewbank and Preston, 1990). Due to the particular vulnerability of children and pregnant mothers to infectious diseases, special funds were allocated to their care. The Children's Bureau was established

[13]The 1918 influenza epidemic briefly interrupted the decline in mortality rate, but it soon recovered back to its original path.

[14]Some vaccines and antitoxin serums were developed prior to 1900 (smallpox, cholera, rabies, and plague), and others were invented or licensed after 1900 (typhoid (1914), diphtheria (1923), tuberculosis (1927)). Some are invented after 1936 (influenza (1945), pertussis (1949), yellow fever (1953) polio (1955)). But these vaccines had relatively smaller impact on mortality compares to the earlier vaccines. The years in the parenthesis are years in which each vaccine was licensed after the Biologics Control Act of 1902.

in 1912, which assisted state health campaigns, and later provided obstetric care, as well as other types of health care, to children and mothers.

The Second Mortality Revolution occurred between 1937 and 1955, and is associated with a more rapid decline in mortality. Mortality declined by roughly 2% per year (Cutler and Meara, 2003; Hansen, 2014; Armstrong, Conn, and Pinner, 1999). Rapid public health improvements in medicine, as well as economic growth, are believed to have contributed to this phenomenon (Cutler and Miller, 2005; Ewbank and Preston, 1990; Lleras-Muney, 2002; Fogel, 1994; McKeown, 1976, etc.). The Second Mortality Revolution is also referred to as the "big medicine" period. It was initiated with a wave of drug innovations: sulfonamides (or sulfa drugs, 1935), penicillin (1941), streptomycin (1943), para-aminosalicylic acid (1944), and isoniazid (1952) (Hansen, 2014). The New Deal (1933-1936), provided funds for more federal involvement in public health. Research institutions such as the National Institutes of Health (NIH) were established to tackle more infectious diseases. A few more vaccines are also invented during this period. Such as influenza (1945), pertussis (1949), yellow fever (1953) polio (1955).

Overall, the First Mortality Revolution focused on reducing the contraction of infectious diseases; while the second focused on developing new drugs to treat symptoms and cure diseases (Connolly, Golden, and Schneider, 2012; Bhalotra and Venkataramani, 2011). As a result, an infected individual would receive better treatment and suffer from less morbidity during the Second Mortality Revolution versus the First Mortality Revolution.

4 Empirical Strategy

4.1 General Model of Disease Burden and Productivity

To examine the long run outcomes of childhood exposure to infectious diseases and individual productivity, I use synthetic panel data and a model that leverages variation at the state and year of birth-census year-gender-race level. Specifically, I estimate a variant of a difference-in-differences model that allows us to compare adult outcomes among those born in a time with high infectious disease prevalence to those born in the same state but in years with a low infectious disease preva-

lence.

The key independent variable in the model is " Predicted Childhood Exposure" to the most common infectious diseases ($PredChildExp$). It is a proxy of childhood disease exposure by cohort and state of birth[15]. The use of this variable allows me to measure the causal impact of changes in disease exposure on outcome variables of interest. Employing this variable allows me to avoid omitted variable bias, such as local economic growth, education expenditure, and other health/health care improvement. After introducing the components in the estimation equation, I will describe this measure in greater detail.

$$\begin{aligned} Y_{btcgr} &= \rho PredChildExp_{bc} + X_{btcgr}\theta \\ &\quad + \delta_{t-c} + \delta_{bt} + \delta_{rc} + \delta_{rb} + \delta_{rt} + \delta_{gc} + \delta_{gb} + \delta_{gt} + \epsilon_{bct}, \end{aligned} \tag{1}$$

where Y_{btcgr} denotes an outcome variable in adulthood for individuals of a specific gender (g), race (r), birth state (b), birth cohort (c), and observed in census year (t). $PredChildExp_{bc}$ is a proxy of childhood disease exposure for individuals who were born in state b, and year c. It also reflects the probability of catching infectious diseases. The details on how to create this variable are in the following section. I will also argue that this variable is orthogonal to the error term ϵ_{bct}. X includes cell-level control variables such as marital status, and average number of children in the household. These variables help control for labor supply choices due to marriage status and family size. In the robustness checks, I also included other state-level controls, such as: per capita income, cohort size, education expenditure per capita, and density of schools, hospitals and doctors. δ_{t-c} are age dummies, δ_{bt} is a vector of birth state by current year interactions that control for the impacts common to people from the same birth state in the current year. The vectors $\delta_{rc}, \delta_{rb}, \delta_{rt}$ represent interaction terms between race and birth cohort, birth state, and census year; and $\delta_{gc}, \delta_{gb}, \delta_{gt}$, are interaction terms between gender and birth cohort, birth state, and census year. In the some regressions, I use gender-/race-specific samples, therefore, depending on the subsample, some of the

[15]One important underlying assumption I make is that the probability of migration is low between the year of conception and age 5.

interaction terms with race or gender should be removed from the regression model. These interaction terms are $\delta_{rc}, \delta_{rb}, \delta_{rt}, \delta_{gc}, \delta_{gb}, \delta_{gt}$. ie. gender dummy and interaction terms with the gender dummy are not included in the gender specific regressions.

Equation (1) describes the basic specification I use in this study. In the robustness checks, I will also include $\delta_b c$, which is the birth state linear trend; and δ_{Rc}, which is the birth region by cohort dummy. I expect the magnitude of the coefficients on *PredChildExp* to reduce in these columns because the state trends are going to capture some of the linear declines in *PredChildExp*.

4.2 Predicted Childhood Exposure to Infectious Diseases

The difficulty in estimating the effect of disease is omitted variable bias. States with greater declines in infectious diseases may also experience greater economic growth, more education investment, or improvements in other aspects of health. In the absence of an exogenous source of variation in disease mortality in the U.S., exposure to disease might be correlated with unobserved state-level time-varying economic events, thereby biasing the coefficients.

This paper overcomes the problem of omitted variable bias by adopting and improving an identification strategy, called "predicted mortality"[16] that was proposed by Acemoglu and Johnson (2007) and refined in Hansen (2014)[17]. A variation of this method is also widely adopted in studies on public health interventions (Bleakley, 2007, 2010, and Bhalotra and Venkataramani, 2011, etc.).

Specifically I create a child specific predicted exposure to infectious diseases based on the state and year in which the child was born. Because morbidity is hard to measure, especially in historical data, mortality is commonly used to proxy for morbidity (Lleras-Muney and Glied, 2008; Almond, Currie, and Herrmann, 2011; Case and Paxson, 2010; Bhalotra and Venkataramani, 2011; Acemoglu and Johnson, 2007). In most of my specifications, I also use mortality rate to proxy for

[16]The "predicted mortality" was originally used by the development literature to better understand the extent to which cross country and sub-country differences in contemporaneous Gross Domestic Product (GDP) relate to public health improvements. Using this measure, Hansen (2014) finds that in the U.S., life expectancy in a given state in a given year is negatively but insignificantly associated with per capita GDP in the same year.

[17]The early specification used in Acemoglu and Johnson (2007) is a long-difference model that compares observations from two cross sections. This method was criticized by Aghion, Howitt, and Murtin (2011) and Bloom, Canning, and Fink (2014) for lack of initial health controls by country. This problem is resolved in both Hansen (2014) and in the current paper.

the prevalence of infectious disease. In the robustness checks, I also explore the differences across diseases that generate the same amount of mortality[18].

The "predicted childhood exposure" reflects the prevalence of multiple common infectious diseases. For each common infectious disease, the variable captures two components. The first component exploits the decline in national mortality from infectious diseases, which is mostly due to medical and drug innovations (CDC, 1999)[19]. The second component is the state's pre-intervention mortality rate. States with higher disease prevalence prior to the medical intervention should have benefited more from it, so this component provides a measure of the intensity of the treatment. For example, when sulfa drugs were invented, the states with high prevalence of pneumonia or scarlet fever exhibited more mortality decline than those without these diseases. This method is similar in spirit to a "shift-share" index, which captures exogenous labor demand shocks (Bartik, 1991; Katz and Murphy, 1992; Blanchard and Katz, 1992)[20].

Specifically,

$$\text{Predicted Annual Mortality of One Disease d : } p^d_{st} = \frac{M^d_t}{M^d_{t_0}} M^d_{s,t_0} \tag{2}$$

$$\text{Predicted Annual Mortality of Multiple Infectious Diseases : } P_{st} = \sum_{d \in D} p^d_{st} \tag{3}$$

$$\text{predicted childhood exposure : } PredChildExp_{bc} = \frac{1}{7} \sum_{i=c-1}^{c+5} P_{bi} \tag{4}$$

The first term in equation (2), $\frac{M^d_t}{M^d_{t_0}}$, represents the ratio of national mortality rates from base year to year t. M^d_t and $M^d_{t_0}$ correspond to the mortality rates from a specific disease (d) in a specific

[18]In section 6.4.1, I also translate these mortalities into a uniform measure of healthy life years lost due to disease. As I will explain in the coming sections, I do not use the loss of healthy life years as the main specification because 1) only a subset of diseases have this measure and 2) the measures are only true at the average level to the sample of population used to calculate this measure. Great caution should be used the interpretation of these results.

[19]The information on the timing of state-level introduction of each technology is not available and would not be useful because it is likely to be endogenous to economic growth of the state, among other factors.

[20]The original shift-share variable combines the magnitude of the national labor demand shocks by industry, and the initial share of each industry in a given state to capture the exogenous labor demand shocks by state.

year (t) and base year (0), respectively. The second term M^d_{s,t_0} is the base year mortality rate from disease d in state s. If the second term is higher, the state is more susceptible to the technological changes relating to disease d. Accordingly, the "predicted annual mortality rate from multiple diseases" in equation (3) is a summation of the predicted mortality rate of each disease. The set of diseases (D) I use to construct the predicted mortalities contributed significantly to child mortality and morbidity in the early 20th century (Cutler and Meara, 2003). They are the most common infectious diseases in those times (McGuire and Coelho, 2011). These diseases include smallpox, dysentery, typhoid, diphtheria, coup, tuberculosis, influenza, pertussis, measles, pneumonia, scarlet fever, and malaria[21]. The fixed base year prior to the birth of all the cohorts in my sample[22], t_0, is set at 1900,

To calculate the average childhood exposure to infectious disease, I average the annual predicted mortalities from early childhood years. I arbitrarily define early childhood to be the year of conception through age 5[23], a total of seven years, because infectious causes of death concentrate within this age group (Figure 4 shows the death from infectious diseases by age group). Therefore, the "predicted childhood exposure", or $PredChildExp_{bc}$, is an average of the mortality of seven consecutive years for each cohort and state of birth. For cohort c born in state b, the average is from year c-1 to c+5 in state b. I refer to this age span as "age -1 to 5" in the rest of the paper. Substituting the terms in equation (2) and (3) into the formula for predicted childhood exposure, we get

$$\begin{aligned} PredChildExp_{bc} &= \frac{1}{7}\sum_{i=c-1}^{c+5} P_{bi} = \frac{1}{7}\sum_{i=c-1}^{c+5}\sum_{d\in D}\frac{M^d_i}{M^d_{t_0}}M^d_{b,t_0} \\ &= \sum_{d\in D}(\frac{1}{7}\sum_{i=c-1}^{c+5} M^d_i)\frac{1}{M^d_{t_0}}M^d_{b,t_0} \end{aligned} \tag{5}$$

[21]The most common infectious causes of death also include diarrhea and fever, but those are usually undiagnosed diseases from unknown causes. Therefore, they are excluded from this exercise.

[22]I will perform robustness checks where the base year is many years earlier than the earliest sample in the regression. A similar technique is used in the constructing of a shift-share instrument to reduce correlation between the constructed shift-share variable and the error term. The method was not used in any of the previous literatures studying disease mortality (ie. Acemoglu and Johnson, 2007; Hansen, 2014; Bleakley, 2007).

[23]This is consistent with the early influence literature.

Let $\overline{M}^d_{child}$ denotes the average childhood mortality of disease d $\frac{1}{7}\sum_{i=c-1}^{c+5} M_i^d$, then

$$\begin{aligned} PredChildExp_{bc} &= \sum_{d\in D}(\frac{\overline{M}^d_{child}}{M^d_{t_0}})(M^d_{b,t_0}) \\ &= \sum_{d\in D}(\text{National Mortality Relative to } t_0 * \text{State Base Year Mortality})^d \quad (6) \end{aligned}$$

The "predicted childhood exposure" exploits both the national mortality decline brought by a broad selection of public health efforts and the approximated intensity of intervention in each state. Its first component ($\frac{\overline{M}^d_{child}}{M^d_{t_0}}$) is a ratio of state mortality in the year 1900 and the year in question. It represents the mortality that the current cohort experiences, relative to the mortality in 1900. In the year 1900, this ratio is one, and when the disease is eradicated, the ratio becomes zero. By construction, $PredChildExp_{bc}$ is unrelated to unobservable state-level time varying events. In the robustness checks, I test the validity of this measure by including additional control variables in the basic specification. Those state-year control variables include: per capita income, cohort size, population size, number of schools per square mile, education expenditure per capita, hospitals per square mile, number of doctors per capita, deaths from other major diseases, percent of urban/manufacture industry in the population, and farm area in acres. These variables are not available in all years for all states, which is why I cannot control for them in the basic specification. Instead, I use them to prove that the index I created in the basic specification do not suffer from omitted variable bias.

5 Data

5.1 Mortality

The construction of "predicted childhood exposure" requires two datasets – the state-level mortality by infectious causes in 1900 and national-level mortality from 1901 to 1960 of the same infectious deaths.

The first data set, the state-level mortality by specific causes in 1900, comes from McGuire

and Coelho (2011). This is an underutilized dataset. The authors carefully assembled these state-level mortalities from the United States census mortality data[24]. The data cover 49 states in 1900. The mortality rate from all infectious causes of death collected by the 1900 census were available through this data set. I choose to use the following 12 common infectious diseases: smallpox, dysentery, typhoid fever, diphtheria, croup, tuberculosis (all forms), pertussis, measles, scarlet fever, malaria, influenza and pneumonia. They capture the most common infectious diseases that could be diagnosed in 1900 (Armstrong, Conn, and Pinner, 1999). And because they are common, they are also well documented in the vital statistic mortality records, which allows me to observe their change throughout the years. As discussed earlier, figure 2 (a) and figure 2 (b) demonstrate the large geographic variation in this data set. Figure 2 (a) maps this geographic variations in diseases prior to the interventions. It shows the considerable variation in the 12 common infectious mortalities in 1900[25]. If we look at the each infectious cause of death, mortality from a single disease also demonstrates considerable geographic variation. For example, figure 2 (b) illustrates the state-level variation of tuberculosis mortality. In this figure, the mortality rate in California is 16 deaths per 10,000 population higher than North Dakota. Therefore, streptomycin, which is effective in treating tuberculosis, should reduce more mortality in California than in North Dakota.

Past literature that studies infectious causes of death in the historical U.S. usually used the vital statistics data set because it was more complete than the census mortality data[26]. I choose to use the data in McGuire and Coelho (2011) because the vital statistics do not provide enough information about the First Mortality Revolution. The vital statistics cover 10 registration states in 1900. The coverage increased to 25 states in 1915 and 48 states in 1936. Because the First Mortality Revolution is from 1900 to 1936, the data in McGuire and Coelho (2011) is a much better data set than the vital statistics for this research. Therefore, I use McGuire and Coelho (2011) in most of my regressions and perform robustness checks with data from vital statistics.

[24]The numbers were originally reported in United States Bureau of the Census 1902b table 19 and 1902c, table 4 and 8, in the form of number of deaths in each state by cause, gender, race, age, and sometimes citizenship.

[25]They are smallpox, dysentery, typhoid fever, diphtheria, croup, tuberculosis (all forms), pertussis, measles, scarlet fever, malaria, influenza and pneumonia

[26]This data is made available through the National Bureau of Economic Research by Grant Miller. http://www.nber.org/data/vital-statistics-deaths-historical/

The second data set, the national mortality data, is from the United States Vital Statistics, made available through the Center of Disease Control and Prevention (Grove and Hetzel, 1968). I digitized the 1901-1960 mortality by infectious causes from these historical vital statistic volumes.

Combining the two data sets allows me to create the predicted annual mortality for the following infectious diseases: smallpox, dysentery, typhoid fever, diphtheria and croup[27], tuberculosis (all forms), pertussis, measles, scarlet fever, malaria, influenza and pneumonia. These data would allow me to calculate predicted mortality (Equation (2)) for 47 states between 1900 and 1960. From these numbers, I calculate "predicted childhood exposure" (Equation (4)) between age -1 to 5 from 1901 to 1955[28].

5.2 Long-run Outcomes Data

The analysis requires information on an individual's year of birth, childhood location, and adult labor market outcomes[29]. This information is available in the Census and the American Community Survey (ACS). The ACS is the successor to the Census long-form and provides similar detailed demographic and economic information. For my analysis, I use the 1950-2010 Census and the 2010 American Community Survey. I restrict the analysis to individuals born in the U.S. between 1901 and 1955, the years of the Mortality Revolution. The individuals must have been living in the U.S. at the time of the interview and was between 30 and 60 years old. Individuals born in North/South Dakota, Alaska, Hawaii, and the District of Columbia are eliminated due to lack of disease mortality data. I drop all individuals for whom information on birth year, birth state, gender, or race is allocated.

The outcome variables cover the topic of education attainment, employment status, and earnings. The outcome variables pertaining to education attainment include indicators for having at-

[27]The term diphtheria and croup is used as a title in the International List, but as deaths from croup are really deaths from diphtheria, the term croup is now seldom seen on death certificates. (Bureau of the Census, 1922)

[28]Considering the drastic increase in pneumonia and influenza deaths during the 1918 Spanish flu pandemic, I also try to exclude this cohort from any mortality measure in the robustness checks. These results are similar to the baseline results reported in the paper, and are available upon request. For a summary description of this measure, see table 1

[29]For details, please refer to Section A1.

tended 12th grade[30], for having attended 4 years of college, and for years of education[31]. The outcome variables about work status include indicators for currently employed, annual earnings, and logged annual earnings in 2011 dollars. They also include measures for earning more than 20,000, 40,000 and 60,000 dollars. An additional outcome variable is Duncan Socioeconomic Index (SEI). The SEI is a measure of occupational status based upon the income level and education attainment associated with each occupation in 1950 (IPUMS[32]). A total of more than 300 occupations are indexed. The measure can capture more between-occupation shifts in skill requirements for jobs (Bleakley, 2010); therefore, it may be interpreted as a better measure of ability than years of education.

To create the sample for analysis, I drop all the individuals for whom education, earnings, or employment are imputed. This group of individuals would later be used to create my main regression sample. A total of 17.7 million individuals are in the main sample. All the variables are created based on the same number of individuals, except for SEI and logged earnings. Because SEI has a higher chance of being missing[33], I use the subset of people in my full sample that reported a valid SEI for this regression. Logged Earnings exclude anyone with zero earnings, therefore, it also is constructed using a subsample of the full sample with non-zero earnings. The individual data are then collapsed into cells defined by gender, race, cohort, state of birth, and year of observation. Then I calculate cell-level means of each outcome variable and control variable taking into account the personal weights. The regressions are later weighted by the total number of individuals represented by the observations in the cell.

5.3 Additional Control Variables

The state time series data I controlled in multiple robustness checks include: deaths from other major diseases, state per capita income, cohort size, population size, number of schools per square

[30]Attended 12th grade does not imply obtaining a high school degree or equivalent. Degree information is not available throughout the sample period.

[31]Years of education is approximated using a method introduced by Jaeger (1997).

[32]$https://usa.ipums.org/usa-action/variables/SEI\#description_section$

[33]If a person failed to report his/her occupation, the SEI variable is considered to be missing. This leads to a higher fraction of missing values in this variable than all of the other outcome variables.

mile, education expenditure per capita, hospitals per square mile, number of doctors per capita, percent urban/manufacture industry population, farm in acres.

The state per capita income variable is defined as state income divided by state population of all ages. The state income from 1919-1960 combines two sources. National Industrial Conference Board. Division of Industrial Economics. (1939) contains information from 1919-1928; and the Bureau of Economic Analysis provides data from 1929-1960.

The cohort size at birth (number of live births per state and year) and additional mortality controls come from multiple volumes of the Vital Statistics of United States. The vital statistics do not include every state in 1900-1936, so the sample size is smaller whenever these variables are included as controls. A set of baseline estimates, which includes the same state and years as the limited vital statistics, is labeled as "basic specification" estimates.

Data on state characteristics are provided by Adriana Lleras-Muneys[34]. The data on hospitals per square mile and number of doctors per capita were assembled from the American Medical Associations American Medical Directory. The the number of schools and education expenditure were collected from various volumes of the Biennial Survey of Education. The size of the manufacture population were from Census of Manufacture. The data on acres of farm land was reported in the Statistical Abstract of the United States for 1910, 1920, 1925, 1930, and 1940. The percent urban population was from decennial censuses. For farm acres and percent urban population, data for years in between was generated using a linear interpolation by state.

5.4 Summary Statistics

Table 1 shows the summary statistics of the main dependent and independent variables in the study. The number of observations – 37870 – indicates the total number of birth state-birth year-current year-race-gender cells. All variables are based on a same number of individuals collapsed to cell level, except the Duncan Socioeconomic Index (SEI) and the logged earnings. These two variables have 36,309 and 36225 cell-level observations, respectively.

Table 1 also shows the range and variation of the key dependent variable, predicted childhood

[34]http://www.econ.ucla.edu/alleras/research/data.html

exposure. Comparing the 1901 to 1955 cohorts, the average decline within a state is ***5.5 deaths per 1,000*** for the predicted childhood exposure. Although the First Mortality Revolution was responsible for 80 percent of these mortality reductions[35], the Second Mortality Revolution had the advantage of a faster rate of decline. Figure 5 shows the value of predicted childhood exposure in a handful of states. The state with the biggest decline in predicted childhood exposure is New Mexico, with a difference of 12.21 deaths per 1,000 state population. The sharp decline in its initial years was the result of the control of smallpox. The sudden increase around 1918 was the result of the Spanish flu pandemic. The state with the biggest decline in predicted childhood exposure is Oklahoma, with a 2.31 deaths per 1,000 decrease. The median state is California, experiencing a 4.6 deaths per 1,000 change in predicted childhood exposure. In the 1901 cohort, the state at the 20th percentile predicted childhood exposure is Oregon, having 3.4 deaths per 1,000; and the 80th percentile state is Virginia, with 7.8 deaths per 1,000. The difference is ***4.4 deaths per 1,000***.

6 Results and Discussions

6.1 Basic Results

Table 2 shows the basic results with a range of fixed effect choices. I gradually added the interaction terms on to the regression until I reach my preferred specification shown in equation (1).

The first column of table 2 starts with a basic set of birth state, census year, cohort dummies, and age dummies[36]. Column (2) replaces the birth state and census year dummies with their interaction terms, which capture variations in labor market conditions, including business cycles, competition, etc. In the preferred specification in columns (3), I add a set of interaction terms with the male dummy and a set of interaction terms with the race dummies, including white, black and all other races. The results are consistently negative in this column, showing that lower infectious disease exposure during childhood leads to better outcomes. The coefficients are significant for most of the

[35]Within the 5.5 deaths per 1,000 decline in predicted childhood exposure, the First Mortality Revolution was responsible for 4.4 deaths per 1,000 on average, while the Second Mortality Revolution was responsible for 1.1 deaths per 1,000.

[36]A common misunderstanding states that one cannot simultaneously control for cohort, census year, and age dummies. This statement is easily falsified with rigorous deduction.

outcomes, with the exception of current employment status, and $>= 4$ Years of College, although the coefficient on the latter is also close to being significant at the 10% level. The magnitude of the results becomes slightly lower than in column (2). This is due to the higher benefits of mortality decline on male and black populations. I will discuss this further in section A3.1.

To interpret the magnitude of the reduced-form estimates above, I put the effects on adult outcome per unit of predicted childhood exposure into perspective. The predicted childhood exposure is not directly comparable to real mortality changes. It is composed of two parts – the effect on adult outcome of a given childhood disease burden and the magnitude of decline of the disease burden during the Mortality Revolution. The literature using this methodology usually focus on the interpretation of the first half – units of outcome change per unit of mortality rate – a number that can be applied to other situations with known infection rates (Bleakley, 2010; Bhalotra and Venkataramani, 2011). But the second half of the parameter, the decline of disease burden over time, is more interesting in this study. Therefore I interpret the coefficients in both ways.

Table 3 summarize the interpretation of the coefficients in the preferred specification (table 2 Column (3)). Using the traditional interpretation, a change in childhood disease exposure corresponding to a move from the 80th to the 20th percentile of the pre-intervention mortality rate, which is roughly 4.4 deaths per 1000, resulted in a 4 percentage point increase in the probability of completing 12th grade, 1 percentage point increase in chance of attending 4 years of college, a 7 percent increase in annual earnings, and an improvement in the person's social-economic standing by 1%.

Comparing the 1901 cohort to the 1955 cohort, the average decline within a state is 5.5 deaths per 1,000 on the predicted childhood exposure. The first row in table 3 multiplies the average predicted mortality decline between 1901 and 1955 cohorts with the coefficients reported in the preferred specification. This number represents the improvement in the outcome variable due to the entire Mortality Revolution. In the second row, I used the census to approximate the total improvement between the 1901 and 1955 cohorts in each outcome variable[37]. The third row then

[37]These are mean differences in weighted averages of outcome variables measured when each cohort is in their 40s.

summarizes the share of the total labor market improvement that resulted from the Mortality Revolution. I obtain this number by dividing the first row by the second row.

I find large and significant improvement in high school attainment and earnings among adults who benefited from the Mortality Revolution during childhood. The effect contributes to 7% to 15% of the changes in various measure of education and earnings between the 1901 and the 1955 cohort. More specifically, comparing the 1955 cohort to the 1901 cohort, an average decline of 5.5 deaths per 1,000 population improves high school completion by 5.5 percentage points. Measured in their 40s, the high school completion rate grew from 25% of the 1901 cohort to 93% of the 1955 cohort. This would imply that the Mortality Revolution accounts for about 8% of the total increase in high school attainment. Similar calculations for years of education and earnings in 2011 dollars gives effects of 14% and 9%, respectively. In their seminal study, Goldin and Katz, 2010 argued that education reform in the 20th century contributed to 15% of the total increase in productivity. This is larger than what I find for infectious disease control, but comparable in magnitude.

6.2 Test for Omitted Variable Bias

The states with higher mortality declines may also have higher state economic growth, more education investment, or improvement in other aspects of health. I argue that the predicted childhood exposure is exogenous to these unobserved state-by-cohort level variations. To test this claim, I control for the likely sources of omitted variable bias using a limited set of years in which those controls are available. Table 4 reports these results. The results show that the measure is robust in controlling for local economic growth, education expenditure, and other health/health care improvement.

Panels A.1, B.1, C.1 and D.1 report the results using the basic specification on a limited set of cohorts (states) for which the control variables are available. They serve as a point of comparison for the other regressions in the same panel. The regression uses the same specification as column (3) in Table 2[38].

[38]The magnitude of the coefficients here are larger than those in column (3) in Table 2 because of the change in the cohorts covered by the sample. The later cohorts have a larger coefficient than the earlier cohorts. This is due to the change in methods of disease control. The details will be discussed in section 6.5.

Panel A.2 controls for local economic growth using state per capita income from 1920-1955. Ordinary Least Squares (OLS) estimates will typically be biased downward because of reverse causality and common shocks to income and health, such as local economic growth. After I control for per capita income in panel A.2, the estimates hardly deviate from the baseline results in panel A.1.

Another source of bias takes place when the states that invest in public health also invest in other public goods, such as education. Panel B.2 tests the validity of the measure against this hypothesis by controlling for state-by-year education expenditure per capita and the number of schools per square mile. The results stay similar to the baseline results in panel B.1.

Alternatively, the lower mortality from infectious disease might be the result of better quality of health care in the state, which simultaneously reduces disease mortality and improves other aspects of health. In panel B.2, I control for number of doctors per capita and number of hospitals per square mile, which are proxies for the provision of health care. In panel C.2, I also control for the change in other major causes of death, such as diabetes, circulatory diseases, cancer and tumors. Both panels provide robust estimates compared to their references.

Panel D tests the validity of using data from McGuire and Coelho (2011). The data in McGuire and Coelho (2011) are assembled from the Census Mortality Module. Some literature critiqued the incompleteness of these data (Crimmins, 1980) because the deaths used to calculate these state-level estimates are not 100% of all deaths. I substituted the base year mortality from McGuire and Coelho (2011) with the vital statistics mortality by cause. I then calculated another version of $PredChildExp$ using the same method, except that the data now cover fewer years in some states. The panel D.1 uses the McGuire and Coelho (2011) data and D.2 uses the predicted mortality from the vital statistics data. The small differences between the two rows are probably due to the missing states in the vital statistics data in early years. But the differences are mostly within one standard deviation.

6.3 Age-Specific Weights

From figure 4 shows that disease exposure at different ages poses different health threats. Exposure under age 1 is very likely to cause severe damage to health, or even mortality, while the degree of severity declines with age. To incorporate this observation into the formula of $PredChildExp$, I take the infectious mortality at age i as a measure of importance of disease environment in age i. I re-weight each age using the following formula:

$$PredChildExp_{bc} = \sum_{i=c-1}^{c+5} (P_{bi} * W_i) \tag{7}$$

where W_i is the infectious mortality in age i divided by total mortality from age -1 to 5. This weight represents how susceptible a child is to shocks at age i. The result of this exercise is in table 7 panel B. The result is consistent with the basic specification.

6.4 Scarring, Selection, and Fertility Response

Disease intervention has two off-setting effects on the health of the population, commonly referred to as scarring and selection effects. Scarring is the long-term effect of the epidemic on survivors' health, which translates into a negative health shock, or morbidity. Selection occurs when the least healthy members of a population are removed through epidemic related mortality, which translates into a positive health shock through mortality. The coefficients in previous tables captures the net effect of selection and scarring. The two effects operate in opposite directions, which means that the coefficients in the basic specification are biased toward zero.

6.4.1 Morality and Morbidity by Type of Disease

Since selection effect is a result of mortality, diseases with a higher chance of fatality should have higher selection effects. To test this hypothesis, I compare the diseases with higher mortality to those with lower mortality. Among the 12 common diseases studied in this paper, smallpox, typhoid, tuberculosis, measles, and scarlet fever have case-fatality ratios[39] of higher than 15% on

[39]Case-fatality ratio is the fatality rate among the infected population.

average, and the rest of the diseases have case-fatality ratios that range from 5% to less than 1%[40]. Table 5 shows that mortality decline in diseases with high mortalities results in lower long-term productivity gain per death prevented[41].

An alternative way to examine the discrepancies between mortality and morbidity is to use the disability weights. Disability weights are estimated by the WHO and mainly used in the calculation of Disability Adjusted Life Year (DALY). A disability weight is a weight factor that reflects the severity of the disease on a scale from 0 (perfect health) to 1 (equivalent to death). It becomes a measure of disease morbidity by translating the number of disease cases into Years Lost due to Disability (YLD). I use these disability weights to re-weight the measure of disease exposure in the following way:

$$PredChildExp_{bc} = \frac{1}{7}\sum_{i=c-1}^{c+5}\sum_{d\in D}(p_{bi}^{d} * W_d) \tag{8}$$

where W_d is the disability weight divided by the case-fatality ratio, both specific to disease d. The case-fatality ratio is taken from the CDC pink books[42]. The reason to incorporate W_d is to translate mortality into life years lost or YLD. The YLD provides a unified measure for all diseases, cohorts, and geographic regions, which is useful for interpretation of the coefficients. Dividing mortality by case-fatality ratio renders the total number of cases. Then, the number of cases is multiplied by disability weights, which produces the disease-specific life years lost. Adding different diseases together provides a YLD measure of all infectious causes for the entire state-cohort.

The WHO provides the disability weights for the following diseases: diphtheria, pertussis, tuberculosis, measles, malaria, and pneumonia. Using only these diseases, I show the estimates

[40]A case-fatality ratio of 15% in this case implies that, given the medical conditions in the early 1900s, the chance of mortality after contracting the disease is 15%. The case-fatality ratios are not very accurate because they are subject to changes in medical conditions, but the distinction between high and low fatality diseases is quite clear.

[41]The magnitude of these coefficients should be carefully interpreted. It does not directly imply that high mortality diseases have little impact on long run outcomes. For example, if the typical case-fatality ratio of a high mortality disease is 15, and of a low mortality disease is 5, then for each death prevented, the high mortality disease has a 70% lower long run effect on percent population earning higher than $20,000. Assuming that each disease infected 100 patients, then the total aggregate long-term effect of high mortality disease on the percent of population earning more than $20,000 is still more than twice the amount compared to low mortality diseases.

[42]These case-fatality ratios usually contain an upper and a lower bound; the average is used here.

from the basic specification in table 7 panel B.1. The results with the re-weights are shown in table 7 panel B.2. The coefficient can loosely be interpreted as the result of losing one life year during each year of early childhood due to disease exposure. However, there are rounding errors and measurement errors in both the case-fatality ratio and the disability weights, so these results should be taken with a grain of salt.

6.4.2 Control for Fertility Response and Selection Effect after Birth

Cohort size is intrinsically connected with the size of the selection effect. Fertility responds to mortality. The death of a child may cause replacement of the child, and expectations about future mortality may cause more total births per family (Preston, 1978; Rosenzweig and Paul Schultz, 1987; Montgomery and Cohen, 1998; Palloni and Rafalimanana, 1999; Bleakley and Lange, 2009). A recent paper by Nobles, Frankenberg, and Thomas (2014) studied the fertility repercussions of the 2004 Indian Ocean tsunami. They found that mothers who lost one or more children in the disaster were significantly more likely to bear additional children. In light of these studies, it is possible that children born in the years with sudden mortality decline would face more domestic or labor market competition, because the survival rate increased.

In table 6, I control for the selection effect and the fertility response in order to reveal the magnitude of the scarring effect. I do so by controlling for the initial cohort size and the amount of attrition within each birth state and cohort. I use the size of the cohort at birth to capture the fertility responses due to lower mortality and the survival rate of infants to adulthood to capture the selection effect after birth. To create this survival rate, we use the census data to estimate the number of people from each birth state and cohort that survived until the census year. The percent of infants who survived until the census survey year is calculated by dividing the survived population size by corresponding cohort size.

Panels 1 B and C show that the effects of childhood exposure to infectious disease, net of fertility response and selection, stayed close to the baseline results. This implies that the scarring effect dominates this period and the selection effect after birth is relatively small.

6.4.3 Age-Specific Dynamic in Scarring and Selection Effects

To study the effects of mortality shocks at different ages of exposure, I substitute the average predicted exposure between age -1 to 5 ($PredChildExp$) in equation (1) with a series of average predicted exposure between age -1 to 0, 1 to 2, 3 to 4, and 5 to 6.

$$Y_{bctgr} = \sum_{i=c-1,c+1,..c+5} \rho_i P_{bi} + X_{bc}\theta + \delta_g + \delta_r + \delta_a + \delta_{bt} + \delta_b c + \epsilon_{bct} \tag{9}$$

The coefficients ρ_i are plotted in figure 7. The results reveal that disease exposures between ages 1 to 4 have a larger negative impact than shocks after age 5. This is consistent with the age-specific characteristics of infectious disease, which mainly harms children under age 5 (figure 4).

One interesting phenomenon is the positive effects on the disease exposure from the year of conception to age 1. One likely hypothesis is that in the earliest years of childhood, all the children at the lowest end of the health distribution fail to survive. This selection effect is so large that it overpowers the scarring effect and results in a positive estimate (Rawlings, 2012; Almond, 2006; Almond, Currie, and Herrmann, 2011; Bozzoli, Deaton, and Quintana-Domeque, 2009). I test this hypothesis by incorporating infant mortality into the regression as an imperfect proxy for the intensity of selection effect *in utero* and in the year of birth. Figure 8 compares the ρ for age -1 to 0 before and after including infant mortality in the control variables. The findings confirm the hypothesis. After imperfectly controlling for the selection effects in the earliest stage of childhood, the coefficients decrease in magnitude and become negative in sign, which is consistent with the sign of the scarring effect.

6.5 Results by First and Second Mortality Revolution

In this section, I compare the effects of medical advancements and public health intervention during the First Mortality Revolution with those during the Second Mortality Revolution.

To examine the differences between the First and the Second Mortality Revolution, I estimated the labor market effect of childhood disease exposure by groups of vintage cohorts. I calculated equation (1) repeatedly, each time using data from 20 consecutive cohorts. The coefficients of

$PredChildExp_{bc}$ from all these regressions are plotted together in Figure 6. For example, the value corresponding to year 1936 is the coefficient ρ in equation (1), estimated based on 1917 to 1936 cohorts.

The figure shows how the key coefficient ρ changed over time. The effect per mortality gain is much stronger in the Second Mortality Revolution, in which people benefited from the availability of new drugs during childhood. The results are consistent across different outcomes[43].

This pattern can be explained by the types of interventions in the two periods. The First Mortality Revolution focused on reducing the contraction of infectious diseases through encouraging healthier habits and introducing chlorinated water, while the second focused on new drugs that treated the symptoms and cured the diseases (Connolly, Golden, and Schneider, 2012; Bhalotra and Venkataramani, 2011). These results suggest that each mortality prevented is accompanied by more productivity gain in the Second Mortality Revolution[44].

This finding also reconciles the discrepancy in the existing literature. Bhalotra and Venkataramani (2011) found that 1 per 1,000 mortality decline caused by sulfa drugs, which were invented during the Second Mortality Revolution, increased family income by 8 percent[45]. On the other hand, Beach, Ferrie, Saavedra, and Troesken (2014) found that a similar mortality change due to water treatment, which took place during the First Mortality Revolution, increased earnings by merely 1%[46]. These past results are consistent with the findings in this paper.

[43]To minimize the impact of the mortality spike in 1918 due to the influenza pandemic, I also provide estimates taking out the 1918 cohort using the same repeat estimation method. The coefficients are plotted in figure A3. The results are similar to those displayed in figure 6.

[44]In general, the case-fatality ratio had decreased during the early 20th century (figure A1). This implies that the difference between the two periods is even bigger than assuming that the case-fatality ratio is unchanged.

[45]Sulfa drugs are very effective in treating pneumonia. Bhalotra and Venkataramani (2011) find that sulfa drugs brought a 0.26 deaths per 1,000 decline, which leads to a 2.05 percentage point increase in probability of completing high school, and 2.11% change in family income.

[46]Beach, Ferrie, Saavedra, and Troesken (2014), however, find that the elimination of typhoid by water chlorination in 75 cities during the First Mortality Revolution brought a 1% increase in earnings and a one-month increase in education attainment for males. The elimination of typhoid resulted in a decrease of roughly 1 death per 1,000. Other water-borne diseases, such as cholera, are also affected. Beach, Ferrie, Saavedra, and Troesken (2014) use typhoid mortality to proxy for the other diseases.

6.6 By Income Distribution

Figure 9 shows an interesting effect of disease exposure on different income groups. The result is counter-intuitive. The richer group seems to suffer a larger loss from the same amount of disease exposure. In figure 9, the y-axis demonstrates results by *ex post* income distribution. Each point is an independent regression, with the outcome variable indicating the probability of falling into an earnings group. For example: the point corresponding to $20,000 is the coefficient in front of PredChildExp, with the outcome variable being earnings more than $20,000.

The effect on the percentage of population earning more than $20,000 was consistently closer to zero and insignificant compared to earning more than $40,000. If I assume that overall social mobility is low, then these results indicate that underprivileged populations, on average, benefited less from the mortality revolutions. Such a result is consistent with the findings in Cutler and Lleras-Muney (2008), where they pointed out that higher socioeconomic status groups are more likely to use newly approved drugs. If high income individuals in the early twentieth century were using more of the drugs and vaccines, they could also have had a bigger response to a unit of mortality decline.

Alternatively, this dynamic can be put in the framework of scarring and selection effect. Consider the following setting: Assume there are two groups of people; one of them is richer, has access to better medical care, and is unlikely to die from infectious diseases; the other group is poorer and these people always die as soon as being affected by disease. Assume disease infection is random, and people cannot recover fully to their original productivity after suffering from the disease. Then the richer group would have a negative change in average productivity, while the poorer group has zero change in mean productivity. A more formal version of this argument is proved mathematically in Appendix Section A2.

The same pattern is also commonly found in other specifications. Figure 6, figure 10, figure A2[47], and figure A3 all show results on the percent earning more than $20,000 and $60,000. Both the level of the coefficient and the change across the 1936 threshold were smaller in the graph

[47]Female subsample does not follow this pattern. This might be the result of household dynamics in labor supply.

corresponding to lower income categories.

The selection and scarring channel I mentioned and the SES channel mentioned in Cutler and Lleras-Muney (2008) complement each other. It is likely that both channels exist.

7 Conclusion

This paper studies the effect of infectious disease control during early childhood on adult labor market outcomes. I exploit the exogenous variation in the public health projects and drug innovations during the Mortality Revolution in the United States, an era with unmatched mortality decline due to infectious disease control. Exploiting cross-state variation in pre-intervention infectious causes of death, along with time variation arising from medical innovations during the Mortality Revolution, I create a measure of disease exposure during early childhood. The results indicate that higher disease exposure in childhood reduces adult education attainment and earnings. The Mortality Revolution reduced infectious mortality and contributed to approximately 10% of the changes in labor market outcomes between the 1900 and the 1960 cohort. These results are robust to controls of local economic development, education expenditure, and improvement in other aspects of health, etc.

These results shed light on a new source of growth during the twentieth century. The contribution by education, although tremendous (Goldin and Katz, 2010), may not be the only source of productivity gain. Health capital accumulation, especially through lower childhood disease exposure, may also be an important pathway for the drastic change in labor force productivity during this era.

This paper also finds that the effect per a unit of mortality decline is stronger in the Second Mortality Revolution (1937-1955), when medication such as penicillin and sulfa drugs were introduced. The United Nations set Millennium Development Goals to reduce mortality in developing countries. As a response, massive amounts of resources were allocated to vaccination, water and sewage treatment, and health education – the same combination used in the First Mortality Revolution. According to this study, the effect per death prevented was stronger in the Second Mortality

Revolution, plausibly due to the stronger effect of drugs on disease morbidity. Putting the worry for antibiotic resistance aside, providing effective drugs should be considered as a key strategy in achieving improved long-term economic outcomes in developing countries.

References

ACEMOGLU, D., AND S. JOHNSON (2007): "Disease and Development: The Effect of Life Expectancy on Economic Growth," *Journal of Political Economy*, 115(6), 925–985.

AGHION, P., P. HOWITT, AND F. MURTIN (2011): "The Relationship Between Health and Growth: When Lucas Meets Nelson-Phelps," *Review of Economics and Institutions*, 2(1).

ALMOND, D. (2006): "Is the 1918 Influenza Pandemic Over? LongTerm Effects of In Utero Influenza Exposure in the Post1940 U.S. Population," *Journal of Political Economy*, 114(4), pp. 672–712.

ALMOND, D., AND K. Y. CHAY (2006): "The long-run intergenerational impact of poor infant health: Evidence from cohorts born during the Civil Rights era," Mimeo, Department of Economics, University of California, Berkeley.

ALMOND, D., AND J. CURRIE (2011a): *Human Capital Development before Age Five*vol. 4 of *Handbook of Labor Economics*, chap. 15, pp. 1315–1486. Elsevier.

——— (2011b): "Killing Me Softly: The Fetal Origins Hypothesis," *Journal of Economic Perspectives*, 25(3), 153–72.

ALMOND, D., J. CURRIE, AND M. HERRMANN (2011): "From Infant to Mother: Early Disease Environment and Future Maternal Health," NBER Working Papers 17676, National Bureau of Economic Research, Inc.

ALMOND, D., AND B. MAZUMDER (2013): "Fetal Origins and Parental Responses," *Annual Review of Economics*, 5(1), 37–56.

ARMSTRONG, G., L. CONN, AND R. PINNER (1999): "Trends in infectious disease mortality in the united states during the 20th century," *JAMA*, 281(1), 61–66.

ASHRAF, Q. H., A. LESTER, AND D. N. WEIL (2009): "When Does Improving Health Raise GDP?," in *NBER Macroeconomics Annual 2008, Volume 23*, NBER Chapters, pp. 157–204. National Bureau of Economic Research, Inc.

BANERJEE, A., E. DUFLO, G. POSTEI-VINAY, AND T. WATTS (2010): "Long-run Health Impacts of Income Shocks: Wine and Phylloxera in Nineteenth-Century France," *The Review of Economics and Statistics*, 92(4), pp. 714–728.

BARKER, D. J. P. (1995): "Fetal origins of coronary heart disease," *BMJ*, 311(6998), 171–174.

BARRO, R. J., AND X. S. I MARTIN (1995): *Economic Growth.*MIT Press.

BARTIK, T. (1991): *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute for Employment Research.

BEACH, B., J. FERRIE, M. SAAVEDRA, AND W. TROESKEN (2014): "Typhoid Fever, Water Quality, and Human Capital Formation," Working Paper 20279, National Bureau of Economic Research.

BHALOTRA, S. R., AND A. VENKATARAMANI (2011): "The Captain of the Men of Death and His Shadow: Long-Run Impacts of Early Life Pneumonia Exposure," IZA Discussion Papers 6041, Institute for the Study of Labor (IZA).

BLANCHARD, O. J., AND L. F. KATZ (1992): "Regional Evolutions," *Brookings Papers on Economic Activity*, 23(1), 1–76.

BLEAKLEY, H. (2007): "Disease and Development: Evidence from Hookworm Eradication in the American South," *The Quarterly Journal of Economics*, 122(1), 73–117.

——— (2010): "Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure," *American Economic Journal: Applied Economics*, 2(2), 1–45.

BLEAKLEY, H., AND F. LANGE (2009): "Chronic Disease Burden and the Interaction of Education, Fertility, and Growth," *The Review of Economics and Statistics*, 91(1), 52–65.

BLOOM, D. E., D. CANNING, AND G. FINK (2014): "Disease and Development Revisited," *Journal of Political Economy*.

BOZZOLI, C., A. DEATON, AND C. QUINTANA-DOMEQUE (2009): "Adult height and childhood disease," *Demography*, 46(4), 647–669.

BUREAU OF THE CENSUS (1922): *Mortality Statistics 1920*. Washington, DC: Bureau of the Census, U.S. Department of Commerce.

CASE, A., AND C. PAXSON (2010): "Causes and Consequences of Early Life Health," NBER Working Papers 15637, National Bureau of Economic Research, Inc.

CDC (1999): "Achievements in Public Health, 1900-1999: Control of Infectious Diseases," *MMWR*, 48(29), pp. 621–629.

CHAY, K. Y., J. GURYAN, AND B. MAZUMDER (2009): "Birth Cohort and the Black-White Achievement Gap: The Roles of Access and Health Soon After Birth," Working Paper 15078, National Bureau of Economic Research.

COMMITTEE FOR THE STUDY OF THE FUTURE OF PUBLIC HEALTH; DIVISION OF HEALTH CARE SERVICES (1988): *The Future of Public Health*. The National Academies Press.

CONNOLLY, C., J. GOLDEN, AND B. SCHNEIDER (2012): "A Startling New Chemotherapeutic Agent: Pediatric Infectious Disease and the Introduction of Sulfonamides at Baltimores Sydenham Hospital," *Bulletin of the History of Medicine*, 86(1), pp. 66–93.

COSTA, D. (2013): "Health and the Economy in the United States, from 1750 to the Present," Working Paper 19685, National Bureau of Economic Research.

CRIMMINS, E. M. (1980): "The Completeness of 1900 Mortality Data Collected by Registration and Enumeration for Rural and Urban Parts of States: Estimates Using the Chandra Sekar-Deming Technique," *Historical Methods*, 13(3), 163, Last updated - 2013-02-23.

CRIMMINS, E. M., AND C. E. FINCH (2006): "Infection, inflammation, height, and longevity," *Proceedings of the National Academy of Sciences of the United States of America*, 103(2), 498–503.

CURRIE, J. (2011): "Inequality at Birth: Some Causes and Consequences," *American Economic Review*, 101(3), 1–22.

CUTLER, D., AND A. LLERAS-MUNEY (2008): *Making Americans Healthier: Social and Economic Policy as Health Policy.*New York: Russell Sage Foundation.

CUTLER, D., AND E. MEARA (2003): *Changes in the Age Distribution of Mortality over the 20th Century.*University of Chicago Press.

CUTLER, D., AND G. MILLER (2005): "The role of public health improvements in health advances: The twentieth-century United States," *Demography*, 42(1), 1–22.

DURAND, J. (1960): "Comment," in *Population and Economic Change in Developing Countries*, ed. by R. A. Easterlin, p. 345. Chicago,1960.

EASTERLIN, R. (1995): "Industrial revolution and mortality revolution: Two of a kind?," *Journal of Evolutionary Economics*, 5(4), 393–408.

ELO, I. T., AND S. H. PRESTON (1992): "Effects of Early-Life Conditions on Adult Mortality: A Review," *Population Index*, 58(2), pp. 186–212.

EPPIG, C., C. L. FINCHER, AND R. THORNHILL (2010): "Parasite prevalence and the worldwide distribution of cognitive ability," *Proceedings of the Royal Society B: Biological Sciences*.

EWBANK, D. C., AND S. H. PRESTON (1990): *Personal Health Behaviour and the Decline in Infant and Child Mortality: the Unites States, 1900-1930.*The Australian National University Printing Service.

FINCH, C. E., AND E. M. CRIMMINS (2004): "Inflammatory Exposure and Historical Changes in Human Life-Spans," *Sci. Aging Knowl. Environ.*, 2004(38), or17.

FOGEL, R. W. (1994): "Economic Growth, Population Theory, and Physiology: The Bearing of Long-Term Processes on the Making of Economic Policy," *The American Economic Review*, 84(3), pp. 369–395.

GOLDIN, C., AND L. KATZ (2010): *The Race Between Education and Technology*. Belknap Press for Harvard University Press.

GROVE, R. D., AND A. M. HETZEL (1968): *Vital Statistics rates in the United States 19401960*. Washington, DC: U.S. Department of Health, Education, and Welfare.

HAINES, M. R. (2001): "The Urban Mortality Transition in the United States: 1800-1940," *Annales de Demographie Historique*, 33(64).

HANSEN, C. W. (2014): "Cause of death and development in the US," *Journal of Development Economics*, 109(C), 143–153.

HOYERT, D. L., K. D. KOCHANEK, AND S. L. MURPHY (1999): "Deaths: final data for 1997.," *National vital statistics reports*, 47(19), pp. 1–108.

JAEGER, D. A. (1997): "Reconciling the Old and New Census Bureau Education Questions: Recommendations for Researchers," *Journal of Business & Economic Statistics*, 15(3), pp. 300–309.

JAYACHANDRAN, S., A. LLERAS-MUNEY, AND K. V. SMITH (2010): "Modern Medicine and the Twentieth Century Decline in Mortality: Evidence on the Impact of Sulfa Drugs," *American Economic Journal: Applied Economics*, 2(2), 118–46.

KATZ, L. F., AND K. M. MURPHY (1992): "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *The Quarterly Journal of Economics*, 107(1), 35–78.

LEBERGOTT, S. (1964): *Manpower in economic growth : the American record since 1800*. New York : McGraw-Hill, Bibliographical footnotes.

LINDEBOOM, M., F. PORTRAIT, AND G. J. VAN DEN BERG (2010): "Long-run effects on longevity of a nutritional shock early in life: The Dutch Potato famine of 18461847," *Journal of Health Economics*, 29(5), 617 – 629.

LLERAS-MUNEY, A. (2002): "Were Compulsory Attendance and Child Labor Laws Effective? An Analysis from 1915 to 1939.," *Journal of Law and Economics*, 45(2), pp. 401–435.

LLERAS-MUNEY, A., AND S. GLIED (2008): "Health Inequality, Education and Medical Innovation.," *Demography*, 45(3), 741–761.

LORENTZEN, P., J. MCMILLAN, AND R. WACZIARG (2008): "Death and development," *Journal of Economic Growth*, 13(2), 81–124.

LUCAS, A. M. (2010): "Malaria Eradication and Educational Attainment: Evidence from Paraguay and Sri Lanka," *American Economic Journal: Applied Economics*, 2(2), 46–71.

MCGUIRE, R. A., AND P. R. P. COELHO (2011): *Parasites, Pathogens, and Progress: Diseases and Economic Development.* Cambridge and London: MIT Press.

MCKEOWN, T. (1976): *The modern rise of population*. Academic Press.

MONTGOMERY, M. R., AND B. COHEN (eds.) (1998): *From Death to Birth: Mortality Decline and Reproductive Change.*Washington DC: National Academy Press.

MORTALITY RATES 1910-1920. (1923): *With Population of the Federal Censuses of 1910 and 1920 and Intercensal Estimates of Population*.

NATIONAL INDUSTRIAL CONFERENCE BOARD. DIVISION OF INDUSTRIAL ECONOMICS. (1939):. New York: National industrial conference board.

NOBLES, J., E. FRANKENBERG, AND D. THOMAS (2014): "The Effects of Mortality on Fertility: Population Dynamics after a Natural Disaster," NBER Working Papers 20448, National Bureau of Economic Research, Inc.

OLMSTEAD, A. L., AND P. W. RHODE (2007): "Not on My Farm! Resistance to Bovine Tuberculosis Eradication in the United States," *The Journal of Economic History*, 67, 768–809.

PALLONI, A., AND H. RAFALIMANANA (1999): "The Effects of Infant Mortality on Fertility Revisited: New Evidence from Latin America," *Demography*, 36(1), pp. 41–58.

PRESTON, S. H. (ed.) (1978): *The Effects of Infant and Child Mortality on Fertility*. New York: Academic Press.

RAWLINGS, S. (2012): "Scarring and Selection Effects of Epidemic Malaria on Human Capital," Economics & Management Discussion Papers em-dp2012-01, Henley Business School, Reading University.

RODRIGUEZ, F., AND J. D. SACHS (1999): " Why Do Resource-Abundant Economies Grow More Slowly?," *Journal of Economic Growth*, 4(3), 277–303.

ROSENZWEIG, M. R., AND T. PAUL SCHULTZ (1987): "Fertility and investments in human capital : Estimates of the consequence of imperfect fertility control in Malaysia," *Journal of Econometrics*, 36(1-2), 163–184.

SHASTRY, G. K., AND D. N. WEIL (2003): "How Much of Cross-Country Income Variation is Explained by Health?," *Journal of the European Economic Association*, 1(2-3), 387–396.

STEIN, A. D., H. X. BARNHART, M. HICKEY, U. RAMAKRISHNAN, D. G. SCHROEDER, AND R. MARTORELL (2003): "Prospective study of protein-energy supplementation early in life and of growth in the subsequent generation in Guatemala," *The American Journal of Clinical Nutrition*, 78(1), 162–167.

TRIVERS, R. L., AND D. E. WILLARD (1973): "Natural Selection of Parental Ability to Vary the Sex Ratio of Offspring," *Science*, 179(4068), 90–92.

TROESKEN, W. (2004): *Water, race, and disease*. MIT Press.

WALDRON, I. (1983): "Sex differences in human mortality: The role of genetic factors," *Social Science & Medicine*, 17(6), 321 – 333.

WEIL, D. N. (2010): "Endemic Diseases and African Economic Growth: Challenges and Policy Responses," *Journal of African Economies*.

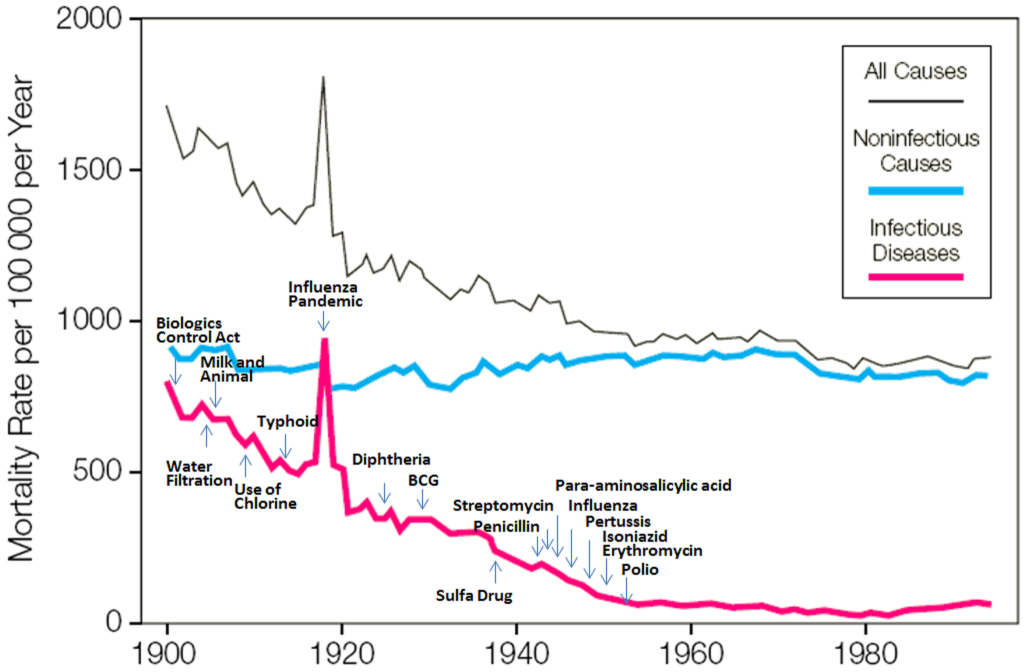
——— (2014): "Chapter 3 - Health and Economic Growth," in *Handbook of Economic Growth*, ed. by P. Aghion, and S. N. Durlauf, vol. 2 of *Handbook of Economic Growth*, pp. 623–682. Elsevier.

WERKER, E., A. AHUJA, AND B. WENDELL (2007): "Male Circumcision and AIDS: The Macroeconomic Impact of a Health Crisis," .

YOUNG, A. (2005): "The Gift of the Dying: The Tragedy of AIDS and the Welfare of Future African Generations," *The Quarterly Journal of Economics*, 120(2), 423–466.

8 Figures and Tables

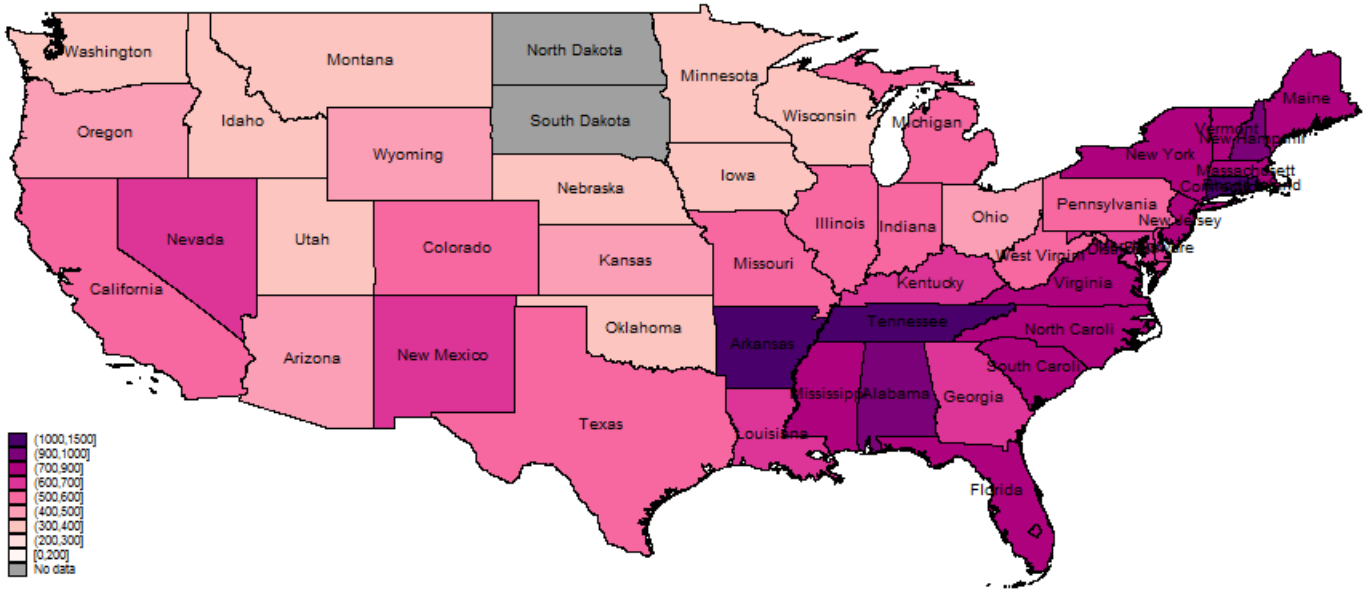
Figure 1: Crude Mortality Rates for all Causes, Noninfectious Causes and Infectious Diseases



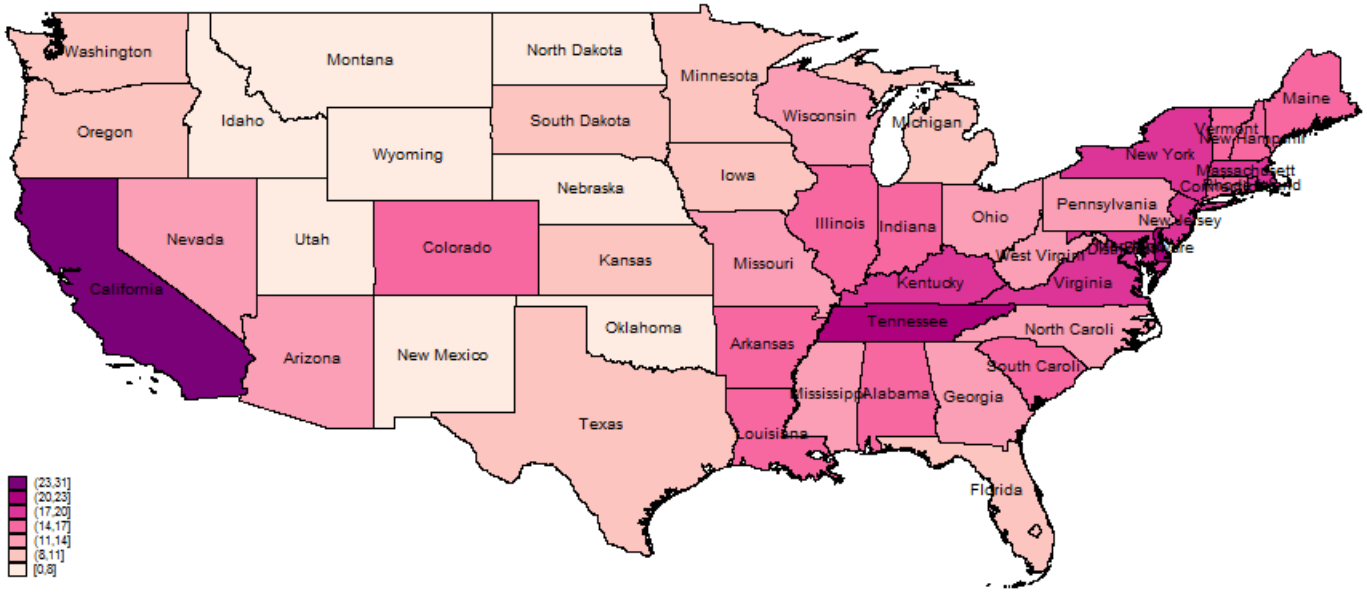
Graph adapted from Armstrong, Conn, and Pinner (1999). "Biologics Control Act" in 1902 starts the national licensing of vaccines and serums. The law was established in response to hundreds of deaths as a result of vaccination. "Milk and Animal" represents the discovery of low temperature slow pasteurization by Milton J. Rosenau in 1906 that gradually became the standard. It also represents the starting point of a series of efforts in animal control (Olmstead and Rhode, 2007). "Water Filtration" indicates the beginning of water filtration and treatment of water in cities (Beach, Ferrie, Saavedra, and Troesken, 2014). "Use of Chlorine" indicates the first continuous municipal use of chlorine in water in Boonton Reservoir in New Jersey, U.S. (American Water Works Association); "Thyphoid", "Diphtheria", "BCG", "influenza", "Pertussis", and "Polio" indicates vaccine licensing years of corresponding disease. "BCG" is the vaccine for tuberculosis, which was effective, but underutilized ("The BCG vaccine" by Neville K. Irvine). The "Influenza Pandemic" is in 1918; The mass production of Sulfa Drug is in 1937 (Bhalotra and Venkataramani, 2011); the use of "penicillin" in March 1942 (CDC, 1999); The use of a series of drugs – "Streptomycin" is in 1943; "Para-aminosalicylic" acid is in 1944; "Isoniazid" and "Erythromycin" are both in 1952;

Figure 2: Infectious Causes of Death by State, 1900

(a) Infectious Causes

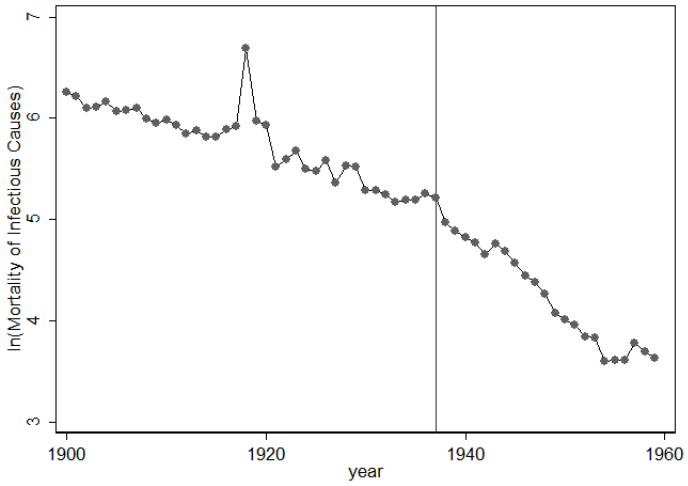


(b) Tuberculosis



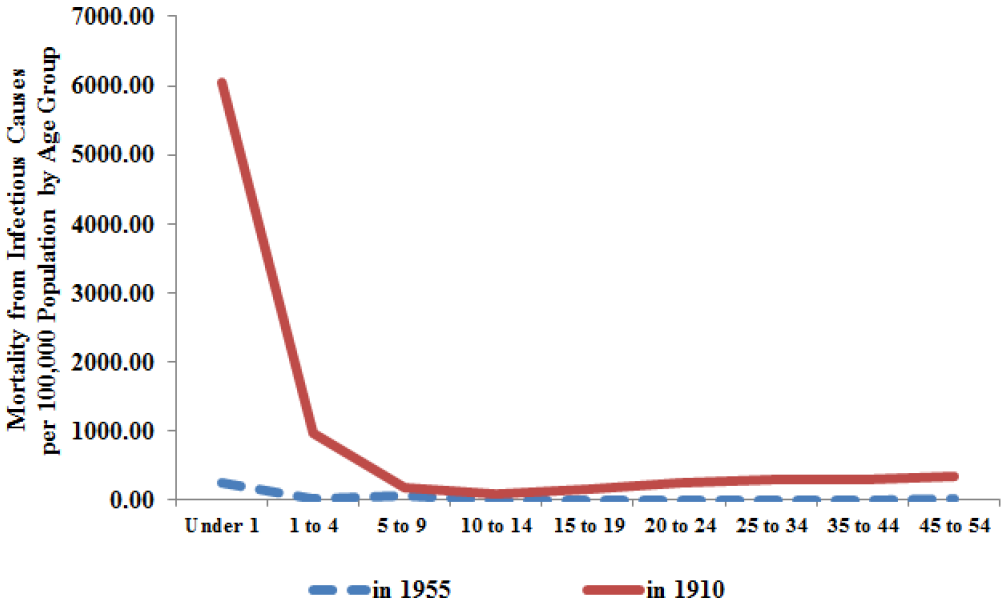
Data from McGuire and Coelho (2011)

Figure 3: Speed of Decline in Mortality Rates in the First and Second Mortality Revolution



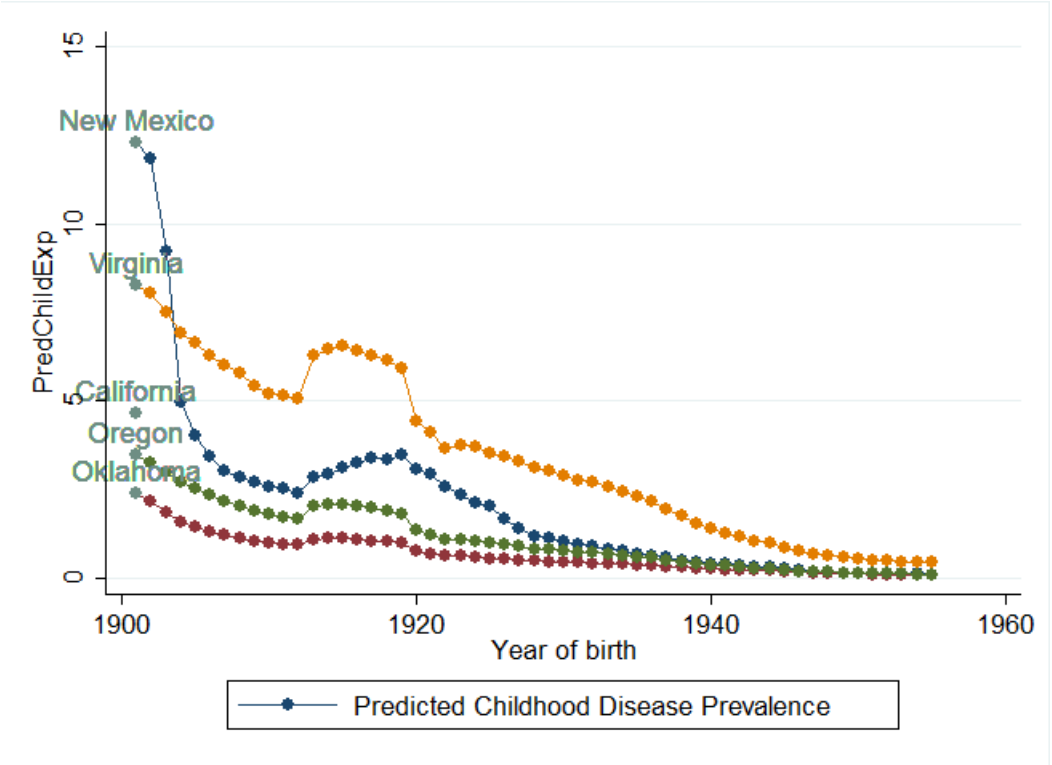
The graphs is log crude mortality rates from the 12 common infectious causes of death that are used in the analysis. Mortality rate defined as mortality per 1000 population in the U.S. The vertical line represents the year 1937. Prior to 1937 is referred to as the First Mortality Revolution, later the Second Mortality Revolution. Data from Grove and Hetzel (1968)

Figure 4: Mortality from Infectious Causes by Age



Data from *Mortality Rates 1910-1920. (1923)*, include the 12 common infectious causes of death used in the analysis.

Figure 5: predicted childhood exposure "*PredChildExp*" in a Selection of States

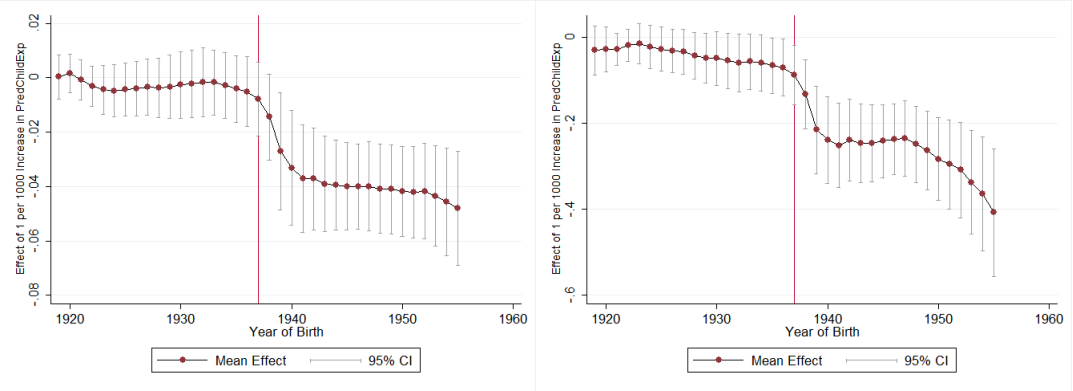


New Mexico is the state with the biggest decline in predicted childhood exposure, while the smallest is Oklahoma. And, the median state is California, experiencing a 4.6 deaths per 1,000 change in predicted childhood exposure. In the 1901 cohort, the 20th percentile state is Oregon, and the 80th percentile state is Virginia. The large decline in New Mexico is due to the national decline in smallpox mortality.

Figure 6: Coefficient using Individuals Born in Moving Cohorts Windows

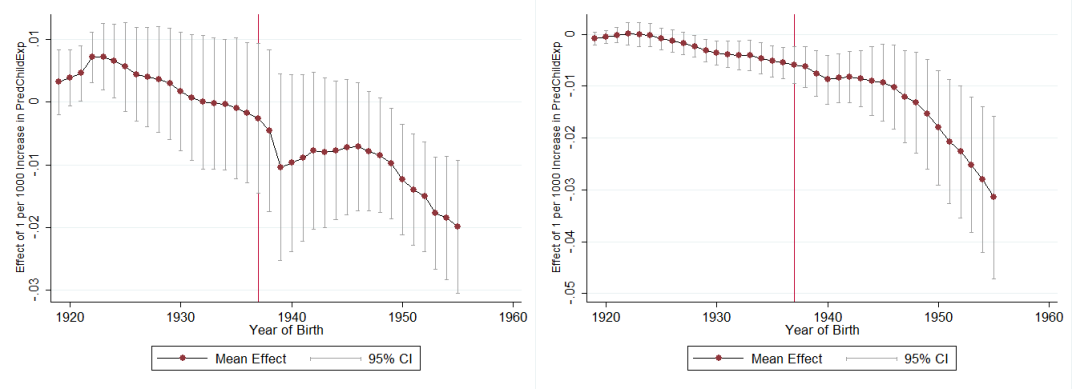
(a) HS or More

(b) Yrs of Education



(c) Earning>=20k

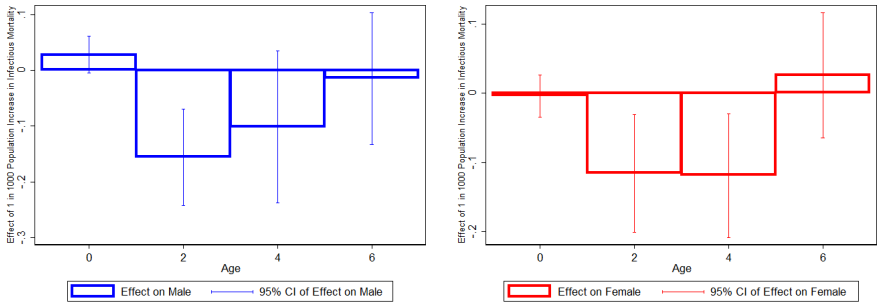
(d) Earning>=60k



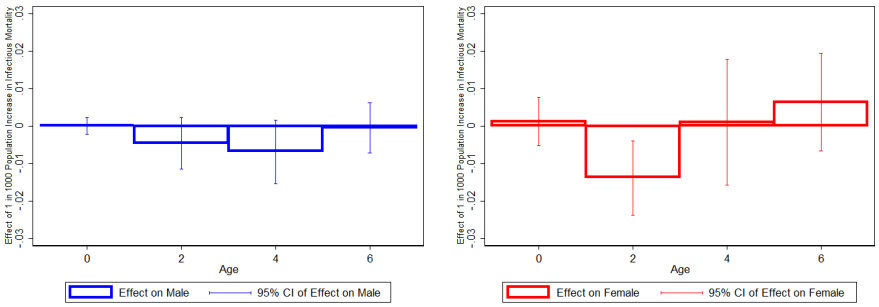
Each point on the graph represents one regression with the same control variables indicated in equation 1. The vertical axis indicates the magnitude of the coefficient of *PredChildExp* multiplies -1. The coefficient corresponding to year of birth "y" is estimated using cohorts born in "y-20" to "y". The vertical line in the graph indicates year 1937, which separates the cohorts into First Mortality Revolution cohorts and the Second Mortality Revolution cohorts.

Figure 7: Effect by Gender and Age of Exposure

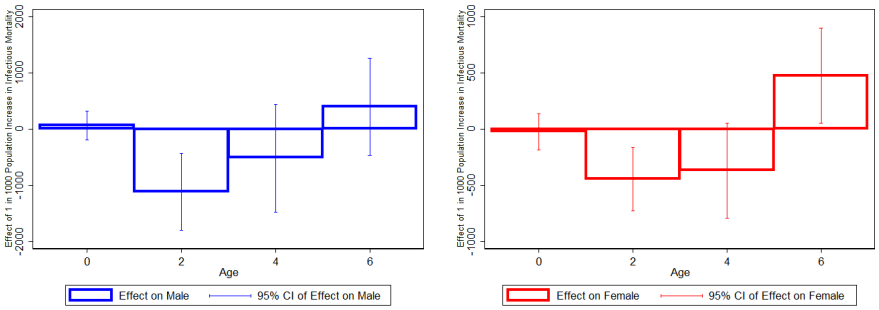
(a) Outcome Variable: Yrs of Education



(b) Outcome Variable: EMP



(c) Outcome Variable: Earnings

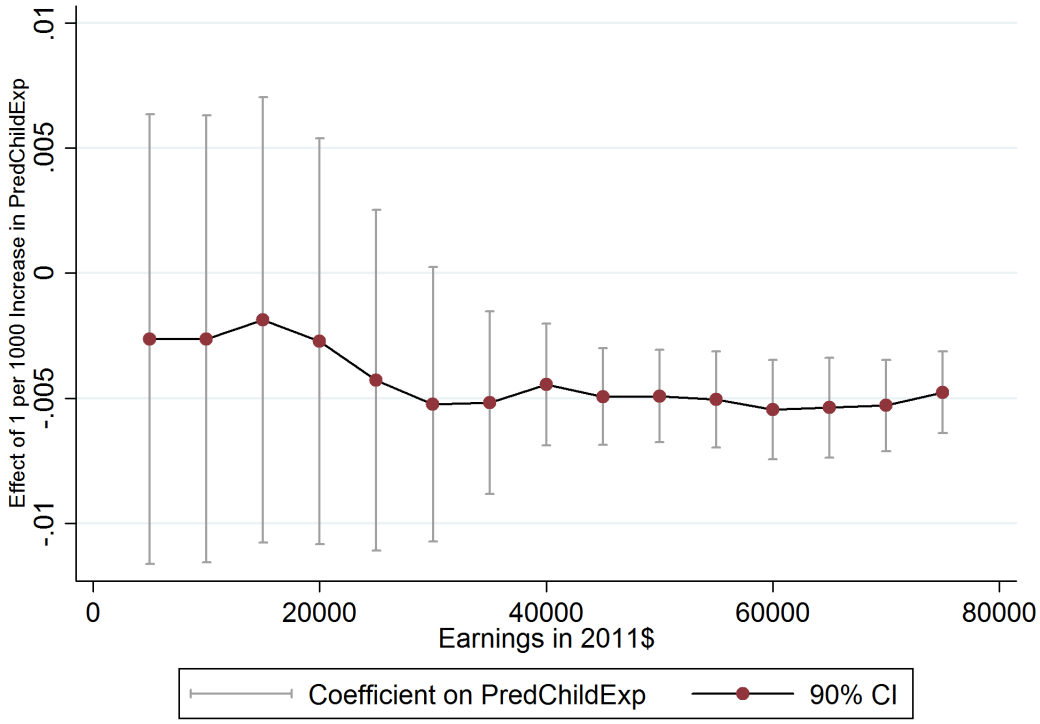


The graph plots coefficients and 95% confidence intervals of ρ_i in equation 7 from age -1 to age 6 in two year age groups. The left column are results based on the male subsample, and the right column is based on the female subsample.

Figure 8: Effect by Gender and Age of Exposure, Control for Selection during Infancy

The graph plots coefficients and 95% confidence intervals of ρ_{c-1} in equation 4. They are effects of disease exposure between age -1 to age 0. In each cell, the left column is based on the basic specification, and the right column adds an additional control of infant mortality.

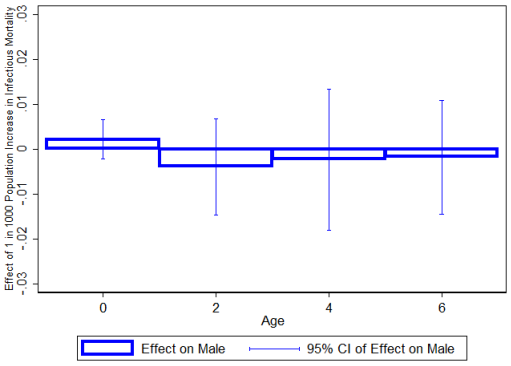
Figure 9: Result on Different Part of the Income Distribution



Each point is an independent regression. For example: the point corresponding to 20,000 is the coefficient in front of PredChildExp, with the outcome variable being earnings more than 20,000.

Figure 10: Effect by Gender and Age of Exposure, by Income Group

(a) Outcome Variable: Earning>=20k



(b) Outcome Variable: Earning>=60k

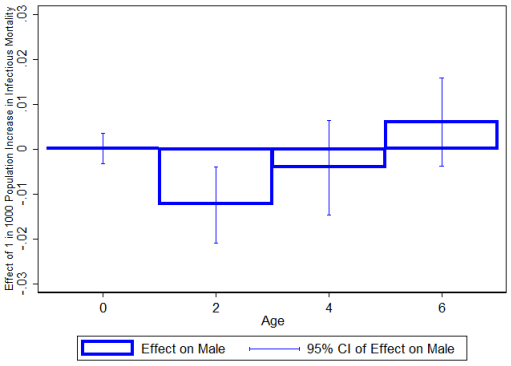


Table 1: Summary Statistics

	Mean	Std	Min	Max
Census year	1974.73	18.52	1940	2010
Year of birth	1930.15	16.14	1901	1955
Age	44.58	8.86	30	60
Fraction Male	0.49	0.50		
Fraction White	0.90	0.30		
Fraction Black	0.10	0.29		
Fraction Married	0.80	0.10		
Number of Children	1.36	0.69	0.00	9.00
Yrs of Education	11.79	1.80	0.00	21.00
Duncan Socioeconomic Index(SEI)	0.44	0.08	0.04	1.00
Annual Earnings(in 2011 dollars)	30155.47	20048.19	0.00	437400.00
Fraction Earning>=20k	0.51	0.26	0.00	1.00
Fraction Earning>=40k	0.30	0.25	0.00	1.00
Fraction Earning>=60k	0.16	0.17	0.00	1.00
ln(Annual Earnings)	10.15	0.62	2.74	13.15
Fraction Employed	0.69	0.23	0.00	1.00
PredChildExp	2.07	1.88	0.03	12.31
PredChildExp no 1918	1.97	1.76	0.03	12.31
PredChildExp >1936	0.21	0.36	0.00	2.82

Notes: Labor market outcomes are taken from the 1940-2000 Census and the 2010 American Community Survey. I eliminate individuals born in the AK, DC, HI, SD, and ND. I collapse the data into cells based on gender, white/black/other race, state of birth, year of birth, and year of observation, weighting each observation by its associated person weight. Each variable has a total of 37870 cells, the cells are also formed using the same set of individuals. The exceptions are ln(earnings) and SEI, which has 36255 and 36309 cells, respectively. The individuals forming the cells for ln(earnings) [SEI] are a subset of individuals used for the other variables who reported a non-zero wage [valid occupation]. Earnings are measured in 2011 dollars. The $PredChildExp$ is calculated from historical Volumes of Vital Statistics of the United States, and from McGuire and Coelho (2011). The details of how to calculate this measure are discussed in the text. These disease mortality measures are matched to individuals in the combined Census/ACS sample by their state and year of birth. The cells are by gender and race, so no minimum and maximum values are reported.

Table 2: Childhood Disease Exposure and the Long-run Outcomes with Different Specifications

	(1)	(2)	(3)
Panel A: Educational Outcomes			
>=12th Grade	-0.021***	-0.012**	-0.010**
	(0.006)	(0.005)	(0.005)
>=4 Yr College	0.001	-0.000	-0.003
	(0.004)	(0.003)	(0.003)
Yrs of Educ	-0.255***	-0.154***	-0.118***
	(0.050)	(0.032)	(0.027)
Panel B: Labor Market Outcomes			
Employed	0.006***	-0.007*	-0.006
	(0.002)	(0.004)	(0.004)
Earnings	-198.351	-567.626***	-569.090***
	(189.394)	(208.475)	(195.872)
ln(Earnings)	-0.048***	-0.024***	-0.016**
	(0.012)	(0.008)	(0.006)
SEI	-0.008***	-0.004***	-0.002*
	(0.002)	(0.001)	(0.001)
Birth State Dummies	Yes	No	No
Census Year Dummies	Yes	No	No
Cohort Dummies	Yes	Yes	No
Age Dummies	Yes	Yes	Yes
Birth State by Census Year	No	Yes	Yes
Interaction Terms with Race	No	No	Yes
Interaction Terms with Gender	No	No	Yes
Birth state cohort Trends	No	No	No
Birth region by cohort	No	No	No

Notes: Labor market outcomes are taken from the 1940-2000 Census and the 2010 American Community Survey. The dataset consists of individuals who were between the ages of 30 and 60 at the time they were observed. I eliminate individuals born in DC, HI, AK, ND or SD. I collapse the data into cells based on gender, white/black/other race, state of birth, year of birth, and year of observation, weighting each observation by its associated person weight. Earnings are measured in 2011 dollars. The details of how to predict the mortalities are discussed in the text. These disease mortality measures are matched to individuals in the combined Census/ACS sample by their state and year of birth. I present the results from population weighted regressions. Standard errors are in parentheses and are clustered by birth state. The interaction terms with race or gender represent three sets of interaction terms–with birth state, birth cohort, and Census Year.* significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

Table 3: Interpret Effects on Adult Outcome

Comparing Across Cohorts: 1901 and 1955

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variables	>= 12th Grade	>=4 Yrs College	Yrs of Educ	Earnings	ln(Earnings)	SEI Score
Real Differences	67.90%	24.42%	4.8	35729.97	114.01%	17.37%
Reduced-form differences	5.50%	1.65%	0.64	3112.47	8.80%	1.65%
% Effect	8%**	7%	14%***	9%***	8%**	10%*

Comparing Across States in 1901: 20/80th percentile comparison

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variables	>= 12th Grade	>=4 Yrs College	Yrs of Educ	Earnings	ln(Earnings)	SEI Score
Reduced-form differences	4%**	1%	0.51***	2489.98***	7.04%**	1.32%*

Notes: The reduced-form differences across cohorts multiplies the difference in predicted childhood exposure between 1901 and 1955 cohorts with the coefficients reported in table 2 column (3). The real differences between 1901 and 1955 cohort are calculated from the census, measured at their 40s. The % effect divides the first row by the second row of numbers. It represents the contribution of mortality revolution to the change in the outcome variable. The reduced-form differences across states with different pre-intervention infections multiplies the 20/80th percentile difference in predicted childhood exposure with the coefficients reported in table 2 column (3).

Table 4: Robustness Checks: Growth, Education, Health, and Alternative Data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	>= 12th Grade	>=4 Yrs College	Yrs of Educ	Emp	Earnings	ln(Earnings)	SEI Score
Panel A: Economic Growth, 1920 to 1955 Cohorts							
A.1: Basic Specification							
PredChildExp	-0.037***	-0.008	-0.268***	-0.013**	-1554.402***	-0.036***	-0.008***
	(0.009)	(0.006)	(0.049)	(0.006)	(247.268)	(0.007)	(0.003)
A.2: Control for State per capita Income (SI)							
PredChildExp	-0.034***	-0.009	-0.254***	-0.011**	-1567.747***	-0.035***	-0.008***
	(0.007)	(0.005)	(0.042)	(0.005)	(247.553)	(0.007)	(0.003)
Panel B: Education and Health Expenditure, 1915 to 1939 Cohorts							
B.1: Basic Specification							
PredChildExp	-0.008	-0.000	-0.091***	-0.002	-516.832**	-0.015**	-0.001
	(0.006)	(0.003)	(0.032)	(0.005)	(224.838)	(0.006)	(0.002)
B.2: Control for Educ Expenditure per Capita and Schools per Square Mile							
PredChildExp	-0.008	-0.000	-0.096***	-0.003	-545.853**	-0.016***	-0.001
	(0.006)	(0.002)	(0.033)	(0.004)	(210.371)	(0.006)	(0.002)
B.3: Control for Doctors per Capita and Hospitals per Square Mile							
PredChildExp	-0.010	0.001	-0.105***	-0.005	-639.485***	-0.020***	-0.000
	(0.007)	(0.002)	(0.035)	(0.003)	(207.523)	(0.006)	(0.002)
Panel C: Other Major Causes of Death, Available States and Years							
C.1: Basic Specification							
PredChildExp	-0.022***	-0.005	-0.179***	-0.006	-966.525***	-0.021***	-0.005***
	(0.008)	(0.004)	(0.040)	(0.006)	(265.801)	(0.007)	(0.002)
C.2:Control for Other Major Causes of Mortality							
PredChildExp	-0.022***	-0.005	-0.177***	-0.006	-959.201***	-0.021***	-0.005***
	(0.007)	(0.003)	(0.038)	(0.006)	(266.207)	(0.007)	(0.002)
Panel D: Alternative Source of Base Year Mortality							
D.1: McGuire and Coelho (2011) as the Base Year Mortality							
PredChildExp	-0.010**	-0.003	-0.117***	-0.005	-565.904***	-0.016**	-0.002*
	(0.005)	(0.003)	(0.026)	(0.004)	(193.882)	(0.006)	(0.001)
D.2: Vital Statistics as the Base Year Mortality							
PredChildExp	-0.007	-0.010**	-0.090	0.004	-240.970	0.002	-0.003
	(0.012)	(0.004)	(0.075)	(0.007)	(458.330)	(0.014)	(0.002)

Notes: Refer to footnote in Table 2. Each large panel includes a baseline regression that follows equation 1, which serves as a reference point for the other rows in the same panel. The rows after the baseline regression add additional variables to the regression, including controls for education investment, health investment at the state-year of birth level. Other major diseases include diabetes, circulatory diseases, cancer and tumors. The main columns in panel A to C(column (1)-(5)) have 27954, 17601, and 20668 observations,respectively. The regressions with ln(Earnings) [SEI score] has 26997[26988], 16802[16781], 29515[29527] observations, in panel A, B, C, respectively.

Table 5: Mortality and Morbidity : High Mortality Diseases

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	>= 12th Grade	>=4 Yrs College	Yrs of Educ	Emp	Earnings	ln(Earnings)	SEI Score
Low Mort Diseases	-0.017**	-0.003	-0.132***	-0.006	-706.121**	-0.014*	-0.003*
	(0.006)	(0.004)	(0.038)	(0.006)	(265.583)	(0.008)	(0.002)
High Mort Diseases	0.007	-0.004	-0.075	-0.002	-170.846	-0.023	0.000
	(0.008)	(0.003)	(0.062)	(0.006)	(279.997)	(0.014)	(0.002)
Obs	37870	37870	37870	37870	37870	36255	36309

Notes: The high mortality diseases include: smallpox, typhoid, tuberculosis, measles, and scarlet fever.

Table 6: Selection and Scarring, Control for Selection Effect

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	>= 12th Grade	>=4 Yrs College	Yrs of Educ	Emp	Earnings	ln(Earnings)	SEI Score
Panel A: Basic Specification							
PredChildExp	-0.029***	-0.009*	-0.214***	-0.007	-1296.998***	-0.029***	-0.009***
	(0.010)	(0.005)	(0.052)	(0.006)	(303.501)	(0.008)	(0.002)
Panel B: Scarring Effect, Control for Percent Live Birth Survived till Survey Year							
PredChildExp	-0.029***	-0.009*	-0.214***	-0.007	-1301.680***	-0.029***	-0.009***
	(0.010)	(0.005)	(0.052)	(0.005)	(303.255)	(0.008)	(0.002)
Selection	0.023	-0.026**	0.063	0.032**	1054.952	0.016	-0.005
	(0.020)	(0.010)	(0.097)	(0.012)	(781.799)	(0.023)	(0.006)
Panel C: Fertility Response, Control for Cohort size at Birth							
PredChildExp	-0.024**	-0.007	-0.187***	-0.006	-1234.841***	-0.027***	-0.007***
	(0.009)	(0.004)	(0.046)	(0.006)	(309.374)	(0.008)	(0.002)
Cohort Size (10k)	-0.003***	-0.001	-0.018***	-0.001**	-41.691	-0.001	-0.001***
	(0.001)	(0.001)	(0.005)	(0.000)	(32.542)	(0.001)	(0.000)
Obs	23343	23343	23343	23343	23343	22554	22557

Notes: Refer to footnote in Table 2. The first row panel is a baseline regression that follows equation 1, which serves as a reference point for the other regressions. The rows after the baseline regression add additional variables to the regression, including the cohort size, the number of people survived to the census year, and the survival rate. The survival rate is defined as the number of population survived to the census year within each cell, divided by the number of live births for the corresponding state and year of birth.

Table 7: Weighted predicted childhood exposure, by Age or Disease Morbidity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	>= 12th Grade	>=4 Yrs College	Yrs of Educ	Emp	Earnings	ln(Earnings)	SEI Score
Panel A: Re-weight by Age							
A.1: Basic Specification							
PredChildExp	-0.010**	-0.003	-0.117***	-0.005	-565.904***	-0.016**	-0.002*
	(0.005)	(0.003)	(0.026)	(0.004)	(193.882)	(0.006)	(0.001)
A.2: Re-weight by Age							
Weight by Age	-0.006*	-0.003	-0.081***	-0.004	-391.249***	-0.013***	-0.002*
	(0.003)	(0.002)	(0.019)	(0.003)	(129.408)	(0.004)	(0.001)
Panel B: Re-weight by Disease Morbidity							
B.1: Basic Specification							
PredChildExp	-0.014**	-0.005	-0.122***	-0.004	-519.775**	-0.009	-0.003
	(0.006)	(0.004)	(0.032)	(0.005)	(250.332)	(0.007)	(0.002)
B.2: Re-weight by Disability Weights (DW)							
Weight by DALY	-0.004***	-0.001	-0.031***	-0.002**	-139.210***	-0.003**	-0.001*
	(0.001)	(0.001)	(0.006)	(0.001)	(41.839)	(0.001)	(0.000)

Notes: Refer to footnote in Table 2. Panel A.1 and B.1 use basic specifications. Panel A.2 and B.2 apply the re-weighting methods introduced in equation 7 and equation 8

A1 Data Appendix

The outcomes data were taken from multiple United States Census Microdata samples for 1940 to 2000, and American Community Survey in 2010. These data are publicly available via the Integrated Public Use Microdata Series USA project (IPUMS). These data sets include: 1940 1% sample ; 1950 1% sample; 1960 1% sample; 1970 1% state samples and 1% metro samples[48]; 1980 1% Labor Market Areas, 1% urban, 1% metro and 5% state samples; 1990 1% metro,1% unweighted, 0.5% Labor Market Areas, and 5% sample; 2000 5%, 1%, and 1% unweighted sample; 2010 1% ACS [49].

[48]Note, this is not a sample only of metro areas

[49]According to the "Selection of the Public-Use-Microdata Samples" section of multiple years of Sample Design page on IPUMS, the procedure for selecting the microdata samples was designed to, "minimize the likelihood that any one case would be selected into more than one public-use microdata sample, and the overlap among the samples may be considered negligible." So combining the samples should not be problematic. However, to properly combine the data, one has to adjust the weights in each year according to the amount of samples used in each year.

We took the following outcome and control variables from the IPUMS:

The basic variables include: individuals‘ race, gender, year of birth, state of birth, and current state of residence. Using these variables, I restrict the analysis to individuals born in the U.S. between 1901 and 1955. The individuals must be living in the U.S. at the time of the interview, and be between 30 and 60 years old. Individuals born in North/South Dakota, Alaska, Hawaii and the District of Columbia are eliminated due to lack of disease mortality data. I drop all individuals for whom information on birth year, birth state, gender, or race is allocated. I also drop all the individuals for whom education (EDUCD in IPUMS), wage (INCWAGE in IPUMS) or employment (EMPSTAT in IPUMS) are imputed.

The individual data are then collapsed into cells defined by gender, race, cohort, state of birth, and year of observation. Then I calculate cell-level means of each outcome variable and control variable taking into account the adjusted personal weights. The regressions are later weighted by the total number of individuals represented by the observations in the cell.

Specific sources and construction of each of our outcome variables is as follows. Inside the parenthesis are the corresponding variable names :

>=12th grade (EDUCD) Attended 12th grade = 1 for those individuals who completed grade 12 and above. It does not imply obtaining a high school degree or equivalent. Degree information are not available throughout the sample period.

>=4 Years of College (EDUCD) Attended 4 years of college = 1 for those individuals who completed 4 years of college. It also does not imply obtaining a bachelor's degree.

Years of education (EDUCD) Years of education is approximated using a method introduced by David Jaeger (Jaeger, 1997).

Earnings (INCWAGE) INCWAGE reports each respondent's total pre-tax wage and salary income - that is, money received as an employee - for the previous year. This measure is inflated to 2011 dollars and then taken logarithm. We calculate earnings in categories, earnings dollars, and logged earnings using this raw variable. The former two do not eliminate those with zero earnings.

Employed (EMPSTAT) Individual employment = 1 if the individual report to be employment.

Marriage Status (MARST) Married=1 if the individual report to be married. Cohabitation status is not surveyed in the censuses.

Number of Children (NCHILD) The number of own children (of any age or marital status) residing with each individual. This variable is capped at 9.

Duncan Socioeconomic Index (SEI) The SEI is a measure of occupational status based upon the income level and educational attainment associated with each occupation in 1950.

All the variables are created based on the same number of individuals, except for SEI. Because SEI have a higher chance of being missing. When a person failed to report their occupation, the SEI variable is set to be missing. This leads to a higher fraction of missing values in this variable than all of the other outcome variables. I use the subset of people in the full sample that reported a valid SEI for this regression.

A2 Model of Scarring and Selection Effect

The model introduced in this section explains the dynamics of the scarring and selection effect on population under different initial health endowments. This model is an extension of the model introduced in Almond (2006) and Bozzoli, Deaton, and Quintana-Domeque (2009).

Let $h_{i,t}$ denote the physiological characteristics, or health, of child i at the beginning of year t. Infant i is born with physiological characteristics $h_{i,0}$. Child i dies in period t if

$$h_{i,t} - v_t \leq z,$$

where v_t is the shock size and z is the mortality threshold. Let F_t denote the distribution of health at the beginning of year t. The mortality at year t is

$$m_t = F_t(z + v_t).$$

After the shock, the health of the survivors recovers by $(1-\theta)v_t$, where $\theta \in [0,1]$. In other words, the permanent effect of the shock on the survivors' health is θv_t. The average health at the end of year t or the beginning of year $t+1$ is

$$\bar{h}_{t+1} = \frac{\int_{z+v_t}^{\infty} h dF_t(h)}{1 - F_t(z+v_t)} - \theta v_t. \tag{10}$$

The partial derivative of $\bar{h}_{t+1}$ over v_t is

$$\begin{aligned}
\frac{\partial \bar{h}_{t+1}}{\partial v_t} &= \frac{-(z+v_t)f_t(z+v_t)}{1-F_t(z+v_t)} - \frac{\int_{z+v_t}^{\infty} h dF_t(h)}{(1-F_t(z+v_t))^2}(-f_t(z+v_t)) - \theta \\
&= \frac{f_t(z+v_t)}{1-F_t(z+v_t)}\left(-z - v_t + \frac{\int_{z+v_t}^{\infty} h dF_t(h)}{(1-F_t(z+v_t)}\right) - \theta \\
&= \frac{f_t(z+v_t)}{1-F_t(z+v_t)}\left(-z - v_t + \bar{h}_t + \theta v_t\right) - \theta.
\end{aligned}$$

Given that the shock size is unobservable, it is useful to rewrite the average health formula in (10)

in terms of mortality

$$\bar{h}_{t+1} = \frac{\int_{F^{-1}(m_t)}^{\infty} h dF_t(h)}{1 - m_t} - \theta(F^{-1}(m_t) - z). \tag{11}$$

Let Let $\Phi_{\mu,\sigma}(\cdot)$ and $\phi_{\mu,\sigma}(\cdot)$ denote the CDF and PDF of a normal distribution with mean μ and standard deviation σ. When $F_t = \Phi_{\mu,\sigma}$,

$$\begin{aligned}
\bar{h}_{t+1} &= \frac{\int_{\Phi_{\mu,\sigma}^{-1}(m_t)}^{\infty} h\phi_{\mu,\sigma} dh}{1 - m_t} - \theta(\Phi_{\mu,\sigma}^{-1}(m_t) - z). \\
&= \frac{\int_{\Phi_{\mu,\sigma}^{-1}(m_t)}^{\infty} h\phi_{\mu,\sigma} dh}{1 - m_t} - \theta(\Phi_{0,\sigma}^{-1}(m_t) + \mu - z) \\
&= \frac{\int_{\Phi_{0,\sigma}^{-1}(m_t)}^{\infty} h\phi_{0,\sigma} dh}{1 - m_t} + \mu - \theta(\Phi_{0,\sigma}^{-1}(m_t) + \mu - z) \\
&= \sigma\frac{\int_{\Phi_{0,1}^{-1}(m_t)}^{\infty} h\phi_{0,1} dh}{1 - m_t} + \mu - \theta(\sigma\Phi_{0,1}^{-1}(m_t) + \mu - z) \\
&= \sigma\frac{\int_{\Phi_{0,1}^{-1}(m_t)}^{\infty} h\phi_{0,1} dh}{1 - m_t} + \mu - \theta(\sigma\Phi_{0,1}^{-1}(m_t) + \mu - z) \\
&= \frac{\sigma}{1 - m_t}\phi_{0,1}(\Phi_{0,1}^{-1}(m_t)) + \mu - \theta(\sigma\Phi_{0,1}^{-1}(m_t) + \mu - z),
\end{aligned} \tag{12}$$

where

$$m_t = \Phi_{0,1}\left(\frac{z + v_t - \mu}{\sigma}\right). \tag{13}$$

Substituting the mortality formula (13) in (12),

$$\begin{aligned}
\bar{h}_{t+1} &= \frac{\sigma}{1 - m_t}\phi_{0,1}(\Phi_{0,1}^{-1}(m_t)) + \mu - \theta(\sigma\Phi_{0,1}^{-1}(m_t) + \mu - z) \\
&= \frac{\sigma}{1 - \Phi_{0,1}\left(\frac{z+v_t-\mu}{\sigma}\right)}\phi_{0,1}\left(\frac{z + v_t - \mu}{\sigma}\right) + \mu - \theta((z + v_t - \mu) + \mu - z) \\
&= \frac{\sigma}{1 - \Phi_{0,1}\left(\frac{z+v_t-\mu}{\sigma}\right)}\phi_{0,1}\left(\frac{z + v_t - \mu}{\sigma}\right) + \mu - \theta v_t
\end{aligned}$$

We are interested in the effect of μ on $\bar{h}_{t+1} - \bar{h}_t$. An initial distribution F_0 with a high μ can be interpreted as the health or income distribution of the sub-population of infants born in high

income families. Similarly, F_0 with a low μ can be interpreted as the health or income distribution of the sub-population of infants born in low income families. The relevant partial derivative is

$$\begin{aligned}\frac{\partial \bar{h}_{t+1}}{\partial \mu} &= 1+\frac{-1}{\sigma}\sigma\left[\left(\frac{\phi_{0,1}\left(\frac{z+v_t-\mu}{\sigma}\right)\phi_{0,1}\left(\frac{z+v_t-\mu}{\sigma}\right)}{\left(1-\Phi_{0,1}\left(\frac{z+v_t-\mu}{\sigma}\right)\right)^2}\right)+\frac{-\left(\frac{z+v_t-\mu}{\sigma}\right)\phi_{0,1}\left(\frac{z+v_t-\mu}{\sigma}\right)}{1-\Phi_{0,1}\left(\frac{z+v_t-\mu}{\sigma}\right)}\right] \\ &= 1-\left[\frac{\phi_{0,1}^2\left(\frac{z+v_t-\mu}{\sigma}\right)}{\left(1-\Phi_{0,1}\left(\frac{z+v_t-\mu}{\sigma}\right)\right)^2}+\frac{-\left(\frac{z+v_t-\mu}{\sigma}\right)\phi_{0,1}\left(\frac{z+v_t-\mu}{\sigma}\right)}{1-\Phi_{0,1}\left(\frac{z+v_t-\mu}{\sigma}\right)}\right] \qquad (14)\end{aligned}$$

Net we prove that the second term in (14) is between 0 and 1, *i.e.,*

$$0<g(x):=\frac{\phi_{0,1}^2(x)}{\left(1-\Phi_{0,1}(x)\right)^2}+\frac{-x\phi_{0,1}(x)}{1-\Phi_{0,1}(x)}<1, \quad \forall x.$$

The first step is to show that it is monotonically increasing in x.

$$\begin{aligned}\frac{\partial g(x)}{\partial x} &= \frac{2\phi_{0,1}^2(x)(-x)\left(1-\Phi_{0,1}(x)\right)^2+\phi_{0,1}^3(x)\,2\left(1-\Phi_{0,1}(x)\right)}{\left(1-\Phi_{0,1}(x)\right)^4} \\ &\quad -\frac{\phi_{0,1}(x)(1-x^2)(1-\Phi_{0,1}(x))+\phi_{0,1}^2(x)x}{\left(1-\Phi_{0,1}(x)\right)^2} \\ &= -\frac{\phi_{0,1}(x)(1-x^2)(1-\Phi_{0,1}(x))+3\phi_{0,1}^2(x)x}{\left(1-\Phi_{0,1}(x)\right)^2}+\frac{2\phi_{0,1}^3(x)}{\left(1-\Phi_{0,1}(x)\right)^3} \\ &= \phi_{0,1}(x)\frac{x^2(1-\Phi_{0,1}(x))^2-3\phi_{0,1}(x)x(1-\Phi_{0,1}(x))+2\phi_{0,1}^2-(1-\Phi_{0,1}(x))^2}{\left(1-\Phi_{0,1}(x)\right)^3},\end{aligned}$$

which is positive for $x\leq -1$. For $x>-1$, we have

$$\begin{aligned}\frac{\partial g(x)}{\partial x} &= \phi_{0,1}(x)\frac{\left[2\phi_{0,1}-x(1-\Phi_{0,1}(x))\right]\left[\phi_{0,1}(x)-x(1-\Phi_{0,1}(x))\right]-(1-\Phi_{0,1}(x))^2}{\left(1-\Phi_{0,1}(x)\right)^3} \\ &= \phi_{0,1}(x)\frac{\left[2\phi_{0,1}-x(1-\Phi_{0,1}(x))\right]\left[\phi_{0,1}(x)-x(1-\Phi_{0,1}(x))\right]-(1-\Phi_{0,1}(x))^2}{\left(1-\Phi_{0,1}(x)\right)^3} \\ &= \phi_{0,1}(x)\frac{\phi_{0,1}(x)\left[\phi_{0,1}(x)-x(1-\Phi_{0,1}(x))\right]+\left[\phi_{0,1}-x(1-\Phi_{0,1}(x))\right]^2-(1-\Phi_{0,1}(x))^2}{\left(1-\Phi_{0,1}(x)\right)^3} \\ &= \phi_{0,1}(x)\frac{\phi_{0,1}(x)\left[\phi_{0,1}(x)-x(1-\Phi_{0,1}(x))\right]+\left[\phi_{0,1}-x(1-\Phi_{0,1}(x))\right]^2-(1-\Phi_{0,1}(x))^2}{\left(1-\Phi_{0,1}(x)\right)^3} \qquad (15)\end{aligned}$$

To prove that (15) is positive for $x > -1$, it is sufficient to prove that

$$\phi_{0,1} - x(1 - \Phi_{0,1}(x)) > 1 - \Phi_{0,1}(x).$$

We have

$$\begin{aligned}
\frac{\partial(\phi_{0,1} - x(1 - \Phi_{0,1}(x)))}{\partial x} &= -x\phi_{0,1}(x) - (1 - \Phi_{0,1}(x)) + x\phi_{0,1}(x) = -(1 - \Phi_{0,1}(x)), \\
\frac{\partial(1 - \Phi_{0,1}(x))}{\partial x} &= -\phi_{0,1}(x). \qquad (16)
\end{aligned}$$

Thus both $\phi_{0,1} - x(1 - \Phi_{0,1}(x))$ and $1 - \Phi_{0,1}(x)$ are decreasing in x. For $x > 1$, $\phi_{0,1}(x) \geq x(1 - \Phi_{0,1}(x)) \geq (1 - \Phi_{0,1}(x))$, thus $1 - \Phi_{0,1}(x)$ decreases faster than $\phi_{0,1} - x(1 - \Phi_{0,1}(x))$. For $0 < x < 1$, $\phi_{0,1}(x) > (1 - \Phi_{0,1}(x))$ because $\phi_{0,1}(1) > (1 - \Phi_{0,1}(0))$. The second step is to show that

$$\lim_{x \to \infty} g(x) = 1,$$

for which we use the L'hospital rule

$$\begin{aligned}
\lim_{x \to \infty} g(x) &= \lim_{x \to \infty} \frac{\phi_{0,1}^2(x) - x\phi_{0,1}(x)(1 - \Phi_{0,1}(x))}{(1 - \Phi_{0,1}(x))^2} \\
&= \lim_{x \to \infty} \frac{-2x\phi_{0,1}^2(x) + x\phi_{0,1}^2(x) - (1 - \Phi_{0,1}(x))(\phi_{0,1}(x) - x^2\phi_{0,1}(x))}{-2(1 - \Phi_{0,1}(x))\phi_{0,1}(x)} \\
&= \lim_{x \to \infty} \frac{-2x\phi_{0,1}(x) + x\phi_{0,1}(x) - (1 - \Phi_{0,1}(x))(1 - x^2)}{-2(1 - \Phi_{0,1}(x))} \\
&= \frac{1}{2} + \lim_{x \to \infty} \frac{-x\phi_{0,1}(x) + (1 - \Phi_{0,1}(x))x^2}{-2(1 - \Phi_{0,1}(x))} \\
&= \frac{1}{2} + \lim_{x \to \infty} \frac{-\phi_{0,1}(x)(1 - x^2) - \phi_{0,1}(x)x^2 - (\Phi_{0,1}(x) - 1)2x}{2\phi_{0,1}(x)} \\
&= \frac{1}{2} - \frac{1}{2} + \lim_{x \to \infty} \frac{-(\Phi_{0,1}(x) - 1)2x}{2\phi_{0,1}(x)} \\
&= 1.
\end{aligned}$$

Thus we have $g(x) \in [0, 1]$ and

$$0 \leq \frac{\partial \bar{h}_{t+1}}{\partial \mu} \leq 1,$$

which can be interpreted as that children from high income families have higher income than those from low income families after the health shock, but not as much as before the health shock.

A3 Result Appendix

A3.1 Gender, Race, and Region

We also examine the effects by gender and/or by race. Table A1 shows the result by gender and race. The education attainment effects on blacks are larger than whites, but their income effects are similar. The story becomes even clearer in table A2, which shows that, on average, children born in southern and non-southern states enjoyed comparable productivity gains from a unit of mortality decline. But the black population in southern states do not benefit as much from the Mortality Revolution compared with the blacks outside of southern states. A similar result was found by Bhalotra and Venkataramani (2011)[50]. They thoroughly examine the historical background faced by the blacks in slavery states and find that the blacks born outside of the South enjoyed similar or larger gains from sulfa drugs access than the whites, whereas the Southern blacks had limited opportunity to translate their physical gains into labor market gains. The authors argued that the likely explanation is the difficulty for blacks to realize human capital improvements in the pre-Civil Right era.

Figure A2 looks at the results in figure 6 by gender. The returns on mortality decline trend similarly between males and females during the First Mortality Revolution. During the Second Mortality Revolution, the return for both genders grew, but the return on males climbs faster for almost all outcome variables. This finding is common in the early childhood shocks literature (Almond, 2006; Bhalotra and Venkataramani, 2011; Stein, Barnhart, Hickey, Ramakrishnan, Schroeder, and Martorell, 2003; Crimmins and Finch, 2006; Lindeboom, Portrait, and van den Berg, 2010; Banerjee, Duflo, Postei-Vinay, and Watts, 2010; Almond and Mazumder, 2013). The biomedical literature also suggests males are more susceptible to environment shocks, particularly infectious diseases, compared with females (Trivers and Willard, 1973; Waldron, 1983; Lucas, 2010). Therefore the effective drugs in the Second Mortality Revolution helped the boys more.

[50]Diffusion of any medical innovation was likely slower among blacks, due to the inferior medical care available to them under segregation (Douglas Almond, Kenneth Y. Chay, and Michael Greenstone 2006, Loudon 1992). But the coefficient reflects the response of blacks to an average state-level decline in mortality, which might not be the perfect measure for the mortality decline among blacks.

A3.2 Including State Time Trends and Region by Cohort Interaction Terms

Table A4 add additional birth state-specific cohort trends and birth region by cohort dummies to the regression. The effect of adding the state-specific cohort trends reduced the size of the coefficients by approximately half, and earnings related variables become insignificantly different from zero. However, one must consider the trade-off of adding birth state-specific cohort trends. It helps to control for the state-specific trend in economic growth. However, because the mortality is also declining linearly (Figure 3), the birth state-specific cohort trends may lead to underestimation of the effect of mortality decline. I believe birth state by census year dummies can sufficiently control for the changes in average income across time, and it is better not to include birth state-specific cohort trends.

The table also include one way of controlling for mean reversion. The result is within one standard deviation comparing to the baseline results. But the significance of most of the coefficients were lost.

A3.3 Urban Penalty

Prior to 1940, the urban mortality rates were much higher than rural rates because infectious diseases spread more rapidly in highly populated places. This phenomenon is also called the "Urban Penalty" (Haines, 2001; Cutler and Miller, 2005). Table A5 studies the differences in outcomes of mortality reduction in urban versus rural areas. The table introduce %urban population by birth state and cohort, and its interaction term with the $PredChildExp$ into the basic specification in equation 1. The coefficient on the interaction term captures the differences between 100% urban population versus none. Results show that urban areas benefit more from mortality decline. For a 1-in-1000 childhood mortality reduction, 1.6 percentage point more children finish 12th grade, and 1.9 percentage point more finish four years of college. The workers' socioeconomic standing also improves. I performed similar analysis with the percent of manufacturing population and farm land in acres. The results are similar to what we see here with the percent of urban population.

A4 Appendix Tables and Graphs

Figure A1: Case-Fatality Ratio for a Selection of Diseases

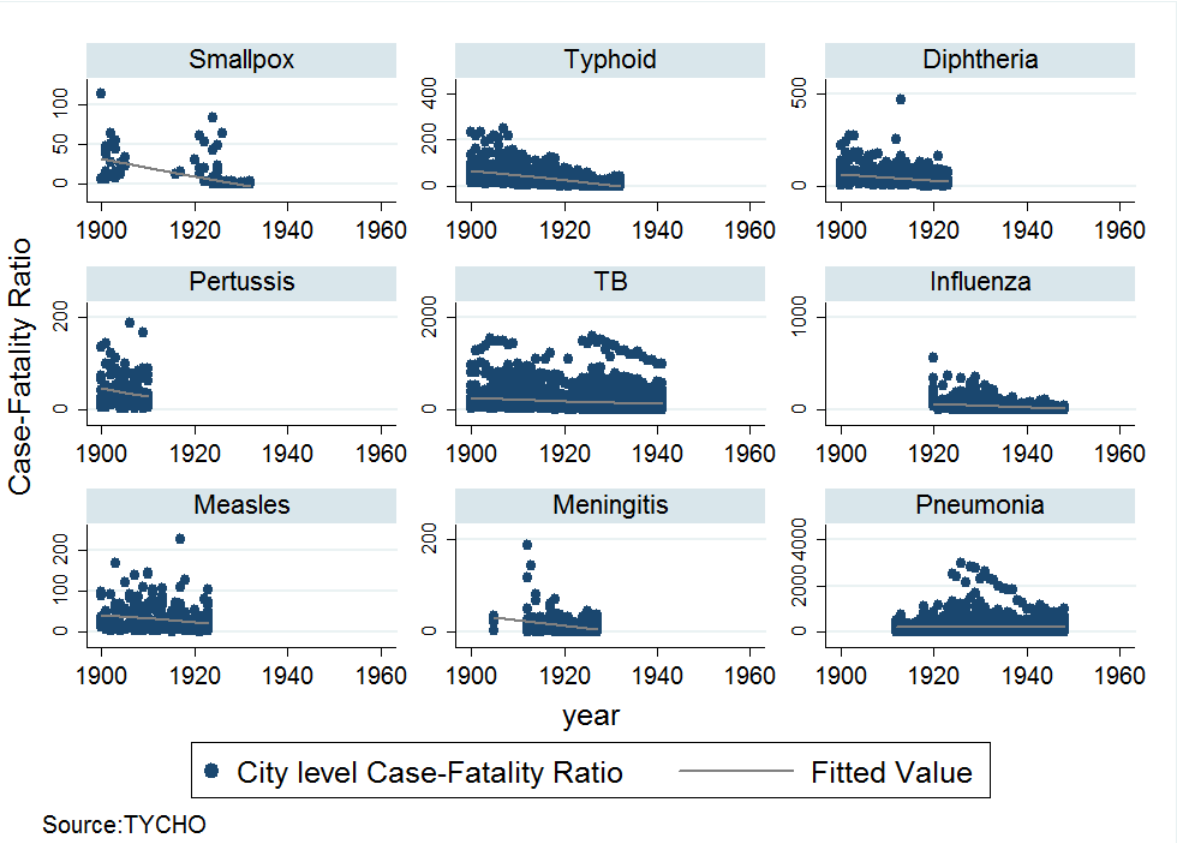
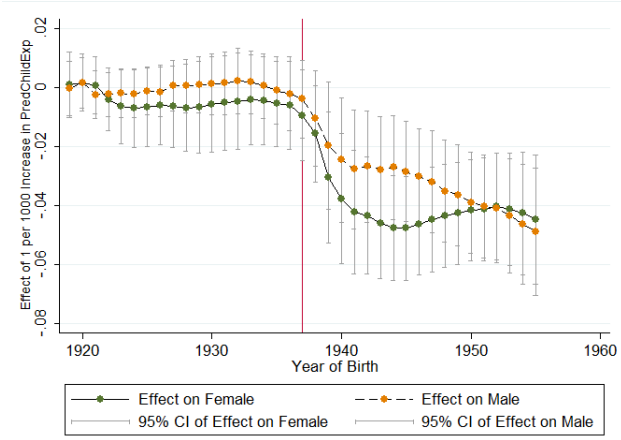
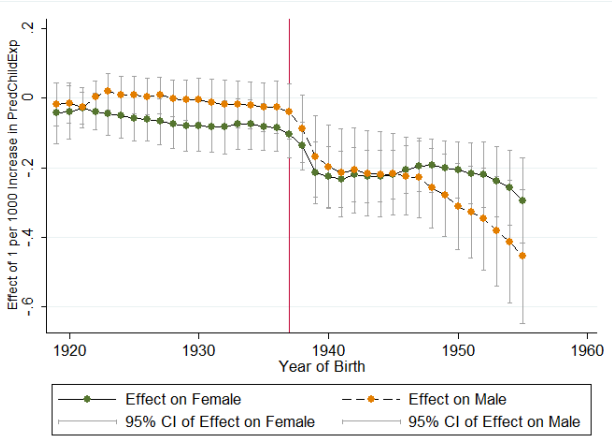


Figure A2: Coefficient using Individuals Born in Moving Cohorts Windows, by Gender

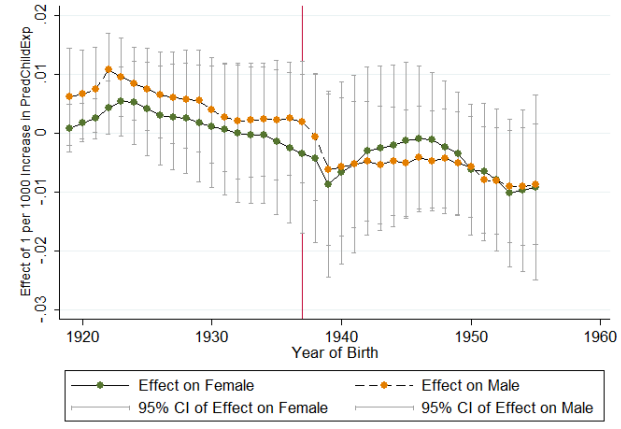
(a) HS or More



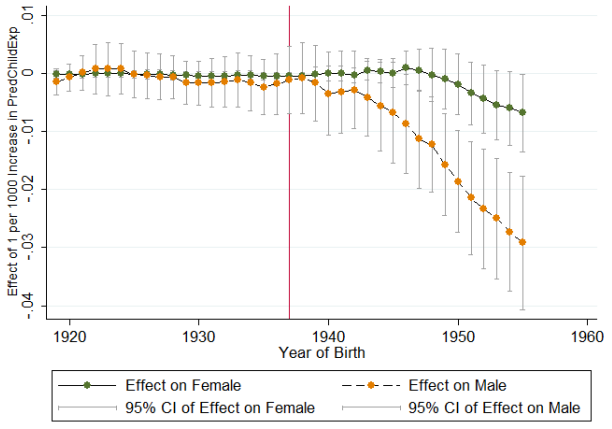
(b) Yrs of Education



(c) Earning>=20k



(d) Earning>=60k

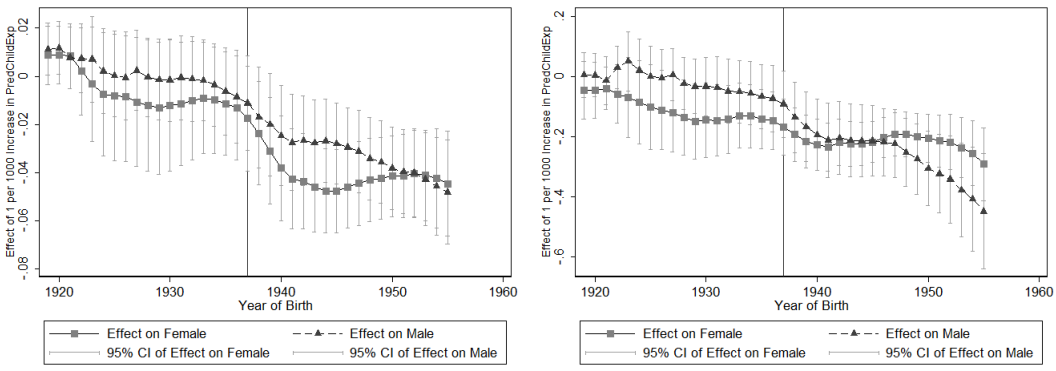


Each point on the graph represents one regression with the same control variables indicated in equation 1. The vertical axis indicates the magnitude of the coefficient of *PredChildExp*. The coefficient corresponding to year of birth "y" is estimated using cohorts born in "y-20" to "y". The vertical line in the graph indicates year 1937, which separates the cohorts into First Mortality Revolution cohorts and the Second Mortality Revolution cohorts.

Figure A3: Coefficient using Individuals Born in Moving Cohorts Windows, Remove 1918 Cohort

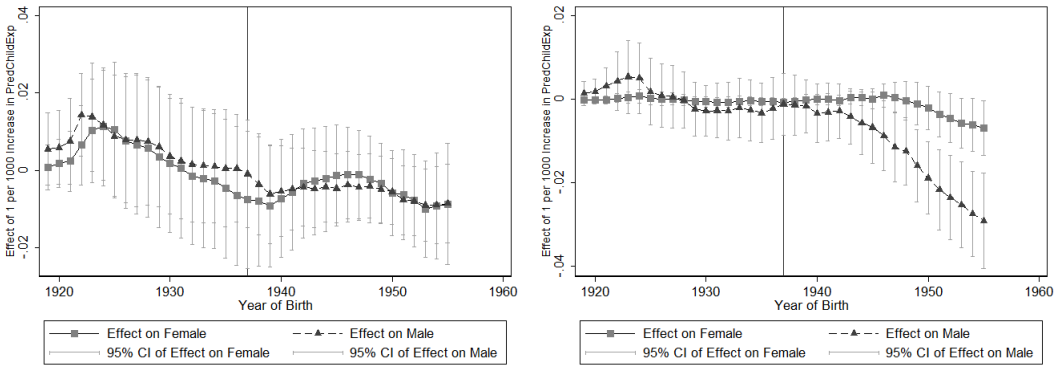
(a) HS or More

(b) Yrs of Education



(c) Earning>=20k

(d) Earning>=60k



Each point on the graph represents one regression with the same control variables indicated in equation 1. The vertical axis indicates the magnitude of the coefficient of *PredChildExp*. The coefficient corresponding to year of birth "y" is estimated using cohorts born between "y-20" and "y", skipping over the 1918 cohort. I do not include the 1918 cohort because of its unusually high mortality rate due to the Spanish Flu Pandemic. The vertical line in the graph indicates year 1937, which separates the cohorts into First Mortality Revolution cohorts and the Second Mortality Revolution cohorts.

Table A1: Childhood Disease Exposure and the Long-run Outcomes by Race and Gender

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	>= 12th Grade	>=4 Yrs College	Yrs of Educ	EMP	Earning >= 20k	Earning >= 40k	Earning >= 60k	SEI Score
Panel A: White Female Subsample								
PredChildExp	-0.011**	-0.004	-0.109***	-0.005	-0.003	-0.001*	-0.001***	-0.004*
	(0.006)	(0.003)	(0.028)	(0.006)	(0.006)	(0.001)	(0.000)	(0.002)
Obs	7958	7958	7958	7958	7958	7958	7958	7905
R-Squared	0.97	0.95	0.97	0.95	0.96	0.97	0.96	0.87
Panel B: White Male Subsample								
PredChildExp	-0.005	0.000	-0.073**	-0.004	0.000	0.000	-0.004*	0.000
	(0.005)	(0.003)	(0.032)	(0.003)	(0.005)	(0.003)	(0.002)	(0.001)
Obs	7965	7965	7965	7965	7965	7965	7965	7964
R-Squared	0.97	0.95	0.98	0.91	0.93	0.98	0.97	0.91
Panel C: Black Female Subsample								
PredChildExp	-0.018**	-0.006***	-0.199***	-0.010*	-0.005	-0.004**	-0.002**	-0.005*
	(0.008)	(0.001)	(0.060)	(0.006)	(0.004)	(0.002)	(0.001)	(0.003)
Obs	6124	6124	6124	6124	6124	6124	6124	5718
R-Squared	0.95	0.75	0.96	0.64	0.92	0.91	0.82	0.90
Panel D: Black Male Subsample								
PredChildExp	-0.020***	-0.002	-0.275***	-0.009**	-0.004	-0.006	-0.003	-0.004**
	(0.007)	(0.002)	(0.062)	(0.003)	(0.006)	(0.005)	(0.002)	(0.002)
Obs	6021	6021	6021	6021	6021	6021	6021	5919
R-Squared	0.95	0.70	0.95	0.75	0.86	0.92	0.86	0.85
Panel E: Other Race Female Subsample								
PredChildExp	-0.022	-0.003	-0.116	-0.023	-0.020	-0.005	-0.005	-0.006
	(0.020)	(0.012)	(0.161)	(0.019)	(0.014)	(0.005)	(0.004)	(0.014)
Obs	4905	4905	4905	4905	4905	4905	4905	4112
R-Squared	0.69	0.52	0.75	0.49	0.52	0.52	0.46	0.46
Panel F: Other Race Male Subsample								
PredChildExp	-0.022	-0.005	-0.123	-0.004	0.001	-0.023**	-0.020*	-0.007
	(0.014)	(0.013)	(0.152)	(0.021)	(0.017)	(0.011)	(0.011)	(0.010)
Obs	4897	4897	4897	4897	4897	4897	4897	4691
R-Squared	0.68	0.51	0.73	0.48	0.52	0.59	0.56	0.51

Table A2: Effects by Race and Region of Birth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	>= 12th Grade	>=4 Yrs College	Yrs of Educ	EMP	Earning >= 20k	Earning >= 40k	Earning >= 60k	SEI Score
Panel A: Southern Birth State Subsample								
PredChildExp	-0.006	-0.003	-0.068**	-0.004**	-0.005***	-0.003**	-0.002*	-0.001
	(0.005)	(0.002)	(0.026)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Obs	13970	13970	13970	13970	13970	13970	13970	13581
R-Squared	0.97	0.90	0.97	0.94	0.96	0.95	0.93	0.93
Panel B:Nonsouth Birth State Subsample								
PredChildExp	-0.005	-0.010***	-0.078**	0.012***	0.016***	-0.001	-0.008***	-0.005**
	(0.006)	(0.004)	(0.035)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Obs	23900	23900	23900	23900	23900	23900	23900	22728
R-Squared	0.96	0.93	0.96	0.96	0.97	0.98	0.96	0.86
Panel C: Southern Birth State Black Subsample								
PredChildExp	-0.004	-0.002	-0.045	-0.002	-0.005*	-0.001	-0.001	-0.000
	(0.005)	(0.003)	(0.026)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Obs	5439	5439	5439	5439	5439	5439	5439	5430
R-Squared	0.97	0.94	0.97	0.98	0.98	0.98	0.97	0.90
Panel D:Nonsouth Birth State Black Subsample								
PredChildExp	-0.005	-0.009**	-0.071*	0.014***	0.019***	0.000	-0.008***	-0.004**
	(0.007)	(0.004)	(0.035)	(0.004)	(0.004)	(0.002)	(0.002)	(0.002)
Obs	10484	10484	10484	10484	10484	10484	10484	10439
R-Squared	0.97	0.96	0.97	0.98	0.98	0.99	0.98	0.89

Table A3: Selection and Scarring, Control for Selection in to the Labor Force

	(1)	(2)	(3)	(4)	(5)
	Earnings	ln(Earnings)	Earning >= 20k	Earning >= 40k	Earning >= 60k
Panel A: Basic Specification					
PredChildExp	-565.904***	-0.016**	-0.003	-0.004***	-0.006***
	(193.882)	(0.006)	(0.005)	(0.001)	(0.001)
Panel B: Control for Fraction in the Labor Force					
PredChildExp	-395.253***	-0.012***	0.000	-0.003**	-0.005***
	(97.829)	(0.004)	(0.003)	(0.001)	(0.002)
in Labor force	33865.703***	0.794***	0.542***	0.341***	0.186***
	(1165.828)	(0.040)	(0.015)	(0.022)	(0.013)
Obs	37870	36255	37870	37870	37870
Panel C: Control for Fraction Employed					
PredChildExp	-397.465***	-0.012***	0.000	-0.003**	-0.005***
	(97.461)	(0.004)	(0.003)	(0.001)	(0.002)
Employed	33580.343***	0.822***	0.527***	0.353***	0.195***
	(1132.536)	(0.044)	(0.015)	(0.020)	(0.012)
Obs	37870	36255	37870	37870	37870

Notes: Refer to footnote in Table 2. The Fraction in the Labor Force [Employed] is defined as the number of population in the labor force [employed] in the census year within each cell, divided by the number of live births for the corresponding state and year of birth.

Table A4: Control for State Time Trends, Region by Cohort Interactions, and Mean Reversion

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	>= 12th Grade	>=4 Yrs College	Yrs of Educ	Emp	Earnings	ln(Earnings)	SEI Score
Panel A: Basic Specification							
PredChildExp	-0.010**	-0.003	-0.118***	-0.006	-569.090***	-0.016**	-0.002*
	(0.005)	(0.003)	(0.027)	(0.004)	(195.872)	(0.006)	(0.001)
Obs	37228	37228	37228	37228	37228	35657	35712
Panel B: Control for Mean Reversion using 1899 Farm Labor Wage							
PredChildExp	-0.005	-0.004	-0.079***	-0.001	-288.308*	-0.008	-0.002
	(0.004)	(0.003)	(0.022)	(0.003)	(163.277)	(0.005)	(0.001)
Obs	37228	37228	37228	37228	37228	35657	35712
Panel C: Birth state cohort Trends							
PredChildExp	-0.011**	0.001	-0.056***	-0.000	-6.421	-0.002	-0.001
	(0.005)	(0.002)	(0.020)	(0.002)	(82.057)	(0.005)	(0.002)
Obs	37228	37228	37228	37228	37228	35657	35712
Panel D: Birth region by cohort							
PredChildExp	-0.005	-0.000	-0.050**	-0.000	-12.388	-0.005	-0.002
	(0.004)	(0.002)	(0.022)	(0.002)	(91.113)	(0.005)	(0.001)
Obs	37228	37228	37228	37228	37228	35657	35712

Notes: The regressions in panel B control for the mean reversion. The additional control variable is the initial state-level farm labor wage in 1899 and multiplied with the national trend of infectious diseases. The state unskilled farm worker wages in 1899 from Lebergott (1964), and is made available from Bleakley, 2010.

Table A5: The Urban Penalty and Childhood Disease Exposure

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	>= 12th Grade	>=4 Yrs College	Yrs of Educ	Emp	Earnings	ln(Earnings)	SEI Score
Panel A: 1915 to 1939 Cohorts Basic Specification							
PredChildExp	-0.008	-0.000	-0.091***	-0.002	-516.832**	-0.015**	-0.001
	(0.006)	(0.003)	(0.032)	(0.005)	(224.838)	(0.006)	(0.002)
Panel B: 1915 to 1939 Cohorts including interaction term with %Urban Population							
PredChildExp	-0.006	0.007***	-0.099**	-0.014***	-949.187***	-0.029***	0.003**
	(0.008)	(0.002)	(0.044)	(0.002)	(172.864)	(0.006)	(0.001)
PredChildExp*%Urban	-0.016*	-0.019***	-0.028	0.027***	925.853***	0.028***	-0.012***
	(0.008)	(0.003)	(0.041)	(0.005)	(231.779)	(0.006)	(0.002)