The Brazilian Bombshell? The Long-Term Impact of the 1918 Influenza Pandemic the South American Way*

Amanda Guimbeau[†], Nidhiya Menon[‡] and Aldo Musacchio[§]

June 17, 2019

Abstract

We study the short-term, medium-term and long-term repercussions of the 1918 Influenza Pandemic on demographic measures, human capital formation, and productivity markers in the state of Sao Paulo, Brazil's financial center and the most populous city in South America. Using historical and contemporary documents, we construct a unique database on socio-economic and health outcomes for all districts in Sao Paulo. Using temporal and spatial variation in district-level numbers of influenza-related deaths for the period 1917-1920 and measures of social and health infrastructure, we find that the 1918 Influenza Pandemic had significant effects on infant mortality and sex ratios at birth in the short-run. We find robust evidence for persisting impacts on health and educational attainment, productivity measures, and development-related indicators in the long-run.

Key Words: 1918 Influenza Pandemic, Demographic Measures, Health, Education, Short-run, Long-run, Sao Paulo Brazil

JEL Codes: N36, O12, I15, J10

Declarations of interest: none

*We thank Dani Castillo, Andre Lanza, Pedro Makhoul, Stephanie Orlic, and Uros Randelovic for excellent research assistance. We thank James Feigenbaum, Ricardo Godoy and Robert Margo for early conversations, and participants in the Ph.D. seminar series at Brandeis for comments and suggestions. This project was made possible by a Provost Research Grant and funds from the Brazil Initiative at Brandeis University. The usual disclaimer applies.

[†]International Business School, MS 032, Brandeis University, Waltham, MA 02453, USA. Email: amanda2016@brandeis.edu

[‡]IZA & Department of Economics, MS 021, Brandeis University, Waltham, MA 02453, USA. Email: nmenon@brandeis.edu

[§]NBER & International Business School, MS 032, Brandeis University, Waltham, MA 02453, USA. Email: aldom@brandeis.edu

1 Introduction

There has been significant interest in the impact of health conditions on social, economic and demographic outcomes since Almond (2006) analyzed the long-term effects of *in utero* exposure to the greatest epidemic of modern history-the 1918 Influenza Pandemic - which recently crossed its centennial. US Census data from 1960-1980 revealed that *in utero* exposure negatively affected education, health, socioeconomic status, employment and earnings of these cohorts. Other studies (Parman, 2013, Karlsson et al., 2014, Percoco, 2016) have attempted to understand the effects of this aggregate health shock, mostly in developed countries. Parman (2013) in particular shows how this health shock played out at the very disaggregated household level, where Census and US military records from the Second World War reveal that parents' actions reinforced the negative impact of the flu on the child that was exposed *in utero*. The 1918 Influenza Pandemic, by its sheer magnitude and features, provides a unique natural experiment to test a range of hypotheses related to the short and long-run consequences of exposure to diseases, and offers an unusual framework to study the effects of an extraordinary mortality shock on demographic and economic outcomes.

In this paper, we study the short-term, medium-term and long-term repercussions of the 1918 Influenza Pandemic on demographic, human capital and productivty outcomes in the State of Sao Paulo, Brazil. Although today Sao Paulo is Brazil's financial center and South America's most populous city, in the early twentieth century, it was far from such. Given the lack of resources for remedial action and the relatively more primitive health care structure, it is likely that the pandemic's immediate and lingering effects were more harmful in poorer nations like Brazil. Our study is unique in that in contrast to the bulk of the literature in this area that has focused on developed countries, we provide a lens to analyze the possible enduring effects of the pandemic in a poor country marked by social inequalities and a nascent health system. Moreover, an issue with studying the consequences of this Pandemic using data from countries in the Northern Hemisphere such as the United States, is that the timing of the Pandemic coincided with the ending years of the First World War. It is thus hard to disentangle the detrimental impacts of the Flu from the overall havoc and widespread destruction caused by the War. Although Brazil did enter the

War in late 1917, it's actual contribution to the Allied war effort was minimal, and the country certainly did not experience the level of physical destruction that countries in the North endured. It is thus more likely that what we measure in the historical Brazilian context is a repercussion of the 1918 Pandemic.

Records reveal that in Sao Paulo (City), the disease caused 5331 deaths in the short period between mid-October and mid-December 1918 (Massad et al., 2007), and infected up to 350,000 people, two-thirds of Sao Paulo's population (Bassanezi, 2013, Barata, 2000, Bertolli, 2003). While in other contexts the Spanish Flu appears to have chosen its victims randomly, the pandemic's duration and intensity differed significantly across geographical markers of districts (the spatial unit of our analysis) in the State of Sao Paulo. We exploit this spatial variation in influenza's incidence rates in our methodology to link the number of influenza-related deaths to a range of outcome variables of interest over time. We accomplish this using detailed district-level historical data from 1912 to 1921 on vital statistics, health and educational infrastructure, and several other controls. We complement this data with other information drawn from official statistical reports to study our outcomes of interest conditional on the social and health framework that would have been in place at the time the disease arrived in Sao Paulo.

We consider three time horizons: 1920 (the short-run), 1940 (the medium-run), and more recent data from 2006-2016 which constitutes the long-run. The sources of the short-run and medium-run data are the Brazilian censuses from 1920 and 1940. The 2006-2016 data are on the FIRJAN Development Index (explained in detail below). Using two stage least squares, we show that normalized respiratory deaths, our proxy for influenza-related deaths during the pandemic, significantly increased infant mortality and still births in 1920, and reduced sex ratios at birth in 1920. As an important demographic input into future fertility, marriage patterns and mortality dynamics, we argue that changes in the sex ratio at birth in 1920 (in favor of girls) led to measurable impacts on future demographics. Furthermore, using categories on cause-of-death from 1927 we find that districts that were more severely impacted by the pandemic faced increased normalized death rates by cause. All else equal, a unit increase in the respiratory deaths per 1000 inhabitants as of 1917-1920 increases the 1927 epidemic/endemic and respiratory deaths per 1000 inhabitants by 2.73 and 0.94 points, respectively. Other 1927 deaths from causes including circulatory or central nervous system sicknesses, and from puerperal sepsis, also rise as a consequence of the negative health shock of 1918. Continuing to focus on the short-run, productivity measures may also be affected since it is clear that both education and health were impacted. We find that short-run productivity, as measured by the volume of coffee, rice and maize per capita, and per establishment in 1920, deteriorated.

We also find that while there is a zero effect for males who can read and write (and those receiving instruction) in 1940, for women aged 5-39 years, an additional respiratory death per 1000 inhabitants leads to a decline of 0.07 points in the share that is literate. Hence, the medium-term analysis reveals that the pandemic led to the deterioration of education at most levels, with potentially a stronger negative impact on women. Considering baseline averages of overall educational attainment in Brazil at the time, these results are economically significant. Moreover, we find that respiratory deaths from the pandemic have a significant impact on the normalized number of inpatient hospital admissions in 1940. Productivity too was impacted over this time horizon. More specifically, the primary sector's output per employee and per establishment fell as of 1940.

We then consider long-term effects, focusing first on educational outcomes. We use the 2016 IFDM Brazilian Education Index to find persisting effects of the pandemic. The results indicate that normalized respiratory flu deaths from close to a century ago still have lasting impacts in some districts. In particular, this variable is a significant predictor of the IFDM Education Index, in a fully saturated model that includes all covariates and fixed-effects. Moreover, an additional unit of respiratory deaths per 1000 inhabitants as of 1920 leads to a sizeable downgrade (of 0.05 points) in the 2016 IFDM Brazilian Health Index, consistent with the intuition that health shocks of this magnitude have contemporaneous as well as consequential impacts from the inter-generational transmission of health and knowledge. Finally, the Influenza of 1918 is found to have an impact on the Employment and Income IFDM Index from 2016 and the IFDM Development Index (which envisages health, education and income) from 2006, 2011 and 2016. A one standard deviation increase in normalized respiratory deaths is associated with a lower Development Index of 0.09

points.

Our paper contributes to growing research that documents the path-dependency of human capital, inequality, poverty and development in response to historical epidemics and shocks (Clay et al., 2018, Alsan, 2015, Bleakley, 2010). This literature focuses on the deep roots of economic development and seeks to understand how past episodes of instability and historical shocks explain variations in contemporaneous growth rates, stocks of civil/social capital and political outcomes. To the best of our knowledge, our paper is the first to document the persistent effects of the 1918 Influenza Pandemic on demographic, health, education, and productivity measures, over the short, medium and long-run, in a developing country. Our study benefits from the fact that Brazil was not a major actor in the First World War, and from the fact that given its location in the Southern Hemisphere, the 1918 Pandemic arrived in Sao Paulo in the Spring. These factors make the context of our study unique, and allow us to distill results from an environment that is different from other research that has concentrated on the developed world in the Northern Hemisphere. Furthermore, we use data at a very granular level to compile a rich historical database that allows us to highlight what economic historians mean when they say that the 1918 Influenza Pandemic forever changed the social, economic and cultural landscape of countries.

The paper is structured as follows: Section 2 discusses relevant literature and Section 3 describes the historical background of the 1918 Influenza Pandemic in Sao Paulo, Brazil. Section 4 describes the data and provides the summary statistics on vital, geographical, demographic and economic characteristics. Section 5 describes the empirical methodology, and Section 6 details the short-term analysis of demographic indicators. Sections 7 and 8 document the persistent effects on human capital formation and productivity, respectively. Section 9 includes robustness and falsification tests and Section 10 concludes.

2 Literature

Several studies explore the historical roots of comparative development and use major demographic events, geopolitical turning points, or changes in the institutional environment, to explain contemporaneous disparities between nations and in-country variation in economic outcomes. A large segment of this literature acknowledges that economic processes are persistent, and that the path dependency of development can be understood only by acknowledging the legacy of past events.[1] In order to understand how diseases in particular can alter developmental paths, shape institutions, and have lasting consequences, Sokoloff and Engerman (2000) relate institutions and factor endowments to the persistence of inequality. Beach and Hanlon (2017) using wind patterns for identification, find that British industrial coal use in 1851-60 has significant mortality effects. These results provide further support for findings in Hanlon (2016) that long-run city growth in Britain in the nineteenth and twentieth centuries was negatively affected by local industrial coal use.

Looking specifically at the Influenza Pandemic of 1918, Almond (2006) evaluates empirically whether early life insults and *in utero* exposure to this health event had repercussions later in life. Using US census microdata from 1960-1980, the study finds that cohorts that were *in utero* during the pandemic have lower socioeconomic status and lower levels of education, health, and employment outcomes later in life. These results support the *fetal origins hypothesis*[2]. In a related paper, Almond and Mazumder (2005), using data from the Survey of Income and Program Participation (SIPP) from 1984-1996, find that birth cohorts *in utero* during the 1918 Pandemic have poorer health outcomes almost 65-80 years after the onset of the event. Nelson (2010) examines the impact of poor fetal health caused by the pandemic on later life outcomes (education, employment and wages) using labor surveys from 1986 to 1998 in Brazil. In keeping with previous work, those who were immediately impacted by this mortality shock suffered in the long-run along these dimensions. Using data from Taiwan, Lin and Liu (2014) conclude that 1918 birth cohorts are shorter as teenagers, have less education, and are more prone to various

[1] Seminal contributions include Huillery (2009) which investigates the long-term impact of colonial public investments in French West Africa; Nunn and Wantchekon (2011) that traces mistrust to the slave trade in Africa; Putterman and Weil (2010) which studies the role of human capital in long-run development; and Acemoglu et al. (2001), Dell (2010), and Glaeser et al. (2004) for work on relating variations in development to the institutional environment in the past. Guiso et al. (2016) shows that Italian cities with self-government in the Middle Ages have relatively high levels of civic capital today, and Rocha et al. (2017) exploits variation induced by the state-sponsored settlement policy during the historical episode of mass migration in Brazil to evaluate path dependency in human capital formation.

[2] The hypothesis that insults to the developing fetus, nutritional and otherwise, have lingering effects; attributed to Barker (1995).

diseases. These results are in line with those in Chul Hong and Yun (2017) which uses data from colonial Korea. Richter and Robling (2013) and Cook et al. (2018) are other studies that extend the time horizon to consider multigenerational effects.

Similar in spirit to our study, the few that measure aggregate effects of the pandemic on earnings, human capital accumulation and economic growth include Karlsson et al. (2014) which finds that regions more burdened by the pandemic had higher future poverty rates. However, there are no effects on earnings. Percoco (2016), using mortality rates across Italian regions finds that exposure to the flu lowered educational attainments for those who were *in utero* and in early childhood. Finally, Bakken and Husoy (2016) match 1912-1920 data on influenza mortality and other demographic estimates to the Norwegian 1960 census and find that exposure to prenatal influenza leads to significant declines in the years of education for men, with a larger effect in the poorest municipalities.

Our study contributes to the literature by evaluating a whole spectrum of outcomes that span demographic measures, educational outcomes, health variables and productivity measures, over three different time-lines: the short-run, the medium-run, and the very long-run. Importantly, we do this in the context of a developing country where previous data limitations have meant that studies can consider only a few outcomes at best. The disaggregated nature of our historical data matched with censuses from 1920 and 1940, and more recent data from the 2000s, enables us to analyze the consequences of the 1918 Influenza Pandemic in Brazil in a more exhaustive manner.

3 Historical Background

Sometimes described as the greatest medical pandemic of modern times, the 1918 Influenza pandemic resulted in global mortality estimates ranging from 20 to 100 million within the span of a few months (Alonso et al., 2016). Even if there is much debate around the precise mortality estimate, the consensus is that the 1918 Influenza Pandemic killed more people than World War I and more than the HIV/AIDS pandemic of the past decades. A century later, this historical event is mostly forgotten; often misleadingly believed to be devoid of important lessons because the high death toll could simply have resulted from the precarious preparedness of most countries at that time. The pandemic reached Brazil in the beginning of Spring, on 14th September 1918. *Demerara*-an English-flagged ship, entered the port of Recife, a northern Brazilian city, and then anchored in the harbor city of Santos[3], Sao Paulo (State) (Massad et al., 2007). The sailors on board *Demerara* returned very sick since they had been directly exposed to the flu in Dakar, Senegal, a stop on their way back from Europe (Alonso et al., 2016). A few days after the arrival of the ship in Santos, the influenza virus penetrated the interior parts of the country and many cases were reported in Rio de Janeiro and Sao Paulo, and in other cities in the northeast. With the intense movement of Brazilians, foreigners and immigrants across the country via ports and a broad railway network, the epidemic's pace and intensity could not be contained and it spread rapidly to urbanized and interior districts.

Evidence suggests that Brazilian authorities did not anticipate the deadly effects of the pandemic. The information that reached the country warning about the severity of the flu overseas did not receive due importance. Authorities had erroneously believed that oceanic distance would prevent the pandemic from reaching the country, and misjudged its seriousness. Hence when it arrived, it easily overwhelmed public authorities. The few measures in place proved feeble when confronted with the outbreak of a severe public health crisis, as seen by newspapers extracts from the time (Figure 1 and Figure 2). Prices of foodstuff (including milk, meat, and lemons) and medicines (mostly quinine-considered to be a powerful drug for any illness) quickly surged as deficiencies set in (Hochman, 2016). Improvised hospitals and healthcare posts had to be established in many cities that were trying to keep up with rising death rates. Sao Paulo (city) reported 116,771 cases of influenza, a prevalence rate of 22.32% (Nelson, 2010), for an estimated population of 523,194 inhabitants, and 5331 deaths in the short period between mid-October and mid-December 1918 (Massad et al., 2007). The Brazilian president Francisco de Paula Rodrigues Alves, newly elected for a second term in 1918, succumbed to the flu in January 1919. Testimonies immediately after the event suggest that in Sao Paulo alone, there could have been 10, 20 or even 50 thousand deaths (Alvarez et al., 2009). Some studies estimate that approximately 350,000 (two-thirds of

[3] A major port that witnessed the entry of thousands of immigrants and volumes of merchandise trade. The population transited daily from the port to the capital, the coffee plantations, colonial settlements, or from it to the outside world (Bassanezi, 2013).

Sao Paulo's population) might have been infected (Bassanezi, 2013, Barata, 2000, Bertolli, 2003). Our data suggest that in 1918 and 1919, when the pandemic was at its peak, the percentage of deaths attributed to influenza-related causes stood at 52.6% and 53.8%, respectively.

To bring some perspective, Pennsylvania, Maryland and Colorado in the US, with influenza deaths of 883.1, 803.6 and 776.5 per 100,000 of the population, respectively, had the highest mortality rates in 1918.[4] Garrett (2007) in an analysis of the economic effects of the Pandemic for the US, explains that the first wave occurred in March 1918 and lasted through summer of 1918. The second wave of fall of 1918 was the worst. As in Brazil, there was much variation in mortality rates across states and cities. Some studies provide much insight into the causes of such variations and one important reason could be that state and city officials took drastic measures to dampen the effects as quickly as possible, such that the last wave of 1919 was less deadly.

The municipality Capital (Sao Paulo, city), for which monthly data is available from 1917-1920 from Annual Statistical Reports, gives a snapshot of the situation[5]. Figure 3 shows mortality patterns in this region. Compared to the same months in 1917, there were 514 and 5274 more deaths in October and November 1918, respectively; with the total number of deaths up by 87.3% in this municipality. In Figure 4, we show the monthly total deaths per 100,000 of the population from January 1917 to December 1919. The extraordinary mortality shock caused by the pandemic is obvious. Figure 5 demonstrates that both neo-natal and post-natal mortality also peaked in October 1918. Figure 6 shows the spatial distribution of influenza-related deaths per 1000 of the population in Sao Paulo (State) in 1918. In Figure 7, we aggregate the district-level data up to the municipality level for the period 1917-1920 and show the spatial variation in normalized influenza-related deaths using the 1920 boundaries for the state of Sao Paulo.

The pandemic resulted in a proportionally higher mortality rate among prime-age adults, creating a 'W'-shaped age-profile distribution of affected groups [6]. A reason proposed for this unusual

[4]Mortality rates reported in (Garrett, 2007) are drawn from *Mortality Statistics in 1920* and include influenza and mortality estimates.

[5]The Capital area is representative of the State of Sao Paulo for the purpose of our analysis and includes the following sub-regions: Se, Mooca, Consolacao, Bom Retiro, Cambuci, Santa Cecilia, Perdizes, Bela Vista, Vila Mariana, Bras, Penha de Franca, Ipiranga, Santana, Lapa, Nossa Senhora do O, Sao Miguel, Butanta, Osasco, Liberdade, Belenzinho, Santa Efigenia.

[6]The 1918 Flu killed relatively younger men and women aged 15-44, unlike other influenza epidemics that

pattern is that young male adults in particular might have been more vulnerable given their relatively higher participation rates in the labor force, and their disproportionate exposure to other diseases such as tuberculosis. Another explanation stemming from the medical literature suggests that the "cytokine storm", causing an excess production of immune cells and their related compounds, cytokines, is evident in healthy young adults aged 20-40 years old during a flu infection, and the strong immune reaction results in premature death (Loo and Gale, 2007, Kobasa et al., 2004). Children and seniors with weaker immune systems are less affected as there is less risk that their immune systems overreact.

This respiratory illness had further debilitating effects on those affected by tuberculosis, renal or heart diseases, and women in the child-bearing age and in the post-partum period. Further, since Brazil was experiencing its first wave of any such pandemic, there was no benefit of prior exposure to dampen the severity of the event (unlike other regions of the world where populations were affected by both spring and fall waves). Reflecting the lack of experience, medical reports from the time were incomplete and imprecise.

The Spanish flu was termed the Democratic flu by the press because it did not spare the country's elites nor the poor layman (Hochman, 2016). Although the "socially neutral" view is widely held, in Sao Paulo, relatively few wealthy elites died of influenza. It was the poor, living in unsanitary conditions and slums that are bereft of basic health services, that were most affected (Bertolli, 2003).[7] Given its sudden, unanticipated arrival in Spring in the Southern Hemisphere, and with its short duration of about ten weeks, previous studies posit random assignment of the infection in Brazil (Nelson, 2010).

In response to the onset of the Pandemic, public health authorities designed several information campaigns to disseminate preventive and curative measures to be taken.[8] These included isola-

typically kill mostly young children and older cohorts resulting in a U-shaped distribution.

[7]Testimonies from doctors reveal that the mortality rate was higher for those living in poverty; working odd hours doing heavy manual tasks under difficult conditions; and for those exposed to unstable climatic conditions (Alvarez et al., 2009).

[8] Failure to comply with these measures may explain why different areas in Sao Paulo were impacted differently by the virus. For instance, Bassanezi (2013) reveals that two municipalities, Campinas and Riberao Preto, that had suffered from yellow fever in the late nineteenth and early twentieth century, acted on these measures propitiously (Bertucci-Martins, 2005). Sorocaba, relatively smaller than Campinas and Riberao Preto, and with less experience in epidemics, struggled more as they failed to adopt the right initiatives.

tion, good personal hygiene, no extra work that could result in extreme fatigue, and quarantine measures in several areas, amongst others. Furthermore, there was confusion everywhere as statements released by medical establishments that conveyed that the epidemic was mostly benign was in dissonance with published data that indicated an exponential growth in the number of victims. Hospitals quickly ran short of medicines and a full blown public health crisis ensued in Sao Paulo as Brazilian authorities struggled to keep pace with the pandemic.

4 Data and Summary Statistics

Using historical and archival records, we construct a unique database on health outcomes related to major disease categories and socioeconomic indicators for districts in the State of Sao Paulo in the early decades of the twentieth century. We complement the data on district-level deaths by cause with measures of infrastructure, demographic information and geographic variables. We accomplish this by digitizing statistical reports and by matching these to Brazilian regional census data from 1920 onwards to obtain a panel of information. Our spatial unit of analysis is at the district level-a disaggregated territorial stratification of the State of Sao Paulo.[9] Use of districts gives us 350 geographical observations for the period 1912-1921, which we combine with other historical and contemporary data from 1872, 1912-1921, 1940, and 2006-2016.

An important consideration while creating our dataset is to ensure that standardized boundaries are carefully tracked over time. The boundaries of many districts and municipalities changed considerably over the time period of our analysis and Figure 8 shows the evolution of borders over time in Sao Paulo from 1872-2016.[10] We use official territorial and administrative maps that provide information on the evolution of districts over time to match the data in a consistent manner.

[9]A district in Brazil is an administrative unit within a municipality. We choose to use district-level data rather than municipality-level data for three reasons. First, the data on vital statistics for 1917-1921 is available at the district-level, and second, using districts provides a larger sample size. Third, use of districts allows a more refined analysis of realities at a very microeconomic level.

[10]The first census of 1872 contains data on 88 municipalities; by the 2010 census reports, the number of municipalities reached 645.

We start by describing the key variables of our analysis and the source from which they are obtained. Table A1 in the Appendix provides details on other variables used in this study.

4.1 Deaths by Cause and Vital Statistics

The Sanitary Authority, created in 1892, was responsible for ensuring the collection of reliable health statistics in Sao Paulo. These statistics are publicly available from 1901-1928 (Bassanezi, 2013).

The Annual Statistical Reports of Sao Paulo contain the vital statistics required for our analysis. More specifically, we obtain deaths by cause for 14 major disease categories, allowing us to construct our key variable of interest: normalized deaths from respiratory infections from 1912-1921, our best proxy for influenza-related deaths.[11] The Annual Statistical Report also provides information on economic and financial statistics, including municipalities' receipts and expenditures which we use as controls in some of our models.

We also obtain the initial shares of influenza-related deaths for 1915 in the *Annual Statistical Report of Sao Paulo (1915)*, for which deaths by cause are available for 180 municipalities only. Further, we use the *Demographic Studies: The Population of Sao Paulo in the last decade: 1907-1916* to obtain pre-pandemic population statistics for most municipalities. The nationwide influenza-related deaths come from the yearly statistical reports of 1917-1921.

For measuring the short-term health effects, we obtain deaths by cause at the district level from the *Annual Statistical Report of 1927*. We also use the normalized number of in-patient hospital and asylum admissions in 1940 (obtained from *Annual Statistical Report of 1940*) as a medium run measure of health. We use per capita municipality expenditures on hospitals and health from the same source as additional controls.

[11] For reasons explained in the next section, we also collect data on mortality caused by other diseases and several other vital statistics. These data are further complemented with information from a special official Sanitary Report in 1918.

4.2 Climate Data

The Annual Statistical Reports of Sao Paulo provide scant statistics on meteorological variations. For instance, the 1917 report (pp.35-37) provides the monthly temperature for less than 65 regions and the monthly precipitation for less than 35 municipalities. The atmospheric pressure, humidity level and detailed air temperature-related statistics are only at the aggregate level. Since our instrumental estimation strategy relies on October's average climatic variations, we refrain from matching arbitrarily these weather conditions since the lack of variation might distort our results. Rather, we compile our climate data using averages from the World Bank Climate Change Knowledge Database, where we are able to extract the disaggregated relevant data for the period 1901-1930, using latitude and longitude coordinates of districts.

4.3 Sanitary and Health Infrastructure

We collect the number of doctors, chemists and midwives, and the number of people with mental and physical deficiencies (all per 1000 inhabitants) as of 1872 and match these to the 1920 district-level data for use as proxies for initial historical health indicators before the pandemic struck. Information on water and sewage systems is obtained from the *Sanitary Service Report of the State of Sao Paulo: Annual Demographic and Sanitary Statistics Section (1920)*. This report is also used to obtain the number of hospitals and old-age homes, and specialized maternity hospitals, for use as controls when the outcomes are child death, or still births or deaths caused by specific diseases. The *1910 Annual Statistical Report of Sao Paulo (Volume II)* is used to obtain the pre-pandemic public expenditures on cleaning, waste disposal, and maintenance of good sanitary conditions in 132 municipalities.

4.4 Geography and Railroads

We control for altitude, latitude and longitude in our models. Altitude and related data are collected from the different volumes of *The Encyclopedia of Brazil Municipalities (IBGE 1957,1958)* for Sao Paulo.

The dummy for the presence of railway in the district is created from the *Secretariat Report of Agriculture, Commerce and Public Works of the State of Sao Paulo; Coffee: Statistics of Production and Commerce (1920).* We know the number of railway companies and stations in 1920, and the year in which the railway network was established. Distance to capital (both in a straight line and more precisely) come from the same 1920 report and from the *Ipeadata* (the Institute of Applied Economic Research in Brazil).

4.5 Productivity

The *1920 Census of Agricultural and Industrial activities* is the source for our productivity measures in 1920. We calculate productivity measures at the municipality level by combining data on establishment size (measured by the number of employees and by size), production of different commodities, and the characteristics of small scale and large scale business from the 1940 census for Sao Paulo.

Data on the number of people employed in schooling activities (school workforce) is sourced from the *1940 census (Parte XVII Sao Paulo, Volume 1).* The *Annual Statistical Report of 1940* is used to obtain the total municipality expenditure on education[12] and the budget share that accrues to school supplies, materials and to teachers.

The 1940 census (*Parte XVII Sao Paulo, Volume 3*) is used to obtain the total expenditure per establishment for the primary sector (and for the subsectors agriculture, farming and livestock), and the number of plows used per establishment, which is a proxy for capital utilization and mechanization. We also compute expenditures per agricultural establishment on new seeds, fertilizers and insecticides, and on salaries and the acquisition of new machinery and animals in 1939. These variables are used as controls in our models.

[12]These data allow us to calculate the share of total municipality expenditures devoted to the Secretary of Education and Public Health.

4.6 Census Data

We use the 1872, 1920, and 1940 censuses for the State of Sao Paulo to estimate the short-term and medium-term effects. The first round of 1872 allows us to obtain pre-pandemic initial conditions. In addition to those noted above, other variables that we obtain from this Census include the share of foreigners, the share of literate people, the share of people who were slaves, population density, and race. Controling for initial conditions is important to ensure that we have a baseline for the degree of development, urbanization, infrastructure, social aspects, health-service, and the sanitary environment before the Flu Pandemic struck.[13]. The 1920 census is used to compute demographic and economic variables.[14] The 1940 census is used to construct medium-run measures, as has been detailed above.

4.7 Contemporaneous Data

To estimate the long-term effects of the 1918 Pandemic on education, health, productivity, income, employment and development, we use the *FIRJAN Municipal Development Index (IFDM)*. This index monitors the socioeconomic development of all Brazilian municipalities on an annual basis. We use the overall index and the sub-indices pertaining to health, education, employment and income from 2006-2016.

4.8 Summary Statistics

Tables 1 and 2 present the summary statistics for the dependent variables and selected explanatory variables in our analysis. Variables in Panel A of Table 1 are the dependent variables in the immediate short-run of the pandemic and include infant mortality, still births, the literacy rate for males and females, and the sex ratio at birth, amongst others. Focusing on a few of these, the mean value for the infant mortality rate is 0.02 and that for still births per total births is 0.05.

[13]Note that we use these measures in proportions.

[14]Amongst others, we calculate sex ratios for Brazilians and foreigners, and literacy rates by age groups and gender.

As expected, male literacy is higher than female literacy. The mean sex ratio at birth in 1920 was 121, with a standard deviation of 82, suggesting that there were highly skewed sex ratios in some districts. Panel B of Table 1 report statistics for the medium-run variables of interest. These include educational measures disaggregated by gender and health measures including hospital admissions. Finally Panel C of Table 1 denote the dependent variables of interest in the long-run. These include the IFDM development indices from three points in time and the 2016 index disaggregated into its components of employment and income, health and education.

Table 2 reports summary statistics for the variables used as controls in the models that follow. These are also arranged by panels that focus on deaths by cause, geography, variables from the 1872 census, and other measures that are used as controls. Our key variable of interest is in Panel A of Table 2–respiratory deaths per 1000 inhabitants, which has a mean of 1.74 and a standard deviation of 2.03 for the period 1917-1920. The average October's temperature was 68.43 Fahrenheit with a standard deviation of 3.76; while the average October's precipitation was 124.15 mm (standard deviation 8.39mm). We note the high standard deviation for health and sanitation municipality expenditure in 1910, and the adult sex ratio in 1920 had a mean of 112 with a standard deviation of 59. Overall, the data provides evidence of the relatively low levels of development of most districts during the time period under analysis. Further, these statistics reveal that there was significant variation in these measures, especially in those pertaining to literacy, health and infrastructure.

5 Estimation Methodology

In Section A1 of the Appendix, we attempt to use the individual-level data from the Integrated Public Use Microdata series (IPUMS) for Sao paulo for the years 1960, 1970 and 1980 to replicate the methodology used in Almond (2006) and Beach et al. (2018), and to understand the possible long-term effects on specific cohorts exposed to the 1918 Influenza Pandemic. Since these regressions cannot be equipped with a robust set of cohort controls given the limitations of the data, we choose to rather collect detailed granular-level outcomes and controls for additional variation from the Annual Statistical Reports from the actual time period when the Pandemic was prevalent as the cohort effects shown in the Appendix might be distorted (given the drastic changes that happened in Sao Paulo during the twentieth century). Thus, we think that relying on cohorts alone might not be sufficient in the Brazilian case and that the compilation of rich data from various historical sources and the use of regressions based on 1920 geographical demarcations are both justified in our analysis.

5.1 Empirical Specification

We examine the impact of the 1918 Flu Pandemic using the following specification:

$$y_d = \beta_0 + \beta_1 Flu_d + \beta_2 X'_d + \beta_3 X'_{d0} + \beta_4 D_d + \epsilon_d \tag{1}$$

where y_d is the outcome of interest for district d in either 1920, 1940, or 2006-2016. Flu_d is the normalized respiratory deaths in district d at time t (sample restricted to 1917-1920).[15] X_d is a vector of district-level controls, including geographic variables such as altitude, latitude, longitude, and other controls including the distance to capital, a dummy for the presence of a railway, and the interactions of these variables.[16] X_{d0} is a vector of initial conditions obtained from the 1872 census that includes the share of foreigners, the share of literate people, the number of doctors, chemists and midwives, population density, the share of people who were slaves (during the pre-abolishment era), the share of people of different races, and employment in different economic sectors. D_d denotes district dummies. Time dummies are used when we stack the data allowing for time variation in the dependent variable. ϵ_d is the idiosyncratic error term. The coefficient of interest is β_1, the impact of normalized respiratory deaths on the outcome of interest.

There are several reasons why respiratory deaths may not be exogenous. First, measurement error. Official published data may understate true death rates, particularly during the peak of the health crisis between October and December 1918. This may not have happened intentionally as in the confusion that followed the initial outbreak that we detail above, doctors may have

[15] When we estimate the short-run impact on 1920 variables, the sample is restricted to 1917-1919.

[16] Other controls are added in 1940, depending on the outcome of interest.

had little time to make accurate entries. Many stricken victims did not report to the doctor and many doctors, in turn, did not always provide unambiguous detailed death reports. Further, as Richter and Robling (2013) note, panicked and overwhelmed doctors probably chose to make optimal use of their time by treating long lines of frail patients, and focusing on curative work rather than working on long descriptive death reports required by government officials. Others probably struggled to provide the right diagnosis and to assign the cause of death. The epidemic required a complex and well-defined health agenda to ensure, amongst many other things, that vital statistics are properly recorded. In the midst of chaos, deaths caused by dengue, cholera, typhoid and other diseases might have been misclassified as influenza-related mortality, and vice versa. The reported number of deaths caused by influenza might not reflect ground realities in Sao Paulo. The resulting measurement error could cause an attenuation bias if the above concerns are not addressed. Figure 9 shows the spatial variation in the normalized number of deaths caused by respiratory infections (left map) and by unknown causes (right map) only in 1918. We note for that specific year, that the maximum value for respiratory deaths per 1000 of the population was 4.5 while that of unknown causes stood at 45. This suggests that it is highly probable that many cases were recorded as deaths caused by unknown diseases given the difficulty of recognizing the symptoms and the public health crisis triggered that year.

Second, omitted variables may be simultaneously correlated with the incidence of flu in the district and with some unobserved characteristics. To control for threats to identification posed by omitted variables related to the mortality gradient by socio-economic class, sanitary conditions, literacy, and nutrition among other variables, we include several controls related to health, demography, human capital, levels of economic development, sanitary infrastructure (including water and sewerage), and the municipality's public health expenditure. Models further include a set of demographic and geographic initial conditions from the 1872 census to control for differences in baseline conditions, and for existing pre-pandemic characteristics.[17] Standard errors

[17] In Table A5 in the Appendix, we show the sample means differences for selected geographic, climatic, demographic, economic and health initial characteristics for regions in our dataset. We calculate the median normalized respiratory deaths for 1917-1920 and then obtain summary statistics for these characteristics for regions with below and above median flu exposure. We note from these results that geography, climate, population density, the dependency ratio and the literacy rate differs significantly between the two groups. The sample mean difference for per capita municipality expenditure on cleaning and waste disposal in 1910 is also significant at the 1 percent leve.

are clustered at the micro/mesoregion level.[18] The underlying intuition is that neighboring districts and municipalities share similar unobservable features; more compellingly, most 1920 census districts belonged to larger stratifications sharing common initial conditions. Micro/mesoregion fixed-effects are also included in all models.[19]

The third reason for why respiratory deaths may not be exogenous is that male and female survivors may have been the strongest members of society by virtue of being resilient to shocks *in utero*, and because female fetuses are less sensitive to *in utero* stressors (Noymer and Garenne, 2000, Hamoudi and Nobles, 2014, Nelson, 2010).[20]

5.2 Instrumental Variables Approach

We use an instrumental variables approach to overcome the issues noted above, relying on the fact that seasonal patterns and average environmental conditions observed across regions in the month of October can explain part of the variation in the prevalence of viral respiratory infections and influenza-related diseases. In doing so, we draw on the medical and epidemiological literature that the incidence of an influenza epidemic depends largely on climatic conditions (Polozov et al., 2008, Tamerius et al., 2013, Slutsky and Zeckhauser, 2018). We rely on (only) October's temperature and precipitation, as average yearly climatic variables could potentially determine historical prosperity and future economic outcomes simultaneously. We posit that October's temperature and rain are valid instruments as they are relevant and satisfy the exclusion restriction. In particular, they are correlated with influenza exposure, and the geographic distribution of these climatic variations in October are plausibly exogenous to measures of economic well-being, educational infrastructure and other long-term trends. Further, they affect outcomes at different

[18] Micro/mesoregions in Brazil are as follows: The mesoregion 'Aracatuba' contains the microregions 'Andradina', 'Aracatuba' and 'Birigui'. Each of these microregions in turn contains various districts/cities. Another example: the mesoregion 'Piracicaba' contains the microregion 'Limeira', 'Piracicaba' and 'Rio Claro', each having their own districts.

[19] Bertrand et al. (2003) note the importance of clustering standard errors at the largest sensible aggregation.

[20] Our instrumental variables approach allows us to overcome the threats posed by measurement errors and the possible attenuation bias that could cause an OLS framework to show zero effects. As for the possible selection effect that could possibly be work with stronger resilient babies with altered heterogeneity at birth following the major health shock, we make the assumption that given the magnitude of this shock, there is a normal distribution of strong and weak young members across the state.

points in time only through their effect on respiratory deaths in the early short-run period. To ensure that this is the case, our regressions also include potential geographical controls for average temperature and rainfall.

Figure 10 shows the spatial variation in average october's temperature and precipitation, respectively. The mean temperature for all districts is estimated to be 68.43 Fahrenheit and the mean precipitation is 124.15 mm. We collect the data for the few observations available in the Annual Statistical Report of 1917 and obtain a mean of 68.54 Fahrenheit and 119.4 mm for October's temperature and precipitation. Figure 11 provides a visual representation of the relationship between October's temperature, October's rainfall, and the median normalized respiratory deaths for the pandemic years of 1918 and 1919. Panel A of Figure 11 shows that regions that registered higher average precipitation levels in October relative to the median in all regions had more respiratory deaths in both pandemic years. Panel B indicates that the prevalence of respiratory deaths was higher in relatively hotter districts in 1918. While this might appear counter-intuitive, the epidemiological literature has documented that most influenza fatalities of the 1918 Pandemic involved complications from secondary bacterial pneumonia caused by bacteria which thrive in hotter climates. In Panel C, we note that regions with higher average temperature had a lower prevalence of deaths caused by unknown causes. This more intuitive statistical relationship once again point to the possibility of measurement error in deaths caused by respiratory deaths.

Panels A and B of Figure 12 portray binscatter plots (with geographical controls) for normalized respiratory deaths and October's rain and temperature, and are a closer approximation of the first-stage results than the correlations depicted in Figure 11. Panel A focuses on the period 1917-1920 while Panel B uses only 1918 data. Both panels show that October's rain and temperature impact predicted normalized respiratory deaths; the empirical results that follow confirm that respiratory deaths are inversely related to temperature and positively related to precipitation.

The first stage regression for respiratory deaths (per 1000) is as follows:

$$Flu_d = \alpha_o + \alpha_1 OctTemp_d + \alpha_2 OctRain_d + \alpha_3 Geog_d^{'} + \alpha_4 Ini72_d + \alpha_5 D + u_d \tag{2}$$

where $OctTemp_d$ and $OctRain_d$ denote average October's temperature and precipitation for each district; $Geog_d$ is a vector of geographical controls and $Ini72_d$ is a vector of baseline socioeconomic characteristics. D are region dummies, and u_d is the standard idiosyncratic error term. Results for the first stage regressions are shown in Table 3.

Table 3 (columns 1-3, for our period 1917-1920) reports that temperature has a negative effect while precipitation has a positive effect on normalized respiratory deaths. In columns (2) and (3), the results remain unaltered with the inclusion of district-level controls. The estimates on temperature and rainfall (along with the full set of controls) explain about 44.3% of the variation in normalized respiratory deaths. Moreover, the F-statistic on identifying instruments are above the rule-of-thumb threshold value of 10. To ensure that our instruments affect our outcomes of interest only through their effects on respiratory deaths in 1917-1920, we carry out in columns 4-6 and in column 7 of Table 3, similar first-stage regressions but on 1913-1915 and only 1914 respiratory deaths, respectively. We obtain evidence that October's mean temperature and precipitation have no statistical impact on our endogenous regressor (except for a positive significant effect of temperature noted in 1914 only, but of lower magnitude compared to the pandemic years, and with a joint F test suggesting that the two instruments are not jointly significant in any year, except for 1917-1920).

October's temperature and rainfall are the instruments in all cases except when the outcome is agricultural productivity. This is because studies have documented the possible relationship between climatic conditions and agricultural output in the Age of Mass Migration (Hatton and Williamson, 1998, Solomou and Wu, 1999). Instead, we use an alternative instrument as discussed below. For the agricultural productivity outcomes, we construct another instrument based on Acemoglu and Johnson (2007) and Percoco (2016). This instrument relies on the baseline regional distribution of deaths by respiratory infections in each district in $t < 1917$, and also leverages the surge in influenza deaths during the pandemic years of 1917-1920. Let $AggFlu_t$ denote the aggregate number of deaths from influenza in Sao Paulo in year t. Let $\bar{s}_{1915d}$ be the proportion of deaths caused by influenza in a baseline period in district d. We choose the year 1915 as baseline as data are available for a representative number of districts in that year. We can reasonably assume that the baseline distribution of deaths by respiratory infections in each district and Sao Paulo's aggregate influenza mortality rate in a later year t, are exogenous. So, as Percoco (2016) shows, this instrument controls for the bias caused by the endogenous spatial distribution of the pandemic. For $t > 1915$, we impute the respiratory deaths as $(\bar{s}_{1915d} \times AggFlu_t)$. We standardize this value by pop_{dt}, the population of district d in year t. Equation 3 shows $IVResp_{dt}$, the constructed instrument for respiratory deaths:

$$IVResp_{dt} = \bar{s}_{1915d} \frac{AggFlu_t}{pop_{dt}} \tag{3}$$

Equation 3 exploits two sources of variation: the cross-sectional variation in the proportion of deaths caused by influenza for each district in 1915, and the time-series variation induced by changes in the aggregate mortality rate. The validity of the instrument rests on the assumption that the district-specific characteristics that determined $\bar{s}_{1915d}$ do not predict the future spatial distribution of the disease. This assumption may be violated however if the baseline proportion in 1915 is correlated with future changes in population, literacy rates, or income, and the district-specific features that determined the baseline weight affect the evolution of social and economic conditions during the pandemic years. To ensure this is not the case, we interact a district's pre-pandemic initial characteristics with year dummies. We do likewise for the geographic variables. Further, we use the per capita public expenditures on cleaning, waste disposal and maintenance of good sanitary conditions available for 132 municipalities in 1910 to ensure the assumption on which the validity of the instrument rests, holds. The first stage regression (as a function of the identifying instrument only) is shown in equation 4, with results reported in Table 4 :

$$Flu_{dt} = \gamma_0 + \gamma_1 IVResp_{dt} + \omega_{dt} \tag{4}$$

Table 4 indicates that this synthetic instrument is strong with F-statistics above the required threshold of 10. Morever, the coefficient on the constructed instrument $IVResp_{dt}$ is positive and significant, and robust to the inclusion of fixed-effects, and district-level geographical, health and initial controls. However, compared to the climatic instrumental variables discussed above, using the predicted influenza mortality rate in equation 3 reduces the sample size considerably. Hence we use this constructed instrument only for agriculture-related outcomes, as was our original intension.

6 Short-term Effects on Demographic Indicators

We begin our analysis on the pandemic's effects by evaluating its immediate impact on infant mortality, sex ratios at birth, and on the normalized number of still births. Panel A in Table 5 shows the short-run impact of normalized respiratory deaths caused by the pandemic on infant mortality (normalized by the number of live births) in 1920. Both OLS and 2SLS methodologies yield positive and significant coefficients in all specifications, with and without the inclusion of region fixed-effects, and initial and region-level controls (including dummies for the presence of water and sewage systems, hospitals, nursing homes, specialized maternity hospitals, geographical controls, literacy rates, and initial and demographic controls). The IV estimates are larger than the OLS results indicating that the latter are downward biased, possibly for the reasons outlined above.[21] In the fully saturated model, a unit increase in normalized respiratory deaths results in a 0.02 point increase in the number of infant death per live birth. Our estimates of the impact of the pandemic on the normalized number of still births–loss of the baby before the 20th week of pregnancy in 1920–are presented in Part B of Table 5. The instrumental variable effects in Panel B indicate that, all else equal, there was a significant increase in the normalized number of still births by 21 log points.

Table 6 reports results for sex ratios at birth in 1920. Controlling for adult sex ratios in 1920, the share of people of different races, the initial normalized number of midwives, and the presence of water, sewage systems, hospitals, nursing homes and specialized maternity hospitals, as well as geographical controls, we find a negative impact on sex ratios, as shown in Panel B, columns (1) to (3). In keeping with the intuition that male fetuses are relatively more vulnerable to shocks as compared to female fetuses, the IV results indicate that sex-ratios declined, that is, relatively

[21] Throughout our analysis, the IV coefficients are larger in magnitude than the OLS estimates. We argue that this is due to measurement error resulting in attenuation bias, and possibly also because of omitted variables bias. We report only 2SLS results from 1940 onwards.

fewer males were born in the immediate aftermath of the pandemic.

7 The Persistent Effects on Human Capital Formation

7.1 Educational Attainment

There is substantial evidence that shocks *in utero* or in very early ages can have long-lasting consequences on educational attainment (Currie and Stabile, 2006, Currie, 2009, Parman, 2013). We test this hypothesis by estimating the impact of influenza deaths on male and female literacy rates for different age groups. As discussed above, we use the 1920 census for short-term effects and the 1940 census for the medium-term effects.[22] For the contemporaneous data, we use the sub-index of the FIRJAN Development Index of 2016 that focuses exclusively on education. Intuitively, we do not expect to find significant findings for 1920 as it is unlikely that literacy rates would have changed in such a short time span as a consequence. We find that this is indeed the case in Table 7 where both the OLS and IV estimates are measured imprecisely.

The Census of 1940 included 7 questionnaires related to literacy. Compared to 1920, however, questions were not limited only to the ability to read and write. We are therefore able to compute average municipality's literacy rate for people receiving some form of instruction. Given the evidence in Musacchio et al. (2014) that Sao Paulo experienced a rapid increase in the number of schools, students and teachers during the 1889-1930 period, we control for the share of total municipality expenditure allocated to education and health in 1940, and the number of teachers (per person aged 10 and above). Other controls for initial conditions, share of people who were slaves, race measures and the urbanization rate as well as controls for altitude are also included in these regressions.

In Table 8, we estimate the impact on the literacy rate of males and females (aged 5 and above and aged 18 and above) in columns (1) to (4), and on the share of people (aged 5-39) who can

[22] The 1920 census reports literacy figures for those who can read and write, but we do not know whether formal schooling took place. Further, the 1920 census provides the number of literate and illiterate individuals at the district level, while the 1940 census provides municipality-level data.

read and write, out of those receiving some form of instruction in 1940 in columns (5) and (6). We find a significant impact on the male literacy rate in column (1). However, we do not find any statistical significance for the female literacy rate in column (2) of Table 8. As a robustness check, we estimated the impact on the literacy rate for males and females aged 18 and above (in columns (3) and (4)) and obtain a positive significant coefficient on males' literacy rates again, but find no significance for female shares. However, this is not the case when considering columns (5) and (6) of Table 8. Here, for females aged 5-39, all else equal, one more respiratory death per 1000 inhabitants leads to a decrease of about 0.07 points. For girls aged 7-14 (results not reported here), the corresponding impact is also negative and significant.

We next analyze the impact on Brazil's IFDM Education Index for each municipality in 2016.[23] The instrumental variables results are as shown in Table 13 (columns (3) and (4) only, Panel A). All else equal, the results indicate that 1917-1920 respiratory deaths cause some municipalities to lag in education. A one unit increase in respiratory deaths per 1000 of inhabitants leads to a 0.03 point decrease in the Education Index, almost a century later. The findings in this section lend credence to the hypothesis that the 1918 Influenza Pandemic has long-lasting implications on human capital formation even today.

7.2 Health Outcomes

We begin our short-term analysis of health outcomes by estimating the impact of our key variable of interest on the 1927 district-level normalized number of deaths caused by diseases. We choose as our dependent variables the number of deaths caused by an epidemic or by endemic diseases; central nervous system diseases, failure or complications of the circulatory system, respiratory diseases, maternal deaths from puerperal sepsis, and finally, skin-related diseases.

[23] The index ranges from 0 (minimum) to 1 (maximum) to classify the level of development in four categories: low (0 to 0.4), regular (0.4 to 0.6), moderate (0.6 to 0.8) and high (0.8 to 1). It includes various aspects of education at the municipality level: attention to early childhood education, primary school drop-out rate, an indicator for the proportion of pupils with a school delay of two years or more (in Brazil, elementary school starts at 6 years old, and primary education should end at age 14), teachers with a high-level of qualifications, average daily classroom hours in elementary school, scores from standardized tests, and the Basic Education Development Index (IDEB). All data were obtained through the National Institute of Studies and Research Anisio Teixeira (INEP) from the Ministry of Education.

The results are shown in Table 9. Both OLS and IV results suggest that districts more severely impacted by the pandemic indeed faced increased deaths (per 1000) seven years after the pandemic years. On average, a unit increase in the respiratory deaths during the pandemic years increases the 1927 epidemic/endemic, respiratory and circulatory deaths per 1000 inhabitants by 2.7, 0.94 and 0.80 points, respectively. Though of lower magnitudes, the estimates for deaths caused by nervous failures, puerperal sepsis and skin-related diseases are also positive and significant. These short-term findings suggest that there might be lasting mortality effects from the 1918 Pandemic.s

For the medium-term analysis, we compute the hospital in-patient admissions per 1000 inhabitants and the normalized number of people with disabilities admitted to nursing homes and asylums in the same year. The results are reported in Table 10. While we find no statistical significance for the nursing homes/asylum admissions, we find that respiratory deaths have a significant impact on the normalized number of in-patient hospital admissions.

In Panel A of Table 13 (columns 5 and 6), we report the IV effects of the pandemic's respiratory deaths on the 2016 IFDM Brazilian Health Index.[24] We find that in both columns, the coefficients on respiratory deaths are negative and highly significant. An additional unit of respiratory deaths per 1000 inhabitants leads to a sizeable downgrade in the index (of 0.05 points, when all controls are included). This means that almost a century after the pandemic, districts/municipalities still face its enduring impacts in various dimensions of health.

8 The Persistent Effects on Productivity

We study other channels through which the 1918 Pandemic could have altered Sao Paulo's development path in this section. In particular, we focus on the agricultural sector and consider different measures of productivity. Our outcomes of interest in the short-term are the 1920 output of coffee, rice and maize, normalized by the total population to obtain per capita values, and also

[24] The index ranges from 0 (minimum) to 1 (maximum) to classify the level of development in four categories: low (0 to 0.4), regular (0.4 to 0.6), moderate (0.6 to 0.8) and high (0.8 to 1).The Health Index includes the number of prenatal visits; deaths due to ill-defined causes; and infant deaths due to preventable causes. The index then encapsulates all these factors such that a negative coefficient implies a downgrade in health-related development. Only Official Public Statistics from the Ministry of Health are used.

by the number of establishments to obtain per establishment volume of output. The results are reported in Table 11. There are immediate sizeable effects on per capita volumes of coffee, rice, and maize in 1920, indicating that the pandemic caused decreases in agricultural productivity. These agricultural changes may have encouraged plantation owners to mechanize more quickly (even more so if the fall in productivity is accompanied by a rise in wages).[25] We note finally that only the coefficient for per establishment volume of rice is significant (at the 10% level), while the other per establishment estimates are imprecisely measured.

Table 12 reports the IV results for aggregate measures of productivity for the primary sector in 1940. To study the medium-term effects, we calculate the value of primary sector's output per employee (columns 1 and 2) and per establishment (columns 3 and 4), in logs. The results indicate that there are still measurable drags on labor productivity from the pandemic.[26] A one unit increase in respiratory deaths (per 1000) leads to a 0.45 log point decline in the value of the primary sector per employee. We obtain a coefficient of similar magnitude for the value per establishment. Relative to the averages for these measures of productivity, these impacts on productivity are significant.[27]

We next estimate the long-term effects of the pandemic on productivity using the Employment and Income IFDM Index for 2016.[28] The results are reported in Table 13 in columns (1) and (2) only. The estimates are negative and significant implying that the pandemic altered labor market dynamics in consequential ways. In sum, these findings indicate that the negative effects of the 1918 pandemic on economic and social indicators are still evident in Sao Paulo today.

[25]Table A7 in the Appendix shows that wages for coffee workers rose.

[26]We have added several additional controls including a proxy for mechanization and capital utilization–the number of plows used per establishment, disaggregated forms of expenditure per agricultural establishment on new seeds, fertilizers and insecticides, on salaries and on the acquisition of new machinery and animals in 1939.

[27]Table A8 in the Appendix shows the different measures of productivity in 1940 for small-scale and large-scale businesses in farming, agriculture, and livestock. The results suggest that by 1940, there had been declines in land productivity as measured by the value of the agricultural sector per hectare of cultivated land. Moreover, the disaggregated analysis indicates that the negative effect on labor productivity is stronger for the agricultural sector relative to livestock and farming.

[28]This index is composed of important dimensions related to local labor and income including indicators for the creation of formal employment; the municipality's ability to absorb local labor, income generation, actual wages in the formal labor market, and the Gini index of income inequality in formal work. The measure ranges from 0 to 1. The data sources used to build the index include the Annual Report of Social Information, General Register of Employment and Unemployment from the Ministry of Labor, and Official IBGE population projections.

Finally Panel B of Table 13 reports impacts on the IFDM Brazilian Municipal Development Index (which includes health, education, employment and income dimensions). We study the impact using the index for 2006 (columns 1 and 2), 2011 (columns 3 and 4), and 2016 (columns 5 and 6). We find that the Development Index for all years analyzed is reduced by approximately 0.04 points for each unit increase in the respiratory deaths per 1000 inhabitants. This finding is robust to the inclusion of region fixed-effects and historical and district-level covariates. These results further underline that the sizeable short-run effects translated into persistent effects on development indicators in the long run in Sao Paulo.

9 Robustness Checks

We conduct specification checks to verify the robustness of the results. First, we test to ensure that the sample of births in the flu-years are representative. We rely on male and female literacy rates in 1872 as a measure of parental characteristics, and compare these to the rates in 1920 in order to rule out selection in parental observables. The test of differences in means between the 1872 and 1920 literacy rates cannot reject that these are the same ($p = 0.41$). We then generate an indicator variable that takes a value of 1 if a district in 1918 had flu exposure above the median value in the sample (that is, if respiratory deaths per 1000 inhabitants exceeded 1.21 in 1918). We then test for differences in sample means in literacy rates across these two groups. The results are reported in Table 14. Again, test results indicate that the samples are comparable along this dimension.

In Table 15, we carry out a falsification test by demonstrating that the normalized number of deaths caused by respiratory infections during a non-pandemic year (1915) has no impact on any outcome variables. All the coefficients are statistically zero for sex ratios and literacy rates. Finally, since part of our empirical strategy relies on the predicted mortality instrument, we show in Table 16 that the baseline weight ($\bar{s}_{1915d}$), the share of total deaths caused by influenza-related causes, has no predictive power for economic variables before the 1918 pandemic. Since we do not have many observables between 1915 and 1920, we use wages in 1915 and still births in 1915 to show that there is no effect on future income or life expectancy. This confirms that long-run effects caused by the computed cross-sectional baseline weight act only through their impacts on normalized respiratory deaths during the pandemic years.

10 Conclusion

We use data from the 1872, 1920, 1940 Brazilian censuses, and 2006-2016 contemporaneous data for more than 300 geographical units in the State of Sao Paulo, Brazil, to study the impact of the 1918 Influenza Pandemic across different time dimensions. We match these data to district-level deaths by cause during the pandemic years (1917-1920) and combine this matched dataset with an array of auxiliary historical and geographical information, drawn from several archival documents. We show that the pandemic caused significant short-run demographic and other impacts that had lasting economic and social repercussions. These, in turn, altered mortality dynamics and employment structures. We then study dynamics of the impact on literacy rates and alternative indicators of educational attainment in the medium and long-terms. Since health is an important input to other forms of human capital, we use different categories of cause-of-death in 1927 and in-patient hospital admissions in 1940 to demonstrate lasting mortality effects from the 1918 pandemic. The sizeable effects on per capita and per establishment volumes of coffee, rice and maize output in 1920, and on aggregate measures of agricultural productivity in 1940, indicate that productivity is another channel through which the pandemic altered Sao Paulo's development path. We conclude by showing that municipalities with greater exposure to the pandemic still lag in health, education, employment and income, and that almost a century later, there is a long-lasting significant drag on development in these areas. Our results underline that historical determinants of development can have lasting effects on modern outcomes, and that a comprehensive understanding of the legacy of health shocks provides scope for targeted present-day policy-making in order to ameliorate some of the negative consequences.

References

Acemoglu, D. and S. Johnson (2007). Disease and development: The effect of life expectancy on economic growth. *Journal of Political Economy 115*(6), 925–985.

Acemoglu, D., S. Johnson, and J. Robinson (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review 91*(5), 1369–1401.

Almond, D. (2006). Is the 1918 influenza pandemic over? long-term effects of in utero influenza exposure in the post-1940 us population. *Journal of Political Economy 114*(4), 672–712.

Almond, D. and B. Mazumder (2005). The 1918 influenza pandemic and subsequent health outcomes: An analysis of sipp data. *American Economic Review 95*(2), 258–262.

Alonso, W., F. Nascimento, R. Acuna-Soto, C. Schuck-Paim, and M. Miller (2016). The 1918 influenza pandemic in florianopolis: A subtropical city in brazil. *Vaccine 29*, B16–B20.

Alsan, M. (2015). The effect of the tsetse fly on african development. *American Economic Review 105*(1), 382–410.

Alvarez, A., A. Carbonetti, A. Carrillo, C. B. Filho, C. Souza, L. Bertucci, and N. Azevedo (2009). The flu far and near: Comparing the 1918 and 2009 pandemics. *Historia, Ciencias, Saude-Manguinhos 16*(4), 1065–1113.

Bakken, M. and S. Husoy (2016). The long term impact of the 1918 influenza pandemic in norway. *Master's Thesis*.

Barata, R. (2000). Cem anos de endemias e epidemias. *Ciencia & Saude Coletiva 5*(2), 333–345.

Barker, D. (1995). Fetal origins of coronary heart disease. *BMJ 311*, 171–4.

Bassanezi, M. (2013). Uma tragica primavera. a epidemia de gripe de 1918 no estado de sao paulo, brasil. *ESTADO DE SAO P*, 73.

Beach, B. and W. Hanlon (2017). Coal smoke and mortality in an early industrial economy. *The Economic Journal 128*(615), 2652–2675.

Beach, B., J.Ferrie, and M.Saavedra (2018). Fetal shock or selection? the 1918 influenza pandemic and human capital development. *National Bureau of Economic Research.* (No. w24725).

Bertolli, F. (2003). A gripe espanhola em sao paulo, 1918: epidemia e sociedade, sao paulo, paz e terra. *Colecao Sao Paulo 5.*

Bertrand, M., E. Duflo, and S. Mullainathan (2003). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics 119*, 249–275.

Bertucci-Martins, L. (2005). Entre doutores e para os leigos: Fragmentos do discurso medico n influenza de 1918. *Historia, Ciencias, Saude-Manguinhos 12*(1).

Bleakley, H. (2010). Malaria eradication in the americas: A retrospective analysis of childhood exposure. *American Economic Journal: Applied Economics 2*(2), 1–45.

Chul Hong, S. and Y. Yun (2017). Fetal exposure to the 1918 influenza pandemic in colonial korea and human capital development. *Seoul Journal of Economics 30*(4), 353–383.

Clay, K., J. Lewis, and E. Severnini (2018). Pollution, infectious disease, and mortality: Evidence from the 1918 spanish influenza pandemic. *The Journal of Economic History 78*(4), 1179–1209.

Cook, C., J. Fletcher, and A. Forgues (2018). Multigenerational effects of early life health shocks. *National Bureau of Economic Research* (No. w25377).

Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature 41*(1), 87–122.

Currie, J. and M. Stabile (2006). Child mental health and human capital accumulation: The case of adhd. *Journal of Health Economics 25*(6), 1094–1118.

Dell, M. (2010). The persistent effects of peru's mining mita. *Econometrica 78*(6), 1863–1903.

Garrett, T. (2007). Economic effects of the 1918 influenza pandemic.

Glaeser, E., R. L. Porta, F. L. de Silanes, and A. Shleifer (2004). Do institutions cause growth? *Journal of Economic Growth 9*(3), 271–303.

Guiso, L., P. Sapienza, and L. Zingales (2016). Long-term persistence. *Journal of the European Economic Association 14*(6), 1401–1436.

Hamoudi, A. and J. Nobles (2014). Do daughters really cause divorce? stress, pregnancy, and family composition. *Demography 51*(4), 1423–1449.

Hanlon, W. (2016). Coal smoke and the costs of the industrial revolution. *NBER* (W22921).

Hatton, T. and J. Williamson (1998). The age of mass migration: Causes and economic impact. *Oxford University Press*.

Hochman, G. (2016). The sanitation of brazil: Nation, state, and public health, 1889-1930. *University of Illinois Press*.

Huillery, E. (2009). History matters: The long-term impact of colonial public investments in french west africa. *American Economic Journal: Applied Economics 2*(1), 176–215.

Karlsson, M., T. Nilsson, and S. Pichler (2014). The impact of the 1918 spanish flu epidemic on economic performance in sweden: An investigation into the consequences of an extraordinary mortality shock. *Journal of Health Economics 36*, 1–19.

Kobasa, D., A. Takada, K. Shinya, M. Hatta, P. Halfmann, S. Theriault, H. Suzuki, H. Nishimura, K. Mitamura, N. Sugaya, and T. Usui (2004). Enhanced virulence of influenza a viruses with the haemagglutinin of the 1918 pandemic virus. *Nature 431*(7009), 703.

Lin, M. and E. Liu (2014). Does in utero exposure to illness matter? the 1918 influenza epidemic in taiwan as a natural experiment. *Journal of Health Economics 37*, 152–163.

Loo, Y. and M. Gale (2007). Influenza: Fatal immunity and the 1918 virus. *Nature 445*(7125), 267.

Massad, E., M. Burattini, F. Coutinho, and L. Lopez (2007). The 1918 influenza a epidemic in the city of sao paulo, brazil. *Medical Hypothesis 68*(2), 442–445.

Musacchio, A., A. Fritscher, and M. Viarengo (2014). Colonial institutions, trade shocks, and the diffusion of elementary education in brazil, 1889-1930. *Journal of Economic History 74*(3), 730–766.

Nelson, R. (2010). Testing the fetal origins hypothesis in a developing country: Evidence from the 1918 influenza pandemic. *Health Economics 19*(10), 1181–1192.

Noymer, A. and M. Garenne (2000). The 1918 influenza epidemic's effects on sex differentials in mortality in the united states. *Population and Development Review 26*(3), 565–581.

Nunn, N. and L. Wantchekon (2011). The slave trade and the origins of mistrust in africa. *American Economic Review 101*(7), 3221–3252.

Parman, J. (2013). Childhood health and sibling outcomes: The shared burden and benefit of the 1918 influenza pandemic. *NBER Working Paper No. w19505*.

Percoco, M. (2016). Health shocks and human capital accumulation: The case of the spanish flu in italian regions. *Regional Studies 50*(9), 1496–1508.

Polozov, I., L. Bezrukov, K. Gawrisch, and J. Zimmerberg (2008). Progressive ordering with decreasing temperature of the phospholipids of influenza virus. *Nature Chemical Biology 4*(4), 248.

Putterman, L. and D. Weil (2010). Post-1500 population flows and the long-run determinants of economic growth and inequality. *The Quarterly Journal of Economics 125*(4), 1627–1682.

Richter, A. and P. Robling (2013). Multigenerational effects of the 1918-19 influenza pandemic in sweden. *Swedish Institute for Social Research 5*.

Rocha, R., C. Ferraz, and R. Soares (2017). Human capital persistence and development. *American Economic Journal: Applied Economics 81*(1), 51–57.

Slutsky, D. and R. Zeckhauser (2018). Sunlight and protection against influenza. *National Bureau of Economic Research No. w24340*.

Sokoloff, K. and S. Engerman (2000). Institutions, factor endowments, and paths of development in the new world. *Journal of Economic Perspectives 14*(3), 217–232.

Solomou, S. and W. Wu (1999). Weather effects on european agricultural output, 1850-1913. *European Review of Economic History 3*(3), 351–373.

Tamerius, J., J. Shaman, W. Alonso., K. Bloom-Feshbach, C. Uejio, A. Comrie, and C. Viboud (2013). Environmental predictors of seasonal influenza epidemics across temperate and tropical climates. *PLoS Pathogens 9*(3), e1003194.

Figure 1: A Newspaper Extract in October 1918

de Outubro de 1918

GAZETA DE NOTICIAS

A GRANDE EPIDEMIA

As providencias do governo de muito pouco valeram até agora

Continuamos entregues á Divina Providencia

: E os casos fataes augmentam :

A gravidade da situação e as providencias do governo

Os mortos de bontem

A Caixa Economica

Providencias do governo

Uma nota da Directoria de Saude Publica

O serviço sanitario na Guanabara

O trabalho no palacio do Cattete

A Gazeta de Notícias critica o governo federal pelo imobilismo diante da disseminação da epidemia (imagem: Biblioteca Nacional)

Notes: This extract conveys that the government was criticized for its lack of involvement during the pandemic in 1918. Translated headline: the government's input of very little value until now. Source: Biblioteca Nacional.

Figure 2: Headline of the Correio da Manha in October 1918

SÃO INNUMEROS OS PEDIDOS DE SOCCORROS AO PALACIO DO CATTETE

O governo faz um appello aos medicos, pharmaceuticos e estudantes, para auxiliarem a Saude Publica nesta difficil emergencia

O governo determinou hontem novas providencias

O caso dos coveiros

O que fez officialmente o Commissariado

O presidente da Republica fala com o sr. Bulhões

A deserção na Saude Publica

Em Nictheroy

Na guarda-moria da Alfandega

A Telephonica soffre o primeiro golpe

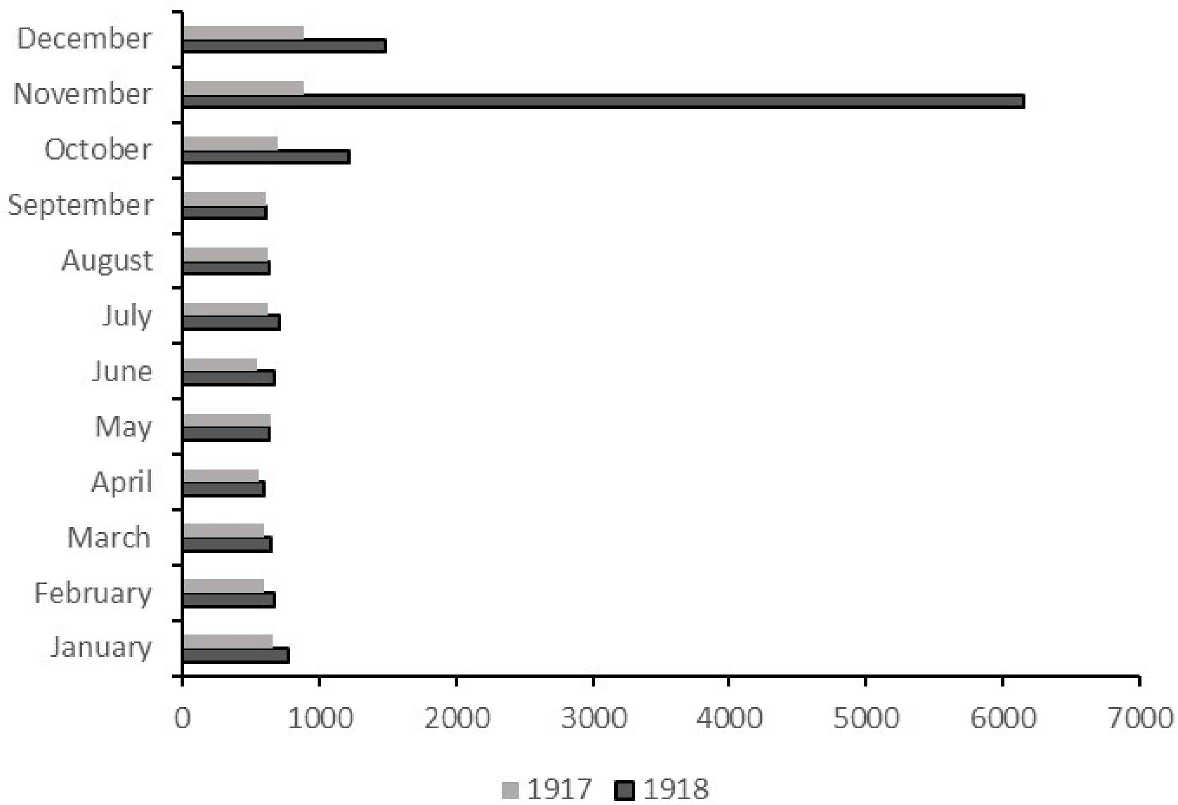
Em Petropolis

VARIAS

Senadores enfermos

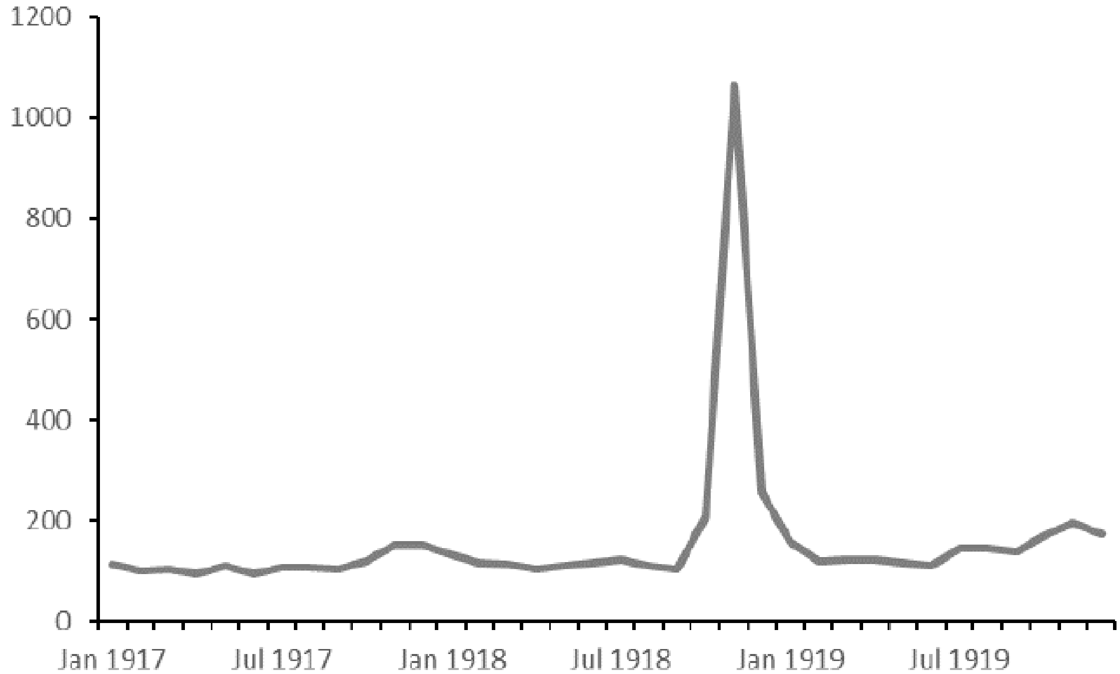
Notes: This extract conveys that the government needed help from medical experts and students at the peak of the Spanish Flu outbreak in October 1918. Translated headline: The government appeals to doctors, chemists and students to help public health during this period of emergency. Source: Collections of the Morning Mail in the Brazilian Digital Newspaper Archive.

Figure 3: Monthly Deaths for 1917 and 1918 for Sao Paulo (City)



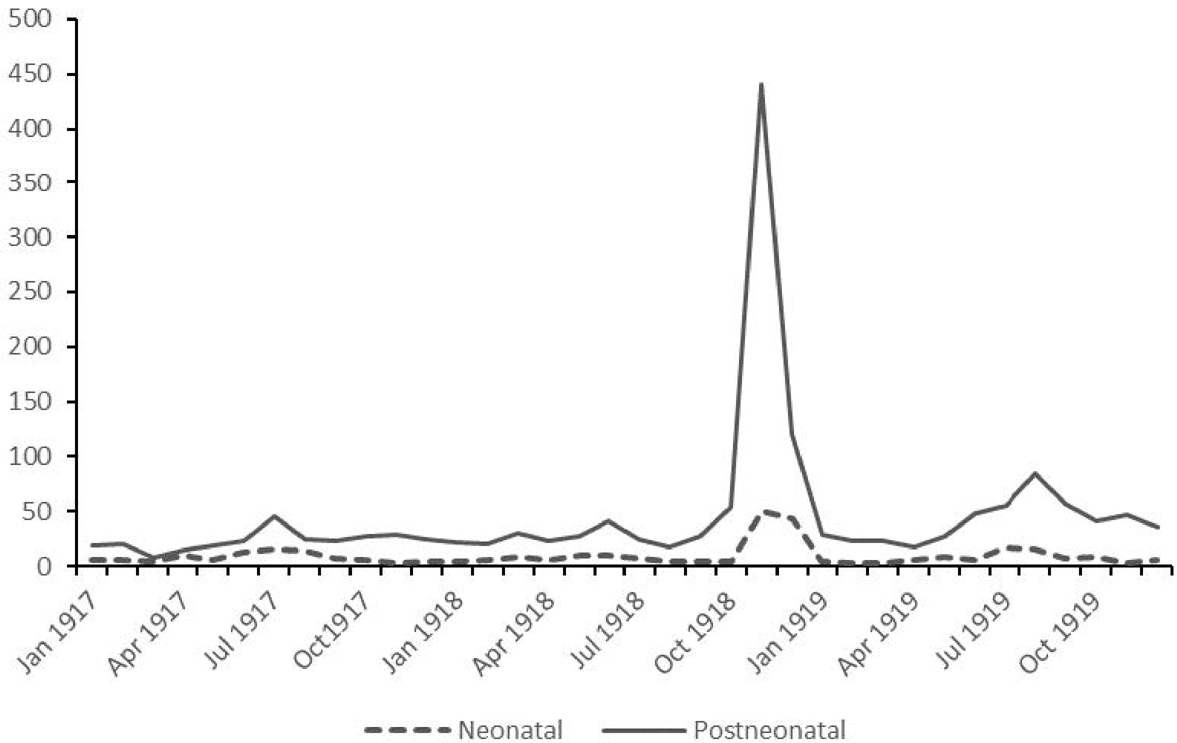
Source: Annual Statistical Reports Vol. 1 (1917, 1918, 1919) for the Sao Paulo (City)

Figure 4: Monthly Total Deaths per 100,000 of the Population (January 1917-December 1919)



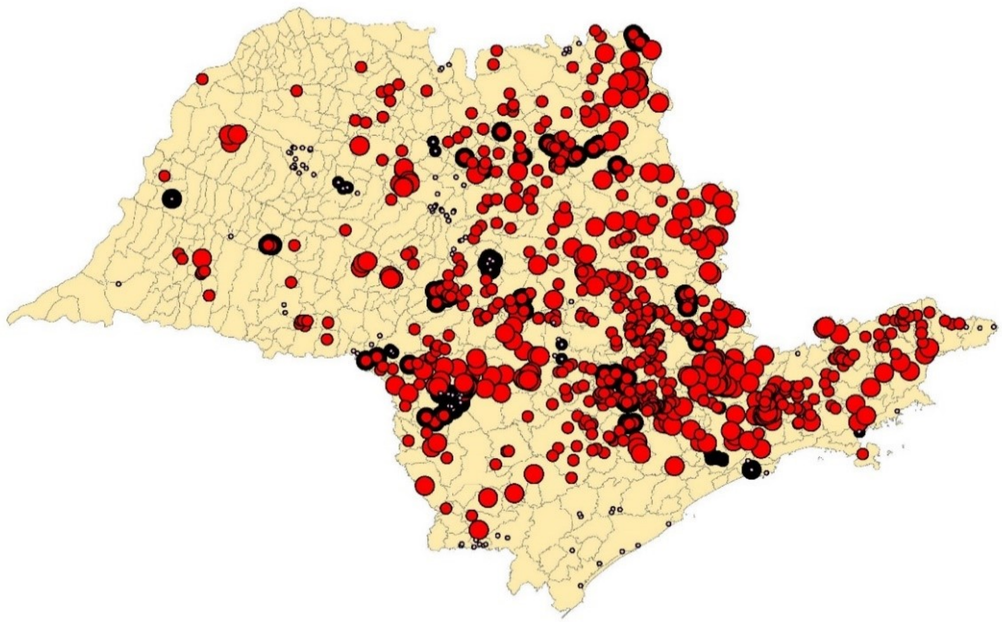
Source: Authors' calculations from the Annual Statistical Reports Vol. 1 (1917, 1918, 1919) for the Municipality Capital (Sao Paulo City)

Figure 5: Monthly Neo-Natal and Post-NeoNatal Influenza-Related Mortality Rate (Deaths per 1000 Live Births for January 1917-December 1919)



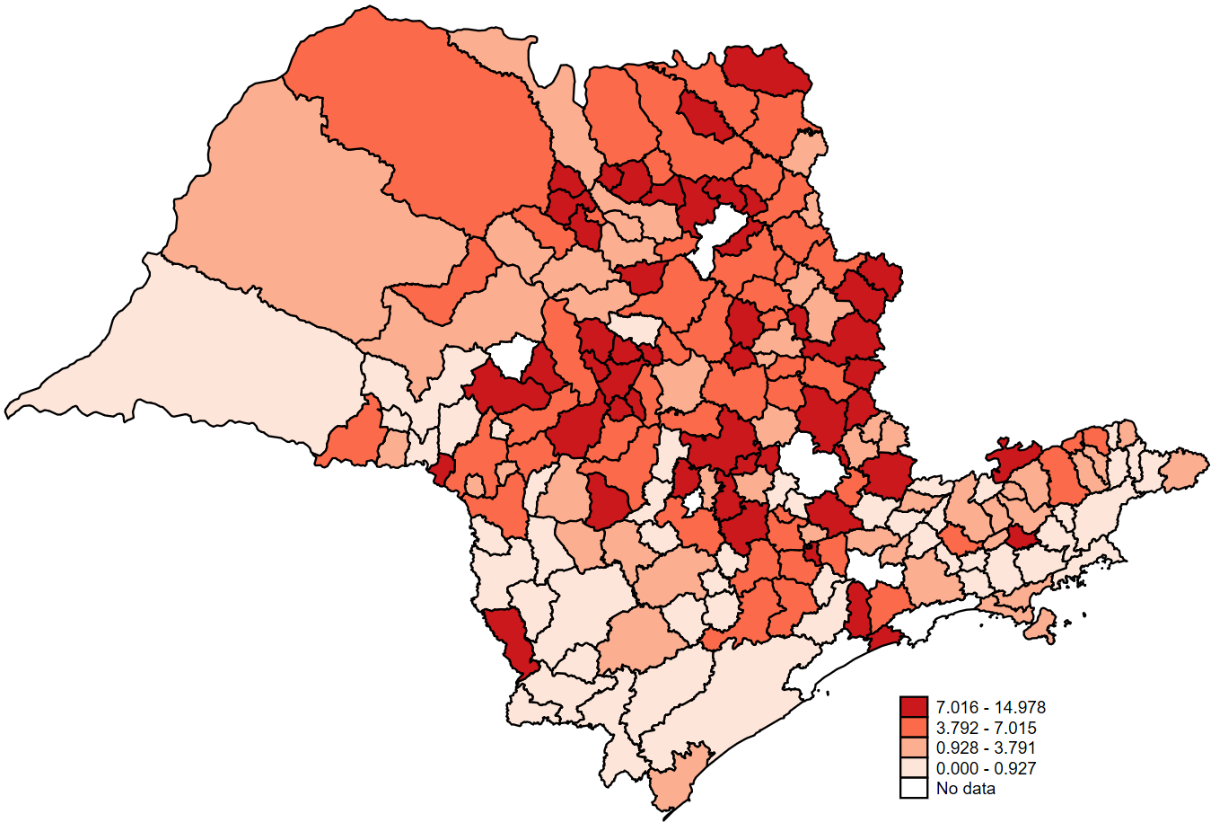
Source: Authors' calculations from the Annual Statistical Reports Vol. 1 (1917, 1918, 1919) for the Municipality Capital (Sao Paulo City)

Figure 6: Prevalence of Influenza-Related Deaths per 1000 of the Population in Sao Paulo (State), 1918



Source: Authors' calculations from the Annual Statistical Report Vol. 1 (1918). The data is at the district level in the report and we use the 2016 boundaries to show the prevalence of normalized deaths in 1918.

Figure 7: The Spatial Variation in Influenza-Related Deaths per 1000 of the Population in Sao Paulo (State), 1917-1920



Source: Authors' calculations from the Annual Statistical Reports Vol.1 (1917, 1918, 1919, 1920). The data is aggregated up to the municipality level for each year during this period. The 1920 boundaries are used to consider the variation.

Figure 8: The Territorial Fragmentation of Sao Paulo, 1872-2016

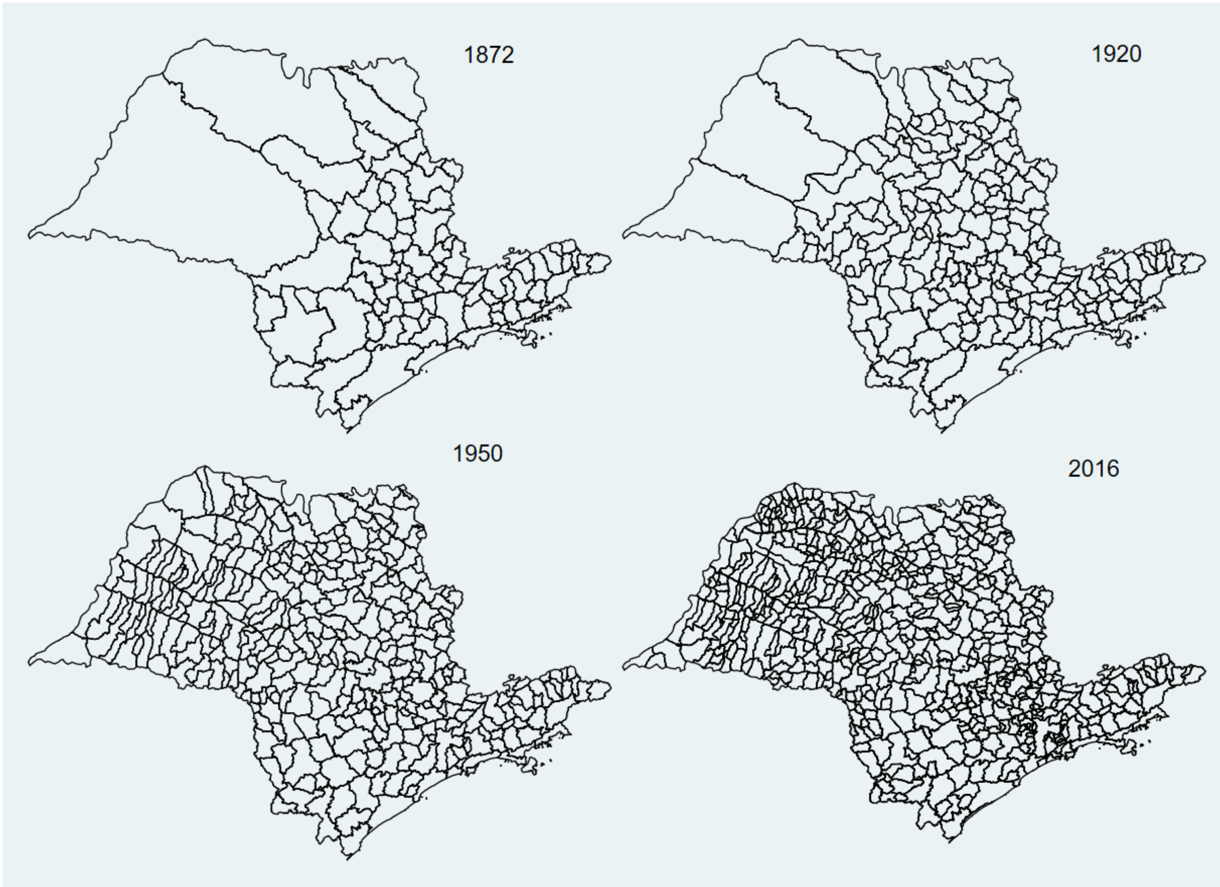


Figure 9: The Spatial Variation in respiratory and unknown Deaths per 1000 of the Population in Sao Paulo (State), 1918

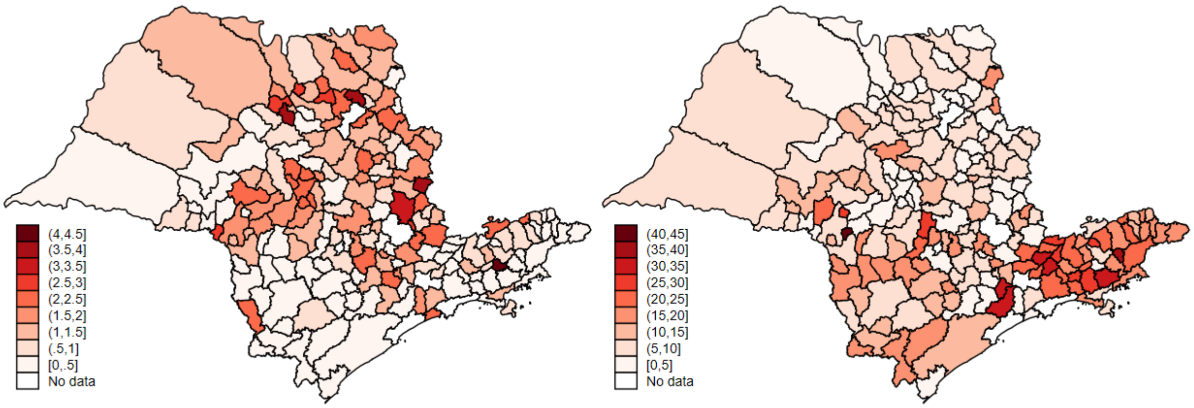
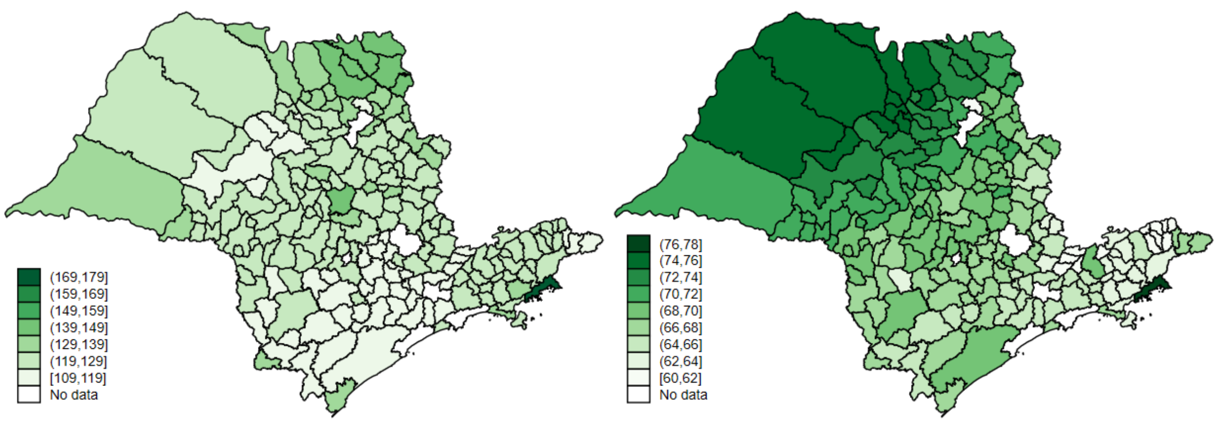


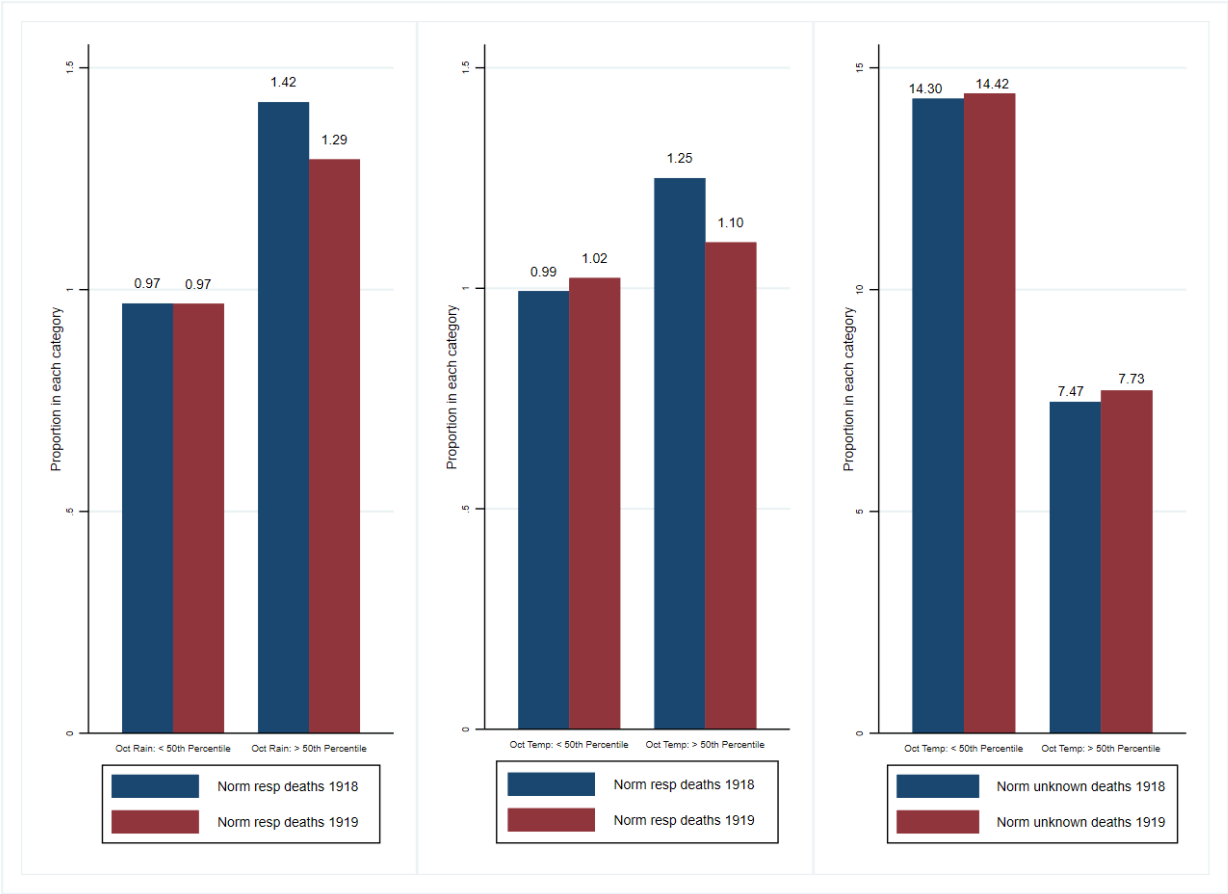
Figure 10: Average October's Rain and Temperature, 1901-1930



(a) Average Precipitation (mm) in October

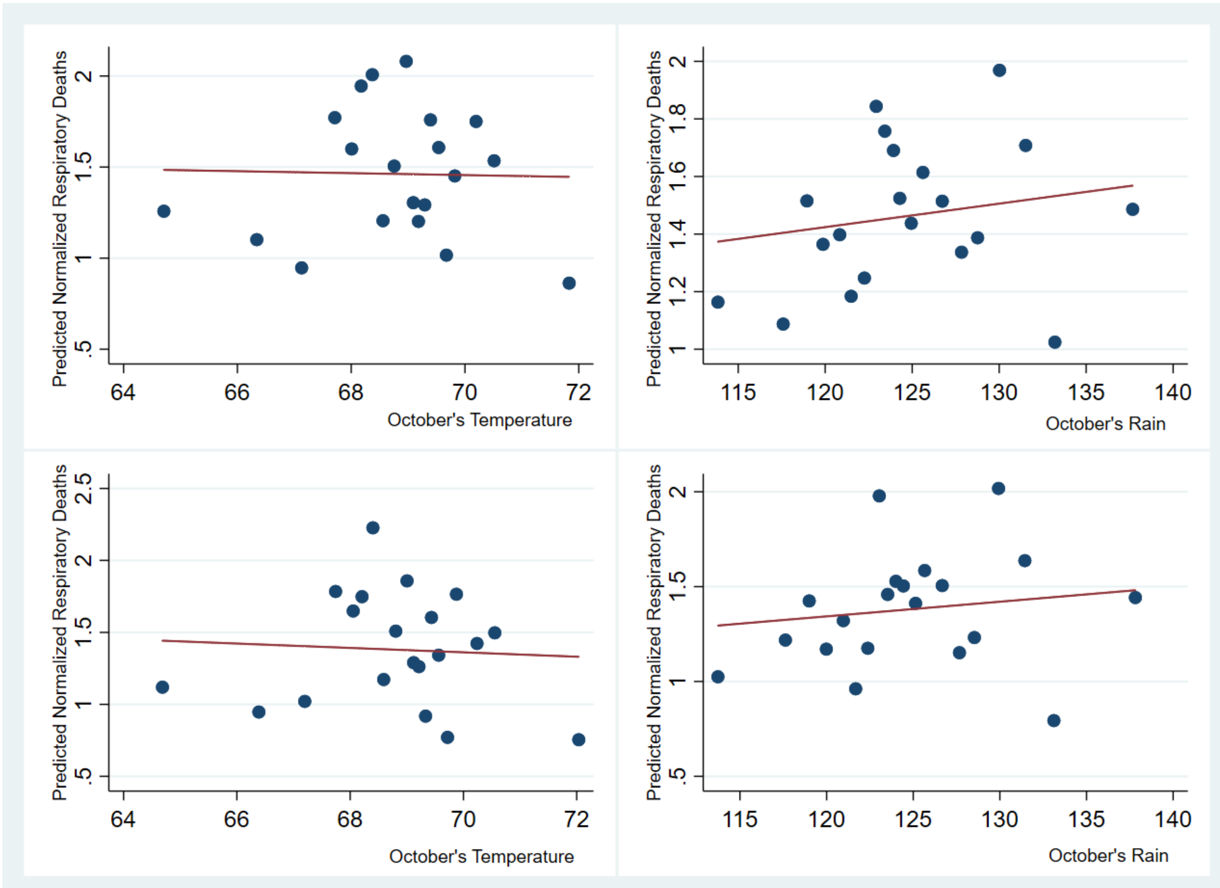
(b) Average Temperature (Fahrenheit) in October

Figure 11: The Distribution of Normalized Respiratory Deaths for 1918/1919 with respect to October's Rain and Temperature



Notes: Panel A (left), Panel B (middle), and Panel C (right). Authors' calculations.

Figure 12: Binscatter Plots for Normalized Respiratory Deaths and October's Rain and Temperature



Notes: Panel A (top: 1917-1920) and Panel B (Bottom: 1918/1919 only). Regressions condition on geographical controls that include altitude, longitude and latitude.

Table 1: Summary Statistics for Dependent Variables

	Mean (1)	Standard Dev (2)
Panel A: Variables for 1920		
Infant mortality (per live births)	0.02	0.03
Still births (per total births)	0.05	0.03
Literacy rate (total)	0.23	0.08
Literacy rate (male)	0.28	0.09
Literacy rate (female)	0.16	0.08
Sex ratio at birth (Male babies/Female babies)	1.21	0.82
Coffee production per capita (tons)	0.31	0.45
Coffee production per establishment (tons)	14.88	14.16
Rice production per capita (tons)	0.29	0.59
Rice production per establishment (tons)	5.58	8.06
Maize production per capita (tons)	1.19	2.13
Maize production per establishment (tons)	19.67	12.68
Panel B: Variables for 1940		
Share receiving instruction and literate:		
Male (age 7-14)	0.82	0.089
Female (age 7-14)	0.84	0.10
Share receiving instruction and literate:		
Male (age 5-39)	0.82	0.09
Female (age 5-39)	0.85	0.35
Literacy rate:		
Male (Age 5 and above)	0.49	0.12
Female (Age 5 and above)	0.35	0.12
Inpatient Hospital Admissions (per 1000)	3.79	15.19
Inpatient Asylum Admissions (per 1000)	6.18	14.48
Value of primary sector per emp. (cr$1000)	1009.03	490.85
Value of primary sector per est. (cr$1000)	9470.19	7815.85
Panel C: Contemporaneous Variables		
IFDM Brazilian Development Index:		
2006	0.75	0.07
2011	0.78	0.07
2016	0.77	0.06
IFDM Brazilian SubIndices 2016:		
Employment and Income	0.55	0.11
Health	0.83	0.08
Education	0.92	0.04

Notes: See Appendix 1 for data description and sources.Authors' calculations.

Table 2: Summary Statistics for Explanatory and Control Variables

	Mean (1)	Standard Dev (2)
Panel A: Deaths by Cause (Per 1000)		
Respiratory	1.74	2.03
Central Nervous System	0.74	0.96
Skin-related	0.14	0.17
Puerperal Sepsis	0.28	0.26
Circulatory system	1.37	5.68
Digestive	3.57	4.37
Panel B: Geography		
Altitude (in meters)	615.56	192.02
Latitude	-22.44	1.09
Longitude	-47.75	1.31
Distance to Capital (in Km)	236.34	113.60
October's 1918 Temperature (Fahrenheit)	68.43	3.76
October's 1918 Precipitation (mm)	124.15	8.39
Panel C: Variables from 1872 Census		
Literacy rate	0.18	0.09
Share of foreigners	0.03	0.034
Share of slaves	0.18	0.10
Share of race Branca	0.53	0.11
Share of race Parda	0.24	0.06
Share of race Preta	0.19	0.08
Share emp. in agriculture	0.64	0.11
Share emp. in manufacturing	0.13	0.05
Share with physical/mental diseases	11.56	6.02
Doctors (per 1000)	0.26	0.36
Chemists (per 1000)	0.18	0.22
Midwives (per 1000)	0.35	0.50
Population density (per sq.Km)	12.55	8.80
Panel D: Other Controls		
Health&sanitation exp. per capita (1910)	199.73	422.09
Dummy for frost intensity (1919)	0.71	0.45
Railway station dummy (1920)	0.70	0.46
Water&sewerage system dummy (1920)	0.20	0.40
No. hospitals&nursing homes (1920)	1.67	1.03
Taxes out of revenue (%,1920)	0.34	0.16
Sex ratio of adults (1920)	1.12	0.59
Share of foreigners (1920)	0.13	0.10
Share of exp. on health&education (1940)	0.61	0.12
Primary sector exp per est. (cr$1000,1940)	7.91	13.36
Agriculture exp per est.(cr$1000,1940)	5.82	7.06
Farming exp per est.(cr$1000,1940)	7.75	8.19
Livestock exp per est. (cr$1000,1940)	11.75	12.23
Number of plows per est.(1940)	1.54	0.63
Share of est. exp on salaries(1940)	0.70	0.41
Share of est. exp on new machinery(1940)	0.04	0.03
Share of est. exp on new animals(1940)	0.08	0.10
Share of est.exp on fertilizers(1940)	0.10	0.09

Notes: See Appendix 1 for data description and sources.Authors' calculations.

Table 3: First Stage Regressions: Dependent variable is normalized respiratory deaths (per 1000 of the population)

	For 1917-1920			For 1913-1915			For 1914 only
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
October's Mean Temperature	-0.245***	-0.246***	-0.252***	-0.320	-0.290	0.0337	0.087**
	(0.056)	(0.056)	(0.056)	(0.361)	(0.464)	(0.338)	(0.030)
October's Mean Precipitation	0.032**	0.032**	0.029**	0.167	-0.001	-0.077	0.019
	(0.013)	(0.013)	(0.013)	(0.128)	(0.109)	(0.0394)	(0.0161)
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	No
Initial Controls of 1872	No	Yes	Yes	No	Yes	Yes	Yes
Geographical Control	No	No	Yes	No	No	Yes	Yes
F-Statistic	10.24	10.24	10.53	1.08	2.69	3.13	4.65
p-value for the F-Stat	[0.000]	[0.000]	[0.000]	[0.342]	[0.070]	[0.045]	[0.012]
Observations	939	937	933	414	413	410	127
R-squared	0.442	0.442	0.443	0.736	0.735	0.735	0.147

Notes: The table reports OLS regressions. In Column (1), we consider both instruments with only municipality fixed effects. In Columns (2) and (3), the controls are added sequentially. These include altitude, population density, share of employment in agriculture, manufacturing and in services/retail, the share of people who were slaves, the share of white people, the share of foreigners, of literate people, the normalized number of people with a mental and physical disease, the number of doctors, chemists and midwives (all refer to the numbers in the given occupation per 1000 of the population). P-values for the F-statistics are reported for the identifying instruments and are in square brackets. $***p < 0.01$, $**p < 0.05$ and $*p < 0.10$.

Table 4: First Stage Regressions on Constructed Instrument : Dependent variable is normalized respiratory deaths (per 1000 of the population)

	(1)	(2)	(3)	(4)
IVResp	0.237***	0.230***	0.223***	0.220***
	(0.043)	(0.051)	(0.049)	(0.069)
Region FE	No	Yes	Yes	Yes
Year FE	No	Yes	Yes	Yes
District-level controls	No	Yes	Yes	Yes
Initial Conditions of 1872	No	No	Yes	Yes
Initial Conditions x Year FE	No	No	Yes	Yes
1910 Heath and Sanitation Exp.	No	No	No	Yes
F-Statistic	30.22	20.81	21.01	10.20
p-value for the F-Statistic	[0.000]	[0.000]	[0.000]	[0.000]]
Observations	853	544	544	353
R-squared	0.053	0.190	0.245	0.367

Notes: The table reports OLS regressions. Standard errors reported in parentheses are clustered at the district level. We use the variation for the period 1917-1920 only. In column (2), we add the district level controls: altitude, latitude, longitude, the distance to the capital, a dummy for railway station, the interaction between distance and the railway dummy and the year effects. In column (3), the initial conditions of 1872 include the number of doctors, chemists and midwives (all refer to the numbers in the given occupation per 1000 of the population); the normalized number of people with a mental and physical disease, the literacy rate and the share of foreigners, and we add the interaction between the initial conditions and the year FE. Finally, in column (4), we add the 1910 health and sanitation expenditure, also interacted with the years. P-values for the F-statistics are reported for the identifying instruments and are in square brackets. $***p < 0.01$, $**p < 0.05$ and $*p < 0.10$

Table 5: The Short-run Effects on Infant Mortality and Still Births in 1920

	(1)	(2)	(3)	(4)
Part A: Dependent Variable: Infant Mortality				
OLS				
Norm. Respiratory Deaths	0.009***	0.008***	0.001***	0.009***
	(0.000)	(0.000)	(0.001)	(0.001)
2SLS				
Norm. Respiratory Deaths	0.017***	0.018***	0.020***	0.019***
	(0.003)	(0.005)	(0.006)	(0.007)
Part B: Dependent Variable: Log Still Births				
OLS				
Norm. Respiratory Deaths	-0.001	0.028	0.063*	0.061**
	(0.023)	(0.019)	(0.029)	(0.029)
2SLS				
Norm. Respiratory Deaths	0.038	0.199	0.193*	0.214**
	(0.123)	(0.130)	(0.102)	(0.084)
Region FE	No	Yes	Yes	Yes
Region-level Controls	No	No	Yes	Yes
Initial Controls of 1872	No	No	No	Yes
Mean for Infant Mortality	23.31	23.31	23.31	23.31
Mean for Still Births	52.54	52.54	52.54	52.54
Observations for Infant Mortality	443	443	305	305
Observations for Still Births	603	603	399	399

Notes: The table shows the OLS results (Panel A) and 2SLS results (in Panel B with IVResp as instrument). Standard errors reported in parentheses are clustered at the mesoregion level. We use the variation for the period 1917-1919 to evaluate the effects on 1920 infant mortality and still biths. In Column (1), we consider the regression without fixed-effects, controls and the initial controls of 1872. In Column (2), we add the region fixed-effects. In Column (3), the controls include a dummy for the presence of water and sewerage systems and one for the presence of hospitals, nursing homes and specialized maternity hospitals; the literacy rates for male and female aged 15 and above, region-level controls including altitude, latitude, longitude, distance to the capital city, a dummy for the presence of railway and the interaction between distance and the railway dummy. In Column (4), we add the initial controls of 1872 and they include the number of doctors, chemists and midwives (all refer to the numbers in the given occupation per 1000 of the population), the normalized number of people with a mental and physical disease, share of people who were slaves, the share of employment in agriculture, manufacturing, and in services/retail, population density. . $^{***}p < 0.01$, $^{**}p < 0.05$ and $^{*}p < 0.10$.

Table 6: The Short-run Effects on Sex Ratios at Birth in 1920

	Dependent Variable: Sex Ratio At Birth		
	(1)	(2)	(3)
Panel A: OLS			
Norm. Respiratory Deaths	-0.054**	-0.048*	-0.044
	(0.023)	(0.023)	(0.026)
Panel B: 2SLS			
Norm. Respiratory Deaths	-0.479**	-0.470***	-0.497***
	(0.221)	(0.143)	(0.173)
Initial Controls of 1872	No	Yes	Yes
Region-level controls	No	No	Yes
Geographical controls	No	No	Yes
Mean of Dependent Variable	1.22	1.22	1.22
Observations	580	580	576

Notes: The table shows the OLS results (Panel A) and 2SLS results (in Panel B with October's rain and precipitation used as instruments). Standard errors reported in parentheses are clustered at the mesoregion level. We use the variation for the period 1917-1920 to evaluate the effects on 1920 sex ratios. In columns (1) we consider the regressions without controls and fixed-effects. In columns (2) and (3), we also include region-level controls (altitude, the sex ratio in 1920 for adults, a dummy for the presence of water and sewerage systems and one for the presence of hospitals, nursing homes and specialized maternity hospitals), and the initial controls include the normalized number of midwives, the shares of people of race Branca, Parda, Preta and Cabocla in 1872, the literacy rate. ***$p < 0.01$, **$p < 0.05$ and *$p < 0.10$.

Table 7: Literacy Rates by Gender and Age Groups in 1920

	All Ages			*Ages 15 and Above*		
	Male	Female	Total	Male	Female	Total
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS						
Norm. Respiratory Deaths	0.010***	0.011***	0.011***	0.012***	0.013***	0.013***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
Panel B: 2SLS						
Norm. Respiratory Deaths	-0.030	-0.046	-0.039	0.017	-0.031	-0.007
	(0.052)	(0.066)	(0.058)	(0.053)	(0.069)	(0.054)
Geographical Control	Yes	Yes	Yes	Yes	Yes	Yes
Initial Control	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dependent Variable	0.281	0.164	0.226	0.418	0.221	0.326
Observations	669	669	669	669	669	669

Notes: The table reports the OLS results in Panel A and the 2SLS results in Panel B. The instruments used are October's rain and October's precipitation. Standard errors reported in parentheses are clustered at the mesoregion level. We use observations for the period 1917-1919. The initial controls of 1872 include altitude, the normalized number of people with a mental and physical disease, the literacy rate, the share of foreigners, the share of workers in agriculture, manufacturing and services, the share of people who were slaves, the share of people of race Branca (white), population density. ***$p < 0.01$, **$p < 0.05$ and *$p < 0.10$.

Table 8: Literacy Rates and Instruction, by Age Group and Gender in 1940

	Literacy Rate: Age 5 +		**Literacy Rate: Age 18 +**		**Instructed and Literate: Age 5-39**	
	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS						
Norm. Respiratory Deaths	0.012**	0.015**	0.013**	0.001*	0.000	0.001
	(0.006)	(0.005)	(0.005)	(0.005)	(0.003)	(0.003)
Panel B: 2LS						
Norm. Respiratory Deaths	0.081***	0.030	0.067***	0.009	-0.012	-0.073**
	(0.022)	(0.022)	(0.024)	(0.027)	(0.028)	(0.031)
Geographical control	Yes	Yes	Yes	Yes	Yes	Yes
Initial Controls	Yes	Yes	Yes	Yes	Yes	Yes
Region-level Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dependent Variable	0.489	0.348	0.521	0.325	0.823	0.830
Observations	674	674	674	674	665	655

Notes: The table reports the OLS results in Panel A and the 2SLS results in Panel B. The instruments used are October's rain and October's precipitation. Standard errors reported in parentheses are clustered at the mesoregion level. The initial controls of 1872 include altitude, the normalized number of people with a mental and physical disease, the literacy rate, the share of foreigners, the share of workers in agriculture, manufacturing and services, the share of people who were slaves and of race Branca (white), population density. We also control for the share of total municipality expenditure allocated to education and health in 1940; the number of teachers per person aged 10 and plus (teachers/Pop. Aged 10 above)*1000. We use observations for the years 1917-1920 only. ***$p < 0.01$, **$p < 0.05$ and *$p < 0.10$.

Table 9: Results for 1927 Deaths by Cause

	Dependent Variable: Deaths by Cause per 1000 of the population					
	Epidemic or Endemic	Central Nervous	Circulatory	Respiratory	Puerperal Sepsis	Skin-Related
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS						
Norm. Respiratory Deaths	1.018***	0.104***	0.178***	0.486***	0.035**	0.007
	(0.225)	(0.023)	(0.024)	(0.091)	(0.014)	(0.008)
Panel B: 2LS						
Norm. Respiratory Deaths	2.727***	0.344***	0.795***	0.936**	0.123***	0.101***
	(0.764)	(0.117)	(0.259)	(0.398)	(0.047)	(0.020)
Geographical control	Yes	Yes	Yes	Yes	Yes	Yes
Initial controls	Yes	Yes	Yes	Yes	Yes	Yes
Region-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dependent Variable	3.507	0.631	1.011	1.459	0.226	0.134
Observations	806	752	777	780	647	356

Notes: The table reports the OLS results in Panel A and the 2SLS results in Panel B. The instruments used are October's rain and October's precipitation. Standard errors reported in parentheses are clustered at the mesoregion level. The initial controls of 1872 include altitude, the normalized number of people with a mental and physical disease, the literacy rate, the share of foreigners, the share of workers in agriculture, manufacturing and services, the share of people who were slaves, the share of people of race Branca (white), population density. We use observations for the years 1917-1920 to evaluate the impact on 1927 deaths by cause. ***$p < 0.01$, **$p < 0.05$ and *$p < 0.10$.

Table 10: Instrumental Variable Results for In-Patient Hospital Admissions (per 1000) in 1940

	Hospital Admissions		Asylum Admissions	
	(1)	(2)	(3)	(4)
Norm. Respiratory Deaths	1.451*	1.816**	-0.825	-1.777
	(0.770)	(0.890)	(0.948)	(1.586)
Region-level controls	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes
Initial controls	No	Yes	No	Yes
Mean of Dependent Variable	3.909	3.903	5.967	5.967
Observations	356	356	281	281

Notes: We use the variation for the period 1917-1920 to evaluate the effects on 1940 Inpatient Hospital Admissions. The controls include altitude, latitude, longitude, distance to the capital city, a dummy for the presence of railway. The initial controls of 1872 include population density, the shares of employment in agriculture, manufacturing, retail/services, the share of people of race Branca, the share of people who were slaves, the share of foreigners, the share of people who were slaves. ***$p < 0.01$, **$p < 0.05$ and *$p < 0.10$.

Table 11: Instrumental Variable Results for Agricultural Productivity in 1920

	Coffee		Rice		Maize	
	Per Capita	Per Est.	Per Capita	Per Est.	Per Capita	Per Est
Norm. Respiratory Deaths	-0.067**	-1.627	-0.137***	-1.804*	-0.295***	-0.962
	(0.031)	(5.133)	(0.043)	(0.931)	(0.101)	(3.272)
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Region-level Controls	Yes	Yes	Yes	Yes	Yes	Yes
Initial Controls of 1872	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dependent Variable	0.319	15.007	0.289	5.572	1.200	19.724
Observations	383	383	401	401	404	404

Notes: The table shows the OLS results (Panel A) and 2SLS results (in Panel B with IVResp as instrument). Standard errors reported in parentheses are clustered at the regional level. We use the variation for the period 1917-1919 to evaluate the effects on 1920 productivity measures. Region-level controls include the geographical characteristicts (altitude, latitude, longitude), distance to capital city, the literacy rates for male and female in 1920, a measure for frost intensity (a dummy that takes a value of one if more than three frost periods were reported). The initial controls of 1872 include the share of foreigners, share of people who were slaves, the share of employment in agriculture, manufacturing, and in services/retail, population density, the share of people of race Branca. ***$p < 0.01$, **$p < 0.05$ and *$p < 0.10$.

Table 12: Instrumental Variable Results for Aggregate Measures for the Primary Sector in 1940

	Value per Employee		Value per Establishment	
	(1)	(2)	(3)	(4)
Norm. Respiratory Deaths	-0.502*	-0.448***	-0.241	-0.472***
	(0.286)	(0.172)	(0.311)	(0.170)
Region FE	Yes	Yes	Yes	Yes
Region-Level Controls	No	Yes	No	Yes
Initial Controls of 1872s	No	Yes	No	Yes
Mean of Dependent Variable (in log)	6.781	6.781	8.839	8.839
Observations	736	468	736	468

Notes: The table shows the 2SLS results (with IVResp as instrument). Standard errors reported are clustered at the microregion level. We use the variation for the period 1917-1920 to evaluate the effects on 1940 agricultural productivity. In Columns (1) and (3), we consider the regressions the value of yield per employee and per establishment with only region FE. In Columns (2) and (4), we add the controls including altitude, latitude, longitude, distance to the capital city. We also control for the 1940 literacy rates for male and female of age 5 and above, and for the 1920 share of foreigners. Further, we include several other 1940 controls that could affect productivity. These controls include the expenditure per establishment by agricultural, farming and livestock establishments; the number of plows per establishment, the amount of expenditure on salaries, acquisition of machinery and of new animals, and on fertilizers per establishment. The initial controls of 1872 include population density, the shares of employment in agriculture, manufacturing, retail/services, the share of people of race Branca, the share of people who were slaves. ***$p < 0.01$, **$p < 0.05$ and *$p < 0.10$.

Table 13: Instrumental Variable Results for the Development Index for 2006-2016

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A:						
Dependent Variable: Brazil's IFDM Sub-Indices for 2016						
	Employment and Income		*Education*		*Health*	
Norm. Respiratory Deaths	-0.054**	-0.050**	-0.030**	-0.029***	-0.047**	-0.049**
	(0.031)	(0.024)	(0.012)	(0.010)	(0.024)	(0.022)
Panel B:						
Dependent Variable: IFDM Development Index						
	2006		*2011*		*2016*	
Norm. Respiratory Deaths	-0.040*	-0.046**	-0.049***	-0.041**	-0.044***	-0.042***
	(0.021)	(0.022)	(0.017)	(0.020)	(0.015)	(0.013)
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Initial Controls of 1872	Yes	Yes	Yes	Yes	Yes	Yes
Region-level controls	No	Yes	No	Yes	No	Yes
Mean of Dependent Variable (Sub-Indices)	0.549	0.549	0.920	0.920	0.832	0.832
Mean of Dependent Variable (Main Index)	0.749	0749	0.779	0.779	0.767	0.767
Observations	760	524	760	524	760	524

Notes: The table shows the 2SLS for the IFDM Sub-indices for 2016 and for the Brazilian Development Index. Standard errors reported in parentheses are clustered at the mesoregion level. We use the variation for the period 1917-1920. In columns (1), (3) and (5), we consider the regressions with only region FE and initial controls. In columns (2), (4) and (6), we add the region-level controls (altitude, latitude, longitude, distance to the capital city). The initial controls of 1872 include the number of doctors, chemists and midwives (all refer to the numbers in the given occupation per 1000 of the population), the normalized number of people with a mental and physical disease, share of people who were slaves, the share of foreigners, the share of people of the race "Branca", the literacy rate, the share of employment in agriculture, manufacturing, and in services/retail, population density. ***$p < 0.01$, **$p < 0.05$ and *$p < 0.10$.

Table 14: Test of Sample Mean Differences in Literacy Rates for Districts in 1918

	Normalized Respiratory Deaths					
	Below Median		*Above Median*			
	Mean	S.Dev	Mean	S.Dev	Diff.	p-value
Literacy Rates:						
<u>Brazilian</u>						
Male	0.257	0.059	0.249	0.007	0.007	0.480
Female	0.174	0.060	0.160	0.007	0.014	0.174
Total	0.217	0.059	0.206	0.073	0.011	0.290
<u>Foreigners</u>						
Male	0.562	0.153	0.524	0.149	0.038	0.083
Female	0.305	0.143	0.271	0.179	0.035	0.155
Total	0.461	0.152	0.425	0.155	0.035	0.120

Table 15: Falsification Test Using 1915 Deaths from Respiratory Infections

	1920	**1940**	
	Sex Ratios	Literacy Rate (5+)	
	(At Birth)	Male	Female
	(1)	(2)	(3))
Norm. Respiratory Deaths	-0.179	0.036	0.020
	(0.440)	(0.035)	(0.028)
Region FE	Yes	Yes	Yes
Region-level controls	Yes	Yes	Yes
Initial controls 1872	Yes	Yes	Yes
Observations	124	119	119

Notes: We use the same controls, fixed-effects and instruments as for the original regressions showing the impact of respiratory deaths during the pandemic years on the different outcomes of interest. ***$p < 0.01$, **$p < 0.05$ and *$p < 0.10$.

Table 16: Robustness Check for the Baseline Weight of the Constructed Instrument

	1915	
	Still Births	Log Wages
	(Per 1000 Live Births)	(Agricultural)
	(1)	(2)
Baseline Weight ($\bar{s}_{d1915}$)	0.000	0.016
	(0.000)	(0.014)
Region FE	Yes	Yes
Region-level controls	Yes	Yes
Initial controls	Yes	Yes
Observations	157	85

Notes: We use the same geographical controls and region-level initial controls as for the original regressions showing the impact of respiratory deaths during the pandemic years on the different outcomes of interest. ***$p < 0.01$, **$p < 0.05$ and *$p < 0.10$.

A1 Appendix

In this section, we aim at estimating the effects of the 1918 Influenza Pandemic on specific cohorts using the IPUMS individual-level data for Sao Paulo for the years 1960, 1970, 1980. In Table A1 and Table A2, we show the cohort regressions for the following baseline specification used in Almond (2006):

$$y_i^c = \beta_0 + \beta_1 I(yob = 1919) + \beta_2 yob + \beta_3 yob^2 + \epsilon_i$$

where y_i^c denotes the census outcome for individual "i" in year 1960, 1970, and 1980. Our two main outcomes of interest pertaining to education are (i) whether the individual has completed high school or more, and (ii) whether the individual is literate. yob denotes the year of birth. We restrict the sample to those born between 1912 to 1922 and characterize the 1919 birth cohort as the one that experienced the greatest exposure to the pandemic (as in Almond (2006)). In other words, those both in 1919 are the "influenza treatment" group. We report the estimates for men, women, and non-whites, and focus on β_1-our estimate of interest, which as Almond explains, measures the departure of outcomes for the 1919 birth cohort from the quadratic cohort trend. While we are able to construct outcomes which closely mirror what Almond used, we are unable to have similar sets of controls and other outcomes of interest which could be used to lend further strength to the hypothesis that the 1919 birth cohort was negatively affected in several other ways. Our results in this setting suggest that there are no effects on high school completion for all groups considered except for men in 1960 (with the effect statistically significant at the 10% level only) and that there are possible selection effects on literacy rates (which mostly goes away by 1980) as evidenced by the positive coefficients for men and women in 1960 and 1970 in Table A1.

Before proceeding with our region-level analysis of the effects of the Pandemic, we make a second attempt at using individual-level data by using Beach et al. (2018) regression analysis. They include individual and parental controls in addition to the cohort that Almond had, and consider the incidence of the flu at a more disaggregated level (cities). We provide a simpler version of their model and assign the same aggregate value of influenza-related normalized deaths in 1918 to people in the census for Sao Paulo only. We focus once again only on education and literacy outcomes. We restrict the sample to those born between 1915 and 1919. The results are presented in Tables A3 and A4.

Table A1: Regression Results for Literacy Rates, Sao Paulo for 1960, 1970 and 1980

	Men			Women			Non-Whites	
	1960	1970	1980	1960	1970	1980	1960	1980
I(yob=1919)	0.025***	0.058***	0.015	0.033**	0.043***	0.015	-0.011	0.018
	(0.009)	(0.009)	(0.010)	(0.013)	(0.011)	(0.012)	(0.028)	(0.027)
yob	2.368**	-2.558**	-0.220	0.831	0.115	-5.944***	1.307	-6.103*
	(0.963)	(1.148)	(1.336)	(1.290)	(1.350)	(1.550)	(2.853)	(3.388)
yob2	-0.001**	0.001**	0.000	-0.000	-0.000	0.002***	-0.000	0.002*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Constant	-2,271.701**	2,450.348**	204.805	-802.201	-113.941	5,691.090***	-1,261.366	5,837.162*
	(923.417)	(1,100.310)	(1,280.369)	(1,236.735)	(1,293.861)	(1,485.584)	(2,735.046)	(3,248.281)
Observations	28,509	25,874	20,031	27,906	26,675	22,577	5,931	4,699
R-squared	0.001	0.002	0.003	0.002	0.002	0.003	0.004	0.009

Robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$ and *$p < 0.10$.

Table A2: Regression Results for High School Completion, Sao Paulo for 1960, 1970 and 1980

	Men			Women			Non-Whites	
	1960	1970	1980	1960	1970	1980	1960	1980
I(yob=1919)	0.011*	0.010	0.003	-0.003	0.002	0.001	0.000	-0.006
	(0.006)	(0.006)	(0.007)	(0.004)	(0.005)	(0.006)	(0.006)	(0.005)
yob	-0.693	-1.710**	-2.687***	-1.469***	-1.020*	-1.409**	0.159	0.154
	(0.593)	(0.743)	(0.873)	(0.399)	(0.597)	(0.684)	(0.569)	(0.644)
yob2	0.000	0.000**	0.001***	0.000***	0.000*	0.000**	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	664.629	1,639.131**	2,574.736***	1,406.947***	976.825*	1,350.588**	-153.191	-147.387
	(568.072)	(711.751)	(836.481)	(382.446)	(571.909)	(656.114)	(545.455)	(617.021)
Observations	27,966	25,904	20,038	27,487	26,691	22,597	5,791	4,700
R-squared	0.000	0.000	0.001	0.002	0.000	0.000	0.001	0.000

Robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$ and *$p < 0.10$.

Table A3: Regression Results for Literacy Rates, Sao Paulo for 1960, 1970 and 1980

	1960	1970	1980
yob	0.020***	0.014***	0.009***
	(0.002)	(0.002)	(0.002)
Respiratory Deaths*yob1915	-6.017**	-3.560***	-2.929***
	(0.898)	(0.934)	(1.080)
Respiratory Deaths*yob1916	-3.330***	-1.725	-0.432
	(1.059)	(1.060)	(1.217)
Respiratory Deaths*yob1917	-2.836**	-3.687***	-2.105
	(1.161)	(1.194)	(1.383)
Respiratory Deaths*yob1918	-5.817***	-4.361***	-3.197**
	(1.290)	(1.331)	(1.547)
Respiratory Deaths*yob1919	-5.973***	-1.431	-1.218
	(1.506)	(1.494)	(1.747)
male	0.217***	0.179***	0.164***
	(0.004)	(0.005)	(0.006)
white	0.152***	0.115***	
	(0.007)	(0.009)	
Indicator if anyone has high school in household	0.297***	0.293***	0.294***
	(0.004)	(0.004)	(0.005)
Resident in city or not	0.018	0.012	
	(0.019)	(0.020)	
Number of own family members in household	-0.000	-0.010***	-0.019***
	(0.002)	(0.001)	(0.001)
No. of own children en in household	-0.026***	-0.014***	-0.023***
	(0.002)	(0.002)	(0.003)
No.of own children under 5 in household	-0.007*	0.018**	0.027
	(0.004)	(0.008)	(0.021)
Whether individual is married	0.072***	0.047***	0.039***
	(0.006)	(0.005)	(0.006)
Total Income	0.000***	0.000***	
	(0.000)	(0.000)	
Constant	-38.806***	-26.719***	-17.596***
	(3.377)	(3.443)	(4.124)
Observations	46,782	44,762	34,413
R-squared	0.133	0.144	0.135

Robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$ and *$p < 0.10$.

Table A4: Regression Results for High School Completion, Sao Paulo for 1960, 1970 and 1980

	1960	1970	1980
yob	0.002**	-0.000	0.002**
	(0.001)	(0.001)	(0.001)
Respiratory Deaths*yob1915	0.133	-0.308	-0.673
	(0.329)	(0.394)	(0.535)
Respiratory Deaths*yob1916	0.082	0.237	-1.318**
	(0.390)	(0.471)	(0.604)
Respiratory Deaths*yob1917	-0.197	0.013	-1.671**
	(0.426)	(0.514)	(0.675)
Respiratory Deaths*yob1918	-0.504	0.079	-2.240***
	(0.471)	(0.565)	(0.756)
Respiratory Deaths*yob1919	0.152	-0.158	-1.792**
	(0.565)	(0.653)	(0.854)
male	0.044***	-0.028***	0.025***
	(0.002)	(0.002)	(0.004)
white	0.006***	0.012***	
	(0.001)	(0.002)	
Indicator if anyone has high school in household	0.342***	0.221***	0.251***
	(0.006)	(0.005)	(0.006)
Resident in City or not	-0.007	-0.023**	
	(0.010)	(0.011)	
Number of own family members in household	-0.004***	-0.004***	-0.011***
	(0.001)	(0.000)	(0.001)
Number of own children in household	0.001	-0.001***	-0.017***
	(0.001)	(0.001)	(0.001)
No.of own children under 5 in household	0.004***	0.017***	0.084***
	(0.001)	(0.002)	(0.009)
Whether Individual is married	0.001	-0.000	0.008***
	(0.002)	(0.002)	(0.002)
Total Income	0.000***	0.000***	
	(0.000)	(0.000)	
Constant	-3.198**	0.421	-3.932**
	(1.250)	(1.435)	(1.951)
Observations	46,098	44,779	34,433
R-squared	0.332	0.350	0.255

Robust standard errors in parentheses ***$p < 0.01$, **$p < 0.05$ and *$p < 0.10$.

Table A5: Differences in Sample Means for Selected Characteristics for Districts With Above and Below Influenza-Related Deaths during 1917-1920

Variables	Below Median Mean	Below Median Std Dev	Above Median Mean	Above Median Std Dev	Difference	
Municipality Expenditure (per capita)						
Cleaning and Waste Disposal	258.89	611.61	152.40	125.80	106.49	***
Maintenance of Sanitary Conditions	110.12	106.07	110.59	124.29	-0.47	
Geographic and Climate						
Altitude (Meters)	624.13	189.49	602.74	197.22	16.39	
Latitude	-22.23	1.05	-22.65	1.09	0.42	***
Longitude	-47.88	1.17	-47.63	1.39	-0.25	***
Distance to Capital	260.06	108.58	209.41	112.51	50.65	***
October's Temperature	20.61	1.98	19.91	2.10	0.70	***
October's Precipitation	124.85	7.54	123.67	9.18	1.18	**
Demographic						
Overall Median Age	20.75	2.92	20.62	2.96	0.13	
Population Density	9.56	6.53	10.62	8.38	-1.06	**
Dependency Ratio	88.91	21.23	89.70	21.66	-0.79	
Sex Ratio	109.93	6.77	108.66	8.32	1.27	**
Share of Slaves	0.15	0.07	0.15	0.08	0.00	
Literacy Rate	0.22	0.11	0.21	0.10	0.02	**
Economic						
Share of Employment in Agriculture	0.64	0.09	0.65	0.10	-0.00	
Share of Employment in Manufacturing	0.13	0.04	0.13	0.04	0.01	**
Share of Employment in Services	0.22	0.09	0.22	0.09	-0.11	
Health						
Physical or mental disease	11.07	5.19	10.80	5.50	0.28	
Doctors	0.22	0.28	0.18	0.24	0.03	*
Chemists	0.18	0.17	0.17	0.17	0.02	
Midwives	0.27	0.33	0.25	0.35	0.02	

Notes: This table shows the summary statistics and sample means differences for selected initial characteristics. The municipality expenditure per capita is for 1910, the geographic and climate variables are contemporary (1917-1920) and all other initial conditions are drawn from the 1872 census. The influenza-related deaths include deaths caused by respiratory and unknown causes. The median normalized deaths is 8.11. ***$p < 0.01$, **$p < 0.05$ and *$p < 0.10$.

Table A6: Data Description and Sources

Data Description	Sources
Panel A: Geography	
Altitude (in M)	The Encyclopedia of Brazil Municipalities (IBGE 1957,1958)
Latitude	The Encyclopedia of Brazil Municipalities (IBGE 1957,1958)
Longitude	The Encyclopedia of Brazil Municipalities (IBGE 1957,1958)
Distance to Capital (in km)	Secretariat Report of Agriculture, Commerce and Public Works, Sao Paulo (1920)
October's 1918 Temperature (Fahrenheit)	World Bank Climate Change Knowledge Database
October's 1918 Precipitation (mm)	World Bank Climate Change Knowledge Database
Panel B: Deaths by Cause (per 1000)	
14 Disease Categories for 1912-1921;1927	Annual Statistical Reports of Sao Paulo, 1912-1921 and 1927
Panel C: Initial Controls for 1872	
Literacy Rate	1872 Demographic Census for Sao Paulo, Brazil
Share of foreigners	1872 Demographic Census for Sao Paulo, Brazil
Share of slaves	1872 Demographic Census for Sao Paulo, Brazil
Shares of race Branca, Parda, Preta	1872 Demographic Census for Sao Paulo, Brazil
Share emp. in agriculture, manufucturing	1872 Demographic Census for Sao Paulo, Brazil
Share with physical/mental diseases	1872 Demographic Census for Sao Paulo, Brazil
Doctors, chemists, midwives (per 1000)	1872 Demographic Census for Sao Paulo, Brazil
Population density (per sq.Km)	1872 Demographic Census for Sao Paulo, Brazil
Panel D: Variables for 1920	
Infant Mortality (per live births)	Annual Statistical Report, Sao Paulo (1920)
Still Births (per total births)	Annual Statistical Report, Sao Paulo (1920)
Total Population	Demographic Census for Sao Paulo, Brazil (1920)
Literacy rate (total)	Demographic Census for Sao Paulo, Brazil (1920)
Literacy rate (male)	Demographic Census for Sao Paulo, Brazil (1920)
Literacy rate (female)	Demographic Census for Sao Paulo, Brazil (1920)
Sex ratio of adults	Demographic Census for Sao Paulo, Brazil (1920)
Sex ratio at 12 months	Demographic Census for Sao Paulo, Brazil (1920)
Share of foreigners	Demographic Census for Sao Paulo, Brazil (1920)
Production per capita for coffee, rice, maize	Census of Agricultural and Industrial Activities (1920)
Production per est. for coffee, rice, maize	Census of Agricultural and Industrial Activities (1920)
Panel E: Variables for 1940	
Share receiving instruction and literate	Demographic Census for Sao Paulo, Brazil (1940)
Share receiving instruction and literate	Demographic Census for Sao Paulo, Brazil (1940)
Literacy rate	Demographic Census for Sao Paulo, Brazil (1940)
Inpatient Hospital Admissions (per 1000)	Annual Statistical Report of Sao Paulo (1940)
Inpatient Asylum Admissions (per 1000)	Annual Statistical Report of Sao Paulo (1940)
Value of primary sector per emp.	Agricultural, Industrial and Commercial Statistics: Census, Sao Paulo Volume 3 (1940)
Value of primary sector per est.	Agricultural, Industrial and Commercial Statistics: Census, Sao Paulo Volume 3 (1940)
Panel F: Contemporaneous Variables	
IFDM Brazilian Development Index 2006-2016	Official website of the Indice FIRJAN de Desenvolvimento Municipal (IFDM)

Continued ...

Table A6 (Continued): Data Description and Sources

Data Description	Sources
IFDM Brazilian sub-indices 2016:	
Employment and Income	Official website of the Indice FIRJAN de Desenvolvimento Municipal (IFDM)
Education	Official website of the Indice FIRJAN de Desenvolvimento Municipal (IFDM)
Health	Official website of the ndice FIRJAN de Desenvolvimento Municipal (IFDM)
Panel G: Other Controls	
Health&Sanitation exp.per capita (1910)	Annual Statistical Report of Sao Paulo Vol. 2 (1910)
Dummy for frost intensity (1919), Railway station dummy	Secretariat Report of Agriculture, Commerce and Public Works of State of Sao Paulo (1920)
Water&sewerage system dummy (1920)	Sanitary Service Report of the State of Sao Paulo (1920)
No. hospitals&nursing homes (1920)	Sanitary Service Report of the State of Sao Paulo (1920)
Taxes out of revenue (%,1920)	Census of Sao Paulo, Brazil (1920)
Share of exp.on health&education (1940)	Annual Statistical Report for Sao Paulo (1940)
Agricultural sector exp. per est.(1940)	Annual Statistical Report for Sao Paulo (1940)
Sectoral exp. per est. (1940)	Annual Statistical Report for Sao Paulo (1940)
Number of plows per est. (1940).	Agricultural, Industrial and Commercial Statistics: Census, Sao Paulo Volume 3 (1940)
Share of est. exp on salaries, machinery, animals etc. (1940)	Agricultural, Industrial and Commercial Statistics: Census, Sao Paulo Volume 3 (1940)

Notes: Authors' compilation.

Table A7: Instrumental Variable Effects for Wages in All Quarters of 1920

	Quarters 1 and 2 (1)	Quarter 2 (2)	Quarter 3 (3)
Dependent Variable:			
Log Wages of Tending 1000 Coffee Plants			
Norm. Respiratory Deaths	0.218**	0.119**	0.144**
	(0.107)	(0.055)	(0.061)
Dependent Variable:			
Log Wages for Treatment of 1000 Coffee Plants			
Norm. Respiratory Deaths	0.133**	0.127**	0.116**
	(0.059)	(0.064)	(0.061)
Region FE	Yes	Yes	Yes
Region-level controls	Yes	Yes	Yes
Initial Controls of 1872	Yes	Yes	Yes
Mean of dep. variable (for tending)	11.325	11.342	11.341
Mean of dep. variable (for treatment)	9.815	9.814	9.837
Observations	324	324	329

Notes: The table reports the 2sls results for log wages. The instrument used is IVResp. Standard errors reported in parentheses are clustered at the mesoregion level. We use the variation for the period 1917-1919 to measure the effects on 1920 wages. The initial controls of 1872 include population density, the shares of employment in agriculture, manufacturing and services/retail, the share of people of race Branca, and the literacy rate. The controls include altitude, latitude, longitude, the distance to capital, a dummy for the presence of a railway station, the interaction between distance and the railway dummy, the literacy rates in 1920 for both Brazilians and foreigners. In all columns, we include mesoregion fixed-effects. ***$p < 0.01$, **$p < 0.05$ and *$p < 0.10$.

Table A8: IV Effects for Productivity Measures at the Sectoral Level in 1940

	Agriculture		**Farming**		**Livestock**	
	Small scale	Large scale	Small scale	Large scale	Small scale	Large scale
	Dependent Variable: Value Per Hectare					
Norm. Respiratory Deaths	-0.067	-0.157	-0.029	-1.191	0.016	-0.047
	(0.057)	(0.131)	(0.033)	(0.975)	(0.011)	(0.039)
	Dependent Variable: Value Per Employee					
Norm. Respiratory Deaths	-0.579	-0.818*	-0.465	-1.017	0.380	-0.433
	(0.411)	(0.496)	(0.315)	(2.840)	(0.354)	(0.448)
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Region-level Controls	Yes	Yes	Yes	Yes	Yes	Yes
Initial Controls of 1872	Yes	Yes	Yes	Yes	Yes	Yes
Observations (for per hectare)	425	355	440	440	422	417
Observations (for per employee)	425	351	436	324	411	405

Notes: The table shows the 2SLS results (with IVResp as instrument). Standard errors reported are clustered at the microregion level. We use the variation for the period 1917-1920 to evaluate the effects on 1940 land/agricultural productivity. The controls including altitude, latitude, longitude, distance to the capital city. We also control for the 1940 literacy rates for male and female of age 5 and above, and for the 1920 share of foreigners. Further, we include several other 1940 controls that could affect productivity. These controls include the expenditure per establishment by agricultural, farming and livestock establishments; the number of plows per establishment, the amount of expenditure on salaries, acquisition of machinery and of new animals, and on fertilizers per establishment. The initial controls of 1872 include population density, the shares of employment in agriculture, manufacturing, retail/services, the share of people of race Branca, the share of people who were slaves. ***$p < 0.01$, **$p < 0.05$ and *$p < 0.10$.