Courant Research Centre

'Poverty, Equity and Growth in Developing and Transition Countries: Statistical Methods and Empirical Analysis'

Georg-August-Universität Göttingen
(founded in 1737)

Discussion Papers

No. 292

The long-term consequences of the global 1918 influenza pandemic: A systematic analysis of census data from 51 countries

Juditha Wójcik, Christian Bommer, Dominik Naeher, Sebastian Vollmer

First version: November 2022
This version: May 2023

Platz der Göttinger Sieben 5 · 37073 Goettingen · Germany
Phone: +49-(0)551-3921660 · Fax: +49-(0)551-3914059

Email: crc-peg@uni-goettingen.de Web: http://www.uni-goettingen.de/crc-peg

The long-term consequences of the global 1918 influenza pandemic: A systematic analysis of census data from 51 countries

Juditha Wójcik, University of Mainz (ewojcik@uni-mainz.de)

Christian Bommer, University of Goettingen (christian.bommer@wiwi.uni-goettingen.de)

Dominik Naeher, University of Goettingen (dnaeher@uni-goettingen.de)

Sebastian Vollmer*, University of Goettingen (svollmer@uni-goettingen.de)

*Corresponding author:
Prof. Dr. Sebastian Vollmer
University of Goettingen
Centre for Modern Indian Studies (CeMIS)
Waldweg 26, 37073 Goettingen, Germany
Email: svollmer@uni-goettingen.de

Abstract

Several country-level studies, including a prominent one for the United States, have identified long-term adverse effects of in-utero exposure to the 1918 influenza pandemic (also known as the Spanish Flu) on economic outcomes. Although the Spanish Flu was a global phenomenon, with an estimated 500 million people infected worldwide, no comprehensive global study on its long-term economic effects exists. We address this gap by systematically analyzing harmonized census data from 51 countries. Using the same empirical approach as previous studies, we find no evidence of consistent long-term effects on educational attainment and employment across countries. Overall, our results are difficult to reconcile with the view that in-utero exposure to the 1918 influenza pandemic was associated with systematic long-term adverse effects on economic outcomes at the population level. A comprehensive set of robustness checks do not alter this conclusion.

Keywords: Spanish Flu, 1918 Influenza Pandemic, Fetal Origins Hypothesis
JEL Classification: I15, N30, O57

1 Introduction

In recent years, research on the impact of health conditions on economic outcomes has expanded to also investigate how conditions before birth affect individuals' life paths, thereby linking fetal shocks to (economic) outcomes in adulthood. The most prominently cited hypothesis in this context is the fetal origins hypothesis first popularized by British doctor David Barker, who postulated that severe health conditions such as heart disease and diabetes in later adulthood could be linked to the in-utero environment (especially maternal nutritional deprivation) to which the fetus was exposed (Barker 1998). Although Barker's hypothesis was initially met with skepticism in the medical literature (Huxley et al. 2002; Adair and Prentice 2004), it was quickly discovered by economists and used to evaluate the economic consequences of various fetal shocks (Almond and Currie 2011).

In a seminal paper, Almond (2006) first assessed the in-utero impact of the 1918 influenza pandemic on later-life economic outcomes. Based on observational data from three census waves in the United States, he compared long-term health and economic outcomes of cohorts exposed to the pandemic with those of surrounding cohorts. Despite the plausibility of the fetal origins hypothesis on an individual level, it is not guaranteed that such a population-level analysis would reveal significant impacts. First of all, the number of affected individuals needs to be sufficiently large to lead to detectable effects at the population level. Moreover, negative health shocks may not only deteriorate average health, but also increase the survival threshold, leading to a so-called ‘culling effect’ that may offset (or even exceed) any negative impacts on average health (Valente 2015). Nevertheless, Almond (2006) found the studied cohorts to be significantly less likely to graduate from high school and to have lower average income, lower socioeconomic status, and higher risk of being disabled (similar results are also reported by Garthwaite 2008 and Fletcher 2014).

Following Almond (2006), a number of other studies have expanded the investigation to other countries. Neelsen and Stratmann (2012) find that male Swiss birth cohorts exposed to the 1918 influenza pandemic feature lower educational attainment and are less likely to be married compared to the common trend. Lin and Liu (2014) find that Taiwanese cohorts exposed to the pandemic feature lower average educational attainment, are smaller during puberty and more susceptible to severe health conditions such as kidney disease and diabetes in later adulthood compared to surrounding cohorts. Karlsson et al. (2012) find Swedish cohorts exposed to the pandemic to experience elevated poverty rates. Nelson (2010) assesses the effect of the 1918 influenza pandemic for six metropolitan areas in Brazil and finds that, on average, cohorts prenatally exposed to the pandemic are less likely to have graduated from college, have

fewer years of schooling, are less likely to be employed, and earn lower average wages.

As these studies argue, influenza is a particularly suitable case for investigating the long-term effects of in-utero environment as exposure is quasi-random. Specifically, influenza is common in human populations and exposure across age cohorts does not seem to be determined by socio-economic characteristics (Neelsen and Stratmann 2012). The global spread of the 1918 influenza pandemic was rapid and unanticipated, causing exogenous variation in fetal health between cohorts exposed to influenza in utero and those born shortly before and shortly after. More virulent forms of influenza characterized by a high number of infections and deaths occur every once in a while, leading to pandemics such as in 1889/90, 1918/19, 1957/8, 1968/9, and 1977/8. The influenza pandemic of 1918, often called the 'Spanish Flu', spread around the globe within a few months, killing a multiple of the casualties of World War I and sparing only a few remote regions.[1] New modes of transportation of the era such as steamships and railways as well as the large movements of troops and civilians during the war facilitated the global spread of the pandemic. In most regions, the diffusion happened along major transportation routes. Coastal countries were typically infected first through incoming ships carrying ill passengers or crews, but even remote areas in sub-Saharan Africa got infected. As influenza was not a reportable disease, patients were not detained and, hence, the pandemic spread largely unhindered (Patterson and Pyle 1991; Killingray and Phillips 2003).

Despite the global spread of the 1918 pandemic, with an estimated 500 million people infected worldwide (Taubenberger and Morens 2006), we are not aware of any previous study investigating the long-term effects of the 1918 influenza pandemic globally or across different world regions. This is particularly relevant because it is possible that single-country studies focusing on countries with significant long-term effects were more likely published than studies on countries without statistically significant effects, leading to an incomprehensive picture and possibly false conclusions regarding the global long-term impacts of the Spanish Flu.

This paper addresses this gap in the literature by systematically analyzing census data from 51 countries to investigate the long-term effects of the 1918 influenza pandemic following an approach akin to those used in previous studies. Our analysis is based on data provided by IPUMS International (Minnesota Population Center 2014). The main advantage over other sources of census data is that IPUMS International provides harmonized indicators that allow for international comparisons across countries and time. Following Almond (2006), we evaluate

[1] As Spain was a neutral power during the war, newspapers were uncensored and, hence, articles of the disease and its spread were common whereas belligerent countries kept taps on their reports to avoid mass panics. This is usually considered the reason why this pandemic is referred to as the Spanish Flu (Killingray and Phillips 2003, Almond 2006).

the performance of the 1919 birth cohort against a yearly quadratic trend with respect to different outcomes in adulthood. Given limited data availability for other indicators, we mainly focus on three dependent variables, namely completion of primary education, completion of secondary education, and employment status at the time of enumeration (general and work-related disability are used as alternative outcomes in a smaller sample as part of our robustness checks).

Overall, we find no evidence of systematic adverse effects of the 1918 influenza pandemic on human capital accumulation or employment for cohorts exposed to the pandemic in utero. Rather, our analysis suggests that for the vast majority of the considered countries there were no significant effects at the population level.[2] Importantly, among those countries for which there are statistically significant effects, we find both beneficial and adverse effects to a similar extent. While this does not necessarily challenge the internal validity of previous studies finding significant adverse effects for individual countries, it does cast doubt on the existence of a systematic adverse effect on economic outcomes on a global level.

Much of our analysis is concerned with verifying the robustness of these findings. In particular, a potential concern with our analysis is that the absence of statistically significant results may be partially due to measurement error stemming from imprecise information in some countries about the time of birth and timing of the pandemic. We carefully address this concern in a series of robustness checks. Most importantly, we show that our key findings also apply when restricting the sample to those censuses for which the expected degree of error is weakly smaller than the one in the analysis of the United States performed in Almond (2006). In addition, we show that our results are robust to different choices in constructing the comparison group, to different assumptions about the exact timing of the pandemic in each country, and to controlling for potentially confounding effects of WWI and of the less lethal first and third waves of the pandemic.

Apart from the question of measurement, another concern may be that contextual features (such as mean educational attainment or employment rates) across countries are too heterogeneous, so that pooling advanced and less advanced countries may bias results towards zero. Such concerns are, however, unwarranted as our conclusions do not simply rely on the reported small and mostly insignificant average weighted coefficients. Rather, we find that the vast majority of census-specific estimates are statistically insignificant.

Our findings may at first appear surprising, given that there is now a well-established

[2] This finding does not rule out the possibility that in-utero exposure to influenza may have had meaningful effects for some individuals.

body of evidence on the fetal origins hypothesis, including evidence on economic outcomes obtained from natural experiments other than the Spanish Flu, such as in-utero effects of adverse weather conditions (Maccini and Yang 2009), natural disasters (Caruso and Miller 2015), maternal malaria (Barreca 2010), and Ramadan observance during pregnancy (Almond and Mazumder 2011). On the other hand, a recent re-evaluation of the 1918 influenza pandemic in the United States by Beach et al. (2022) shows that using a refined approach that controls for selection effects arising from the coincidence of the pandemic and World War I (WWI) conscription causes the effects identified in Almond (2006) to become statistically insignificant.[3]Similarly, Helgertz and Bengtsson (2019) show that a more careful analysis of the 1918 influenza pandemic in Sweden in fact provides no evidence of long-run adverse effects of fetal exposure, and Brown and Thomas (2018) provide critical reviews of the results from Brazil, Switzerland, and Taiwan mentioned above. Moreover, as noted by Beach et al. (2022), there also exists a substantial body of evidence suggesting that early life disadvantages do not automatically and necessarily lead to worse adult economic outcomes, providing a possible explanation for the absence of systematic long-term adverse effects for cohorts exposed in utero to health shocks like the Spanish Flu (see also Heckman 2006). While our study aims at contributing in a different direction – reviewing the external validity of previous findings rather than their internal validity, – this pattern of mixed evidence may be seen as making our results appear less puzzling.[4]

The rest of the paper is structured as follows. The next section provides a brief historical background of the pandemic. Section 3 describes our data and identification strategy. Sections 4 present our main findings, including a replication of some of the key results from Almond (2006) using our dataset. Section 5 discusses the robustness and limitations of our results. Section 6 concludes.

2 Historical Background: The 1918 Influenza Pandemic

The 1918 influenza pandemic is commonly thought of as having occurred in three waves, the first wave being a precursor to the deadly second wave and receiving only minor public

[3] When examining the effects of in-utero influenza exposure on socioeconomic status (SES), these authors show that the adult SES deficit is reduced when characteristics such as race, birthplace, and parents' birth countries are controlled for, and becomes statistically insignificant when household fixed effects are included. One of their conclusions is that, "*Replicating Almond's state-level dose-response analysis, we find no evidence in census data that influenza exposure reduced adult SES*." (Beach et al. 2022, p. 1964).

[4] Also note that, given the concerns about the validity of the cross-cohort identification strategy raised by Beach et al. (2022), we go beyond our baseline analysis (which is indeed designed to mirror the empirical strategy of Almond 2006) and also consider a set of carefully designed robustness checks (including several ways of controlling for the potentially confounding effects of WWI), without altering the main conclusions.

attention in 1918 (Patterson and Pyle 1991).[5] It is usually assumed that the virus of the first wave mutated leading to a much more virulent and deadly virus of the second wave, terming it 'the pandemic' (Killingray and Phillips 2003). The third wave is usually described as a mild aftermath of the second wave or as "episodic and scattered winter outbreaks" (Patterson and Pyle 1991, p.4) commonly observed after epidemics and without larger impact on mortality trends. A striking characteristic of the 1918 influenza pandemic is the unusually high mortality rate among young adults observed in many countries (Johnson and Mueller 2002).

The literature typically cites Brest in France in August 1918 as the most likely point of origin of the mutated virus, at the time a major port of entry for American troops joining the war in Europe. From there, ships and trains carrying troops and cargo spread the virus around the globe within months. The British ship 'HMS Mantua' arriving in Freetown, Sierra Leone, on August 15, 1918, with 200 sick sailors brought influenza to West Africa. At the end of September 1918, 3 percent of the population of Sierra Leone are estimated to have died from influenza. From Freetown, the virus spread south along the coast and into the continent. Two other ships carrying soldiers back from France brought the disease to Cape Town and influenza quickly spread into southern and central Africa (Killingray and Phillips 2003). Simultaneously, an increased number of deaths from influenza was observed in Boston, USA, where the pandemic spread across the country within two months from east to west (Killingray and Phillips 2003). From Brest in France, influenza spread north, south, and east, infecting all of Europe within weeks, including remoter regions such as Iceland. In mid-October 1918, the pandemic peaked in Europe and reached as far east as Russia and Hungary. Via ships as well as the Trans-Siberian railroad, influenza transmitted into Asia. Latin America and Africa were primarily infected through major sea ports. By January 1919, the pandemic had circled the globe and reached all but a few remote regions that escaped the pandemic through rigorous maritime quarantines such as northern and eastern Iceland, American Samoa, and St. Helena (Patterson and Pyle 1991, Killingray and Phillips 2003).

Table A1 in Online Appendix A provides an overview of the starting dates of the second wave in each country as found in the literature. The table draws extensively from Patterson and Pyle (1991) and a number of additional sources describing the gradual expansion of the pandemic. In line with the historical events described above, there was relatively little variation in starting dates, with most countries experiencing a start in September or October of 1918

[5] Some authors also describe selected outbreaks occurring in 1920 based on calculations of excess mortality (Johnson and Mueller 2002, Chowell et al. 2010, Chowell et al. 2011, Ansart et al. 2009). Johnson and Mueller (2002) themselves, however, suggest that these outbreaks might have been a single, unrelated epidemic caused by a different strain of the virus.

(except for Spain and Portugal with a reported start in August and a few countries with a start in November). We do not attempt to report end dates given conceptual problems associated with defining the end of the second wave (historical sources, in general, do not provide information on when exactly the last patient in a country was infected). With that said, reconstructed mortality patters suggest that nation-wide deaths typically started to increase several weeks after the first reported cases and (nearly) decreased to pre-pandemic levels within three to four months from the initial increase.

While the accurate death toll of the Spanish Flu remains unknown, multiple sources (Killingray and Phillips 2003, Johnson and Mueller 2002, Patterson and Pyle 1991) agree that mortality rates were highest in Africa and Asia,[6] with India suffering one of the highest influenza-specific mortality rates of up to 6.7 percent.[7] Fiji, Botswana, and Ghana encountered death rates in the vicinity of 5 percent, Tonga of 10 percent, and Western Samoa even 25 percent.[8] Markedly higher mortality rates are reported for indigenous populations such as the Maori in New Zealand, the Aborigines in Australia, the Inuit in Canada, and Native Americans in the United States.[9] On the other hand, Northern America, Europe, and Australia experienced much lower mortality rates of about 0.5 percent. More specifically, Patterson and Pyle (1991) deduce 1.9 – 2.3 million deaths in Africa (14.2 – 17.7 per thousand), 19 – 33 million deaths in Asia (19.7 – 34.2 per thousand), 2.3 million in Europe (4.8 per thousand) and 766,000 – 966,000 deaths in Latin America (8.4 – 10.6 per thousand). The United States is estimated to have had 550,000 - 675,000 deaths (Crosby 2003, Killingray and Phillips 2003). In Canada, influenza spread from one coast to the other within a month, with one in six Canadians contracting the disease and 30,000 - 50,000 dying (Herring and Sattenspiel 2003).

Overall, Patterson and Pyle (1991) estimate global mortality at 30 million or a rate of 16.6 per thousand worldwide. Other figures from previous studies cited by Patterson and Pyle (1991) vary between 15 – 100 million deaths and rates between 8.3 – 55.2 per thousand, though the extent or completeness of these previous studies is unclear. Killingray and Phillips (2003) agree with 30 million deaths but caution that this is only a rough estimate given the lack of data for larger areas and populations. Johnson and Mueller (2002) estimate 50 million deaths but admit that this might be "as much as 100 percent understated" (Johnson and Mueller 2002, p.115).

[6] While for China there is little evidence, Killingray and Phillips (2003) quote a source suggesting that the Chinese mortality rate was about 1 percent.

[7] The Indian population was likely weakened by food shortages due to rationing and large exports by the British as well as by malaria (Killingray and Phillips 2003).

[8] In contrast, US-controlled Eastern Samoa escaped influenza through a maritime quarantine (Killingray and Phillips 2003).

[9] Among the Canadian Inuit, influenza death rates were so high that entire villages seized to exist (Johnson 2003).

Despite the limited accuracy of these estimates, it is clear that even the lower bounds of these estimates suggest that the Spanish Flu had a relevant death toll across all regions of the world, motivating a global analysis.

3 Data & Identification Strategy

Following Almond (2006), exposure to influenza should specifically affect those in-utero in 1918 during the second deadly wave, such that the cohort born in 1919 harbors the majority of the prenatally exposed. Our identification strategy therefore focuses on measuring how key post-influenza outcomes (primary education, secondary education, and employment status at time of census) for the 1919 birth cohort differ from the trend of surrounding cohorts that were not affected in-utero by the influenza pandemic.

To this end, we downloaded 107 census datasets from 51 countries collected between 1960 and 1990 provided by IPUMS International (Minnesota Population Center 2014).[10] 1960 is the first year for which IPUMS International provides census data and coincides with the year of the earliest census sample used in Almond (2006). In order to limit the possible bias arising from mortality-related attrition, our main analysis focuses on the first available census in each country and we do not consider censuses collected after 1990.[11]

IPUMS International provides harmonized educational attainment as a categorical variable consisting of 'less than primary education completed', 'primary education completed', 'secondary education completed' and 'higher education completed' based on the United Nations definition of six years of primary schooling and six years of secondary schooling (Minnesota Population Center 2014).[12] This categorical variable is one of the most widely available measures among IPUMS International censuses and was thus chosen to generate binary indicators for completed primary and secondary education, respectively.[13] Similarly, we generate a binary indicator for employment status that equals one if a respondent is employed

[10] The datasets were downloaded on August 31st, 2022. With the exception of India (for which IPUMS International provides survey data), downloaded datasets for all countries were census datasets. For simplicity, we therefore use the term census data for our sample in the remainder of this article.

[11] The definition of what constitutes the first census is outcome-specific, as not every outcome is available in every census. The choice of the 1990 cutoff is motivated by the desire to limit mortality-related attrition. When estimating effects on employment status, we further constrain the analysis to censuses collected until 1976 to rule out retirement-related attrition. In particular, this approach ensures that none of the individuals born in 1912 has reached an age of 65 years, a common retirement age in many countries.

[12] For example, a respondent with eleven years of schooling would be reported as having only a primary school degree as opposed to a nearly completed secondary degree. For some datasets, university completion pools those with university and technical degrees.

[13] We code these variables such that they capture whether at least the stated level of education was completed. For instance, all persons with completed secondary education are coded as having completed primary education, too.

at the time of census enumeration, and zero for those inactive or unemployed. In addition, two measures of disability status are considered as additional outcomes. First, we create a binary measure for self-reported general disability. Second, we code a binary variable equaling one if a respondent reported being unable to work due to a disability ('work disability').[14] Finally, we use several other outcome variables capturing respondents' income and education when replicating some of the key findings from Almond (2006). As the disability measures and income-related variables are only available in a small subset of the considered censuses, our main analysis focuses on primary education, secondary education, and employment status.[15] Disability is considered as an alternative outcome in a smaller sample as part of the robustness checks. Because data availability for income is even more limited, we only consider income when replicating some of the key findings for the U.S. from Almond (2006).

For most of the analysis, we report results for two different samples. The 'full sample' comprises the full set of 51 census samples from the 51 countries for which census datasets were downloaded. Our 'core sample' comprises a subset of 36 census samples from 36 countries for which the available information on the time of birth is at least as precise as the information for the United States in the dataset used by Almond (2006). Specifically, notice that in order to correctly identify individuals that were exposed in-utero to the 1918 influenza pandemic, information on the precise time of birth is crucial. For the United States, Almond (2006) uses information on both age (reported in integers as of March 31 of the census year) and quarter of birth. This information allows for deriving the year of birth for each respondent. Almond (2006) then sets his treatment indicator (capturing in-utero exposure to the Spanish Flu) equal to one for all individuals born in 1919. Importantly, this approach is not free of measurement error (in the following called "misclassification error"). As described by Almond (2006, p. 673), "[…] the pandemic struck without warning in October 1918 and had largely dissipated by the beginning of 1919 […]." Accordingly, the United States faced a high-intensity exposure period of approximately three months during the last quarter of 1918. Despite this fact, Almond's approach codes individuals born in the last quarter of 1919 as having been exposed in utero, although they were conceived during the first quarter of 1919 when the second wave had already largely receded.[16] Similarly, those born in October to December 1918 were exposed in utero

[14] This indicator represents a subcategory of the employment status as it gives a reason for why a respondent is inactive. Analogous to employment status, we limit the analysis to censuses collected until 1976 when discussing work disability in order to rule out retirement effects.

[15] All outcome variables only have a very limited number of missing observations if they are included in a census. Table A2 in Online Appendix A provides an overview on missing data by census.

[16] Given that pregnancy duration varies naturally around approximately nine months (counting from time of conception), it is likely that at least some respondents who were born in September 1919 were not conceived

but are not coded as such.[17] This results in a misclassification of six birth months.[18]

Our 'core sample' is constructed so that it only comprises those censuses for which the implied degree of misclassification is weakly smaller than that of Almond's approach for the United States. While Online Appendix B provides a more detailed description of our method, we outline the main rationale here. First, it is worth noting that due to the way viruses spread within populations, nation-wide mortality rates typically do not rise immediately after the first case occurs within a country. For instance, Patterson and Pyle (1991) report that the Spanish Flu had already arrived in the United States by early September – approximately one month before national mortality rates started rising, as described by Almond (2006). We therefore add a delay of one month to the start dates shown in Table A1 (see Online Appendix A) before we calculate the degree of misclassification in any of the censuses. Second, we assume a duration of three months before mortality rates fall to (near) pre-pandemic levels, given the previously discussed mortality patterns.[19] Third, while the United States are the only country with a variable for quarter of birth in their census sample, for all other countries, we can either approximate the birth year by subtracting a respondent's age in years from the census year or use the exact birth year if reported. Although it may seem that approximating birth years with information on age and census year is only a second-best solution, it is important to note that the misclassification error does not depend on how many respondents are allocated to the wrong birth year but rather on the misclassification of respondents as exposed in-utero vs. not in-utero. As we demonstrate in Online Appendix B, using age and census year can result in more precise classifications than using exact birth years, depending on the timing of the pandemic and the timing of census enumeration. Our strategy is therefore to use the more precise alternative if a census reports both age and exact birth year.[20] If both alternatives come with the same misclassification error, we use the specification based on exact birth year. Finally, for each census (given the enumeration date and reported local start of the second wave and associated

during the fall of 1918. Nevertheless, for simplicity reasons (and thus counting in Almond's favor), we assume that those born in September 1919 were correctly classified as exposed in-utero. We show below that slightly shifting/extending the exposure period has no effect on our conclusions.

[17] More precisely, the cohort born in October 1918 was partially exposed in-utero and partially during the neonatal period. It is beyond the scope of this paper and the capabilities of its authors to provide a medical discussion on how consequential this distinction is. Instead, we assume for the main analysis that those born in the first month of the exposure period were exposed in-utero, and show in a robustness check that conclusions stay the same if we instead assume that they were not exposed in-utero.

[18] It is debatable whether January 1919 should be counted as part of the exposure period given that mortality rates were much lower than those of fall 1918 but not yet fully back to pre-pandemic levels. As we show in a series of robustness checks, the conclusions of this article do not depend on such nuances.

[19] We check the robustness to deviations from these assumptions in sensitivity analyses reported further below.

[20] Age in years is included in all censuses (though sometimes derived from exact birth dates). However, we do not use the variable age for the census France 1962 and all Greek censuses because age was only estimated in these datasets. Table A2 in Online Appendix A shows which census datasets also report the birth year.

exposure period), we count the number of misclassified birth months as was described above for the United States. For each country, we include the first census that reaches a misclassification error of six months or less.

In addition, it is possible that both the original birth year information and our constructed birth year proxy (census year minus age) are subject to measurement error stemming from poor reporting or data collection methods. To address this issue, we make use of Myers' Blended Index of Digit Preference (Hobbs 2004) to systematically check for evidence of heaping in our birth year variable in all datasets. As described in Hobbs (2004), the index increases with the extent of heaping and ranges from zero (no digit preference) to 90 (complete preference for single digit). Hence, for our analysis, small values of Myers' Index close to zero are preferred.[21] Moreover, we limit the analysis to native-born respondents (if nativity status is reported) to increase the probability that respondents were indeed exposed to the pandemic during the lethal second wave in each country.

Table 1 provides summary statistics for the datasets used in the main analysis. Of the 51 censuses in the full sample, 11 stem from Africa, 11 from Asia, 10 from Europe, 17 from Latin America, and 2 from Northern America.[22] Overall, the average percentage of population covered by the censuses is 6.5% (6.4% across the 36 censuses included in the core sample). Data on secondary education are available for all 51 censuses. Data on primary education are unavailable for Austria, St. Lucia and Switzerland due to peculiarities of enumeration methods and local education systems that prevented the creation of harmonized indicators for primary education.[23] Employment data are only available for 26 countries, in part because we restrict the employment sample to only those censuses collected before 1977 to minimize bias from retirement as explained above. As reported at the bottom of Table 1, values for Myers' Index are overall rather low, with a mean of 14.2 in the full sample and 12.0 in the core sample (recall that this indicator ranges from zero (no digit preference) to 90 (complete preference for single digit)).

(Table 1 here)

[21] We calculate Myers' Index using the Stata command "myers". We limit the calculation to the birth year range 1910 to 1919 as the specified range must span a multiple of 10.

[22] Although we limit attention to the first census per country, in rare cases more than one dataset per country is used across our analyses, because not every outcome is available in each census. For example, the first Mexican census (conducted in 1960) does not report employment data, such that the first Mexican census for the employment outcome was collected in 1970 (please see Table A2 in Online Appendix A for details on census-specific data availability).

[23] This is also true for the 1971 East Germany census. However, we are able to use the 1981 census instead.

Our analysis follows a two-step procedure. First, as in Almond (2006), we estimate (separately for each country) the long-term effects of in-utero exposure to the Spanish Flu by calculating deviations of the set of binary dependent variables from the respective squared cohort trend. Each outcome is regressed on a constant, an indicator for being born in 1919 capturing in-utero exposure to the influenza pandemic, and the squared cohort trend. Formally, our regression model for each country can be written as

$$y_i = \beta_0 + \beta_1 * YOB_{i,1919} + \beta_2 * YOB_i + \beta_3 * YOB_i^2 + \varepsilon_i, \qquad \text{(Equation 1)}$$

where y_i is one of the three binary dependent variables described above, YOB_i is the (estimated) year of birth, and $YOB_{i,1919}$ is the binary exposure indicator which equals 1 for individuals born in 1919, and 0 otherwise. As in Almond (2006), we focus on cohorts born between 1912 and 1922 as our baseline specification and estimate Equation (1) separately for men and women.[24] All specifications are estimated using Ordinary Least Squares with heteroskedasticity-robust standard errors.

The coefficient of interest is β_1. It should be noted that this type of analysis constitutes an intent-to-treat approach as we lack individual data on actual influenza exposure. Nevertheless, given the findings of previous studies, one would expect the 1919 birth cohort to be worse off, on average, compared to the general trend. Specifically, therefore, β_1 would be expected to be negative when the dependent variable measures educational attainment or employment status, but positive when the dependent variable captures disability. In total, we obtain k coefficients of interest β_1 and corresponding standard errors, with k being the number of countries (censuses) included in that specification.

In the second step of the analysis, we use these k estimated coefficients and standard errors to perform random-effects meta-analyses allowing us to describe overall patterns across outcomes and world regions.[25] We thus mimic a scenario where each of our k coefficients stems from a unique study of in-utero impacts of the Spanish Flu, and we collectively analyze these to obtain a combined average effect assuming that our coefficients represent a random sample of all possible study effects. Notably, this approach allows for between-study heterogeneity and thus accounts for the fact that the impact of the Spanish Flu might differ across countries due to unobservable characteristics. With this approach, each of our estimates is weighted by the

[24] We also show that our main results are robust to estimating a more general trend for cohorts born between 1910 and 1928, which centers the influenza cohort in the middle of the interval and gives equal weight to outcomes of those born before and after 1919 (see Section 5).

[25] Random-effects meta-analyses are conducted with the Stata command 'metareg', using the method of moments to estimate between-census variance.

inverse of its total variance comprising the sum of the estimation variance within each study (i.e., the sampling error of each census-specific estimate) and the systematic variance between studies (Lipsey and Wilson 2001). The weighted average of the k coefficients yields the so-called average weighted coefficient for a given dependent variable. To facilitate a better understanding of the effect size implied by an average weighted coefficient, we use the same weights to also calculate and report the average of the mean of the respective dependent variable across censuses.

4 Results

We start by replicating some of the key results from Almond (2006). Both our analysis and the results presented by Almond (2006) rely on IPUMS census data, although there are small differences in terms of sample sizes for 1960 and 1980 (see Table 2) due to additional sample refinements imposed by Almond (2006). Larger sample size differences exist for 1970 as IPUMS International only provides a one percent sample, whereas Almond (2006) could rely on a three percent sample. Table 2 reports both Almond's original estimates (in odd-numbered columns) and our estimates (even-numbered columns) for the three U.S. censuses in 1960, 1970, and 1980 considered in Almond (2006). Except for a few instances, the replication appears to work well across years and considered variables.[26]

(Table 2 here)

Table 3 reports our main results when expanding the analysis to the global level. For each of the three outcomes, the table reports average weighted coefficients by region both for the full sample and for our core sample (recall that the core sample comprises only those countries for which the degree of measurement error due to imprecise information on the time of birth is at most as large as in Almond 2006). Panel A in Table 3 reports the results for the probability for men to obtain primary education. For the full sample, the all-regions average weighted coefficient (based on a total of 48 census-specific estimates) is small and positive, but statistically insignificant at conventional levels. Importantly, only 7 of the 48 census-specific regressions yield negative and statistically significant coefficients at the 10% level, while 15 estimates are positive and significant. These results appear inconsistent with the existence of a

[26] As mentioned in Section 3, most of these variables are only available in a small subset of censuses, so that we only use them in the case of the U.S. for the purpose of replication, but not in our main analysis. The outcome "high school graduate" is however equivalent to "completed secondary education" used in our main analysis.

systematic negative effect of in-utero influenza exposure on primary education. This interpretation is further supported by the fact that, within regions, none of the average weighted coefficients is statistically significantly negative. To increase confidence that these findings are not driven by measurement-related attenuation bias, Panel A+ in Table 3 reports the results when we limit the analysis to those censuses fulfilling the requirements for our core sample. Again, we do not find evidence for adverse in utero effects, with the majority of country-level estimates remaining statistically insignificant.[27]

(Table 3 here)

Next, we turn to the effect of the 1918 influenza pandemic on secondary education among men, as reported in Panels B and B+ in Table 3. Again, we find no evidence of a systematic negative effect. In particular, the all-regions average weighted coefficient is statistically insignificant in both the full sample and the core sample, and only 7 of 51 country-level estimates in the full sample (6 of 36 estimates in the core sample) are negative and statistically significant at the 10% level. Similarly, none of the world regions features a statistically significant average weighted coefficient.

Finally, Panels C and C+ in Table 3 report the results for male employment status at the time of census collection. Once more, any evidence of a systematic adverse impact of the 1918 influenza pandemic seems lacking. The only negative result that is statistically significant at the 10% level is the average weighted coefficient for Latin America in the full sample. However, the effect size (-0.2 percentage points) is very small, given that, on average, male employment rates were 91% in Latin America. Moreover, the negative coefficient for Latin America seems to be driven by a single country (Bolivia), providing no evidence of a systematic adverse effect.

Figure 1 illustrates our main findings by showing estimated effect sizes and confidence intervals for the countries in our core sample. Most of the depicted estimates are not statistically significant (indicated by the gray shading), and among the significant estimates (shaded in black) positive coefficients are more frequent than negative coefficients. Consequently, the country-level findings clearly do not support the existence of systematic long-term adverse effects of in-utero exposure to the pandemic on economic outcomes across countries.

(Figure 1 here)

[27] As shown in Table C1 in Online Appendix C, sample sizes in the census datasets are typically large while standard errors for effect estimates are small. Power limitations are therefore unlikely to drive these findings.

It is possible that these results may not apply to women, as the underlying mechanisms and incentives may differ. For instance, average female employment rates in the full sample tend to be substantially lower than those of men (28% vs. 89%) and are likely influenced by family planning considerations. We therefore re-estimate all models using female rather than male respondents. The results are reported in Table C2 in Online Appendix C (country-level estimates for the core sample are reported in Figure C1). Despite the potential differences in underlying mechanisms, the results for men are largely confirmed by the analysis for women. In particular, the all-regions average weighted coefficients are never statistically significant and, among the region-specific estimates, only Africa features significant results for primary education but with absolute effect sizes close to zero.

Finally, we consider the possibility that, while there is no evidence of systematic adverse effects on economic outcomes, the 1918 influenza pandemic may have nevertheless exhibited long-term health impacts. To this end, we re-estimate our main analysis using general disability and work disability as alternative outcomes. The results are reported in Table C3 in Online Appendix C. Note that these results should be interpreted with caution as data on disability are only available for up to 14 countries. With this caveat in mind, we do not observe any statistically significant increases in disability rates. In fact, some of the weighted average coefficients are significantly negative (suggesting a reduction in disability rates).

Overall, our results therefore appear to be difficult to reconcile with the view that the 1918 influenza pandemic exhibited systematic long-term adverse effects across countries on cohorts who were exposed to the pandemic in utero. While there are a few census datasets for individual countries for which we find significant differences between the influenza cohort and surrounding cohorts, the vast majority of our results are insignificant and, among the significant estimates, the beneficial effects outnumber adverse effects. Moreover, average weighted coefficients within and across regions tend to be small in magnitude and provide no indication of economically meaningful impacts.

In what follows, we discuss a series of robustness checks to corroborate this finding and to address potential identification concerns.

5 Robustness and Limitations

Table 4 reports the results of several robustness checks. The table is organized in a similar way as our main results in Table 3, except that Table 4 only reports the all-region average weighted

coefficients and the columns correspond to different specifications (i.e., robustness checks) rather than region-specific estimates.

(Table 4 here)

Column (1) in Table 4 reports the results when the considered time window of birth cohorts is extended from our baseline period 1912-1922 (also studied in Almond 2006) to the longer period 1910-1928, which centers the influenza cohort in the middle of the interval and gives equal weight to outcomes of those born before and after 1919. Similar to the results presented in Table 3, the all-region average weighted coefficients for education are always statistically insignificant (across both the full sample and the core sample). The main qualitative difference to our baseline results is that the negative coefficient for employment is now statistically significant (at the 5% level in the full sample and 10% level in the core sample). However, the effect size continues to be very small, corresponding to a reduction of 0.4 percentage points at a mean employment rate of 90%. Moreover, the overwhelming majority of country-specific estimates do not reach statistical significance (i.e., there are only four countries with significantly negative coefficients for employment), making the existence of a systematic adverse effect rather unlikely.

Column (2) in Table 4 reports the results when the considered period of birth cohorts is restricted to 1912-1919. The motivation behind this specification is that our baseline period includes WWI as a major disruptive event. In particular, declines in fertility during the war period (compared to pre-war levels) and a subsequent increase in fertility rates were observable in many European countries (Vandenbroucke 2014). This fertility response implies that post-WWI cohorts may differ systematically from those born during the war. We therefore explore the robustness of our main results to the exclusion of cohorts born after 1919 (i.e., those that were not in utero during WWI). As shown by the results in column (2) of Table 4, the results remain very similar to our baseline estimates and show no statistically significant effects of the Spanish Flu on education and employment outcomes.

A different view on the role of WWI is that those born before 1919 are war survivor cohorts and are therefore potentially an imperfect control group, either because they too were exposed to various forms of hardship in-utero (e.g. famine), biasing our results towards zero as the control group cannot be considered ‘untreated’, or because those with poor health did not

survive, making the pre-war cohorts unusually healthy (positive selection).[28] To guard against this type of bias, we next exclude all European countries due to the overall strong impact of WWI and their high number of casualties. The results are reported in column (3) of Table 4 and largely confirm our baseline findings, as we do not find statistically significant negative effects on any of the outcomes. Moreover, in column (4) of Table 4, we exclude all countries that are classified as belligerent and focus on non-belligerent (i.e. 'neutral') countries only.[29] While this causes a substantial reduction in sample size, the obtained results lend support to our previous findings. In particular, effect estimates are statistically insignificant in all regressions and effect sizes are either positive or virtually zero. We thus conclude that, despite the theoretical challenges posed by WWI for our identification strategy, it is unlikely that WW1 drives our results.

Another possible concern is that our findings may be driven by attenuation bias stemming from measurement error in the reported time of birth. In part, our main analysis already alleviated this concern by showing that our key findings are largely persistent when focusing on datasets for which the implied degree of misclassification due to imprecise information on exposure status is limited and at most as high as in Almond (2006). As described earlier, it is further possible to evaluate the robustness of results with regards to age heaping using Myers' Blended Index of Digit Preference. We therefore test the robustness of our findings to the exclusion of censuses with relatively high index values. Specifically, columns (5) and (6) in Table 4 report the results when focusing on censuses with index values below 10 and 5, respectively. In both cases, there are no statistically significant average weighted coefficients for primary and secondary education, and numbers of individual countries with significantly negative results remain very small. While in column (6) the negative coefficient for employment is now statistically significant at the 10% level, its magnitude is not only very small (corresponding to a decrease of 0.7 percentage points in employment at a mean rate of 88%) but also entirely driven by a single country (Greece). Overall, this strongly suggests that our findings are not driven by measurement error in birth timing stemming from poor reporting or data collection methods.

While the literature generally identifies the second wave of the Spanish Flu pandemic as the lethal one, most countries experienced an earlier first and/or later third wave as well. These waves are described as much less lethal and comparable to seasonal Flu outbreaks

[28] Given the null findings of our study, the former effect would have to outweigh the latter in order to explain our results.

[29] Belligerent countries are all those that officially declared their participation in WWI. Following Kruizinga (2014), non-belligerent (neutral) countries in our sample are: Argentina, Chile, Mexico, Spain and Switzerland.

without larger impact on mortality trends. To check whether our main results are biased by comparing our treatment cohort to a trend incorporating (potentially) exposed cohorts, we re-estimate our analysis when including two dummies, one for being born in 1918 and one for being born in 1920. Thus, in this robustness check we explicitly control for the two cohorts surrounding the 1919 influenza cohort. The results are reported in column (7) of Table 4. The results for primary and secondary education continue to show no statistically significant estimates for the all-region average weighted coefficients. The only qualitative difference to our baseline results is that the negative coefficient for employment is now statistically significant at the 10% level in the full sample. However, the magnitude of the effect remains very small (corresponding to a decrease of 0.5 percentage points in employment at a mean rate of 89%) and becomes statistically insignificant when focusing on our core sample (see column (7) of Panel C+ in Table 4). Moreover, very few of the country-specific estimates (6 of 26 in the full sample and 4 of 20 in the core sample) are negative and statistically significant, casting further doubt on the existence of a systematic global adverse effect.

Finally, one may be concerned that the high number of insignificant estimates is driven by sample size limitations. To address this issue, we re-run our main analysis using all available census years rather than only the first census year per country (we still exclude census years after 1990 for primary and secondary education and those after 1976 for employment, for the reasons stated earlier). The results are reported in Table 5, which is organized in the same way as Table 3.[30] While several region-specific average weighted coefficients that were statistically insignificant in Table 3 are now significant, many of them are *positive* with small effect sizes.[31] Moreover, the all-regions average weighted coefficients remain small and similar to our main estimates from Table 3. The only outcome for which the all-region estimate is negative and statistically significant is employment. However, similar to before, the effect size is very small and only few of the individual census-specific estimates are significantly negative at the 10% level. Specifically, for the full sample there are only 16 out of 100 such estimates for primary education, 17 out of 107 estimates for secondary education, and 6 out of 38 estimates for employment. Sample size considerations are thus unlikely to be a major concern and key results are robust to the inclusion of additional census years.

(Table 5 here)

[30] Individual census-specific results are reported in Table C1 in Online Appendix C.

[31] Note that Table 5 reports consistent negative effects on education for Northern America which are driven by the US estimates, in line with the findings in Almond (2006). These results only transpire to the meta-analysis level when all available census samples of Northern America collected between 1960 and 1990 are used (last column of Table 5), but not when only the earliest census is used (last column of Table 3).

The results of a number of additional robustness checks are provided in Online Appendix C. In Table C4, we verify the robustness of our main results to moderate changes in the assumed timing of the 1918 influenza pandemic across countries. Specifically, recall that the selection of our core sample is based on the implied misclassification error assuming that the exposure period started one month after the pandemic reached a country and lasted for 3 months. The insignificant results in Table C4 show that it does not matter for our key findings whether we instead assume that the exposure period started in the same month or two months after the first case. Similarly, results remain insignificant in case the exposure period lasted four months rather than three (regardless of whether we assume that the exposure period started immediately with the first case or one month after).

In addition, the results in Table C5 show that our results are largely robust to changes in the error tolerance threshold underlying the selection of our core sample. Specifically, the lack of significantly negative average weighted coefficients in this table indicates that our key findings remain intact when the permitted maximum number of wrongly classified months is reduced from 6 months (corresponding to the implied error in Almond 2006 and in our baseline specification) to any smaller integer, including 0. Together, the results in Tables C4 and C5 show that our main findings are overall not sensitive to the specific assumptions underlying the construction of the core sample, which greatly increases our confidence that the findings are not driven by these assumptions.

Finally, we test how the results for our core sample change if we assume that respondents born in the first month of the exposure period were not exposed in-utero. For this purpose, we repeat the exercise from Table C5 using this alternative assumption (see Table C6). While in this scenario there are no censuses with zero misclassification error, we find no evidence for statistically significant negative average weighted coefficients across the various misclassification tolerance thresholds.

In summary, the robustness checks confirm our main finding of a lack of systematic adverse effects of in-utero exposure to the 1918 influenza pandemic on a global level. We therefore conclude that it is unlikely that the Spanish Flu exhibited systematic long-term adverse effects across countries on human capital accumulation and employment propensities for cohorts exposed to the influenza pandemic in utero.

6 Conclusion

We set out to study the long-term effects of the 1918 influenza pandemic on a global level,

thereby complementing the insights from previous single-country studies. Overall, we were unable to detect a clear pattern regarding the effects of in-utero exposure to the deadly second wave of the pandemic on a population level. A series of robustness checks show that this finding remains intact when varying the construction of the comparison group, adjusting the made assumptions about the exact timing of the pandemic in each country, and excluding countries likely to be highly affected by WWI. The latter point is important, given that the concurrence of the influenza pandemic and WWI has been a key point of criticism in past studies (Brown and Thomas 2018, Beach et al. 2020). Moreover, the fact that increasing the number of countries, years, and censuses beyond our preferred specification (core sample) does not alter the key findings strongly suggests that our results are not driven by sample size limitations.

Overall, therefore, we conclude that there is little empirical evidence to support the view that the 1918 influenza pandemic exhibited systematic long-term adverse effects across countries on economic outcomes for cohorts exposed to the pandemic in utero. This also suggests that existing single-country studies reporting statistically significant adverse effects may have limited external validity and are unable to provide a comprehensive picture of the long-term impacts of the Spanish Flu on a global level.

Nevertheless, some limitations of this study may arise from unobserved heterogeneity. While the harmonization effort of IPUMS International made this research possible, the very nature of harmonization also introduces a certain measurement error. Despite largely comparable educational categories, some underlying country-specific educational standards remain. Differences in the employment and disability indicators are possible as well but less likely. Apart from the question of measurement, it may also be argued that overall education levels across countries are too heterogeneous, such that pooling advanced and less advanced countries may bias results towards zero. Such concerns are, however, unwarranted as our conclusions do not simply rely on the reported small and mostly insignificant average weighted coefficients. Rather, we find that the vast majority of census-specific estimates are statistically insignificant, which raises our confidence in the overall results.

A second possible limitation of this study arises from measurement error in the definition of exposed versus unexposed cohorts. As discussed above, the influenza literature is not always unanimous in the definition of waves, their duration, and their timing. Hence, measurement error is possible and augmented by the fact that limited data availability forces us to rely on yearly rather than monthly birth data. While our results should therefore be interpreted with caution, it is worth noting that our methodology allows us to quantify the expected degree of misclassification for each census, with our core sample analysis limiting the attention to datasets

with errors as low as (or better than) the censuses used by Almond (2006). Moreover, further restricting the error tolerance does not lead to differential results.

A possible reason for insignificant results could also be selective mortality. In Bombay, India, stillbirths are reported to have risen by 50 percent during the height of the influenza pandemic (Ramanna 2003, p.89). If fetuses in bad health were more likely to be stillborn, the resulting population of surviving fetuses would display a positive selection in terms of health and, therefore, negative effects would only surface if the effects of health shocks overcompensate those of the positive selection (Almond and Currie 2011). Similarly, disentangling the effects of in-utero shocks from those occurring during infancy might prove difficult. For example, Echeverri (2003) states that influenza also increased the death rate among post-partum women in Spain due to puerperal septicaemia which could mean that infants that survived fetal exposure to influenza could still be affected by influenza albeit in an indirect manner. In this case, influenza would not operate through the channel proposed by the fetal origins hypothesis but through economic deprivation during childhood. That said, as discussed in detail in Section 2, mortality rates, while subject to considerable uncertainty, were likely to be low in most European and Northern American countries, and the large number of casualties was driven by high infection rates rather than low survival probability conditional on being infected. The absence of economically relevant adverse effects on our main outcomes is however highly robust across world regions. Although it is not possible to fully rule out selection effects using the data at hand, the overall homogeneity of findings raises confidence that selective mortality is not a major concern.

Despite these potential limitations, our study provides a fresh perspective on the fetal origins hypothesis. At the very least, the lack of a systematic adverse effect across countries documented in this study raises concerns about the widely-held view that health shocks such as the Spanish Flu are associated with important long-term adverse effects on economic outcomes of those exposed in utero.

Declarations of interest: none.

Funding: none.

References

Adair LS, Prentice AM. A critical evaluation of the fetal origins hypothesis and its implications for developing countries. Journal of Nutrition 2004;134(1): 191-193.

Almond D. Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population. Journal of Political Economy 2006;114(4): 672–712.

Almond D, Currie J. Killing Me Softly: The Fetal Origins Hypothesis. Journal of Economic Perspectives 2011;25(3): 153–72.

Almond D, Mazumder B. Health capital and the prenatal environment: the effect of Ramadan observance during pregnancy. American Economic Journal: Applied Economics 2011;3(4): 56-85.

Ansart S, Pelat C, Boelle PY, Carrat F, Flahault A, Valleron AJ. Mortality Burden of the 1918–1919 Influenza Pandemic in Europe. Influenza and Other Respiratory Viruses 2009;3: 99–106.

Barker DJ. In Utero Programming of Chronic Disease. Clinical Science 1998;95(2): 115–128.

Barreca AI. The Long-Term Economic Impact of In Utero and Postnatal Exposure to Malaria. Journal of Human Resources 2010;45(4): 865–892.

Beach B, Brown R, Ferrie J, Saavedra M, Thomas D. Reevaluating the Long-Term Impact of In Utero Exposure to the 1918 Influenza Pandemic. Journal of Political Economy 2022;130(7): 1963-1990.

Brown R, Thomas D. On the Long Term Effects of the 1918 US Influenza Pandemic. BREAD Working Paper No. 569. 2018.

Caruso G, Miller S. Long run effects and intergenerational transmission of natural disasters: A case study on the 1970 Ancash Earthquake. Journal of Development Economics 2015;117: 134-50.

Chowell G, Viboud C, Simonsen L, Miller MA, Acuna-Soto R. Mortality Patterns Associated with the 1918 Influenza Pandemic in Mexico: Evidence for a Spring Herald Wave and Lack of Preexisting Immunity in Older Populations. The Journal of Infectious Diseases 2010;202 (4): 567–575.

Chowell G, Viboud C, Simonsen L, Miller MA, Hurtado J, Soto G, Vargas R, Guzman MA, Ulloa M, Munayco CV. The 1918–1920 Influenza Pandemic in Peru. Vaccine, Historical Influenza Pandemics 2011;29: B21–26.

Crosby AW. America's Forgotten Pandemic: The Influenza of 1918, vol. 2. Cambridge University Press; 2003.

Echeverri B. Spanish Influenza Seen from Spain. In: Killingray D, Phillips H (Eds), The Spanish Influenza Pandemic of 1918-1919: New Perspectives, vol. 1. Routledge Chapman & Hall; 2003, p. 173–190.

Fletcher JM. Examining the Long Term Mortality Effects of Early Health Shocks. US Census Bureau Center for Economic Studies Paper CES-WP-14-19; 2014.

Garthwaite C. The Effect of In-Utero Conditions on Long Term Health: Evidence from the 1918 Spanish Flu Pandemic. Working paper; 2008. Available from: http://www.kellogg.northwestern.edu/faculty/garthwaite/htm/fetal_stress_garthwaite_053008.pdf [accessed: February 2nd, 2020].

Heckman JJ. Skill formation and the economics of investing in disadvantaged children. Science 2006; 312(5782): 1900-1902.

Helgertz, J., Bengtsson, T. The Long-Lasting Influenza: The Impact of Fetal Stress During the 1918 Influenza Pandemic on Socioeconomic Attainment and Health in Sweden, 1968-2012. Demography 2019;56: 1389-1425.

Herring D, Sattenspiel A, Sattenspiel L. Death in Winter: Spanish Flu in the Canadian Subarctic. In: Killingray D, Phillips H (Eds). The Spanish Influenza Pandemic of 1918-1919: New Perspectives, vol. 1. Routledge Chapman & Hall; 2003, p. 156–172.

Hobbs FB. Age and Sex Composition. In: Swanson DA, Siegel JS (Eds). The Methods and Materials of Demography. Emerald Group Publishing; 2004, p. 125–174.

Huxley R, Neil A, Collins R. Unravelling the fetal origins hypothesis: Is there really an inverse association between birthweight and subsequent blood pressure?. Lancet 2002; 360(9334): 659-665.

Johnson NP. The Overshadowed Killer: Influenza in Britain in 1918-19. In: Killingray D, Phillips H (Eds). The Spanish Influenza Pandemic of 1918-1919: New Perspectives, vol. 1. Routledge Chapman & Hall; 2003, p. 132–155.

Johnson NP, Mueller J. Updating the Accounts: Global Mortality of the 1918-1920 "Spanish" Influenza Pandemic. Bulletin of the History of Medicine 2002;76(1): 105–115.

Karlsson M, Nilsson T, Pichler S. What Doesn't Kill You Makes You Stronger? The Impact of the 1918 Spanish Flu Epidemic on Economic Performance in Sweden. IFN Working Paper 911; 2012.

Killingray D, Phillips H. Introduction. In: Killingray D, Phillips H (Eds). The Spanish Influenza Pandemic of 1918-1919: New Perspectives, vol. 1. Routledge Chapman & Hall; 2003, p. 1–26.

Kruizinga S. Neutrality. In J. Winter (Ed.). The Cambridge history of the First World War, vol. 2. Cambridge: Cambridge University Press; 2014, p. 542–75.

Lin MJ, Liu EM. Does in Utero Exposure to Illness Matter? The 1918 Influenza Epidemic in Taiwan as a Natural Experiment. Journal of Health Economics 2014;37: 152–163.

Lipsey M, Wilson D. Practical meta-analysis. Thousand Oaks, CA: Sage; 2001.

Maccini S, Yang D. Under the weather: Health, schooling, and economic consequences of early-life rainfall. American Economic Review 2009;99(3): 1006-26.

[dataset] Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 6.3. Minneapolis: University of Minnesota; 2014.

Neelsen S, Stratmann T. Long-Run Effects of Fetal Influenza Exposure: Evidence from Switzerland. Social Science & Medicine 2012;74(1): 58–66.

Nelson RE. Testing the Fetal Origins Hypothesis in a Developing Country: Evidence from the 1918 Influenza Pandemic. Health Economics 2010;19: 1181–1192.

Patterson KD, Pyle GF. The Geography and Mortality of the 1918 Influenza Pandemic. Bulletin of the History of Medicine 1991;65(1): 4–21.

Ramanna M. Coping with the Influenza Pandemic: The Bombay Experience. In: Killingray D, Phillips H (Eds). The Spanish Influenza Pandemic of 1918-1919: New Perspectives, vol. 1. Routledge Chapman & Hall; 2003, p. 86–98.

Taubenberger JK, Morens DM. 1918 Influenza: The Mother of All Pandemics. Emerging Infectious Diseases 2006;12(1): 15–23.

Valente C. Civil conflict, gender-specific fetal loss, and selection: A new test of the Trivers-Willard hypothesis. Journal of Health Economics 2015;39: 31–50.

Vandenbroucke G. Fertility and Wars: The Case of World War I in France. American Economic Journal: Macroeconomics 2014;6(2): 108–136.

Table 1: Summary statistics

	Full sample	Core sample
Countries	51	36
Africa	11	7
Asia	11	8
Europe	10	6
Latin America	17	13
Northern America	2	2
Censuses reporting outcome[a]		
Primary education	48	34
Secondary education	51	36
Employment	26	20
Census characteristics		
Mean coverage[b]	6.5%	6.4%
First year	1960	1960
Last year	1990	1990
Mean Myers' Index[c]	14.2	12.0

All reported statistics are calculated using only the first census per country (men only). [a] Number of censuses providing information. [b] Coverage is the percentage of a country's population included in the census sample of a given year (mean calculated across censuses). [c] Myers' Blended Index of Digit Preference (using original birth year information or census year minus age as birth year proxy depending on which of these alternatives has a smaller expected error; mean calculated across censuses).

Table 2: Replication of key results from Almond (2006)

Outcome	1960 (original)	1960 (replicated)	1970 (original)	1970 (replicated)	1980 (original)	1980 (replicated)
	(1)	(2)	(3)	(4)	(5)	(6)
High school graduate	-0.021***	-0.021***	-0.020***	-0.015**	-0.014***	-0.014***
	(0.005)	(0.005)	(0.003)	(0.005)	(0.003)	(0.002)
Years of education	-0.150***	-0.155***	-0.176***	-0.133***	-0.117***	-0.119***
	(0.038)	(0.037)	(0.023)	(0.038)	(0.019)	(0.018)
Total income	-573*	-585*	-1236***	-1124**	-1065***	-475**
	(295)	(285)	(253)	(415)	(191)	(174)
Wage income	-812***	-700**	-875***	-1207**	-688***	-630***
	(261)	(253)	(233)	(378)	(179)	(188)
Poor	0.010**	0.010*	0.009***	0.006	0.006***	0.005**
	(0.005)	(0.005)	(0.002)	(0.004)	(0.002)	(0.002)
N[1]	114,031	119,539	308,785	109,775	471,803	479,885

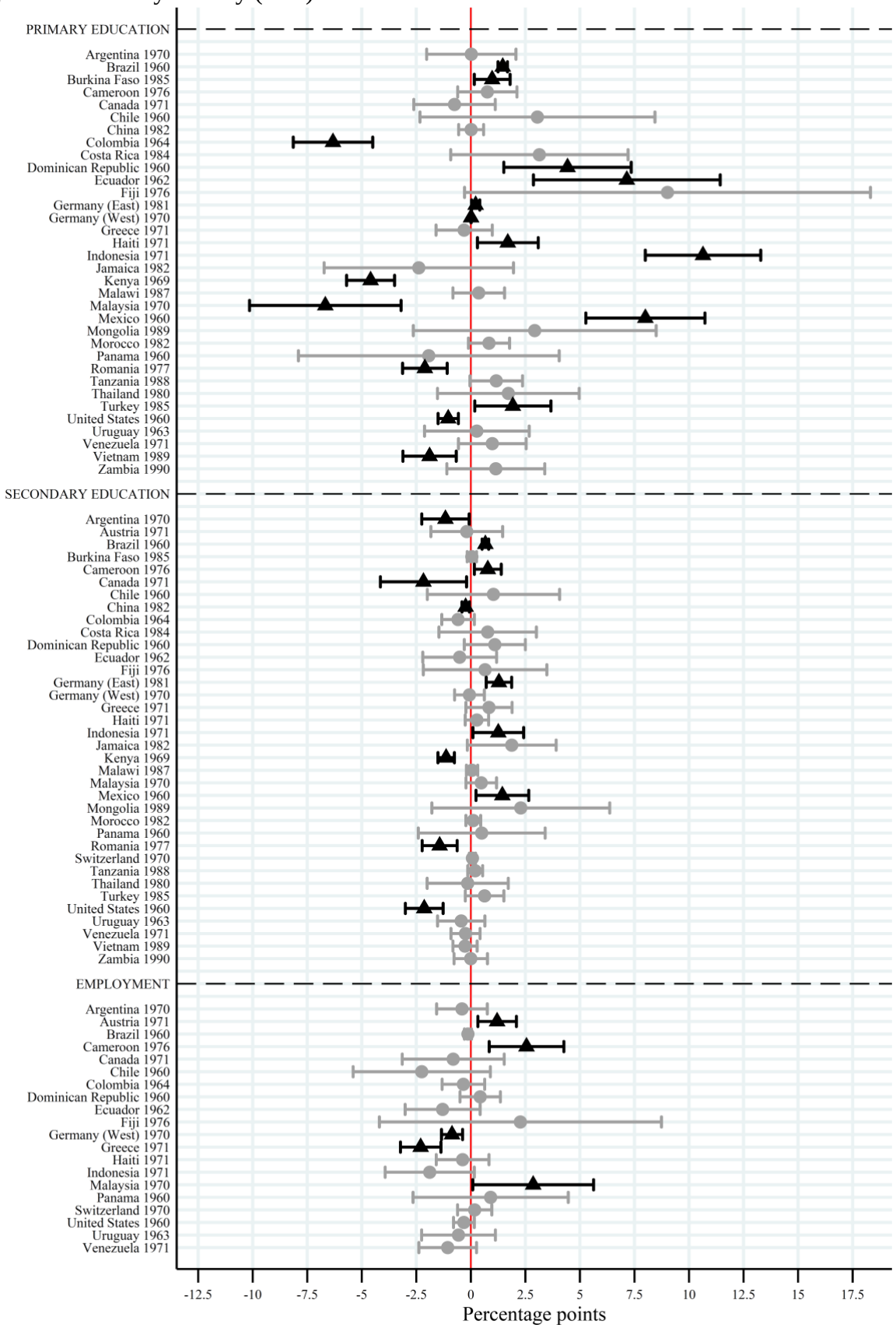
[1]Number of observations refers to outcomes "high school graduate" and "years of education". All results are for men only. We define "high school graduate" as anyone finishing at least grade 12. Years of education is top-coded at 18 years. Total income and wage income measure (wage) income for those receiving any (wage) income at all in 2005 USD. "Poor" is a binary equal to 1 if a respondent is below 150% of the poverty line.

Table 3: Weighted effects on education and employment (men)

	All regions	Africa	Asia	Europe	Latin America	Northern America
Primary education						
Panel A: Full sample						
Average weighted coefficient	0.006	-0.000	0.044*	0.000	0.015	-0.010
	(0.005)	(0.006)	(0.024)	(0.002)	(0.009)	(0.003)
Weighted average of mean of DV	0.31	0.05	0.29	0.62	0.27	0.91
Countries	48	11	11	8	16	2
Positive significant coefficients	15	1	5	2	7	0
Negative significant coefficients	7	2	2	1	1	1
Panel A+: Core sample						
Average weighted coefficient	0.002	0.001	0.015	-0.001	0.013	-0.010
	(0.005)	(0.008)	(0.019)	(0.003)	(0.011)	(0.003)
Weighted average of mean of DV	0.33	0.06	0.29	0.77	0.27	0.91
Countries	34	7	8	4	13	2
Positive significant coefficients	10	1	2	2	5	0
Negative significant coefficients	6	1	2	1	1	1
Secondary education						
Panel B: Full sample						
Average weighted coefficient	0.001	-0.001	0.016	0.001	0.002	-0.021
	(0.002)	(0.002)	(0.011)	(0.002)	(0.002)	(0.005)
Weighted average of mean of DV	0.11	0.01	0.05	0.33	0.04	0.30
Countries	51	11	11	10	17	2
Positive significant coefficients	9	2	4	1	2	0
Negative significant coefficients	7	2	1	1	1	2
Panel B+: Core sample						
Average weighted coefficient	0.000	-0.000	0.001	0.001	0.002	-0.021
	(0.001)	(0.002)	(0.002)	(0.004)	(0.002)	(0.005)
Weighted average of mean of DV	0.12	0.01	0.03	0.49	0.04	0.30
Countries	36	7	8	6	13	2
Positive significant coefficients	5	1	1	1	2	0
Negative significant coefficients	6	1	1	1	1	2
Employment						
Panel C: Full sample						
Average weighted coefficient	-0.003	0.026**	-0.001	-0.002	-0.002*	-0.003
	(0.002)	(0.010)	(0.013)	(0.006)	(0.001)	(0.003)
Weighted average of mean of DV	0.89	0.85	0.82	0.91	0.91	0.87
Countries	26	1	4	5	14	2
Positive significant coefficients	4	1	1	2	0	0
Negative significant coefficients	3	0	0	2	1	0
Panel C+: Core sample						
Average weighted coefficient	-0.003	0.026**	0.006	-0.004	-0.002	-0.003
	(0.002)	(0.010)	(0.019)	(0.007)	(0.001)	(0.003)
Weighted average of mean of DV	0.89	0.85	0.82	0.90	0.92	0.87
Countries	20	1	3	4	10	2
Positive significant coefficients	3	1	1	1	0	0
Negative significant coefficients	2	0	0	2	0	0

Average weighted coefficients are derived from a random-effects meta-analysis as described in Section 3. Stars indicate statistical significance (* = 10%, ** = 5%, *** = 1%). The weighted average of the mean of the dependent variable (DV) is calculated using the same weights as for the average weighted coefficients to ensure comparability. n/a = not available. Single country-level coefficients are counted as significant if corresponding p-values are smaller than 0.1.

Figure 1: Effects by country (men)



Note: Grey circles show insignificant estimates. Black triangles show estimates that are significant on a 10% level. Bars are 90% confidence intervals.

Table 4: Key robustness tests (men)

	1910 – 1928 (centered)	1912 – 1919 (no post-war)	Excluding Europe	Neutral only	Myers < 10	Myers < 5	Controlling for 1918 and 1920
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Primary education							
Panel A: Full sample							
Average weighted coefficient	0.005	0.004	0.013*	0.027	-0.002	-0.004	0.002
	(0.005)	(0.009)	(0.007)	(0.019)	(0.002)	(0.003)	(0.004)
Weighted avg. of mean of DV	0.33	0.30	0.25	0.38	0.47	0.43	0.31
Countries	48	48	40	4	22	13	48
Positive significant coefficients	13	10	13	1	3	0	12
Negative significant coefficients	7	8	6	0	3	2	8
Panel A+: Core sample							
Average weighted coefficient	0.001	-0.004	0.007	0.037	-0.003	-0.004	0.001
	(0.004)	(0.005)	(0.006)	(0.029)	(0.002)	(0.003)	(0.005)
Weighted avg. of mean of DV	0.34	0.32	0.27	0.35	0.50	0.43	0.33
Countries	34	34	30	3	18	12	34
Positive significant coefficients	9	5	8	1	2	0	8
Negative significant coefficients	6	7	5	0	3	2	7
Secondary education							
Panel B: Full sample							
Average weighted coefficient	0.001	0.001	0.002	0.004	-0.001	-0.002	0.000
	(0.002)	(0.003)	(0.002)	(0.004)	(0.002)	(0.002)	(0.001)
Weighted avg. of mean of DV	0.11	0.10	0.05	0.25	0.16	0.17	0.11
Countries	51	51	41	5	23	14	51
Positive significant coefficients	7	9	8	2	2	1	6
Negative significant coefficients	7	8	6	1	5	4	5
Panel B+: Core sample							
Average weighted coefficient	-0.000	-0.002	0.000	0.001	-0.002	-0.003	0.000
	(0.001)	(0.002)	(0.001)	(0.005)	(0.002)	(0.002)	(0.001)
Weighted avg. of mean of DV	0.13	0.12	0.05	0.29	0.21	0.17	0.12
Countries	36	36	30	4	20	13	36
Positive significant coefficients	3	3	4	1	1	0	4
Negative significant coefficients	6	8	5	1	5	4	4
Employment							
Panel C: Full sample							
Average weighted coefficient	-0.004**	0.002	-0.003	-0.001	-0.005	-0.007*	-0.005*
	(0.002)	(0.003)	(0.002)	(0.004)	(0.003)	(0.003)	(0.003)
Weighted avg. of mean of DV	0.90	0.89	0.89	0.91	0.88	0.88	0.89
Countries	26	26	21	4	13	8	26
Positive significant coefficients	1	3	2	0	1	0	3
Negative significant coefficients	4	2	1	0	2	1	6
Panel C+: Core sample							
Average weighted coefficient	-0.004*	-0.001	-0.002	-0.001	-0.005	-0.007*	-0.004
	(0.002)	(0.002)	(0.002)	(0.004)	(0.003)	(0.003)	(0.003)
Weighted avg. of mean of DV	0.90	0.88	0.89	0.92	0.88	0.88	0.89
Countries	20	20	16	3	12	8	20
Positive significant coefficients	1	1	2	0	1	0	2
Negative significant coefficients	4	2	0	0	2	1	4

Average weighted coefficients are derived from a random-effects meta-analysis as described in Section 3. Stars indicate statistical significance (* = 10%, ** = 5%, *** = 1%). The weighted average of the mean of the dependent variable (DV) is calculated using the same weights as for the average weighted coefficients to ensure comparability. n/a = not available. Single country-level coefficients are counted as significant if corresponding p-values are smaller than 0.1.

Table 5: Weighted effects using all available censuses (men)

	All regions	Africa	Asia	Europe	Latin America	Northern America
Primary education						
Panel A: Full sample						
Average weighted coefficient	0.007**	-0.001	0.035**	-0.002	0.014**	-0.006**
	(0.003)	(0.005)	(0.014)	(0.002)	(0.005)	(0.002)
Weighted av. of mean of DV	0.34	0.05	0.27	0.65	0.29	0.93
Countries	48	11	11	8	16	2
Censuses	100	14	23	15	42	6
Pos. significant coefficients	31	2	10	2	17	0
Neg. significant coefficients	16	4	3	3	3	3
Panel A+: Core sample						
Average weighted coefficient	0.005	-0.003	0.020	-0.002	0.012**	-0.007***
	(0.004)	(0.006)	(0.013)	(0.003)	(0.006)	(0.001)
Weighted av. of mean of DV	0.33	0.06	0.27	0.76	0.26	0.92
Countries	34	7	8	4	13	2
Censuses	60	9	17	5	24	5
Pos. significant coefficients	18	1	5	2	10	0
Neg. significant coefficients	12	3	3	2	1	3
Secondary education						
Panel B: Full sample						
Average weighted coefficient	0.001	-0.001	0.011**	0.001	0.002	-0.014***
	(0.001)	(0.001)	(0.005)	(0.002)	(0.001)	(0.003)
Weighted av. of mean of DV	0.13	0.01	0.04	0.38	0.05	0.43
Countries	51	11	11	10	17	2
Censuses	107	14	23	21	43	6
Pos. significant coefficients	20	2	9	3	6	0
Neg. significant coefficients	17	2	2	3	4	6
Panel B+: Core sample						
Average weighted coefficient	-0.000	-0.000	0.004	-0.000	0.001	-0.016***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
Weighted av. of mean of DV	0.14	0.00	0.03	0.56	0.04	0.40
Countries	36	7	8	6	13	2
Censuses	65	9	17	10	24	5
Pos. significant coefficients	10	1	4	1	4	0
Neg. significant coefficients	13	1	2	2	3	5
Employment						
Panel C: Full sample						
Average weighted coefficient	-0.005**	0.026**	-0.012	-0.008	-0.004***	-0.003
	(0.002)	(0.010)	(0.017)	(0.007)	(0.001)	(0.002)
Weighted av. of mean of DV	0.87	0.85	0.82	0.89	0.88	0.88
Countries	26	1	4	5	14	2
Censuses	38	1	5	7	22	3
Pos. significant coefficients	4	1	1	2	0	0
Neg. significant coefficients	6	0	1	3	2	0
Panel C+: Core sample						
Average weighted coefficient	-0.004*	0.026**	-0.011	-0.004	-0.004**	-0.003
	(0.002)	(0.010)	(0.023)	(0.007)	(0.002)	(0.002)
Weighted av. of mean of DV	0.88	0.85	0.82	0.90	0.90	0.88
Countries	20	1	3	4	10	2
Censuses	24	1	4	4	12	3
Pos. significant coefficients	3	1	1	1	0	0
Neg. significant coefficients	4	0	1	2	1	0

Average weighted coefficients are derived from a random-effects meta-analysis as described in Section 3. Stars indicate statistical significance (* = 10%, ** = 5%, *** = 1%). The weighted average of the mean of the dependent variable (DV) is calculated using the same weights as for the average weighted coefficients to ensure comparability. n/a = not available. Single country-level coefficients are counted as significant if corresponding p-values are smaller than 0.1.