The International Epidemiological Transition and the Education Gender Gap

Online Appendix

Mariko J. Klasing Petros Milionis

University of Groningen*

January 2020

This appendix includes supplementary material to the paper and is meant to be published online. Section A provides a list of the countries included in our sample. Section B outlines the data sources and descriptions for all the variables that we employ in our analysis. Section C presents the information based on which the grouping of the infectious diseases into the vaccine-preventable and non-vaccine-preventable groups was made. Section D describes a series of robustness checks related to the composition of our main sample. Section E explores the robustness of our baseline results to prior trends and alternative time periods. Section F describes a series of robustness checks related to our regression specification. Section G describes the estimation results from a multi-period panel analysis.

*Faculty of Economics & Business, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands. E-mail Addresses: m.j.klasing@rug.nl, p.milionis@rug.nl.

Section A: List of Countries in our Sample

Our main sample consists of the following 75 countries:

Algeria, Argentina, Australia, Austria, Bangladesh, Barbados, Belgium, Belize, Bolivia, Brazil, Bulgaria, Canada, Chile, China, Colombia, Costa Rica, Czech Republic, Denmark, Dominican Republic, Ecuador, Egypt, El Salvador, Finland, France, Germany, Greece, Guatemala, Guyana, Haiti, Honduras, Hungary, Iceland, India, Indonesia, Iran, Iraq, Ireland, Italy, Jamaica, Japan, Korea Republic, Malaysia, Mauritius, Mexico, Morocco, Myanmar, Netherlands, New Zealand, Nicaragua, Norway, Pakistan, Panama, Paraguay, Peru, Philippines, Poland, Portugal, Romania, Russian Federation, Slovakia, South Africa, Spain, Sri Lanka, Sweden, Switzerland, Taiwan, Thailand, Trinidad and Tobago, Tunisia, United Kingdom, United States, Uruguay, Venezuela, Vietnam, Yugoslavia.

Section B: Data Description and Sources

This section provides details on all the variables that we employ in our empirical analysis. It explains how each variable is constructed and based on which sources.

Country Panel

Average years of schooling for each gender and age-cohort are taken from Barro and Lee (2013). We focus on the individuals that were between 5 and 9 years old in 1940 and 1980. They correspond to the cohort that started school around those years and we measure their educational attainment 10 years later. Thus, the 1940 values reflect the schooling levels of individuals who were 15-19 years old in 1950, while the 1980 values reflect the schooling levels of individuals who were 15-19 years old in 1990.

Primary schooling completion rates are also from Barro and Lee (2013). They correspond to the share of a given school cohort that eventually completed primary education. The respective cohorts we focus on are the individuals that were 5 to 9 years old in 1940 and 1980 respectively, measured 20 years later at the of age 25 to 29 years.

Secondary schooling completion rates are also from Barro and Lee (2013). They correspond to the share of a given school cohort that eventually completed secondary education. The respective cohorts we focus on are the individuals that were 5 to 9 years old in 1940 and 1980 respectively, measured 20 years later at the age of 25 to 29 years.

Tertiary schooling completion rates are also from Barro and Lee (2013). They correspond to the share of a given school cohort that eventually completed tertiary education. The respective cohorts we focus on are the individuals that were 5 to 9 years old in 1940 and 1980 respectively, measured 20 years later at the age of 25 to 29 years.

Life expectancy at birth in 1940 is primarily based on information reported in the 1948, 1949-1950, 1951 editions and the retrospective section of the 1967 edition of the Demographic Yearbook published by the United Nations. In all cases we use the data reported in the most recent edition. These data are supplemented with the information provided by Preston (1975) and by the Federal Security Agency in their Summary of International Vital Statistics 1937-44. In cases where data are not available exactly for the year 1940, we obtain an estimate for 1940 by linearly interpolating the data between the closest year before and after 1940. For countries that were already involved in World War II by 1940 we interpolate the value for 1940 using life expectancy data for the closest available year before and after the war. This way our life expectancy figures in 1940 are not affected by the war but reflect the normal mortality environment in each country. Life expectancy at birth data for 1900 and 1930 were collected from the Human Mortality Database (www.mortality.org) and the Human Life Table Database (www.lifetable.de). Life expectancy at birth data for 1950 and all subsequent years were obtained from the electronic version of World Population Prospects of the UN Population Division.

Life expectancy at age 5, 10 and 20 in 1940 and 1980 is drawn from the same sources as the life expectancy at birth data. These are specifically the UN Demographic Yearbooks for 1940 and the World Population Prospects database for 1980. Missing values are interpolated using the same procedure as for the life expectancy at birth data.

Life expectancy over the age interval 0-15 in 1940 and 1980 are based on survival rates during that particular age interval and are drawn from the same sources as the life expectancy at birth data. Missing values are interpolated using the same aforementioned procedure. Survival rates are multiplied by 15 so that they reflect life expectancy over that age interval.

Life expectancy over the age interval 15-60 in 1940 and 1980 are based on survival rates during that particular age interval and are drawn from the same sources as the life expectancy at birth data. Missing values are interpolated using the same aforementioned procedure. Survival rates are multiplied by 45 so that they reflect life expectancy over that age interval.

Mortality rates in 1940 by disease are obtained from Acemoglu and Johnson (2007), who report mortality rates separately for 13 infectious diseases. They reflect the number of deaths per 100 population. For the year 1980, we do not use actual mortality rates, but potential ones. Following Acemoglu and Johnson (2007) we take the potential mortality rates in 1980 to be zero

for all infectious diseases. This way our mortality variable does not reflect the actual changes in mortality rates in each country, but the changes that would have potentially resulted from the introduction of IET-related medical advances that provided effective control for these 13 infectious diseases. We furthermore construct three alternative variants of this variable where the potential mortality rates in 1980 are based either on the actual U.S. mortality rates in 1980, on the average mortality rates across other health-frontier countries in 1980, or on the country-specific rates in 1940 scaled down by the average rate of decline in mortality from the respective diseases observed at the global level between 1940 and 1980.

Overall mortality corresponds to the sum of the potential mortality rates from each of the 13 infectious diseases described above.

Mortality from vaccine-preventable diseases (VP group) corresponds to the sum of the mortality rates from the following seven diseases: diphtheria, influenza, measles, pneumonia, smallpox, tuberculosis, whooping cough.

Mortality from non-vaccine-preventable diseases (NVP group) corresponds to the sum of the mortality rates from the following six diseases: cholera, malaria, plague, scarlet fever, typhoid fever, typhus.

Mortality from the 10 remaining diseases corresponds to the sum of the mortality rates from the following ten diseases: cholera, diphtheria, influenza, measles, plague, scarlet fever, smallpox, typhoid fever, typhus, whooping cough.

Mortality from bacterial diseases corresponds to the sum of the mortality rates from the following nine diseases: cholera, diphtheria, plague, pneumonia, scarlet fever, typhoid fever, tuberculosis, typhus, whooping cough.

Mortality rates from diarrhea in 1940 were taken from World Health Organization (1951) and Federal Security Agency (1947). In six cases data for exactly the year 1940 are not available. In these cases, we use the closest available year in the interval 1936-1943.

Maternal mortality ratios in 1940 were taken from World Health Organization (1951), Federal Security Agency (1947) and the 1951 and 1957 editions of the UN Demographic Yearbook. In case of revisions in the UN data, we use the most recently reported values. Missing values are interpolated using the two closest data points before and after 1940 assuming a linear time trend. The 1980 maternal mortality ratios are taken from the 1981, 1982, 1983 and 1985 editions of the UN Demographic Yearbooks. Data are available for 63 out of the 75 countries in our data set and are by definition zero for the male population.

Mortality rates from cancer and cardiovascular diseases for each gender in 1940 and 1980 are based on the causes-of-death statistics reported in the UN Demographic Yearbook. Inspecting all editions of the yearbook published until 2006, we collected data on deaths rates from cancer and cardiovascular diseases by gender for as many years during the interval 1950 to 2004 as possible. The 1980 values were correspondingly obtained from linear interpolation between the available data points. The 1940 figures were obtained from backward extrapolation of the 1950 figures based on the average annual growth rates of mortality from the respective diseases, specific for each country and gender, observed during the 1950s.

Gender-specific mortality rates in 1940 are constructed in two ways. The first is to use the information provided in the Vital Statistics of the United States for 1940 to obtain ratios of female-to-male mortality for each of the 13 infectious diseases covered by Acemoglu and Johnson (2007). We then use these ratios to scale the population mortality rates that we have for each of these diseases in all countries of our sample. The second is to use the gender-specific mortality data reported by Preston et al. (1972) for a smaller sample of 22 countries in order to obtain country-specific ratios of female-to-male mortality. The available data allow us to compute such ratios for three categories of infectious and parasitic diseases: tuberculosis, influenza/pneumonia/bronchitis and other infectious and parasitic diseases. We use the obtained ratio for tuberculosis to scale the population mortality rates for tuberculosis from Acemoglu and Johnson (2007), the obtained ratio for influenza/pneumonia/bronchitis to scale the population mortality rates for influenza and pneumonia, and the obtained ratio for other diseases to scale the population mortality rates for the remaining 10 diseases covered by Acemoglu and Johnson. As with the other mortality variables, whenever the data do not correspond exactly to the year 1940, we interpolate the missing values between the closest observations available before and after 1940 assuming a linear trend. For 1980 we assume again that the mortality rates for both groups of diseases were zero for men and women in all countries.

GDP per capita data are taken from the Maddison Project.

Institutional quality reflects the level of the executive constraints indicator reported in the POLITY IV data set in 1950 and 1990. For countries in our sample that were not independent in 1950 we use instead the figure from the first year after independence.

Urbanization rates in 1950 and 1990 reflect the share of a country's population residing in urban areas. It is calculated as the ratio of the urban population figures reported in the UN World Urbanization Prospects database over the total population figures reported in the World Population Prospects database.

The **services sector employment** variable reflects the share of the employed and self-employed population that was active in the services sector of the economy in 1950 and in 1990.

The 1950 values are computed based on the 1955 edition of the UN Demographic Yearbook. The 1990 values are based on the information reported in the ILOSTA database of the International Labour Organization.

The **compulsory schooling** variable reflects the years of compulsory schooling in each country in 1950 and 1990. The 1950 values are collected from successive editions of the UNESCO International Yearbook of Education between 1950 and 1960. The 1990 values are taken from the 1991 edition of the UNESCO World Education Report.

Female voting rights are measured with a dummy variable which is coded as 1 if women in that year had the right to vote in a given country and as zero otherwise. The coding of this variable is based on the information provided in Ramirez et al. (1997) who document when between 1890 and 1990 women acquired suffrage in 133 countries.

The **divorces rate** variable reflects the ratio of the number of divorces relative to the number of marriages in 1950 and 1990. The 1950 values are computed based on the information reported in the 1951, 1952, 1955 and 1957 editions of the UN Demographic Yearbook. The 1990 values are computed based on the information reported in the 1992, 1994 and 1996 editions of the UN Demographic Yearbook.

Labor force participation rates in 1950 and in 1990 correspond to the ratio of the total size of the economically-active population relative to the population age 15 and over and are computed separately for men and women. To construct the 1950 figures we collected information on the economically-active and total population by gender from the 1948 and the 1949-1950 editions of the UN Demographic Yearbook. The total population figures reported in these sources are scaled down to reflect the population of age 15 and over by multiplying them with the respective shares of the population age 15 and over in 1950, reported in the UN World Population Prospects database. To construct the 1990 figures we combine information on the size of the economically active population reported in the ILOSTA database with information on the size of the population age 15 and over reported in the UN World Population Prospects database.

Fertility rates in 1950 and in 1990 are measured as the number of live births per 1,000 population and are taken from the UN Population Prospects database.

The **abortion law** variable is an indicator variable reflecting the number of cases in which a country's law in 1960 permitted abortions. The indicator is taken from Bloom et al. (2009).

Disease incidence rates correspond to the total number of cases in a country of a given disease, as reported in the WHO World Health Statistics Annuals, divided by the country's population, taken from the Maddison Project. The specific disease incidence rate variable that

we employ in our regression analysis corresponds to the average incidence rate across the following 8 diseases: cholera, diphtheria, malaria, measles, plague, smallpox, typhoid fever and whooping cough in 1962 and 1980. We focus on the change in disease incidence rates between 1962 and 1980, as 1962 is the earliest year for which we were able to find comprehensive information on cases for most of the infectious diseases that we consider in our analysis. Incidence data for the remaining 5 infectious diseases in our sample are not available.

Case fatality ratios correspond to the total number of deaths in a country from a given disease relative to the total number of cases, provided that the number of cases is positive. Otherwise the ratio is taken to be zero. The data for both cases and deaths from each disease are taken from the WHO World Health Statistics Annuals. The variable that we employ in our regression analysis corresponds to the average case fatality ratio across the 8 diseases: cholera, diphtheria, malaria, measles, plague, smallpox, typhoid fever and whooping cough in 1962 and 1980. We focus on the change in case-fatality ratios between 1962 and 1980, as 1962 is the earliest year for which we were able to find comprehensive information on cases for most of the infectious diseases that we consider in our analysis.

Height data for males and females are taken from NCD Risk Factor Collaboration (2016). The data correspond to the mean height of 18-year old males and females born in each country in 1940 and 1980 respectively.

The **post-demographic transition** dummy is coded based on the criteria provided by Cervellati and Sunde (2011). Specifically a country is classified as having gone through the demographic transition if by 1940 life expectancy at birth exceeded 50 years, there was a sustained decline in fertility, or the crude birth rate had fallen below 30/1000. Based on this classification the post transitional countries consist of Argentina, Australia, Austria, Belgium, Bulgaria, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Poland, Portugal, Romania, Russian Federation, Slovakia, Spain, Sweden, Switzerland, U.S.A., United Kingdom, Uruguay and Yugoslavia.

U.S. States Panel

Average years of schooling for each gender and age-cohort are constructed from the 1% U.S. census files provided by IPUMS (Ruggles et al., 2019). We focus on individuals that were between 5 and 9 years old in 1940 and 1980 and calculate their years of schooling from the educational attainment of the male and female population aged 15-19 in 1950 and 1990 respectively. IPUMS reports educational attainment of individuals in twelve discrete categories. We convert these

intervals into years of schooling based on the structure of the US education system. Specifically we take category *no schooling* to imply 0 years, category *nursery school to grade 4* to imply 4 years, category *grade x* to imply x years, category *x years of college* to imply $12 + x$ years and category *5+ years of college* to imply 18 years. We then average individual years of schooling at the state level separately for the respective male and female census population.

Life expectancy at birth in 1940 is taken from the State and Regional Life Tables 1939-41 published by the Federal Security Agency and made available through the website of the Centers for Disease Control and Prevention (www.cdc.gov). They are only available for the white population. The figures for 1980 are taken from the U.S. Decennial Life Tables for 1979-81, also available through the website of the CDC. The figures reported are broken down by gender and race. We use the figures for whites to ensure comparability with the 1940 data.

Mortality rates in 1940 by disease were constructed based on the number of deaths from each cause that occurred in each state, broken down by gender and race, as reported in the 1940 Vital Statistics of the United States. We use the figures for whites in line with the available life expectancy data. To convert the number of deaths from each cause into death rates, we divide them with the size of the white male and female population in each state. This can be obtained indirectly from the number of deaths from all causes and the all-causes death rates by gender and race reported in the same publication.

Section C: Disease Classification

As part of our analysis we separate the 13 infectious diseases, for which we have mortality data, into two groups: a group of vaccine-preventable diseases (VP group) and a group of non-vaccine-preventable diseases (NVP group). The assignment of each disease to one of these two groups is based on whether the observed reductions in mortality rates following the IET can be attributed, at least partially, to the introduction and diffusion of vaccines. For this purpose we consider the nature of each disease and the key medical breakthroughs that were important for its prevention and treatment over our period of interest. Below we summarize this information for each of the 13 diseases.

Cholera: Cholera is an acute diarrheal disease caused by the bacterium Vibrio cholerae and is spread by drinking contaminated water or eating contaminated food. The first effective vaccine against cholera was developed in 1896 by Wilhelm Kolle. However, the vaccines available since then, including the modern ones developed in the 1990s, provide only limited immunity. This is

typically for no more than one year (Parish, 1965). Given that, the main method of controlling cholera is the provision of clean drinking water and good sanitation facilities (WHO, 2010). Since cholera-induced diarrhea results in severe dehydration within 3-4 hours, the development of more effective oral rehydration therapy in the 1950s provided a major breakthrough in the treatment of cholera. Furthermore, antibiotics developed in the 1940s shorten the course of the disease and reduce the symptoms (Kiple, 2003). We, thus, assign cholera to the group of non-vaccine-preventable disease given the ineffectiveness of the available vaccines.

Diphtheria: Diphtheria is caused by the Corynebacterium diphtheriae and is spread by air or touch. Kitasano and Behring developed in 1890 the first antitoxin from the serum of animals that had been infected with the toxins secreted by the bacilli. They showed that this antitoxin could be used to treat diphtheria patients. In 1913 Behring reported first trials of successful immunization with toxin-antixtoxin mixtures (Baker & Katz, 2004). However, even though toxin-antitoxin mixtures had been successfully used for vaccination in some U.S. cities in the 1920s, it was generally difficult to employ these mixtures. This was because it was unclear how toxins and antitoxins should be balanced to induce an effective and at the same time safe immunization (Kiple, 2003). The key event in the development of safe vaccines was the development of a chemically modified toxin, called toxoid, by Leon Ramon of the Pasteur Institute. By using heat and formalin, the toxin's toxicity was reduced without affecting its immunizing power. This vaccine was approved in France in 1927 and made compulsory for children in France in 1938 (Parish 1965), but in other countries the vaccine was only adopted much later. Large-scale vaccination at the global scale did not start until after World War II with the licensing of the combined diphtheria-tetanus vaccine in 1947 (Baker & Katz, 2004). Vaccination campaigns for diphtheria were facilitated with the development of the jet injector in the 1960s and the invention of freeze drying during WWII, which greatly increased the storability of vaccines and facilitated vaccine transportation (Kiple, 2003). Given that large-scale vaccination campaigns outside of France did not start before the mid 1940s, we assign diphtheria to the group of vaccine-preventable disease.

Influenza: Influenza is caused by three strands of the influenza virus, types A, B, and C. It is passed by breath-born droplets, such as sneezing and coughing. The type A virus was first isolated and identified in 1934 and type B in 1940 (Plotkin et al., 2012). The first vaccine was developed during World War II to protect soldiers and production was expanded to civilian markets in 1946 (Hoyt, 2006). While the development of antibiotics in the 1940s also played a role in reducing mortality from secondary bacterial infections, the treatment of influenza is largely symptomatic (Kiple, 2003) and the primary method of control is vaccination. Hence, we assign influenza to the group of vaccine-preventable diseases.

Malaria: Malaria is caused by parasites transmitted by the female mosquito of the Anopheles

type. The disease transmission mechanism and the role of the Anopheles mosquito was first established in the 1890s. Since then the strategy generally applied to control malaria has been the control of mosquitoes. This became much more effective with the discovery of DDT residual insecticides in the late 1940s (Kiple, 2003). Effective vaccines still do not exist to this date. Given this, we assign malaria to the group of non-vaccine-preventable diseases.

Measles: Measles is a viral infection, principally occurring among children. The disease is transmitted by air and through touch. First steps in the development of the measles vaccine were the discovery that the disease was caused by a virus in 1911. Still the isolation of the measles virus did not occur until 1954 (Plotkin et al., 2012). Measles vaccines were first produced in 1958 and licensed in 1963. A single dose of the measles virus vaccine provides life-long immunity in over 95% of cases (Kiple, 2003).With this in mind, we assign measles to the group of vaccine-preventable diseases.

Plague: Plague is caused by the bacterium Yersinia pestis and is transmitted to humans through the bite of infected fleas. First vaccines were developed in the late 1890s, but the exact nature of the disease and the role of the flea in transmitting the disease from infected rats and other wild rodents were not fully studied until 1914 (Parish, 1965). After this was established, the elimination of rats and fleas became the key step in the prevention of plague. While vaccines developed in the 1890s were effective in reducing incidences of plague and plague mortality, protection only lasted for a few months (Parish, 1965). Thus, the vaccine was never routinely used to control the disease (Plotkin et al., 2012). With the development of antibiotics, especially streptomycin, in the 1940s, vaccination became of secondary importance (Kiple, 2003). Given the limited protection afforded by vaccines and their limited use following the invention of streptomycin, we assign plague to the group of non-vaccine-preventable disease.

Pneumonia: Pneumonia is caused by a variety of causative agents, including viruses and bacteria. The identification of these causative agents began with the isolation of the Streptococcus pneumoniae bacterium by Louis Pasteur in 1880. For viral pneumonia, viral vaccines, in particular influenza vaccines developed in the 1940s, have been effective in preventing infections (Ruuskanen et al., 2011). These became available for civilian markets in 1946 (Hoyt, 2006). For bacterial pneumonia, which is caused by the different kinds of bacteria, both vaccines and antibiotics have been used for disease control. A vaccine against bacterial pneumonia was introduced already in 1945. Yet, they were not widely used, as their arrival coincided with the discovery of penicillin and other antibiotics, which were more effective for reducing mortality from bacterial pneumonia (Kiple, 2003). Vaccines against bacterial pneumonia were, however, further developed in the late 1970s and early 1980s and this resulted in the licensing of more effective vaccines. The final breakthrough in controlling the disease was the development of pneumococcal conjugate vaccines in 2000 (Klein & Plotkin, 2007). With all this in mind, we

assign pneumonia to the group of vaccine-preventable diseases, as vaccines did play some role as a method of disease control.

Scarlet Fever: Scarlet fever is caused by certain strains of group A Streptococcus bacteria. The bacteria were first isolated in 1887. Efforts to produce a streptococcus vaccine followed, but the results were not satisfactory. To this date no effective vaccine exists (Plotkin et al., 2012). However, the disease is effectively treated with antibiotics, including penicillin, developed in the 1940s (Kiple, 2003). Given this, we assign scarlet fever to the group of non-vaccine-preventable diseases.

Smallpox: Smallpox, which has been eradicated since the 1970s, was an infectious disease caused by the viruses Variola major and Variola minor. An effective vaccine was developed in 1798 by Edward Jenner, which was derived from cowpox, a related disease found among animals (Kiple, 2003). However, large-scale vaccination was not possible until the 1950s when free-drying, invented in the 1940s, was adopted for mass production of vaccines (Fenner, 2011). Until then the storability of the vaccine was very limited and temperature fluctuations frequently caused the vaccine to lose its potency. The invention of the jet injector in the 1960s was another important development that significantly facilitated the administration of the vaccine (Plotkin et al., 2012). The combination of these innovations were instrumental in the eradication of the disease. We, therefore, assign smallpox to the vaccine-preventable diseases.

Tuberculosis: Tuberculosis is caused by the bacterium Mycobacterium tuberculosis and is transmitted through air. The bacterium was discovered in 1882 by Robert Koch. The first effective vaccine was developed by the French Pasteur Institute and became available for use in 1921 (Parish, 1965). This is the Bacillus Calmette-Guerin (BCG) vaccine, which is still in use today. However, mass vaccination did not start until after the late 1940s when the safety of the vaccine was established and UNICEF in cooperation with the Danish and Swedish Red Cross teams started mass vaccination campaigns (Kiple, 2003). Until then, the vaccine had been primarily used in France where by 1933 about 20% of all children were vaccinated. Outside of France, though, only 500,000 children had been vaccinated at that time (Parish, 1965). This changed with the mass post-war vaccination campaigns that began in the 1950s in Europe and in the 1960s in the developing world (Fine, 1995). In 1948-1951 alone 30 million people were tuberculin tested and 14 million were vaccinated. These campaigns were extended further under a joint UNICEF/WHO initiative which led to the vaccination of an additional 60 million people by 1956 (Comstock, 1994). Given the sensitivity of the original BCG vaccine to sunlight and temperature fluctuations, it could only be stored for 10 days after initial preparation (Bonah, 2005). This changed with the invention of freeze-drying in the 1950s which allowed for the vaccine to be stored under various climatic conditions for prolonged periods of time and to be transported easily over long distances (Parish, 1965). Another major step forward in the fight

against tuberculosis was the development of antibiotics, in particular streptomycin, which provide an effective treatment for most cases of the disease. With all this in mind, we assign pneumonia to the group of vaccine-preventable diseases, as the post-war spread of the BCG vaccine played an important role in controlling the disease. Given that mass vaccination campaigns did not start before the late 1940s, we code tuberculosis as a vaccine-preventable disease.

Thyphoid Fever: Typhoid fever is caused by the bacterium Salmonella typhi and is spread through contaminated water or food. Vaccines were developed before World War I, but they were not very effective (Plotkin et al., 2012). With the introduction of antibiotics in the 1940s, which provided effective treatment, the use of vaccines became of secondary importance in the control of the disease (Kiple, 2003). Improved hygiene has also limited incidences of the disease. Hence, we assign typhoid fever to the group of non-vaccine-preventable diseases, given that vaccines had already been available long before 1940 and were not effective.

Typhus: Typhus is caused by Rickettsia bacteria transmitted by the human body louse Pediculus humanus corporis. The role of the louse was established in 1909 and the bacterium first demonstrated in 1916 (Parish, 1965). First attempts in developing a vaccine to be used in humans were made in the 1930s. This was in the form of the Weigl vaccine, which used killed bacteria, and the Cox vaccine, which used live bacteria (Lindenmann, 2002). They were albeit difficult to manufacture and turned out to not be very effective. With the introduction of antibiotics, which provides effective treatment for typhus, and the development of DDT, which allowed for very efficient delousing, vaccines fell into disuse (Lindenmann, 2002). Public health measures such as good hygiene and sanitation also played an important role in the control of the disease. Thus, we assign typhus to the group of non-vaccine-preventable diseases, given that vaccines were already available before 1940 and not much used after the introduction of broad-spectrum antibiotics.

Whooping Cough: Whooping cough, or pertussis, is caused by the bacterium Bordetella pertussis and occurs primarily among young children. The causative agent was first isolated in 1900. The first effective vaccine was developed in the 1930s following some earlier attempts already in the 1920s. However, these early vaccines had very limited efficacy and became obsolete with the introduction of the vaccines developed by Kendrick and Eldering (Parish, 1965; Granstroem, 2011). First large-scale clinical trials were conducted in the U.S. in the late 1930s and early 1940s, which demonstrated the effectiveness and safety of the vaccine (Parish, 1965). Following that, the American Academy of Pediatrics approved the vaccine for routine use in 1943. However, it was not until the later 1940s that vaccines for whooping cough were widely used when the combination vaccine with diphtheria and tetanus, DPT, was licensed (Parish, 1965). Based on this information, we assign whooping cough to the group of vaccine-preventable diseases.

Section D: Robustness Checks to Sample Composition

For our main analysis we have focused largely on the sample of 75 countries reported in Section A. In this sample we have included all countries for which we were able to obtain data on mortality and life expectancy in 1940 from the sources listed in Section B. In this section we provide some additional robustness checks to ensure that our main results do not hinge on the composition of that country sample. These are motivated by several potential concerns.

First of all, a crucial assumption behind our estimation is that our potential mortality instrument can be reasonably excluded from our second-stage specification. Practically, this means that the potential changes in mortality from infectious diseases that occurred between 1940 and 1980, as a result of the IET, were exogenous to the evolution of educational attainment for men and women in our sample of countries. While this assumption seems plausible for most countries in our sample, it may not hold for the countries that were actively involved in the development of the key medical innovations that led to the IET. This group of countries includes France, Germany, Switzerland, the United Kingdom and the United States. With that in mind, in column 1 of Table A1, we omit these countries from the sample.[1] As we can see, this modification of our country sample does not affect our main findings. The estimated coefficients for all variables are very similar to those reported in Table 3 of the paper and we continue to observe a stronger effect of the potential mortality reductions from vaccine-preventable diseases on the life expectancy of females compared to males.

[Insert Table A1 around here]

A related concern has to do with the quality of the data. As for some of the 75 countries in our sample the data may not be very reliable, in columns 2 to 4 of Table A1 we omit one by one certain groups of countries for which data quality may be questionable. Specifically, we omit African and Middle Eastern countries in column 2, small Latin American and Caribbean countries in column 3, and Eastern European countries in column 4.[2] In all cases we can see from the resulting estimates that omitting these groups of countries does not alter our findings, neither in the first stage nor in the second stage. Thus, the quality of the data does not appear to be a major concern for our estimates.

Rather than excluding groups of poorer countries, in column 5 we do the opposite and exclude all countries that were already developed in 1940.[3] This is to ensure that the observed positive

[1] Omitting these countries also takes care of the concern that medical research at that time may have been biased toward finding vaccines against the diseases most prevalent in these countries.

[2] The omission of Eastern European countries also takes care of the concern that the expansion of education in Eastern Europe may have been driven by the policies of communist regimes in the post-war period.

[3] We classify as developed all countries in which in 1940 less than 25% of the labor force was employed in agriculture. Using an alternative criterion based on GDP per capita, as in Acemoglu and Johnson (2007), does not alter the set of countries that are classified as developed.

relationship between life expectancy and educational attainment is not driven by the experiences of a few rich countries. The obtained estimates in this case again eliminate this concern. Both in the first stage and in the second stage the estimates are similar to those of Table 3.

As a final check regarding the sample composition, we provide in column 6 the OLS estimates for the relationship between life expectancy at birth and average years of schooling for all 145 countries that are covered in the Barro-Lee data set over the period from 1950 to 1990. Taking 1950 as a starting year allows us to broaden our sample and include many more developing countries. Although in this case we can not perform our 2SLS estimation, the resulting OLS estimate is not very different from that reported for our main sample in column 7 of Table 3. Hence, these results suggest that the relationship between life expectancy and years of schooling is not fundamentally different for developed and developing countries.

Section E: Robustness Checks for Prior Trends and Alternative Sample Periods

An important concern regarding the results presented so far is that they may not necessarily reflect the effect of health improvements triggered by the IET, but pre-existing trends in health or education. To check that this is not the case, in Table A2 we present a series of additional robustness checks to account for prior trends in the data and vary the sample period over which we estimate the relationship between life expectancy and years of schooling.

In column 1 we lag our life expectancy data by 40 years. Looking at the first-stage estimates, we see no evidence for a negative relationship between changes in potential mortality rates after 1940 and life expectancy changes before 1940. In fact, the baseline coefficient estimates are positive. This is not surprising as countries that had a higher disease burden in 1940 naturally must have experienced smaller life expectancy increases in the previous 40 years. At the same time, the obtained coefficients for the interaction terms with the female dummy are both statistically insignificant. This is important because it implies that there was no difference in the evolution of life expectancy between men and women prior to 1940 in countries with different disease environments in 1940. In other words, while in countries with a high disease burden in 1940 women experienced subsequently larger life expectancy increases compared to men, prior to 1940 there were no such systematic differences. Looking at the second-stage estimates, furthermore, we see that the coefficient on life expectancy has the wrong sign. If anything, these estimates suggest that countries with larger gains in life expectancy prior to 1940 experienced smaller increases in educational attainment after 1940.

[Insert Table A2 around here]

While these results indicate an important discontinuity in the evolution of life expectancy across countries after 1940, with the onset of the IET, they could also be driven by a subsequent convergence of life expectancy in less developed countries to pre-existing trends. To ensure that our results are not driven by this type of convergence dynamics, in column 2 we estimate our main 2SLS specification using 40-year lagged values for both life expectancy and potential mortality. This corresponds to a falsification test for the hypothetical scenario that the IET took place after 1900 rather than 1940[4]. Looking at the obtained estimates we see that changes in potential mortality are not related in any way to changes in life expectancy between 1900 and 1940. Moreover, the changes in life expectancy predicted by our first stage regression can not account for the subsequent changes in schooling after 1940 either. These results together suggests that the IET-related health improvements are not related to any prior trends in mortality or life expectancy.

In column 3 we proceed to test for prior trends in educational attainment by lagging our average years of schooling variable by 40 years. If post-1940 health improvements can predict the rise in schooling observed between 1900 and 1940, this would indicate that our previously estimated relationship may be spurious and that the post-1940 increases in educational attainment would have occurred even in the absence of the IET. Reassuringly, we find the relationship between the two to be weak and statistically not different from zero. To account directly for pre-existing trends in the schooling data, we also include lagged schooling as a control variable in both the first-stage and second-stage estimation in column 4. In both cases we find lagged schooling to be statistically insignificant and to not affect our main coefficients of interest. Taken together, the results of column 1 to 4 strongly refute the possibility of any pre-existing trends in the life expectancy and schooling data to be driving our results.

Beyond concerns about pre-existing trends, one potential drawback of focusing our analysis on the period from 1940 to 1980 is that our starting period coincides with the outbreak of World War II, which created major disruptions in the health and education systems of many countries. While this choice was driven by considerations related to data availability and comparability with previous work in the literature, as mentioned in Section 4 of the paper, in column 5 we show that this is not crucial. Specifically the column presents the estimates for our main specification when we use 1930 as the starting period instead of 1940. While the country sample that we are left with in this case is substantially smaller, both the first and the second stage estimates that we obtain are not very different from those in our baseline estimation. This suggests that our conclusions do not hinge on taking 1940 as the starting point.

[4]Specifically the lagged values of potential mortality are based on imputed values for each country's disease mortality rates in 1900, extrapolated backward from 1940 based on the US the trends. For 1940 we take these values to be zero for all countries.

In columns 6 to 8 of Table A2, we also consider more recent changes in health and education. In particular, in column 6 we regress changes in educational attainment between 1980 and 2000 on changes in life expectancy over the same period, using still potential mortality reductions between 1940 and 1980 as our instrument. In this case we find the first stage estimates to be negative, but largely statistically insignificant. As a result, the 2SLS estimation does not perform well, even though in the corresponding unreported OLS regression changes in life expectancy and changes in schooling between 1980 and 2000 are still positively related.

In column 7, we relate changes in life expectancy between 1940 and 1980, instrumented with changes in potential mortality over the same period, to changes in educational attainment between 1980 and 2000. In this case, the first stage estimates are identical to our baseline specification, but the second-stage estimates are again statistically insignificant. Finally, in column 8, we relate changes in life expectancy between 1980 and 2000, instrumented with changes in potential mortality between 1940 and 1980, to changes in educational attainment between 1940 and 1980. The first stage estimates are weak, just like in column 6. In the second-stage we find a strong positive coefficient, but this should be interpreted with caution considering the weak first-stage estimation.

These patterns are not surprising. Improvements in health that started with the IET continued also after 1980. Given that the changes in potential mortality between 1940 and 1980 can still somewhat explain the changes in life expectancy that occurred after 1980, it is no wonder that the post-1980 changes in life expectancy predicted by the potential mortality instrument can also explain changes in educational attainment prior to 1980. As time goes by and public health across countries improves, though, the initial variation in the disease environment, captured in our instrument, becomes less relevant for explaining the cross-country differences in life expectancy. Furthermore, if educational attainment is primarily driven by current health conditions, we should expect at most a weak relationship between post-1980 changes in educational attainment and changes life expectancy prior to 1980, as predicted by variation in the disease environment in 1940. Additionally, as we can see from columns 6 and 8, our potential mortality instrument does not explain the differential improvements in life expectancy between men and women after 1980. While differential changes in health may still have contributed to reductions in the education gender gap even after 1980, our estimation strategy does not allow us to precisely estimate this effect.[5]

[5] In Section G of the appendix we provide some additional evidence regarding these points by using a multi-period panel with decadel observations that extends all the way to the year 2000 rather than a two-period panel.

Section F: Robustness Checks on Regression Specification

In our analysis so far we have focused on a panel regression setup that combines observations for males and females from all countries and employs gender-country fixed effects to account for differences in educational attainment across genders in each country. As we explained in Section 3 of the paper, we choose this regression specification given our interest in explaining differential changes in schooling between 1940 and 1980 with differential changes in life expectancy between men and women. In this section we focus on assessing the robustness of our baseline results to using alternative regressions setups which conduct the estimation separately for males and females or employ different sets of fixed effects.

In order to facilitate the comparability of the estimates across these different regression setups, for this part of the analysis we standardize all mortality variables to have a zero mean and a unitary standard deviation. Using these standardized potential mortality rates Table A3 explores the robustness our main 2SLS estimates to changes in our regression setup. Table A4 does the same for the reduced form estimates. In both tables the exact details of the regression specification are indicated at the top of each column. In all cases we use for the estimation the main potential mortality instrument for which the mortality rates in 1980 are set to zero in all countries.

[Insert Table A3 around here]

In columns (1) and (2) we repeat the estimations of columns (2) and (3) from Table 3 of the paper with standardized mortality figures while still employing gender-country and year fixed effects. The resulting estimates confirm all the conclusions that we reached in Section 5 of the paper regarding the magnitudes of the estimated effects and their statistical significance. As we can see in panel A, reductions in potential mortality from both vaccine preventable and non-vaccine preventable diseases triggered increases in life expectancy. Females gained on average more in terms of life expectancy from these reductions and this difference is clearly driven by mortality reductions from vaccine-preventable diseases. The interaction effect of the female dummy with potential mortality from the VP group is clearly negative and statistically significant, while the corresponding interaction effect for NVP group is clearly weaker and not statistically different from zero.

Panel B reports the corresponding second-stage estimates regarding the relationship between changes in life expectancy and years of schooling. The obtained coefficients are identical to the estimates reported in Table 3. In Table A4 we also report for comparison the corresponding reduced-form relationship between changes in potential mortality and changes in schooling. Columns (1) and (2) repeat the estimation in columns (9) and (10) of Table 3 with standardized mortality figures and echo the results reported there. Improvements in the mortality environment are clearly associated with increases in average years of schooling. Moreover, this effect is

estimated to be stronger for women and is clearly driven by reductions in potential mortality from vaccine-preventable infectious diseases.

[Insert Table A4 around here]

In the next four columns we report the estimates of our main specification but now estimated as a system of seemingly unrelated regressions for the sub-samples of male and the female observations. Columns (3) and (4) reports the 2SLS estimates for the sub-sample of males, first considering mortality from all diseases and then splitting it between the VP and the NVP groups of diseases. Columns (5) and (6) do the same for the sub-sample of females. Having the potential mortality figures standardized, we can easily compare the estimates across columns. Comparing the estimates between columns (3) and (5), we see that the obtained effect of the mortality reductions from all infectious diseases was stronger for females, with the standard chi-squared test clearly rejecting the null of the two coefficients being equal. Conducting the same test for the corresponding coefficient estimates in columns (4) and (6), we find the effect of mortality reductions to be clearly larger for females in the case of VP diseases, but not in the case of NVP diseases. The chi-squared test rejects the null in the former, but not in the latter case.

Looking at the same columns, (3) to (6), in Table A4 we see the corresponding estimates for the reduced-form relationships between potential mortality and schooling, estimated separately for males and female. The pattern that we see and the results from the chi-squared tests for the equality of the coefficient estimates in the male and female regression equations matches the observed ones based on the 2SLS estimations in Table A3. The estimated coefficients of potential mortality from all diseases in columns (3) and (5) of Table A4 are statistically different from each other and larger in absolute terms for the female sample. This is also case for the coefficients of potential mortality from VP diseases in columns (4) and (6). For the estimated coefficients of potential mortality from NVP diseases, though, we find this not to be the case.

In columns (7) and (8) we go back to perform our estimation in a pooled sample of males and females, but we now change the employed set of fixed effects to country and gender-year fixed effects. This allows us to control for differential trends in the evolution of life expectancy and schooling between men and women. This regression setup is challenging for our data, as our potential mortality instrument does not vary across genders. Nevertheless, the obtained estimates are very similar to the magnitudes reported in the previous columns. The estimates for the interaction effects of the female dummy with potential mortality from all diseases in column (7) and with potential mortality from VP diseases in column (8) are slightly larger compared to those of columns (1) and (2). Yet, they marginally fall below conventional levels of statistical significance, with the p-values being 0.11 and 0.18 respectively.

Looking at the reduced form estimates in Table A4, we see a similar pattern in columns (7) and (8). Even when we include gender-year fixed effects, we see again that reductions in potential

mortality are associated with increases in schooling and this particularly the case for mortality from VP diseases. The estimated interaction effect with the potential mortality from all diseases in column (7) and with potential mortality from VP diseases in column (8) are again larger compared to those of columns (1) and (2) of the table. In this case, they are both statistically significant at conventional levels.

Columns (9) and (10) present the estimates for another alternative regression setup where we employ country-year and gender fixed effects. This allows us to account for country-specific trends in the evolution of life expectancy and schooling. Because our potential mortality figures are not gender-specific, in this case we can not estimate the baseline effect of potential mortality in each country, as it is absorbed in the country-year fixed effect. Still it is possible to estimate the interaction effect between potential mortality and the female dummy. In line with our previous estimates, we find the estimated interaction coefficient, reported in column (9), to be negative, statistically significant and of similar magnitude as in column (1). In column (10) we see the same pattern for the estimated interaction coefficients of potential mortality from VP and NVP diseases with the female dummy. The former is statistically significant and the latter is not, with their magnitudes being similar to those estimated in column (2).

Turning to the corresponding reduce-form estimates in columns (9) and (10) of Table A4 leads to similar conclusions. Even when we include country-year fixed effects, the magnitudes and the statistical significance of the estimated coefficients for the interaction effects with the female dummy are very similar to those shown in columns (1) and (2). The similarity of our estimation across the different regression setups also applies to the 2SLS estimates reported in panel B of Table A3. The estimated coefficients are very stables across the different columns and support our conclusions regarding the nature of the relationship between life expectancy and years of schooling. This suggests that our empirical findings are robust to the particularities of the regression specification.

Section G: Multi-Period Panel Analysis

The results of the regression analysis presented in the paper and in the previous sections of the appendix are based on estimating a long-differences panel with two time periods. This two-period regression setup was chosen in order to focus on the long-run effects of the IET-related health improvements on educational attainment and to facilitate the comparability of our results with previous work in the literature. In this section, we extend our analysis to a multi-period regression setup in order to study more carefully how the relationship between health and education evolved

over time. For this purpose we expand the coverage of our data to construct a panel data set with decadal observations covering the period from 1940 to 2000.

The main equation that we want to estimate in this case is:

$$AYS_{gct} = \alpha_t \cdot LE_{gct} + \mu_{gc} + \gamma_t + u_{gct},$$

which is similar to equation (1) in the main body of the paper. As before, AYS_{gct} denotes the average years of schooling of a given cohort of gender g in country c which started school in year t and LE_{gct} denotes the life expectancy at birth of gender group g in country c in year t. The specification also includes gender-country fixed effects, μ_{gc}, and year fixed effects γ_t. The main difference compared to equation (1) is that we now allow the effect of life expectancy on schooling, α_t, to vary over time.

As in our main analysis, the above equation is estimated using two-stage least square (2SLS), where life expectancy is instrumented based on the following first-stage specification:

$$LE_{gct} = \sum_{s=1950}^{2000} \beta_s \cdot \sum_d M_{dc} \times I_t^s + \sum_{s=1950}^{2000} \beta_s^f \cdot I_g^f \cdot \sum_d M_{dc} \times I_t^s + \eta_{gc} + \delta_t + \varepsilon_{gct}.$$

The subscripts g, c and t again denote gender, country, and year, while the subscript d denotes different infectious diseases. Thus, M_{dc} is the mortality rate from infectious disease d in country c in 1940, at the start of the panel period, in negative numbers.[6] I_g^f is an indicator variable for all female observations and I_t^s is an indicator variable which equals one if $t = s$. Moreover, η_{gc} and δ_t denote gender-country and time fixed effects. The estimated coefficients β_s and β_s^f in this specification reflect the effect of the initial mortality environment on the life expectancy of males and females relative to the initial year 1940. The sum of mortality rates from the 13 infectious diseases $\sum_d M_{dc}$ is later on split into a group of vaccine-preventable diseases (VP) and a group of non-vaccine-preventable diseases (NVP), as in our main analysis.

Table A5 presents the 2SLS estimation results for the above-described multi-period panel specification. Panel A presents the estimates for the first stage regressions and panel B the corresponding estimates for the second-stage regressions. As in our main analysis, the mortality rates from infectious diseases in 1940 correspond to the whole population of each country. In column (1) we estimate the effect of the sum of initial mortality from all 13 infectious diseases without including any additional controls apart from the fixed effects. This effect is allowed to vary over time and across genders. Subcolumn (1a) presents the estimates for the baseline effect in each year and subcolumn (1b) the corresponding interaction effects with the female dummy. The baseline effect estimates are negative and statistically significant in all years. They also

[6] Using negative numbers leads to a negative relationship between changes in mortality rates and changes in life expectancy, as in the two-period panel of our baseline analysis.

increase in magnitude over time. This implies that countries that faced a larger disease burden in 1940 experienced larger increases in life expectancy subsequently and these increases were realized over several decades. The estimates for the female interaction effect are also negative in all periods and increase in absolute value over time, becoming statistically significant after 1970. This implies that the life expectancy gap between men and women increased gradually during the post-war decades.

[Insert Table A5 around here]

Turning to the second-stage estimation in panel B, we see that the relationship between changes in life expectancy and average years of schooling is positive and statistically significant in all years. Thus, life expectancy gains clearly contributed to increases in schooling. In this case the coefficient estimates do not follow a clear trend. This suggests that were no systematic differences in the relationship between life expectancy and educational attainment across different time periods. This observation, combined with the differential time paths of life expectancy of women and men, highlights why female schooling increased faster during the post-war period, leading to the eventual closing of the initial gender gap in education. It was largely the result of the relatively larger gains in female life expectancy.

In columns (2) and (3) of Table A5, we add controls to capture differences in initial health and initial schooling conditions across countries. This is in order to account for the possibility that the realized gains in life expectancy and schooling for different countries may have been a result of different initial conditions that were correlated with the initial mortality environment, as emphasized by Bloom et al. (2014). As a proxy for the initial health conditions in 1940 we use life expectancy at age 20.[7] As a proxy for initial schooling conditions in 1940 we use the average years of schooling of men and women who were 5-9 years old in 1940. Both variables are interacted with a dummy variable that equals one in the post-IET period, namely after 1950, to ensure that we allow the effects of these variables to vary over time.

Looking at the first-stage estimates in this case we see that initial health and initial schooling conditions are negatively correlated with life expectancy changes post 1940. This implies that countries that were initially better off in terms of health or schooling experienced on average smaller gains in life expectancy than countries that were less well off in 1940. Controlling for the initial health or initial schooling conditions also weakens the magnitudes of the baseline mortality effects, reported in subcolumns (2a) and (3a). This is not surprising as our measure of initial heath is correlated with mortality from infectious diseases in 1940 and therefore absorbs part of the variation present in the mortality rates. Nevertheless, controlling for initial health or initial schooling conditions does not affect our conclusions regarding the differential effects of the

[7] Using life expectancy at age 5 in 1940 or life expectancy at birth in 1900, as Acemoglu and Johnson, yields similar results.

potential mortality improvements on female and male life expectancy, as shown in subcolumns (2b) and (3b).

Turning to the second-stage estimates, we see that the effects of initial health and initial schooling are both negative, although for initial health the effect is statistically insignificant. This again suggests some degree of convergence taking place in educational attainment among countries that were initially more and less developed. Controlling for this effect in the second stage, however, we still observe a positive and significant relationship between life expectancy and schooling from 1950 onward.

In column 4 of Table A5 we split the mortality rates into mortality from disease of the VP group and the NVP group. Looking at the first-stage estimates, we see that an initially higher disease burden from both groups of infectious diseases was associated with significant increases in life expectancy relative to 1940, with the effect sizes increasing over time. Moreover, we see again a clear differential effect for females from 1970 onward. This effect, however, is statistically significant only for mortality from vaccine-preventable diseases, as in the case of our main analysis based on the two-period panel. The second-stage estimates in column 4 are very much aligned with those of columns 1, 2 and 3. Increases in life expectancy give rise to increases in schooling with the estimated magnitude of the effect not varying much over time.

Table A6 repeats the 2SLS estimations of Table A5 with the difference that the employed mortality rates in 1940 are now gender-specific. These gender-specific rates, constructed in the way explained in Section 6.2 of the paper, are imputed based on the gender ratios in mortality observed for each disease in 1940 in the U.S. Looking at the estimates of column 1, we see a familiar pattern. Both the estimates for the baseline mortality effect and its interaction effect with the female dummy are negative and statistically significant. Moreover, the estimated magnitudes for the interaction effects are larger than in column 1 of Table A5 and significant in all years. This corroborates the findings of Tables 4 and 10 in the paper, which already indicated that accounting for initial mortality differences between men and women leads to a more pronounced differential effect of IET-related mortality improvements on female life expectancy.

[Insert Table A6 around here]

This pattern is also visible even when we control for initial health and initial schooling conditions, as we do in columns (2) and (3). In column 4 we alternatively include as a control a gender time-trend. The inclusion of this gender–specific time trend filters out common trends in mortality rates across countries within the same gender group and partially weakens the estimated magnitudes for the interaction effect of mortality with the female dummy. Nevertheless, we still see a clear differential effect of the potential mortality improvements on female life expectancy after 1970. In column 5 we split again the mortality rates into mortality from diseases of the VP group and diseases of the NVP group. The estimated effects are similar to those of column 4 in

Table A5 with the difference that the estimated magnitudes for the female interaction terms are again larger in absolute terms.

Turning to panel B where we report the second stage estimates, we see similar patterns as in Table A5. Across all columns we see that the effect of life expectancy on average years of schooling is positive. The coefficient estimates also do not vary much across periods and they are overall similar to the ones we found in the two-period panel analysis. Therefore, we can say that our estimates from the multi-period panel analysis also confirm the importance of the IET in raising female schooling relative to that of males in the post-war era. This relative rise in female schooling can be clearly associated with the larger improvements in female health that were realized over this period. Also the observed differential effect between males and females can be largely attributed to the more effective control of mortality from the group of vaccine-preventable infectious diseases.

References

Acemoglu, Daron, and Simon Johnson. 2007. "Disease and Development: The Effect of Life Expectancy on Economic Growth." *Journal of Political Economy*, 115(6): 925–985.

Baker, Jeffrey P., and Samuel L. Katz. 2004. "Childhood Vaccine Development: An Overview." *Pedriatric Research*, 55(2): 347–356.

Barro, Robert J., and Jong Wha Lee. 2013. "A New Data Set of Educational Attainment in the World, 1950-2010." *Journal of Development Economics*, 104: 184–198.

Bloom, David E., David Canning, and Guenther Fink. 2014. "Disease and Development Revisited." *Journal of Political Economy*, 122(6): 1355–1366.

Bloom, David E., David Canning, Guenther Fink, and Jocelyn Finlay. 2009. "Fertility, Female Labor Force Participation, and the Demographic Dividend." *Journal of Economic Growth*, 14(2): 79–101.

Bonah, Christian. 2005. "The 'Experimental Stable' of the BCG Vaccine: Safety, Efficacy, Proof, and Standards, 1921-1933." *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(4): 696–721.

Cervellati, Matteo, and Uwe Sunde. 2011. "Life Expectancy and Economic Growth: The Role of the Demographic Transition." *Journal of Economic Growth*, 16(2): 99–133.

Comstock, George W. 1994. "The International Tuberculosis Campaign: A Pioneering Venture in Mass Vaccination and Research." *Clinical Infectious Diseases*, 19(3): 528–540.

Federal Security Agency. 1947. *Summary of International Vital Statistics, 1937-1944.* Washington, D.C.:U.S. Public Health Services, National Office of Vital Statistics.

Fenner, Frank. 2011. "Smallpox Eradication: the Vindication of Jenner's Prophecy." In *History of Vaccine Development.* , ed. Stanley A. Plotkin, Chapter 5, 27–32. New York, NY:Springer.

Fine, Paul E. M. 1995. "Bacille Calmette-Guerin Vaccines: A Rough Guide." *Clinical Infectious Diseases*, 20(1): 11–14.

Granstroem, Marta. 2011. "The History of Pertussis Vaccination: From Whole-Cell to Subunit Vaccines." In *History of Vaccine Development.* , ed. S. Plotkin and B. Fantini. Paris:Elsevier.

Hoyt, Kendall. 2006. "Vaccine Innovation: Lessons from World War II." *Journal of Public Health Policy*, 27(1): 38–57.

International Labour Organization. 2018. "Database of Labour Statistics (ILOSTAT)." www.ilo.org/ilostat.

Kiple, Kenneth F., ed. 2003. *The Cambridge Historical Dictionary of Diseases.* Cambridge:Cambridge University Press.

Klein, Jerome O., and Stanley A. Plotkin. 2007. "Robert Austrian: 1917-2007." *Clinical Infectious Diseases*, 45(1): 2–3.

Lindenmann, Jean. 2002. "Typhus Vaccine Developments from the First to the Second World War." *History and Philosophy of the Life Sciences*, 24(3-4): 467–485.

Maddison, Angus. 2013. "The Maddison-Project." http://www.ggdc.net/maddison/maddison-project/home.htm.

Marshall, Monty G., Ted Robert Gurr, and Keith Jaggers. 2017. "Political Regime Characteristics and Transitions, 1800-2016." *Polity IV Project, Center for Systemic Peace.*

NCD Risk Factor Collaboration. 2016. "A Century of Trends in Adult Human Height." *eLife*, 5: e13410.

Parish, Henry J. 1965. *A History of Immunization.* London:E. & S. Livingstone Ltd.

Plotkin, Stanley, Walter Orenstein, and Paul Offit. 2012. *Vaccines.* . 6 ed., Saunders.

Preston, Samuel H. 1975. "The Changing Relation between Mortality and Level of Economic Development." *Population Studies*, 29(2): 231–248.

Preston, Samuel H., Nathan Keyfitz, and Robert Schoen. 1972. *Causes of Death: Life Tables for National Populations.* New York, NY:Seminar Press.

Ramirez, Francisco O., Yasemin Soysal, and Suzanne Shanahan. 1997. "The Changing Logic of Political Citizenship: Cross-National Acquisition of Women's Suffrage Rights, 1890 to 1990." *American Sociological Review*, 62(5): 735–745.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. 2019. "IPUMS USA: Version 9.0."

Ruuskanen, Olli, Elina Lahti, Lance C. Jennings, and David R. Murdoch. 2011. "Viral Pneumonia." *Lancet*, 377(9773): 1264–1275.

UNESCO. 1991. "World Education Report." Paris.

UNESCO. Various Editions. "International Yearbook of Education." Gevena.

United Nations. Various Editions. "Demographic Yearbook." New York, NY.

United Nations Population Division. 2018*a*. "World Population Prospects." esa.un.org/unpd/wpp/.

United Nations Population Division. 2018*b*. "World Urbanization Prospects." esa.un.org/unpd/wup/.

U.S. Census Bureau. 1940. *Vital Statistics of the United States.* Washington, D.C.:United States Government Printing Office.

World Health Organization. 1951. *Annual Epidemiological and Vital Statistics, 1939-46.* Geneva:World Health Organization.

World Health Organization. 2010. "Cholera Vaccines: WHO Position Paper." *Weekly Epidemiological Record*, 85(13): 117–128.

World Health Organization. Various Editions. "World Health Statistics Annuals." Geneva.

Table A1: Robustness Checks to Sample Composition

	(1)	(2)	(3)	(4)	(5)	(6)
Sample	Omit countries involved in drug development	Omit African & Middle Eastern countries	Omit small Latin American countries	Omit Eastern European countries	Omit developed countries	All Barro-Lee countries
Panel A: 1st stage results	**Life expectancy at birth**					
Mortality, VP-group	-10.592***	-11.440***	-11.692***	-11.672***	-5.921***	
	[2.451]	[2.555]	[2.507]	[2.504]	[2.120]	
Female x Mortality, VP-group	-4.853**	-5.154*	-4.852**	-4.623*	-4.553**	
	[2.453]	[2.655]	[2.437]	[2.580]	[2.164]	
Mortality, NVP-group	-17.218***	-18.758***	-21.431***	-18.093***	-12.567***	
	[4.665]	[4.588]	[5.960]	[4.692]	[4.757]	
Female x Mortality, NVP-group	-5.704	-5.341	-6.569	-5.858	-5.820	
	[6.856]	[6.811]	[9.123]	[6.854]	[7.017]	
Panel B: 2nd stage results	**Avg. years of schooling**					
Life expectancy at birth	0.074*	0.115***	0.138***	0.099**	0.094**	0.152***
	[0.040]	[0.0425]	[0.042]	[0.043]	[0.046]	[0.016]
Observations	280	272	280	268	248	580
Countries	70	68	70	67	62	145
Effective F-Statistic	7.63	8.53	8.86	8.30	4.18	-

Notes: This table explores the robustness of our baseline estimates to changes in the sample composition. Column (1) omits France, Germany, Switzerland, the United Kingdom, the United States. Column (2) omits Algeria, Egypt, Iran, Iraq, Morocco, South Africa, Tunisia. Column (3) omits Barbados, Belize, Guyana, Jamaica, Trinidad and Tobago. Column (4) omits Bulgaria, Czech Republic, Hungary, Poland, Romania, Russian Federation, Slovakia, Yugoslavia. Column (5) omits Argentina, Australia, Belgium, Canada, Denmark, France, Germany, the Netherlands, New Zealand, Sweden, Switzerland, the United Kingdom, the United States. Column (6) presents OLS estimates for the full Barro-Lee sample of countries. All regressions include gender-country and year fixed effects. Heteroskedasticity robust standard errors clustered at the gender-country level are reported in brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table A2: Robustness Checks for Prior Trends and Alternative Sample Periods

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Specification	Lagged life expectancy 1900/1940	Lagged mort. and life exp. 1900/1940	Lagged schooling 1900/1940	Controlling for lagged schooling	Sample Period 1930/1980	Sample Period 1980/2000	Lead schooling 1980/2000	Lead life expectancy 1980/2000
Panel A: 1st stage results	**Life expectancy at birth**							
Mortality, VP-group	9.285**	6.993	-11.538***	-11.520***	-11.737***	-5.647***	-11.538***	-5.647***
	[4.541]	[8.066]	[2.410]	[2.379]	[2.763]	[1.069]	[2.410]	[1.069]
Female x Mortality, VP-group	-5.605	-8.479	-4.959**	-4.544*	-6.770**	-0.711	-4.959**	-0.711
	[4.511]	[7.206]	[2.440]	[2.418]	[2.862]	[1.139]	[2.440]	[1.139]
Mortality, NVP-group	38.058**	0.726	-18.534***	-18.282***	-19.895**	-1.983	-18.534***	-1.983
	[17.504]	[0.691]	[4.588]	[4.742]	[8.227]	[2.773]	[4.588]	[2.773]
Female x Mortality, NVP-group	12.061	0.340	-5.618	-5.628	-1.874	-0.807	-5.618	-0.807
	[22.652]	[0.961]	[6.799]	[7.110]	[9.863]	[3.755]	[6.799]	[3.755]
Lagged schooling				0.582*				
				[0.344]				
Panel B: 2nd stage results	**Avg. years of schooling**							
Life expectancy at birth	-0.158***	-0.066	0.032	0.122***	0.199***	0.040	0.034	0.329***
	[0.061]	[0.049]	[0.023]	[0.040]	[0.045]	[0.072]	[0.027]	[0.111]
Lagged schooling				-0.230				
				[0.153]				
Observations	128	128	300	300	156	300	300	300
Countries	32	32	75	75	39	75	75	75
Effective F-Statistic	3.02	5.92	9.17	8.60	8.53	4.16	9.17	4.16

Notes: This table provides a comparison of the estimates for our baseline specification when testing for the existence of prior trends and looking at alternative sample periods. The exact modification to the variables and the sample period considered in each specification are indicated at the top of each column. All regressions include gender-country and year fixed effects. Heteroskedasticity robust standard errors clustered at the gender-country level are reported in brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table A3: Life Expectancy and Educational Attainment: 2SLS Estimates with Different Specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Sample Composition	Both Genders		Males		Females		Both Genders			
Fixed Effects	Gender x Country and Time		Country & Time				Gender x Time and Country		Country x Time and Gender	
Panel A: 1st stage results	**Life expectancy at birth**									
Mortality, overall	-3.749***		-3.749***		-5.168***		-3.764***			
	[0.628]		[0.408]		[0.471]		[0.638]			
Female x Mortality, overall	-1.420**						-1.660		-1.400***	
	[0.624]						[1.034]		[0.154]	
Mortality, VP-group		-2.854***		-2.854***		-4.080***		-2.947***		
		[0.596]		[0.405]		[0.449]		[0.609]		
Female x Mortality, VP-group		-1.227**						-1.330		-1.065**
		[0.603]						[1.009]		[0.414]
Mortality, NVP-group		-2.006***		-2.006***		-2.614***		-2.024***		
		[0.497]		[0.480]		[0.560]		[0.481]		
Female x Mortality, NVP-group		-0.608						-0.622		-0.798
		[0.736]						[0.734]		[0.516]
Panel B: 2nd stage results	**Avg. years of schooling**									
Life expectancy at birth	0.115***	0.114***	0.117*	0.117*	0.117**	0.116**	0.119***	0.114***	0.112***	0.110***
	[0.043]	[0.041]	[0.066]	[0.064]	[0.057]	[0.056]	[0.045]	[0.042]	[0.032]	[0.032]
Observations	300	300	300	300	296	300	300	300	300	300
Countries	75	75	75	75	74	75	75	75	75	75
Effective F-Statistic	15.21	9.17	13.61	8.15	13.73	8.46	13.6	9.18	20.81	9.09

Notes: This table explores the robustness of our baseline 2SLS estimates for the relationship between changes in life expectancy at birth and average years of schooling between 1940 and 1980 to variation in our regression specification. Panel A shows the results of the first-stage estimation where life expectancy is instrumented by potential mortality rates from different groups of infectious diseases. Panel B shows the results of the second-stage estimation. In all cases we employ the potential mortality instrument where the 1980 mortality rates are set to zero in all countries. For comparability across columns all mortality variables are standardized to have a zero mean and a unitary standard deviation. Columns (1) and (2) present the estimates for our preferred specification using gender-country and year fixed effects. Columns (3) to (6) present the estimates for our main specification estimated separately first for males and then for females using country and year fixed effects. Columns (7) and (8) present the estimates for our main specification using country and gender-year fixed effects. Columns (9) and (10) present the estimates for our main specification using country-year and gender fixed effects. Heteroskedasticity robust standard errors clustered at the gender-country level are reported in brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table A4: Life Expectancy and Educational Attainment: Reduced-Form Estimates with Different Specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Sample Composition	Both Genders		Males		Females		Both Genders			
Fixed Effects	Gender x Country and Time		Country & Time				Gender x Time and Country		Country x Time and Gender	
Panel A: 1st stage results	**Avg. years of schooling**									
Mortality, overall	-0.458*		-0.458***		-0.586***		-0.440*			
	[0.254]		[0.171]		[0.173]		[0.237]			
Female x Mortality, overall	-0.128**						-0.227**		-0.157***	
	[0.050]						[0.114]		[0.048]	
Mortality, VP-group		-0.405*		-0.405**		-0.508***		-0.373*		
		[0.240]		[0.178]		[0.176]		[0.222]		
Female x Mortality, VP-group		-0.103**						-0.209*		-0.131***
		[0.051]						[0.111]		[0.050]
Mortality, NVP-group		-0.186		-0.186		-0.259		-0.189		
		[0.171]		[0.220]		[0.237]		[0.154]		
Female x Mortality, NVP-group		-0.073						-0.070		-0.063
		[0.058]						[0.073]		[0.050]
Observations	300	300	300	300	296	300	300	300	300	300
Countries	75	75	75	75	74	75	75	75	75	75

Notes: This table explores the robustness of our baseline reduced-form estimates for the relationship between changes in life expectancy at birth and average years of schooling between 1940 and 1980 to variation in our regression specification. In all cases we employ the potential mortality instrument where the 1980 mortality rates are set to zero in all countries. For comparability across columns all mortality variables are standardized to have a zero mean and a unitary standard deviation. Columns (1) and (2) present the estimates for our preferred specification using gender-country and year fixed effects. Columns (3) to (6) present the estimates for our main specification estimated separately first for males and then for females using country and year fixed effects. Columns (7) and (8) present the estimates for our main specification using country and gender-year fixed effects. Columns (9) and (10) present the estimates for our main specification using country-year and gender fixed effects. Heteroskedasticity robust standard errors clustered at the gender-country level are reported in brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table A5: Multi-Period Panel Estimates with Population-Wide Mortality

	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(4c)	(4d)
Panel A: 1st Stage	**Life expectancy at birth**									
Explanatory Variable	Mort.	MortxFem.	Mort.	MortxFem.	Mort.	MortxFem.	M-VP	M-VPxFem..	M-NVP	M-NVPxFem.
Year 1950	-6.320***	-1.547	2.130	-3.983*	-3.767	-1.172	-4.788**	-1.237	-11.235***	-3.045
	[1.887]	[1.950]	[2.412]	[2.223]	[2.383]	[2.116]	[1.989]	[2.175]	[3.138]	[3.887]
Year 1960	-8.966***	-2.037	0.011	-4.826**	-6.413**	-1.662	-7.251***	-1.777	-14.533***	-3.294
	[2.117]	[2.131]	[2.424]	[2.271]	[2.540]	[2.247]	[2.148]	[2.293]	[4.531]	[6.004]
Year 1970	-11.350***	-3.174	-2.334	-6.185***	-8.797***	-2.799	-9.568***	-3.117	-17.291***	-3.450
	[2.009]	[2.038]	[2.153]	[2.221]	[2.318]	[2.084]	[2.063]	[2.204]	[4.596]	[6.458]
Year 1980	-13.162***	-5.072**	-4.232*	-7.737***	-10.610***	-4.697**	-11.538***	-4.959**	-18.534***	-5.618
	[2.247]	[2.207]	[2.321]	[2.337]	[2.534]	[2.224]	[2.425]	[2.455]	[4.616]	[6.841]
Year 1990	-15.780***	-5.750**	-6.803***	-8.599***	-13.227***	-5.375**	-14.491***	-5.657**	-20.038***	-6.205
	[2.318]	[2.263]	[2.486]	[2.533]	[2.542]	[2.229]	[2.626]	[2.628]	[4.767]	[6.936]
Year 2000	-17.972***	-5.799**	-9.491***	-8.582***	-15.419***	-5.424**	-17.186***	-5.670**	-20.517***	-6.426
	[2.407]	[2.360]	[2.917]	[2.784]	[2.573]	[2.282]	[2.852]	[2.868]	[5.054]	[7.369]
Control Var.	None		Initial LE		Initial Sch.		None			
	-		-0.563***		-0.665***		-			
			[0.088]		[0.175]					
Panel B: 2nd Stage	**Avg. years of schooling**									
LE0 x Year 1940	0.141***		0.101		0.053		0.134***			
	[0.042]		[0.094]		[0.041]		[0.039]			
LE0 x Year 1950	0.172***		0.265***		0.149***		0.162***			
	[0.048]		[0.054]		[0.043]		[0.045]			
LE0 x Year 1960	0.200***		0.293***		0.172***		0.188***			
	[0.053]		[0.062]		[0.045]		[0.049]			
LE0 x Year 1970	0.198***		0.290***		0.165***		0.187***			
	[0.062]		[0.068]		[0.053]		[0.057]			
LE0 x Year 1980	0.166**		0.239***		0.129**		0.153**			
	[0.067]		[0.071]		[0.056]		[0.062]			
LE0 x Year 1990	0.157**		0.203**		0.115*		0.143**			
	[0.078]		[0.082]		[0.065]		[0.070]			
LE0 x Year 2000	0.229***		0.296***		0.184**		0.191**			
	[0.089]		[0.093]		[0.074]		[0.079]			
Control Variable	None		Initial LE		Initial Sch.		None			
	-		-0.223		-0.547***		-			
			][0.190]		[0.105]					
Observations	1,050		742		1,050		1,050			
Countries	75		53		75		75			
Cragg-Donald F-Statistic	13.374		6.302		14.866		7.136			

Notes: This table presents 2SLS estimates of the relationship between changes in life expectancy at birth and average years of schooling based on a multi-period panel. Panel A shows the results of the first-stage estimation where life expectancy is instrumented by mortality rates from different groups of infectious diseases in 1940 interacted with a period dummy. Panel B shows the results of the second-stage estimation where we allow the effect of life expectancy on educational attainment to vary across periods. Columns marked with the same number in Panel A but a different letter provide estimates from the same regression for the different variables indicated at the top. All regressions include gender-country and year fixed effects. Heteroskedasticity robust standard errors clustered at the gender-country level are reported in brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table A6: Multi-Period Panel Estimates with Gender-Specific Mortality

	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)	(5c)	(5d)
Panel A: 1st Stage	**Life expectancy at birth**											
Explanatory Variable	Mort.	MortxFem.	Mort.	MortxFem.	Mort.	MortxFem.	Mort.	MortxFem.	M-VP	M-VPxFem.	M-NVP	M-NVPxFem.
Year 1950	-5.591***	-3.457*	1.818	-3.935	-3.335	-2.380	-6.581***	-1.221	-4.193**	-2.809	-10.515***	-4.800
	[1.676]	[2.060]	[2.125]	[2.397]	[2.108]	[2.277]	[2.146]	[3.823]	[1.751]	[2.342]	[2.918]	[3.884]
Year 1960	-7.927***	-4.710**	-0.040	-5.492**	-5.671**	-3.633	-8.916***	-2.474	-6.348***	-4.156*	-13.619***	-5.462
	[1.881]	[2.272]	[2.144]	[2.508]	[2.249]	[2.438]	[2.331]	[4.017]	[1.899]	[2.479]	[4.208]	[6.046]
Year 1970	-10.022***	-6.632***	-2.087	-7.734***	-7.766***	-5.556**	-11.011***	-4.396	-8.369***	-6.405***	-16.237***	-5.910
	[1.791]	[2.193]	[1.906]	[2.470]	[2.058]	[2.285]	[2.262]	[3.877]	[1.834]	[2.390]	[4.260]	[6.557]
Year 1980	-11.626***	-9.279***	-3.754*	-10.067***	-9.370***	-8.202***	-12.615***	-7.043*	-10.093***	-9.150***	-17.453***	-8.256
	[2.004]	[2.372]	[2.050]	[2.604]	[2.252]	[2.428]	[2.459]	[3.995]	[2.148]	[2.652]	[4.277]	[7.004]
Year 1990	-13.940***	-10.724***	-6.020***	-11.776***	-11.684***	-9.647***	-14.929***	-8.487**	-12.672***	-10.838***	-18.924***	-8.927
	[2.071]	[2.427]	[2.192]	[2.821]	[2.264]	[2.428]	[2.499]	[3.985]	[2.328]	[2.835]	[4.421]	[7.066]
Year 2000	-15.881***	-11.319***	-8.378***	-12.494***	-13.625***	-10.242***	-16.870***	-9.082**	-15.024***	-11.629***	-19.421***	-9.106
	[2.154]	[2.511]	[2.571]	[3.063]	[2.299]	[2.468]	[2.549]	[3.986]	[2.527]	[3.079]	[4.701]	[7.473]
Control Variable	None		Initial LE		Initial Sch.		Gender Time-Trend		None			
	-		-0.559***		-0.664***		1.305		-			
			[0.088]		[0.175]		[1.507]					
Panel B: 2nd Stage	**Avg. years of schooling**											
LE0 x Year 1940	0.142***		0.101		0.053		0.156***		0.134***			
	[0.043]		[0.095]		[0.041]		[0.044]		[0.039]			
LE0 x Year 1950	0.173***		0.267***		0.150***		0.198***		0.162***			
	[0.049]		[0.054]		[0.043]		[0.049]		[0.045]			
LE0 x Year 1960	0.201***		0.295***		0.175***		0.231***		0.189***			
	[0.053]		[0.062]		[0.046]		[0.053]		[0.049]			
LE0 x Year 1970	0.198***		0.292***		0.167***		0.233***		0.188***			
	[0.063]		[0.069]		[0.054]		[0.062]		[0.058]			
LE0 x Year 1980	0.167**		0.242***		0.131**		0.206***		0.154**			
	[0.068]		[0.072]		[0.057]		[0.066]		[0.062]			
LE0 x Year 1990	0.157**		0.205**		0.117*		0.202***		0.143**			
	[0.079]		[0.083]		[0.066]		[0.077]		[0.071]			
LE0 x Year 2000	0.231**		0.300***		0.187**		0.281***		0.192**			
	[0.090]		[0.094]		[0.075]		[0.087]		[0.079]			
Control Var.	None		Initial LE		Initial Sch.		Gender Time-Trend		None			
	-		-0.226		-0.552***		-0.454*		-			
			[0.192]		[0.105]		[0.244]					
Observations	1,050		742		1,050		1,050		1,050			
Countries	75		53		75		75		75			
Cragg-Donald F-Stat.	13.226		6.243		14.672		9.853		7.166			

Notes: This table presents 2SLS estimates of the relationship between changes in life expectancy at birth and average years of schooling based on a multi-period panel. Panel A shows the results of the first-stage estimation where life expectancy is instrumented by gender-specific mortality rates from different groups of infectious diseases in 1940 interacted with a period dummy. Panel B shows the results of the second-stage estimation where we allow the effect of life expectancy on educational attainment to vary across periods. Columns marked with the same number in Panel A but a different letter provide estimates from the same regression for the different variables indicated at the top. All regressions include gender-country and year fixed effects. Heteroskedasticity robust standard errors clustered at the gender-country level are reported in brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.