Health shocks and Human Capital Accumulation: the Case of Spanish Flu in Italian Regions*

Marco Percoco
PAM , CERTeT and Dondena Center
Università Bocconi
via Rontgen 1
20121 Milano, Italy
marco.percoco@unibocconi.it

December 2014

Abstract

The impact of health on economic development is a hotly debated issue in the economics literature, with most scholars supporting the idea that the diffusion of diseases is detrimental to development. In this context, pandemics are an important case study given their exogenous nature, which makes identification of the impact of diseases on development clearer than in other cases such as malaria or smallpox. In this paper we focus on Spanish flu in Italy, one of the countries with the highest mortality rate due to the pandemic. By exploiting the regional variation in mortality and focusing on the hypothesis of the fetal origins of cognitive abilities, we have estimated the long-run consequences of influenza exposure in terms of human capital accumulation. We have found an average reduction of 0.3-0.4 years of schooling for the cohort born in 1918-1920. This result points to a small but persisting effect of health shocks on regional productivity through a variation in the rate of accumulation of human capital.

Keywords: Spanish flu, Human Capital, Regional Development.

(*) The author wishes to thank an Editor, two referees, Simone Ghislandi and Martin Karlsson for useful comments. The usual disclaimer applies. Financial support from the Italian Ministry of Education and Research through the FIRB grant "Social and spatial interactions in the accumulation of civic and human capital" is gratefully acknowledged.

1. Introduction

The effect of health on economic development and growth is much debated in the international literature. A large amount of literature, recently summarized by Acemoglu and Johnson (2007), suggests that mortal diseases and pandemics have a positive economic impact because they contract the labor supply and consequently increase real wages and improve living conditions. Furthermore, Acemoglu and Robinson (2012) have recently argued that an extreme health shock like the Black Death may be considered as a critical juncture in history, i.e. "a major event or confluence of factors disrupting the existing economic or political balance in society. […] On the one hand it can open the way for breaking the cycle of extractive institutions and enable more inclusive ones to emerge, as in England. Or it can intensify the emergence of extractive institutions, as was the case with the Second Serfdom in Eastern Europe" (p. 101). Therefore, health shocks may impose sizeable economic and social changes with persisting effects across decades or centuries. However, it should be stated that the view reported in Acemoglu and Johnson (2007) does not argue in favour of human disasters as a road to prosperity. Rather it is argued that, for the survivors, living standards may improve as a consequence of the rebalance of labor market power between laborers and landowners.

Weil (2007) points out that there is a large body of research that has found a positive effect of health, as measured by life expectancy increase, on growth. The magnitude of the elasticity from macroeconomic data lies in the interval 0.06-0.57, depending on other covariates and on the sample used (Bloom et al., 2004). A crucial role in this framework is played by epidemics and pandemics, which often change life expectancy only marginally in the long run but nevertheless have important economic effects.[1]

To overcome problems of weak causality in the form of endogeneity, the recent literature has focused on health shocks as sources of exogenous variations in health status. Bleakley (2010) and Percoco (2011) have studied episodes of malaria eradication in the Americas and in Italy, finding that cohorts born immediately after the eradication in areas with high malaria morbidity increase their educational attainments.

Whilst malaria eradication can be conceived as a positive health shock, another strand in the literature has focused on pandemics as shocks that negatively affect individuals' health. In this regard, Almond (2006) has proposed an interesting analysis of the impact of Spanish flu on human capital accumulation in the US, finding a strong and negative correlation between the diffusion of the disease and educational outcomes of the cohort of individuals born in 1919.

In this paper we start from this evidence and study the case of Italy, where the flu pandemic of 1918-1920 was even more severe than in the US. In doing so, however, we argue that the spatial variation in the mortality rate adds significant heterogeneity to the cross-cohorts analysis, so that in our analysis the place where an individual was born during the pandemic is of considerable importance for his/her long-run educational outcomes. Our approach hence assigns an explicit role to space in the economics of influenza in terms of the quality of the environment in which an individual was born.

In particular, we assume that exposure to the pandemic *in utero* or during childhood

[1] For a more comprehensive review of the literature on the topic, see Almond and Currie (2011a).

reduced the cognitive capabilities of children, leading to a decrease in the ability to accumulate human capital in the long run. In the case of Italy, we will show that the impact of the disease was heterogeneous across space: that is, children born during the influenza pandemic received less schooling than the following and preceding cohorts, and the magnitude of such effects depended on a measure of child exposure during the first four years of life, as well as on the region of birth.

In this paper, we propose the hypothesis that spatial inequality in terms of health outcomes during an extreme shock may have persisting effects on human capital accumulation processes of regions. Health inequality across regions in the specific case of the Spanish flu may have been caused by several phenomena (Gatrell and Elliott, 2009; Kulkarmi and Subramanian, 2010). First, inequality in the accessibility to health care facilities may have caused substantial heterogeneity in the mortality rates across locations depending on the availability of medicines and services. Second, conditions of diffused material deprivation may have imposed a sizeable burden in case of severe health shock since populations at risk might be weak to respond to the disease. Third, given economic structures, regions may differ in terms of types of jobs and ultimately of life styles. This feature in its turn may affect exposure to diseases.

Our empirical strategy consists in measuring the pandemic both through dummy variables as proposed by previous literature and by using mortality rates across regions and time. Both ordinary least squares and instrumental variable estimates are presented. We have found that exposure to the disease decreased educational attainments by 0.3-0.4 years of schooling, and that this estimate is robust to the inclusion of several confounding factors and different model specification. Our estimates also point to considerable heterogeneity in the effect of the flu, ranging from -0.15 to -0.44 years of schooling.

This paper builds on a growing body of literature on the effect of Spanish flu. However, it adds to existing works in several ways. First, it proposes estimates that remove the strong assumptions of the econometric frameworks proposed so far. In fact, most of the literature, by relying on the so-called ’fetal origins’ hypothesis assumes that the effect of the influenza is evident only for children born during the pandemic, whereas it is assumed to be null for children born before the pandemic but alive during the epidemic. We propose a slightly different approach and find that long-run effects of health shocks are significant also for individuals exposed to the influenza during childhood. Second, we account for several potential confounding factors, such as the effects of World War I and mortality for related diseases, and test the robustness of results by considering several functional forms and estimation procedures. Third, we propose estimates of the impact of the flu pandemic in Italy, the most heavily affected country in the West after Spain.

Finally, it must be stated that the hypothesis of the effect of the Spanish flu on cognitive capabilities of children cannot be tested explicitly with our data, so that we cannot disentangle this channel of transmission from other possible mechanisms. However, in this paper we will refer to this workable hypothesis to justify the assumed relationship between exposure to the disease and education.

The remainder of the paper is organized as follows. In section 2 we review the literature on the economics of the Spanish flu pandemic and set out our main arguments which will be the basis upon which we build the empirical model described in the third section. Section 4 contains baseline results along with robustness checks. Section 5 concludes.

2. The Spanish flu pandemic and its long-run impact

Pandemics have accompanied the pattern of human evolution from antiquity until modern civilization. Some of those pandemics have dramatically changed the fortunes of cities and nations. Historical chronicles indicate the Black Death of the fourteenth century and the Justinian plague of the sixth as the two most deadly pandemics in human history. However, Langford (2002), on the basis of a literature review, argues that the Spanish flu pandemic was probably more dramatic (in terms of loss of human lives) than the Justinian plague, and hence ranks second for mortality.

The 1918-1920 influenza pandemic was immediately referred to as "Spanish flu" because it first appeared in Spain in 1918 and rapidly spread around the world, probably being transmitted by armies engaged in World War I (Crosby, 1989). Soldiers living in poor sanitary conditions were prolific disseminators of the influenza, especially when they returned home during periods of leave from the war front. In addition, the end of the war, with subsequent mass returns of soldiers to their homelands, was probably one of main causes of the rapid geographical diffusion of the influenza (Tognotti, 2002).

The final death toll of the flu pandemic is still disputed; however, recent estimates put the cost of the pandemic in terms of human lives at 50 million. The geography of the death toll is nevertheless extremely uneven as mortality varied significantly across regions (figure 1).

[Figures 1, 2, 3; table 1]

Italy had 390,000 deaths, with a death rate (number of deaths per 1,000 individuals) of 10.7, the second highest after Spain. The pandemic lasted in Italy for the period 1918-1920, with a peak in 1918, the last year of the war (figure 2). The temporal pattern of mortality was similar across regions, although the mortality rate varied (table 1). Figure 2 plots mortality rates in 1918 and shows higher values in Southern regions, far from the war front. Interestingly, figure 3 shows an abrupt increase in the number of influenza deaths, and a similarly abrupt decrease to pre-1918 levels after 1920. Consequently, the flu pandemic can be considered a sort of natural experiment.

Taken together, the information in figure 3 and in table 1 supports the view that the pandemic can be considered a natural experiment with significant spatial variation in its effects, which is a desirable feature for our econometric framework[2].

The incidence of mortality differed significantly across age, cohorts with peaks in childhood and old age and in the 20-40 age range (Langford, 2002). Perhaps, the high mortality rate in the 20-40 age is due to the war and this is evidence of the fact that World War I is a serious confounding factor in our analysis. Unfortunately, we do not have age-specific death rates for Italy, although we can reasonably assume that the influenza had the same pattern as in the rest of the world.

The W-shaped mortality pattern is of great importance for understanding the impact of the disease. With specific reference to the Spanish flu pandemic, Brainerd and Siegler (2002) and Garret (2009) have found evidence corroborating this view for US states, the latter finding a 2-3% increase in the average wage due to influenza mortality.

This strand of literature hypothesizes that influenza (or mortal diseases in general) has an impact on wages, through mortality, because it generates a contraction in labor supply.

[2] In this paper, however, we will relax the assumption of strong exogeneity of the pandemic by making use of instrumental variable estimates.

However, mortality and poor health conditions may also have a temporary or permanent impact on the rate of human capital accumulation. Health, in fact, may influence the decision to invest in education in many ways, among the most important being a change in the rate of return on human capital investment due to a change in life expectancy and a change in the cognitive abilities of individuals, especially children. In other words, an health shock may affect both the labor supply and total factor productivity.

Figure 4 plots regional growth rates in Italy over the years 1920-1930[3] against the influenza mortality rate at its peak. It shows a strikingly negative correlation which supports (although in a very simple and descriptive way) our view that a health shock affects development also through a decrease in total factor productivity.

In the case of persistent and endemic diseases, such as malaria, it is difficult to disentangle the effect of a change in the rate of return on human capital investment from a change in cognitive abilities (Percoco, 2011), whereas in the case of a natural experiment like the Spanish flu pandemic, it is possible to identify the impact on cognitive abilities because the change in mortality is only transitory and does not significantly affect life expectancy and hence the decision to invest in education.

The theoretical rationale for considering the impact of the Spanish flu pandemic on cognitive abilities and ultimately on human capital accumulation relies on the so-called 'fetal and infant origins' hypothesis, which holds that the quality of the environment in which mothers and children live during pregnancy or during early childhood is crucial in explaining outcomes later in life (Almond and Currie, 2011b).

[3] The source for regional GDP over the time period considered is Daniele and Malanima (2007).

The first evidence concerning the importance of the persistent effects of shocks during childhood was provided by Stein et al. (1975), who found adverse health outcomes for Dutch children born during the famine and Nazi occupation. Barker (1990) has systematized the medical evidence available to date, arguing in favor of the 'fetal and infant origins' hypothesis of human development.

From an economic point of view, this hypothesis implies that the technology of skills formation has two inputs, among others, that are imperfect substitutes: investment in education and cognitive capabilities (Cunha and Heckman, 2008; Cunha et al., 2010). A temporary shock occurring in the first input is more easily absorbed by human beings than a shock in the cognitive abilities, as in the case of severe diseases *in utero* or, in our case, during childhood. Hence, if the Spanish flu pandemic was a shock to the capabilities of children to accumulate human capital, then we should observe a persisting effect in terms of educational attainments. It should be noted that assuming the 'fetal and infant origins' hypothesis as a starting point is of particular relevance in the case of the Spanish flu pandemic because of the high mortality rate among children (Somogyi, 1967).

Relying on the 'fetal and infant origins' hypothesis, Almond (2006) has found that the Spanish flu epidemics had a dramatic impact on the lives of children born in 1919 along several dimensions of human life, such as income, education, and well-being in general. Similar studies have been conducted for the case of Brazil (Nelson, 2010), Switzerland (Neelsen and Stratmann, 2011), Taiwan (Lin, 2011). All of them find a significant and negative effect of the influenza on educational or socio-economic outcomes.[4]

[4] Kelly (2011) is a related paper on a case of severe influenza in the UK in the 1950s.

The aforementioned literature is in general supportive of the long-run and persistent effect of the flu pandemic on individuals' outcomes, although it should be stated that some recent works have strongly criticized this approach because it often fails to recognize the selectivity and heterogeneity of the underlying process of infection (Brown, 2011).

In what follows, we provide evidence on the effect of the pandemics on educational attainments by also recognizing the spatial heterogeneity of the infection across Italian regions.

3. Methodology and data

As stated in the previous sections, in order to estimate the impact of the Spanish flu pandemic, we rely on the 'fetal and infant origins' hypothesis. To this end, we propose an econometric framework that makes use of individual-level data.

We used data from the Bank of Italy Survey of Household Income and Wealth (SHIW): in particular, the waves conducted in the years 1977, 1978, 1980, 1981, 1982, 1983, 1984, 1986, 1987. We restricted the sample to the cohorts of people born between 1910 and 1930.

The SHIW contains only levels of education in terms of school attainment. We then calculated years of schooling by approximation on the basis of each level of education attainment. We defined less than primary as zero years, primary as five years, middle as eight years, secondary as thirteen years, and college as eighteen years. However, we also used as a dependent variable the probability of attaining an upper-secondary diploma or a university degree.

Data on mortality are from ISTAT (1957), in which regional time series are reported.[5] To compute mortality rate, we have used SVIMEZ (2011) data on census population for the years 1901, 1911, 1921, 1931. Inter-census data have been obtained by means of interpolation.

Before discussing our empirical approach, let us consider figure 5, in which we have plotted the average years of schooling by cohorts. It seems that our prior is qualitatively verified because, in correspondence with the pandemic, i.e. during the years 1918-1920, a decrease in years of schooling is observed. Furthermore, figure 6 shows that this drop was in common to the three macro-areas of Italy (North, Center and South), albeit with different magnitudes. It should also be noted that figures 5 and 6 show that a linear trend across cohorts seems to fit the pattern of education across cohorts remarkably well.

Our general econometric specification builds on the previous work by Almond (2006) and is specified formally as:

$$school_{ick} = \alpha + \beta year_c + \gamma \text{influenza}_{\text{ck}} + \delta region_k + \varepsilon_{ick} \tag{1}$$

Where $school_{ick}$ is a measure of educational attainment for individual *i* born in year *c* in region *k*, $year_c$ is a time trend across cohorts, and $region_k$ denotes a full set of region-specific fixed effects. Variable $\text{influenza}_{\text{ck}}$ is a measure of influenza pandemic at cohort and/or regional level.

We have several candidates for our treatment variable, i.e. $\text{influenza}_{\text{ck}}$. First, the pandemic

[5] Appendix reports descriptive statistics and correlations for main variables.

flu effect, as in Almond (2006), can be estimated by means of a time dummy variable for the year 1919, indicating a departure from the cohort trend. However, as is evident from figure 3, the flu pandemic lasted for three years in Italy, with a peak in 1918. For this reason, we have made use in our analysis of three dummy variables, one for each year of the pandemic.

The second variable of potential interest is the mortality rate for influenza in the region and year of birth of individual *i*. However, the disadvantage of using this variable is that it approximates the exposure of individual *i* only with the death rate in the year of birth, implying that it is zero for the following years and neglecting the relevance of health shocks during childhood (e.g. the assumption is that a child born in 1917 had almost no exposure to Spanish flu because the mortality rate in 1917 was about 0.04, whereas in 1918 it was 8.08). To circumvent this limitation, we have used a third variable: the sum of regional mortality rates for influenza over five years. Thus, for an individual born in year *c* in region *k* it is defined as:

$$(2) \qquad \text{influenza}_{\text{ck}} = \sum_{t=c}^{c+4} mortality_{tk}$$

Measure in equation (3) indicates that the impact of the flu pandemic on individual *i* is given by the sum of mortality rates during the first four years of life. The rationale behind this variable for disease exposure is the evidence that influenza can have the worst impacts in case of severe episodes during the first four years of life, especially through the onset of complications (Rudan et al. 2008). It should be noted that we do not have an individual level measure of exposure to the pandemic. Rather, we assume that being born or growing during a major health shock (measured in terms of mortality rates) may have significant impact on the accumulation of human capital of individuals and hence of regions. In other terms, our treatment variable is measured at regional level, whereas the outcome is measured at individual level.

Equation (1) is similar in spirit to the one estimated in Almond (2006), although it differs from it in two important ways. First, we do not consider a quadratic and cubic time trend because it does not effectively approximate the evolution of schooling across the years 1910-1930, as shown by figure 5. Second, we introduce region-specific fixed effects in order to take account of regional disparities in the country. Although we have introduced such innovations in the baseline specification, in what follows we test the robustness of our results to changes in model specification.

The advantage of using a measure like (2) is that it considers the impact of pandemic flu not only during the year of birth, but also in the following years. It thus furnishes a more comprehensive assessment of periods of severe pandemic.

4. Results

We begin our empirical analysis by estimating versions of equation (1) in which Spanish flu is measured by means of time dummy variables. Model (1) in table 2 considers as a treatment variable a dummy for the year of birth 1919. Estimates indicate lower educational attainment of magnitude 0.429, i.e. individuals born in 1919 have about 0.43 years of schooling less than individuals in other cohorts. Models (2) and (3) estimate the impact of pandemic exposure in terms of the probability of obtaining a high-school diploma or a university degree. Estimates of logit models are not satisfactory in terms of parameters significance, indicating that the Spanish flu pandemic did not have a significant impact on higher education, although this can be also due to the very low number of individuals with higher education.

In models (4) - (7) we consider years of schooling as an outcome variable and add region-specific time (i.e. years of birth) trends in order to take account of potential heterogeneity among educational policies across regions. The introduction of a region-specific time trend is of particular importance in this case because of the large disparities characterizing Italian socio-economic geography. In fact, in the aftermath of unification of the country, regional disparities in terms of human capital were particularly marked (Gagliardi and Percoco, 2011), and several policies were implemented nation-wide to enhance education. However, regions may have reacted differently to those interventions, and hence region-specific time trends may capture this eventual effect.

In model (4) we consider the year 1918 (the peak of the pandemic) as a treatment variable and cannot find any statistically significant effect, as also in the case of 1920. By contrast, a variable indicating whether the individual was born in one of the years between 1918 and 1920 seems to show a relatively good fit.

[Table 2]

As discussed in the previous sections, however, the impact of the pandemic was heterogeneous across regions, so that the mildly satisfactory results reported in table 2 may hide significant differences across space. To test this hypothesis, models (1)-(3) in table 3 report different versions of equation (1) where the influenza death rate in the year and region of birth is used as an explanatory variable. Notably, none of the specifications reports a significant effect of the influenza pandemic in explaining human capital accumulation.

Models (4)-(6) use as a treatment variable the variable proposed in equation (2) measuring total exposure during childhood. Model (4) in table 3 reports a point estimate of -0.032, implying that a three standard deviation increase in total exposure will result in lower educational attainments by an order of magnitude of -0.2 years of schooling. Models (5) and (6) are logit models that estimate the probability of obtaining a high-school diploma and a university degree respectively. Both models point to a negative impact of child exposure in terms of both outcome variables, with very similar effects in terms of coefficient estimates. Combining the results in table 3 with data on total exposure in Italian regions shows that, for an individual born in 1918, the impact of the Spanish flu pandemic in terms of years of schooling ranges from -0.15 in Veneto to -0.44 in Sardegna, which bears out the importance of our initial hypothesis on location-specific impacts on health. Figure 7 reports the spatial variation in average loss of years of schooling for an individual born in 1918 by using the point estimate of -0.032 (positive values in the map indicate a loss). Interestingly, most of the impact of the pandemic is located in the South, although the average loss in Sicilia is very close to the impact estimated in Lombardia.

[Table 3; figure 7]

The results presented in table 3 confirm that the impact of the flu pandemic was highly significant for children aged from 0 to 4. To test the robustness of our results, in table 4 we present estimates of several sensitivity checks. In particular, in model (1) we have added region-specific time trends (to account also for the potential differential effect of schooling reforms taking place across years) in the specification with the number of years of schooling as a dependent variable, finding no significant change in the coefficient of interest.

Spanish flu spread in the years following World War I; hence this event may be a serious confounding factor in our analysis. Figures 8 and 9 show simple correlations between the mortality rate of soldiers during the war[6] and the mortality rate for influenza and the cumulative mortality rate for influenza, respectively. In both cases, there seems to be no correlation, so that the coefficient of our interest does not seem to be affected by omitted variables related to World War I. However, in order to test more explicitly for this case, in model (2) we have considered region-specific time trends with the regional mortality rate of soldiers, finding non-significant changes in the coefficient associated with the treatment variable of interest. Taken together, these results indicate that the war may be thought to be the transmission channel for the diffusion of the flu (according to the literature), however, there is no correlation between the impact of the war and the long run effect of the influenza.

[Table 4; figures 8 and 9]

---

[6] The mortality rate is defined as the number of deaths among soldiers during the war on the total male population. The source for this variable is Mortara (1925).

By relying on individual-level regressions and using the effect of a natural experiment as an explanatory variable, we are confident about the exogeneity in the estimation of the effect of the flu pandemic. However, omitted variables correlated with $influenza_{ck}$ may still impose a bias on the estimate of γ. To overcome this problem, we have relied on an instrumental variable approach where $influenza_{ck}$ has been instrumented with a variable defined as:

$$\inf\_cond_{ck} = \frac{\varpi_{1910k} \inf luenza_c}{population_{ck}} \tag{3}$$

where ω_{1910k} is the share of deaths from influenza in 1910 in region *k*, $influenza_c$ is the total (i.e. nationwide) number of deaths from influenza in year c, and $population_{ck}$ is total population in region *k* in year *c*.

The instrument in equation (3) has been proposed by Acemoglu and Johnson (2007). It serves to eliminate the possible bias imposed by an endogenous spatial distribution of the flu pandemic. In other words, the instrument in (3) assumes as mildly exogenous both the national mortality rate and the initial regional distribution of influenza mortality.

In this case too, as reported in table 4, in models 3 and 4 our estimates prove to be largely robust to model specification, thus confirming the negative effect of the Spanish flu pandemic on human capital. Finally, in model (5), we have added death rates for bronchitis and pneumonia in order to account for the possibility of other deaths being caused by the worsening of health due to the Spanish flu. Also in this case, our results are largely confirmed.

Panels B and C in table 4 report estimates (probit) of the probability of obtaining a high-school diploma and a university degree respectively. The estimated coefficients are mildly significant, indicating that the influenza pandemic had only a medium effect on higher education.

5. Conclusion

The literature has identified human capital as one of the most important drivers of regional development. However, most of this literature considers education to be the outcome of individuals' investment decisions, possibly stimulated by public interventions. In this paper we have proposed evidence on the dependence of human capital accumulation on the ability of individuals to accumulate such a production input after a health shock.

More in particular, we have studied the case of the Spanish flu epidemic in Italy as an example of exogenous change in (unobserved) cognitive abilities of individuals. Our empirical framework builds on the 'fetal and infant origins' hypothesis that exposure to severe diseases *in utero* or during childhood has dramatic impacts in terms of outcomes during adulthood. In addition to the existing literature, we have exploited the spatial variation of mortality rates to take account of heterogeneity in the burden of the disease.

By using individual-level data we have shown that individuals exposed to the disease *in utero* or during the first four years of life had fewer schooling years, and that the burden of the disease significantly varied across regions, ranging from -0.15 to -0.44 years of schooling. This result is robust to several sensitivity checks and points to an important role of health in human capital accumulation. It also suggests that education has, amongst other things, long-run determinants able to influence, albeit marginally, the process of regional development.

References

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

Acemoglu, D. and J.A. Robinson (2012), *Why nations fail. The origins of power, prosperity and poverty*, London: Profile Books.

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

Almond, D. and Currie, J. (2011a), Human Capital Development Before Age Five, *Handbook of Labor Economics, Volume 4b*, Orley Ashenfelter and David Card, editors, Elsevier.

Almond, D. and Currie, J. (2011b), Killing Me Softly: The Fetal Origins Hypothesis, *Journal of Economic Perspectives*, 25(3):153-172.

Barker, D.J. (1990), The fetal and infant origins of adult disease, *British Medical Journal*, 301(6761):1111.

Bleakley, H. (2010a), Malaria eradication in the Americas: a retrospective analysis of childhood exposure, *American Economic Journal: Applied*, 2(2):1-45.

Blomquist, G. (2006), Measuring quality of life, R. Arnott and D. MacMillen (eds), *A Companion to Urban Economics*, New York: Wiley.

Bloom, D.E., Canning, D., Sevilla, J. (2004), The effect of health on economic growth: a production function approach, *World Development*, 32(1):1-13.

Brainerd, E. and Siegler, M. (2002), The economic effects of the 1918 influenza epidemic, mimeo.

Brown, R. (2011), The 1918 U.S. Infuenza Pandemic as a Natural Experiment, Revisited, Duke University, mimeo.

Crosby, A. (1989), *America's forgotten pandemic: the influenza of 1918*, Cambridge: Cambridge University Press.

Cunha, F. and Heckman, J. (2008), Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation, *Journal of Human Resources*, 43(4):738-782.

Cunha F., Heckman, J., and Schennach, S. (2010), Estimating the technology of cognitive and noncognitive skill formation, *Econometrica*, 78(3):883-931.

Daniele, V. and P. Malanima (2007), Il prodotto delle regioni e il divario Nord-Sud in Italia (1861-2004), *Rivista di Politica Economica*, 2007(Marzo-Aprile):267-315.

Gagliardi, L. and Percoco, M. (2011), Regional disparities in Italy over the long run: the role of human capital and trade policy, *Region and Development*, 33:81-105.

Garrett, T.A. (2009), War and Pestilence as Labor Market Shocks: U.S. Manufacturing Wage Growth 1914-1919, *Economic Inquiry*, 47(4):711-725.

Gatrell, A.C. and S.J. Elliott (2009), *Geographies of health. An introduction*, Oxford: Wiley-Blackwell.

ISTAT (1957), *Cause di morte, 1887-1955*, Rome: ISTAT.
Kelly, E. (2011), The scourge of the Asian flu: in utero exposure to pandemic influenza and the development of a cohort of British children, *Journal of Human Resources*, 46(4):669-694.
Kulkarmi V.S. and S.V. Subramanian (2010), Social perspectives on health inequalities, in T. Brown, S. McLafferty and G. Moon (eds), *A companion to health and medical geography*, Oxford: Wiley-Blackwell.
Langford, C. (2002), The age pattern of mortality in the 1918-1919 influenza pandemic: an attempted explanation based on data for England and Wales, *Medical history*, 46(1):1-20.
Lin, Y. (2011), Does in utero exposure to illness matter? The 1918 influenza pandemic in Taiwan as a natural experiment?, mimeo.
Moretti, E. and M. Neidell (2011), Pollution, Health, and Avoidance Behavior: Evidence from the Ports of Los Angeles, *Journal of Human Resources*, 46(1):154-175.
Mortara, G. (1925), *La salute pubblica in Italia durante e dopo la Guerra*, Bari: Laterza.
Neelsen, S. and Stratmann, T. (2011), Long-run effects of fetal influenza exposure: Evidence from Switzerland, *Social Science and Medicine*, Forthcoming
Nelson, R.A. (2010), Testing the fetal origins hypothesis in a developing country: evidence from the 1918 influenza pandemic, *Health Economics*, 19(10):1181-1192.
Percoco, M. (2011), The fight against geography: malaria and economic development in Italian regions, FEEM Note di Lavoro 2011.007.
Rudan, I., Boschi-Pinto C., Biloglav Z., Mulholland K., and Campbell H. (2008), Epidemiology and etiology of childhood pneumonia, *Bulletin of World Health Organization*, 86(5):408-416.
Stein, Susser, Saenger, Marolla (1975), *Famine and human developmnt: the Dutch hunger winter 1944-1945*, Oxford: Oxford University Press.
Somogyi, S. (1967), *La mortalità nei primi cinque anni in Italia*, Palermo: Ingrana.
SVIMEZ (2011), *150 anni di statistiche italiane: Nord e Sud 1861-2011*, Bolona: il Mulino.
Tognotti, E. (2002), *La "spagnola" in Italia. Storia dell'influenza che fece temere la fine del mondo (1918-1919)*, Milano: Francoangeli.
Weil, D.N. (2007), Accounting for the effect of health on economic growth, *Quarterly Journal of Economics*, 30:1265-1306.

Table 1: The death toll of the Spanish flu across regions

Region	Total mortality (1918-1920)	Mortality rate in 1918
Piemonte	27,783	6.279
Valle d'Aosta	27,783	6.279
Liguria	10,679	6.413
Lombardia	43,695	7.234
Veneto	18,033	3.697
Emilia Romagna	24,483	6.555
Toscana	26,616	7.754
Marche	10,678	6.953
Umbria	7,130	7.969
Lazio	19,478	9.497
Abruzzi	16,113	9.532
Molise	16,113	9.532
Campania	33,159	8.491
Puglia	22,659	9.204
Basilicata	6,194	10.683
Calabria	18,319	10.528
Sicilia	33,740	8.046
Sardegna	11,491	10.911

Notes: The number of deaths for influenza are from ISTAT (1957). Population estimates for the computation of mortality rates are interpolation of census population between 1901 and 1931 as reported in SVIMEZ (2011).

Table 2: The effect of influenza pandemic on education (Dichotomous treatment variable)

	(1) Years of schooling	(2) Probability to obtain a high schoold degree (Logit)	(3) Probability to obtain a university degree (Logit)	(4) Years of schooling	(5) Years of schooling	(6) Years of schooling	(7) Years of schooling
Treat. = 1919	-0.429**	-0.077	-0.128*		-0.435**		
	(-2.298)	(-1.111)	(-1.661)		(-2.340)		
Treat. = 1918				0.101			
				(0.699)			
Treat. = 1920						-0.302*	
						(-1.861)	
Treat. = 1918-1920							-0.275**
							(-2.898)
Region-specific time trend	NO	NO	NO	YES	YES	YES	YES
Observations	24,858	24,858	24,858	24,858	24,858	24,858	24,858
(Pseudo) R-squared	0.139	0.021	0.026	0.140	0.141	0.141	0.141

Notes: Estimates with a dichotomous treatment variable indicating a departure from the trend. OLS estimates in models (1), (4), (5), (6) and (7). Logit estimates in models (2) and (3). Standard errors clustered by region. Robust t-statistics in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Table 3: The effect of influenza pandemic on education (Continuous treatment variable)

	(1) Years of schooling	(2) Probability to obtain a high school degree (Probit)	(3) Probability to obtain a university degree (Probit)	(4) Years of schooling	(5) Probability to obtain a high school degree (Probit)	(6) Probability to obtain a university degree (Probit)
Mortality rate	0.000	-0.014	-0.015			
	(0.0107)	(-1.212)	(-1.147)			
Cumulative mortality rate				-0.032**	-0.016***	-0.019***
				(-2.743)	(-2.692)	(-2.627)
Region-specific time trend	NO	NO	NO	NO	NO	NO
Observations	24,858	24,858	24,858	24,858	24,858	24,858
R-squared	0.139	0.025	0.027	0.139	0.022	0.021

Notes: Estimates with a continuous treatment variable indicating regional mortality rate in the year of birth in models (1)-(3) and the cumulative mortality rate in models (4)-(6). OLS estimates in models (1) and (4); Logit estimates in models (2), (3), (5) and (6). Standard errors clustered by region. Robust t-statistics in parentheses (z-statistics in the case of logit estimates). Significance: *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Robustness checks

	(1) Baseline specification with region-specific time trend	(2) Baseline specification with region-specific time trend and control for war deaths	(3) IV estimates	(4) IV estimates with region-specific time trend	(5) IV estimates with additional controls
	Panel A: Years of schooling (least square estimators)				
Influenza	-0.032**	-0.033**	-0.036***	-0.036**	-0.044***
	(-2.777)	(-2.865)	(-2.610)	(-2.497)	(-3.569)
Soldiers death rate		11.974***			
		(6.902)			
Bronchitis					0.080
					(0.728)
Pneumonia					0.004
					(0.292)
Region-specific time trend	YES	YES	NO	YES	YES
	Panel B: Probability to obtain a high school degree (probit estimators)				
Influenza	-0.015**	-0.015**	-0.016***	-0.015**	-0.022***
	(-2.561)	(-2.566)	(-2.624)	(-2.350)	(-3.726)
Soldiers death rate		30.200***			
		(14.48)			
Bronchitis					0.028
					(0.841)
Pneumonia					0.008
					(1.393)
Region-specific time trend	YES	YES	NO	YES	YES
	Panel C: Probability to obtain a university degree (probit estimators)				
Influenza	-0.017**	-0.017**	-0.014*	-0.012	-0.027***
	(-2.450)	(-2.452)	(-1.912)	(-1.633)	(-3.694)
Soldiers death rate		22.770***			
		(10.40)			
Bronchitis					0.076**
					(2.106)
Pneumonia					0.001
					(0.138)
Region-specific time trend	YES	YES	NO	YES	YES

Notes: Estimates with continuous treatment variables indicating the cumulative regional mortality rate for influenza, bronchitis and pneumonia. OLS estimates in models (1), (2) and (5) and IV estimates in models (3) and (4). Standard errors clustered by region Robust t-statistics in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1.

Appendix

Table A1: Definition of main variables and descriptive statistics

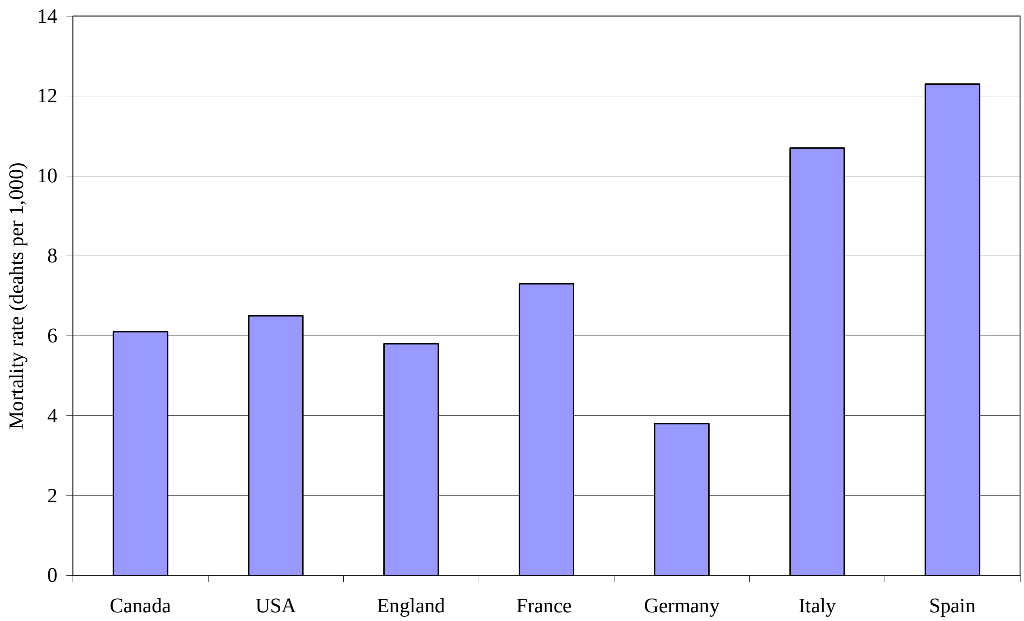
Variable	**Definition**	**Mean (st. dev.)**
Years of schooling	Years of schooling as from education attainment	6.519 (5.917)
High school degree	Probability to obtain a high school degree. Discrete variable (1 indicates the individual holds at least a high school degree)	0.222 (0.416)
University degree	Probability to obtain a university school degree. Discrete variable (1 indicates the individual holds a university degree)	0.148 (0.355)
Influenza mortality	Ratio between the death for influenza and total population in a given region	0.416 (1.210)
Cumulative mortality	4 year sum of influenza mortality rates in a given region	2.049 (2.873)

Table A2: Correlations

	Years of schooling	High school degree	University degree	Influenza mortality	Cumulative mortality
Years of schooling	1				
High school degree	0.886***	1			
University degree	0.826***	0.781***	1		
Influenza mortality	0.0066	0.0019	0.0048	1	
Cumulative mortality	-0.053***	-0.041***	-0.037***	0.422***	1

Notes: Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Figure 1: The death rate of the Spanish flu in selected countries



Source: Johnson and Mueller (2002)

Figure 2: Mortality rates for influenza in 1918 in Italian regions

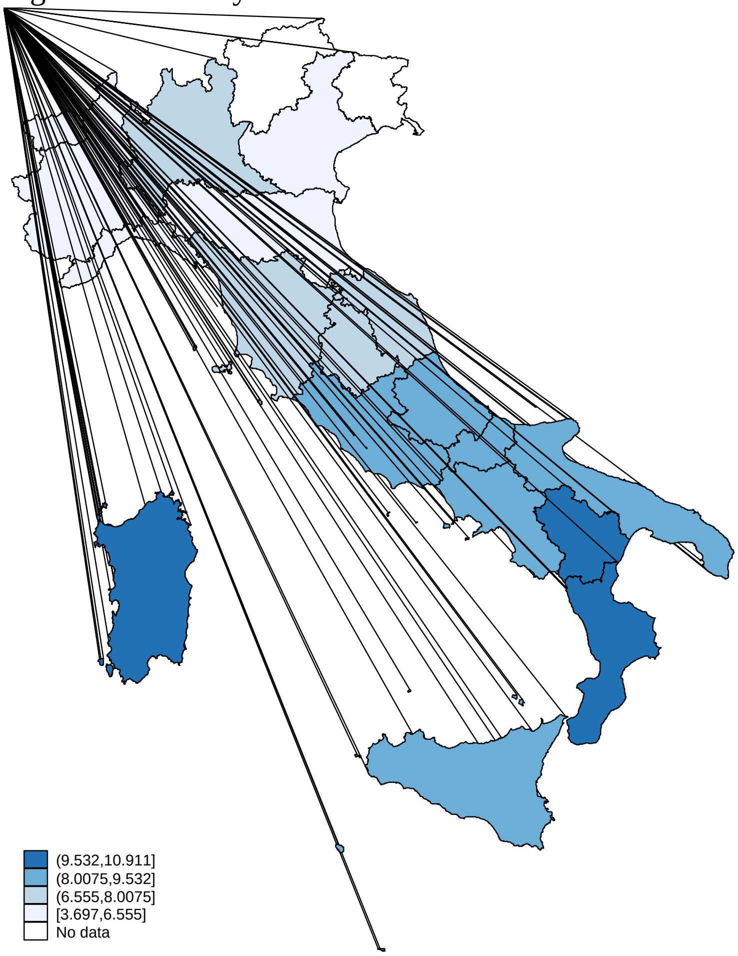
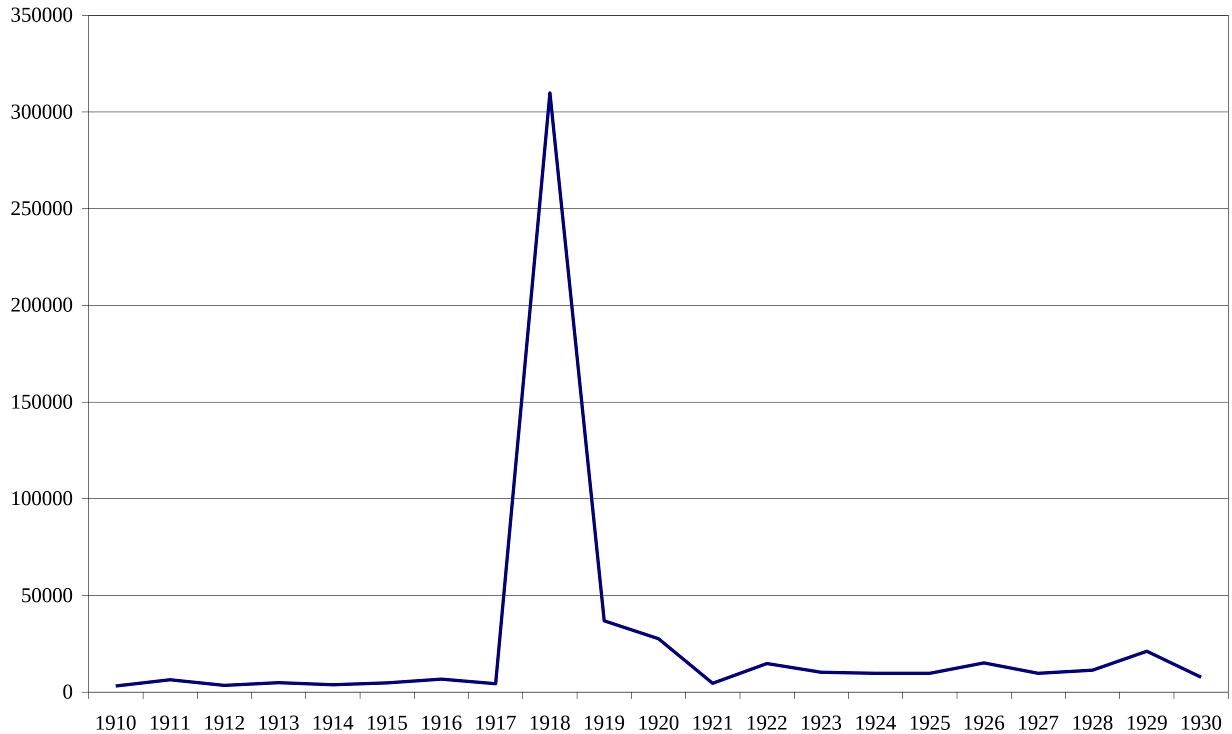


Figure 3: Deaths for influenza in Italy (1910-1930)



Source: ISTAT (1957)

Figure 4: Spanish flu and regional growth

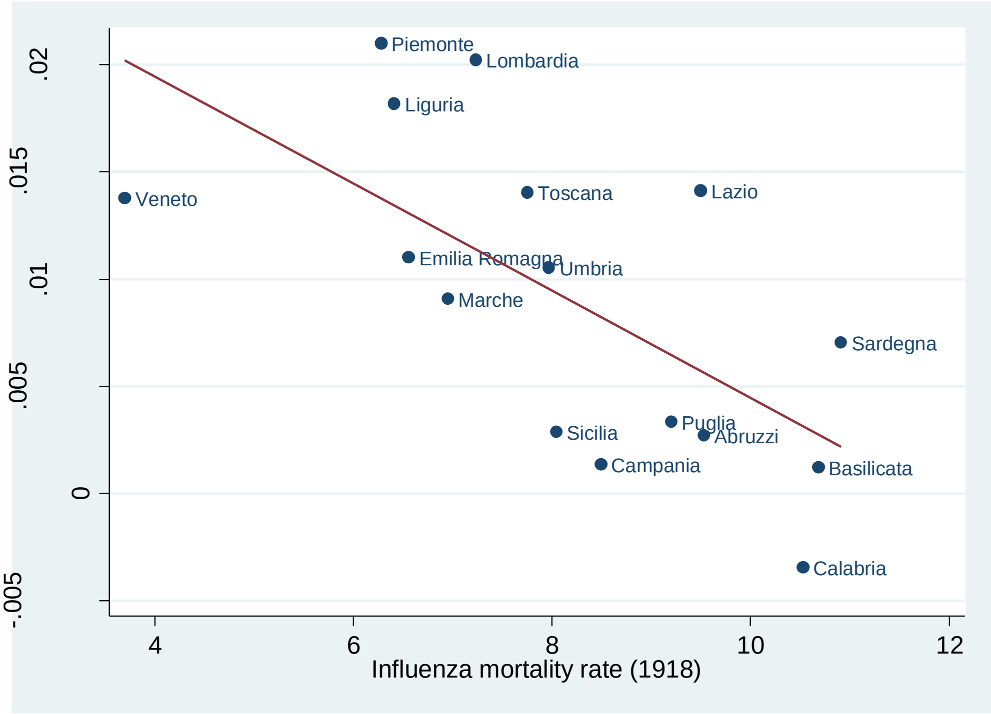


Figure 5: Average years of schooling by cohort

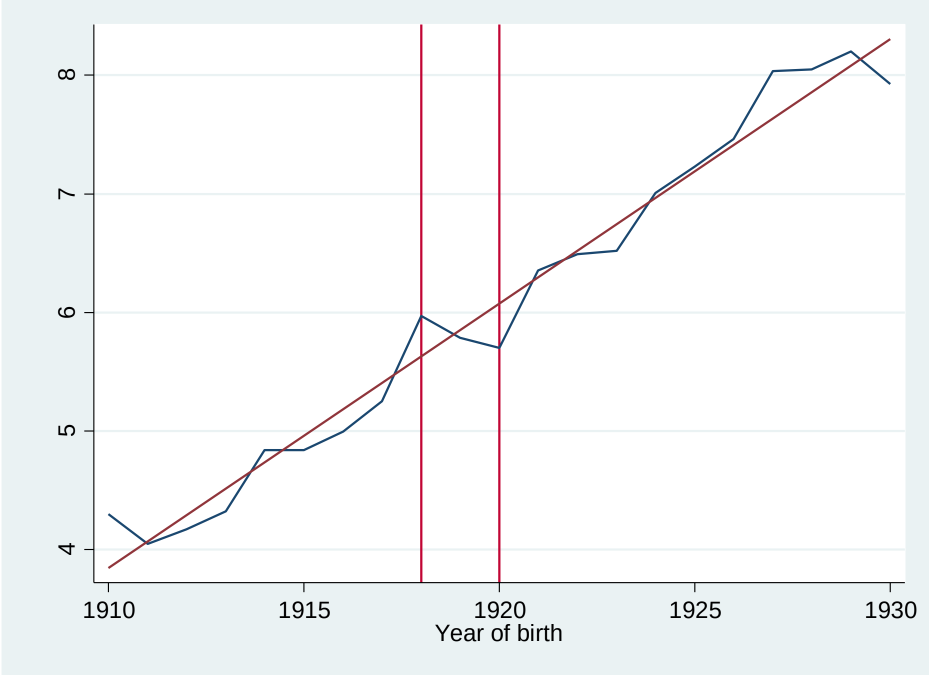


Figure 6: Average years of schooling by cohorts and geographical area

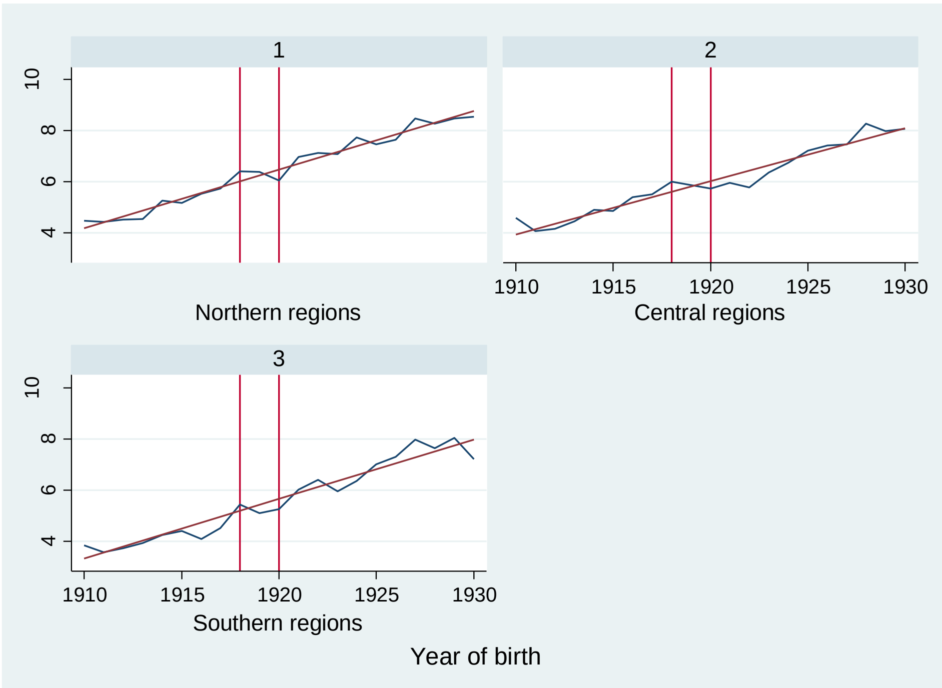


Figure 7: Average loss in years of schooling in Italian regions for an individual born in 1918

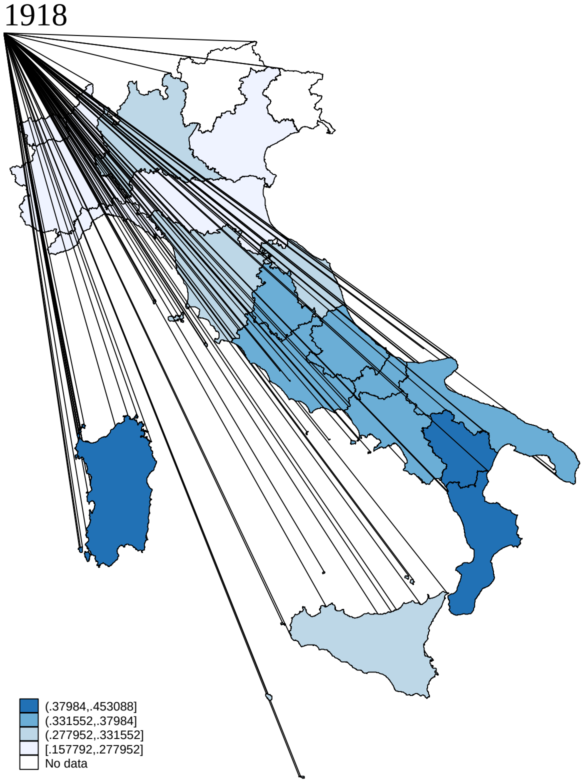


Figure 8: World War I and the Spanish flu

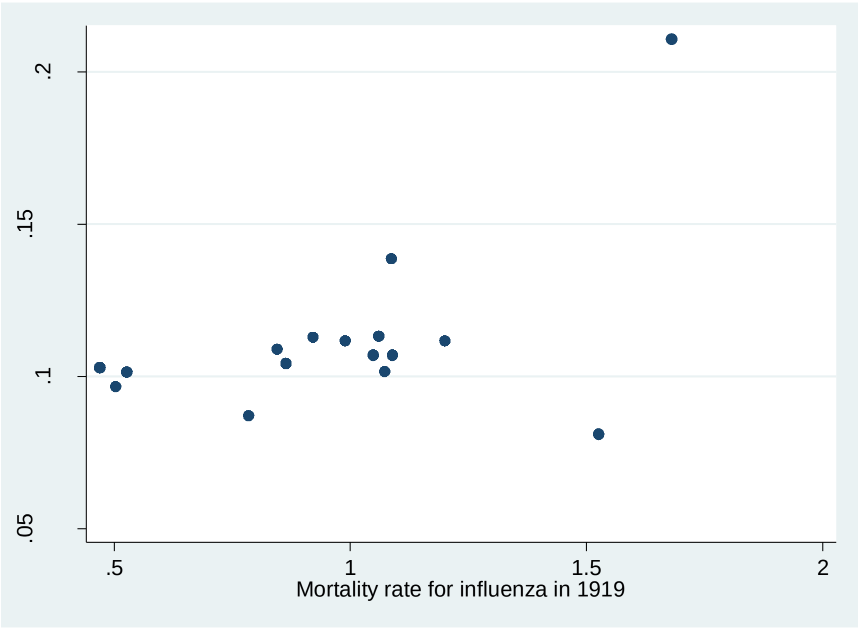


Figure 9: World War I and the Spanish flu