Spanish Flu and Human Capital Accumulation in Italian Regions*

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

PRELIMINARY DRAFT: COMMENTS ARE WELCOME

December 2011

Abstract

The impact of health on economic development is a hotly debated issue in the economics literature, with most of the scholars supporting the idea that the diffusion of diseases is detrimental for development. In this context, pandemics are an important case study given their exogenous nature which makes the identification of the impact of diseases on development more clear than in other cases such as malaria or smallpox. In this paper we focus on the 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 Martin Karlsson for useful comments. The usual disclaimer applies.

1. Introduction

Living in a comfortable and healthy place is probably one of the highest ambitions of human beings, up to the point that they may want to trade higher utility against lower wages. Quality of life is in fact often considered as an argument in the utility function and most of the research has been directed towards the definition of variables proxing the value of amenities (Blomquist, 2006).

An apparently neglected issue in the regional economics literature is the quality of health of population living in given areas. Health in fact depends on individual characteristics and life style but also on the quality of natural and social environment in which individuals live. In other words, features of places may shape human health and this, in turn, may have significant economic impacts. Healthy individuals are in general more productive and hence healthier places are expected to exhibit higher productivity levels, *ceteris paribus*.

It is not an easy task to estimate the effect of environmental quality on human health since causality is often difficult to find. Among the most recent works, Chay and Greenstone (2003) and Moretti and Neidell (2011) have provided convincing evidence on the adverse impact of atmospheric pollution on mortality and respiratory diseases by using spatial variations in pollution intensity to find causality[1].

To overcome problems of weak causality, recent literature has focused on health shocks as a source of exogenous variation in health status. Bleakley (2010) and Percoco (2011) have studied episodes of malaria eradication in the Americas and in Italy and found that cohorts born right after the eradication in areas with high malaria morbidity increased educational

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

attainments.

Whilst malaria eradication can be thought to be a positive health shock, another strand of literature has focused on pandemics as shocks affecting negatively individuals health. On this point, Almond (2006) has proposed an interesting analysis of the impact of the 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 pandemic flu of 1918-1920 has been 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 in which an individual was born during the pandemic is of considerable importance for her long run educational outcomes. Our approach hence assigns an explicit role to space in the economics of the influenza in terms of quality of the environment in which an individual was born.

In particular, we assume that exposure to the pandemic *in utero* or during childhood has reduced cognitive capabilities of children, leading to a decrease into the ability to accumulate human capital in the long run. In the case of Italy, we will show that the impact of the disease has been heterogeneous across the space; that is children born during the pandemic influenza had less schooling than the following and preceding cohorts and that the magnitude of such effects depends on a measure of child exposure during the first four years of life as well as on the region of birth. We have found that the exposure to the disease has 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 an considerable heterogeneity into 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 the Spanish flu, however, it adds to the 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 pandemic flu in Italy, the most heavily bitten country in the West after Spain.

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

2. The Spanish flu pandemic and its long run impact

The effect of health on economic development and growth is lively debated in the international literature. In reviewing the literature, Weil (2007) points out how this large body of research has found a positive effect of health, as measured by life expectancy increase, on growth. The magnitude of the elasticity lies in the interval 0.06-0.57, depending on other covariates and on the sample used (Bloom et al., 2004). A particular role in such framework is played by epidemics and pandemics which often change only marginally life expectancy in the long run but have nevertheless important economic effects.

Pandemics have accompanied the pattern of human evolution from ancient time to modern civilization. Some of those pandemics have dramatically changed the fortune of cities and nations. Historical chronicles indicate in the Black Death of the XIV century and in the Justinian plague of the sixth century the two most deadly diseases in human history. However, Langford (2002), on the basis of a literature review, argues that the Spanish flu has been probably more dramatic (in terms of loss of human lives) than the Justinian plague, hence ranking second for mortality.

The 1918-1920 influenza pandemic has been immediately referred to as "Spanish flu" as it seems that it firstly appeared in Spain in 1918 and rapidly spread all around the world probably through armies employed in World War I (Crosby, 1989). Soldiers living in poor sanitary conditions were fertile soil for the spatial diffusion of the flu, especially during periods of leave from the war front they used to come back to home, hence they have been a privileged vector of the pandemic. In addition, the end of the war with subsequent mass returns of soldiers to home was probably one of the channels of the fast geographical diffusion of the influenza (Tognotti, 2002).

The final death toll of the flu is still disputed; however, recent estimates indicate in 50 million deaths the cost of the pandemic in terms of human lives. The geography of the death toll is nevertheless extremely jeopardized since mortality varied significantly among countries (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. In Italy the pandemic lasted 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 across areas.

Interestingly enough figure 2 points to a sudden increase in the number of deaths for influenza and a similar decrease to pre-1918 levels after 1920.

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

The incidence of mortality differed significantly across age cohorts with a peak in the age 20-40 (Langford, 2002). 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 patter is of great importance for understanding the impact of the disease. There is, in fact, a large literature recently summarized by Acemoglu and Johnson (2007), pointing at a positive economic impact of mortal diseases and pandemics through a contraction of labor supply and a consequently increase in real wages and improvement in living conditions. With specific reference to the Spanish flu, Brainerd and Siegler (2002) and Garret (2009) have found evidence corroborating this view for US states, the latter finding a 2-3% increase in average wage due to the influenza mortality.

This strand of literature hypothesizes that the flu (or in general mortal diseases) has an impact on wage, through mortality, as it generates a contraction in labor supply. However, mortality and poor health conditions may also have a temporary or permanent impact in the rate of accumulation of human capital. In particular, health may influence the decision to invest in many ways, among the most important being a change in the rate of return of human capital investment through a change in life expectancy and through a change in cognitive abilities of individuals (children). In other words, an health shock may affect both labor supply and total factor productivity. In figure 3 we have plotted regional growth rates in Italy over the years 1920-1930[2] against the influenza mortality rate at its peak and found a strikingly negative correlation, supporting (although in a very simple way) our view of a health shock affecting 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 of human capital investment from a change in cognitive abilities (Percoco, 2011), whereas in the case of a natural experiment as the Spanish flu, it is possible to identify the impact on cognitive abilities because the change in mortality is only transitory and does not affect significantly life expectancy and hence the decision to invest in education.

The theoretical rationale for considering the impact of the Spanish flu on cognitive abilities and ultimately on human capital accumulation relies on the so-called “fetal and infant origins”, that is the hypothesis for which the quality of the environment in which mothers

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

and children live during pregnancy or during early childhood is crucial in explaining outcomes later in life (Almond and Currie, 2011b).

The first evidence of the relevance 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 so far and has argued 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 absorbed by human beings more easily 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 has been a shock in 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 because of the high mortality rate among children. On this point, Somogyi (1967) has estimated 21.050 deaths of children over the period 1901-1915 and 55.253 over the years 1916-1920.

By relying on the "fetal and infant origins" hypothesis, Almond (2006) has found that the epidemics of Spanish flu has had a dramatic impact in the lives of children born in 1919, along several dimensions of human life such as income, education, 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 finding a significant and negative effect of the influenza on educational or socio-economic outcomes[3].

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

In this paper we will 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, to estimate the impact of the Spanish flu we rely on the "fetal and infant origins" hypothesis and hypothesize that children exposed to the influenza exhibit lower educational outcomes. To this end, we will 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

[3] A related paper on a case of severe influenza in the UK in the1950s is Kelly (2011).

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.

Before discussing our empirical approach, let us consider figure 4 in which we have plotted the average years of schooling by cohorts. It seems that our prior is qualitatively verified since in correspondence with the pandemic, i.e. during the years 1918-1920, a decrease in years of schooling is observed. Furthermore, figure 5 shows that this drop was in common all across the three macro-areas of Italy (North, Center and South). It should be also noted that figures 4 and 5 show that a linear trend across cohorts seems to fit remarkably well the pattern of education across cohorts.

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

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

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.

As for our treatment variable we have several candidates. First, the pandemic 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 it is evident from figure 2, the pandemic flu lasted for three years in Italy, with a peak in 1918. For this reason, in our analysis we have made use 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 has almost not had exposure to the Spanish flu since 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 that is the sum of regional mortality rates for influenza over five years, that is, for an individual born in year *c* in region *k* it is defined as:

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

Measure in equation (3) indicates that the impact of the pandemic flu on individual i is given by the sum of mortality rates during the first 4 years of life. The rationale for such variable for disease exposure is given by the evidence that influenza can have the worst impacts in case of severe episodes during the first 4 years of life, especially through the emergence of complications Rudan et al (2008).

Equation (1) is similar in spirit with the one estimated in Almond (2006), although it differs in two important ways. First, we do not consider a quadratic and cubic time trend as it does not approximate effectively the evolution of schooling across the years 1910-1930, as shown by figure 4. Second, we introduce region-specific fixed effects in order to take into account regional disparities in the country. Although we have introduced such innovations in the baseline specification, in what follows we will test the robustness of our results to changes in model specification.

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

Data on mortality are from ISTAT (1957) in which regional time series are reported. 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 through interpolation.

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 probability to obtain a high school or a university degree. Estimates of logit models are not satisfactory in terms of parameters significance indicating that the Spanish flu has not had a significant impact on higher education.

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

In model (4) we consider the year of 1918 (the peak of the pandemic) as a treatment variable and could not find any statistically significant effect, as also in the case of 1920, whereas 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 has been heterogeneous across regions, so that the mildly satisfactory results reported in table 2 may hide significant differences across the space. To test this hypothesis, models (1)-(3) in table 3 estimates 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 specification reports a significant effect of pandemic influenza 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 to obtain a high school 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. By combining results in table 3 with data on total exposure in Italian regions, we have that for an individual born in 1918, the impact of the Spanish flu in terms of years of schooling ranges from -0.15 in Veneto to -0.44 in Sardegna, prompting to the relevance of our initial hypothesis on location-specific impacts on health.

[Table 3]

Results presented in table 3 confirm the view that the impact of the pandemic flu has been highly significant for children from 0 to 4 years. 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 in the specification with the number of years of schooling as a dependent variable and found no significant changes in the coefficient of interest.

Spanish flu spread in the years following World War I, hence this event could be a serious confounding factor in our analysis. In order to test for this case, in model (2) we have considered region-specific time trends with regional mortality rate of soldiers and found non significant changes in the coefficient associated to the treatment variable of interest[4].

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

[Table 4]

By relying on individual-level regressions and by 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. However, omitted variables correlated with influenza_{ck} may still impose a bias to 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:

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

Where ω_{1910k} is the share of deaths for influenza in 1910 in region k and influenza_c is the total (i.e. nation-wide) number of deaths for influenza in year c.

Also in this case, as reported in table 4 in models 3 and 4, our estimates prove to be largely robust to model specification, hence confirming the negative effect of the Spanish flu on human capital.

Finally, it must be stated that our results are robust to the inclusion of mortality for other diseases such as bronchitis and pneumonia since those variables did not experience any sharp increase during the years of the pandemic, hence pointing to a correct coding of deaths under the category “influenza”.

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 as the outcome of individuals investment decision, 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 production input after a health shock.

More in particular, we have studied the case of the Spanish flu in Italy as an example of exogenous change into (unobserved) cognitive abilities of individuals. Our empirical framework builds on the "fetal and infant origins" hypothesis for which the exposure to severe diseases *in utero* or during childhood has dramatic impacts in terms of outcomes during adulthood. In addition to existing literature, we have exploited the spatial variation into mortality rates to take into account 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 have less schooling years and that the burden of the disease significantly varies 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 for human capital accumulation and also that education has, among others, long run determinants able to influence, although 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.

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.

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.

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 (Logit)	(3) Probability to obtain a university degree (Logit)	(4) Years of schooling	(5) Probability to obtain a high school degree (Logit)	(6) Probability to obtain a university degree (Logit)
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

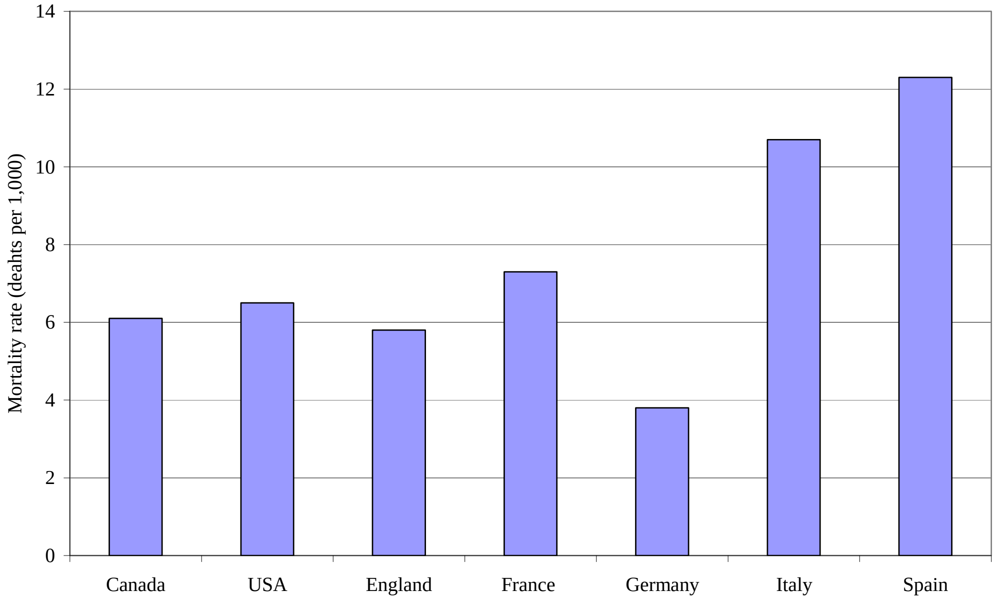
Notes: Estimates with a continuous treatment variable indicating regional mortality rate in the year of birth in models (1)-(3) and the cumulative mortality rate in models (4)-(6). OLS estimates in models (1) and (4); Logit estimates in models (2), (3), (5) and (6). Standard errors clustered by region. 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) Years of schooling with region-specific time trend	(2) Years of schooling with region-specific time trend and control for war deaths	(3) Years of schooling (IV estimates – second stage)	(4) Years of schooling (IV estimates – second stage)
Cumulative mortality	-0.033**	-0.033**	-0.037***	-0.037***
	(-2.833)	(-2.912)	(-2.716)	(-2.591)
Soldiers death rate		15.722***		
		(6.986)		
Region-specific time trend	YES	YES	YES	YES
Observations	24858	24858	24858	24858
R-squared	0.141	0.141	0.139	0.141

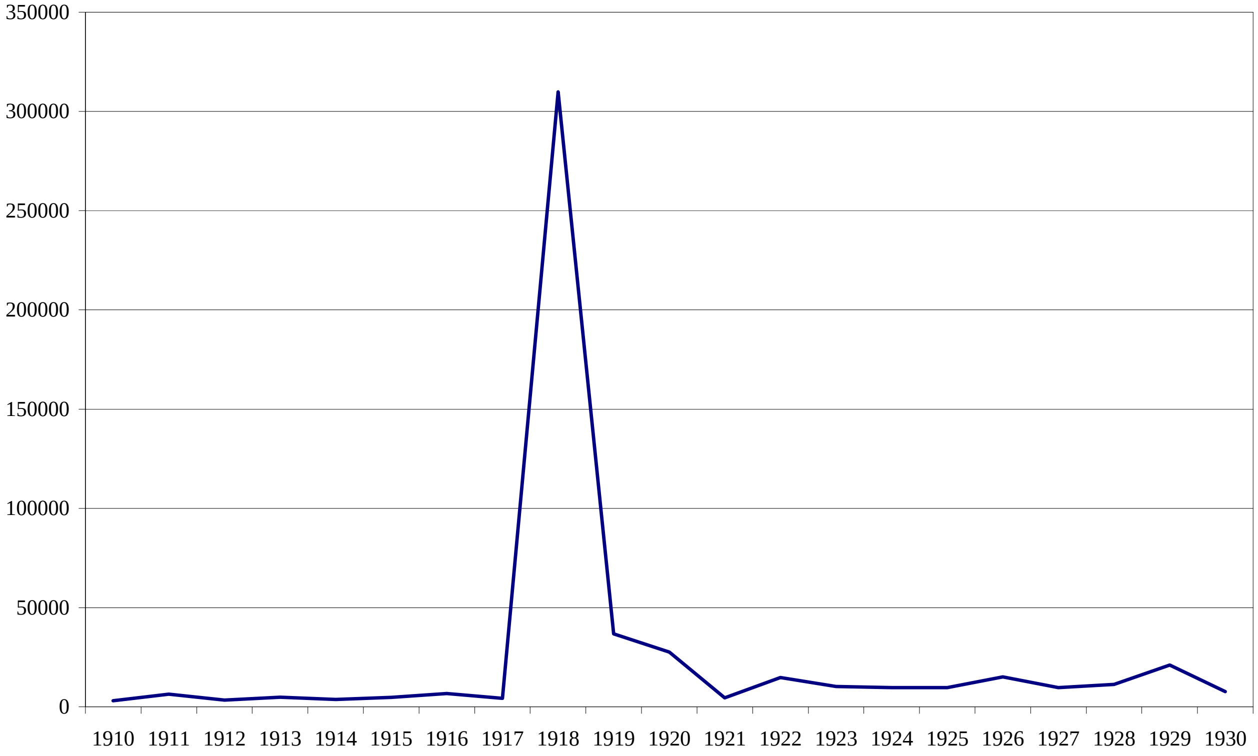
Notes: Estimates with a continuous treatment variable indicating the cumulative regional mortality rate. OLS estimates in models (1) and (2) and IV estimates in models (3) and (4). Standard errors clustered by region Robust t-statistics in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1.

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



Source: Johnson and Mueller (2002)

Figure 2: Deaths for influenza in Italy (1910-1930)



Source: ISTAT (1957)

Figure 3: Spanish flu and regional growth

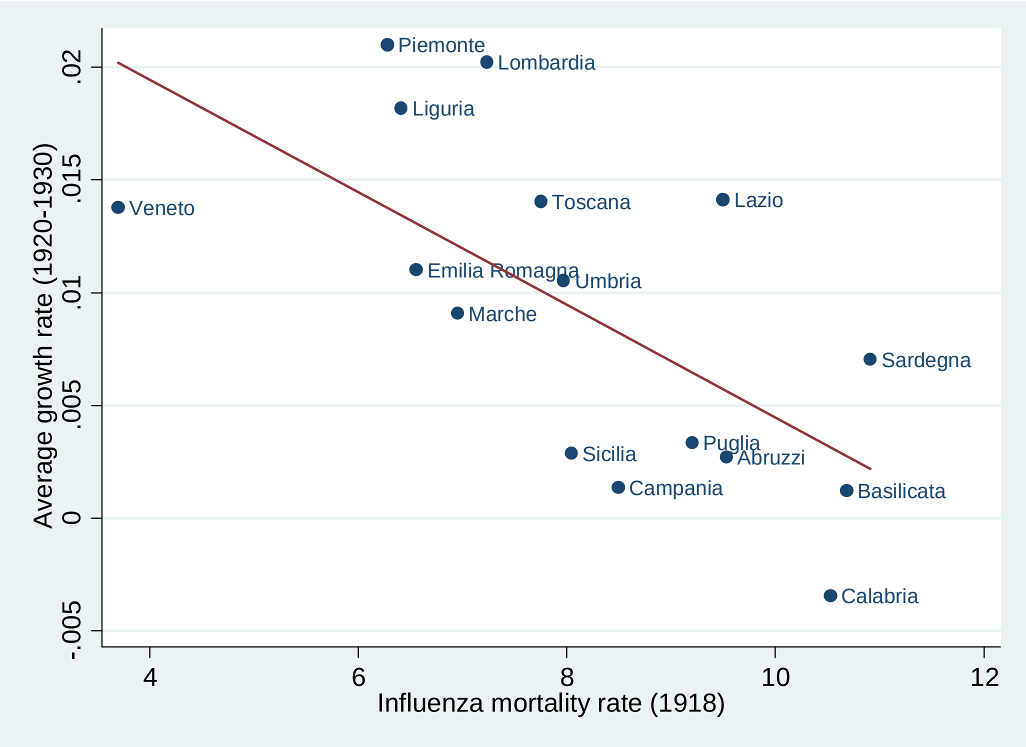


Figure 4: Average years of schooling by cohort

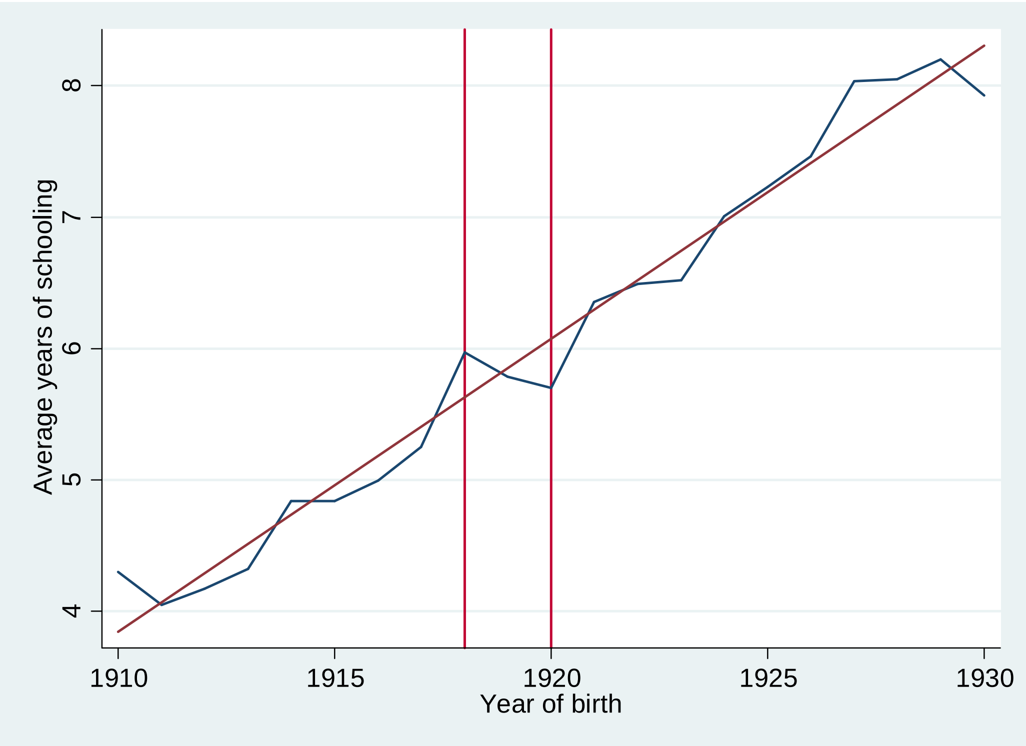


Figure 5: Average years of schooling by cohorts and geographical area