How Positive Was the Positive Check? Investment and Fertility in the Aftermath of the 1918 Influenza in India*

Dave Donaldson[1] and Daniel Keniston[2]

[1]Stanford University
[2]Yale University

PRELIMINARY AND INCOMPLETE

Abstract

The 1918 influenza epidemic struck India when the subcontinent was mired in its long-term Malthusian equilibrium of low population growth and stable per-capita consumption. Its terrible death toll left survivors with additional agricultural land, which we show they rapidly put to agricultural use with no decrease in yields. We explore the extent to which this increased per-capita wealth gave rise, over the ensuing decades, to heightened investments in both child quantity as well as child quality (as measured by the literacy rate). These responses are important because they lie at the heart of standard Malthusian (or "unified growth") models, which hold that increases in wealth at or near the Malthusian equilibrium should be allocated entirely towards additional fertility.

*This version: November 2014. We thank Sunny Lee, Andrew Mollerus, Fatima Aqeel, Ting Wang, and Naureen Rashid for excellent research assistance. We are very grateful to Prashant Bharadwaj, Latika Chaudhury, Arvinda Guntupalli, and Takashi Kurosaki for their generosity in sharing data.

1 Introduction

Households' fertility and investment choices lie behind the radical transformations that have created the global distribution of population and prosperity. Malthusian and unified growth theories provide formal frameworks to explain these choices and their effects on aggregate growth and productivity. Yet as these models make clear, the relationships between wealth, population, fertility, and growth are complex and non-monotonic, making empirical study of these phenomena challenging. In this paper we use a unique historical event—the 1918 influenza in India—to gain insights into the forces driving long run trends in economic growth and population.

The influenza arrived in India in June of 1918, peaked in November, and had largely dissipated by early 1919. During the 6 month period of its greatest extent, over 11 million people died, and millions of others were sickened. Yet the intensity of the influenza varied greatly by district, with the most severely struck districts registering mortalities of over 15% and the least severely struck showing no increases in mortality over baseline. This variation, which we show to be driven largely by exogenous factors, induces a corresponding variation in the subsequent amount of agricultural land per capita by district. Thus, as in Young (2005), the survivors became more wealthy in per capita terms. It is this exogenous decrease in population, and subsequent increase in income, that allows us to explore additional insights into Malthusian dynamics.

In a purely Malthusian equilibrium the short-term increase in per-capita wealth would lead only to a rapid increase in population and ultimately a return to the equilibrium level of per capita income. In particular, in Malthus' original theory and many of the neo-Malthusian unified growth theories inspired by it, there should be no effect on the amount invested in each child and hence no increase in human capital. This stark prediction is central to explaining why, in unified growth models, temporary increases in per-capita wealth prior to the industrial revolution did not induce future increases in productivity and (ultimately) the demographic transition at a much earlier stage in history.

These predictions differ substantially from classical models of demographic and fertility choice such as Becker and Lewis (1973). In this class of model parents value both quantity and the 'quality' (or human capital) of their children, although if the elasticity of (say) quantity with respect to income is substantially larger than that of quality then the 'observed' elasticity with respect to quality may appear to be zero or even negative. In this view there is no discrete change in parental preferences or binding constraints between the Malthusian era and the modern growth era, and the drivers of parental education changes over time must lie elsewhere, for instance in skill-biased technical change.

The 1918 influenza epidemic provides a unique opportunity to observe how an economy plausibly in a Malthusian population equilibrium react to an exogenous decrease in population and hence increase in wealth per capita. India's growth rate prior to 1921 had been low, 0.4% annually since 1891. Literacy rates were just 7.2%. Estimates of the influenza's death toll range from 10.9 million (Hill 2011) to 22.5 million (Kingsley 1951) out of a 1911 population of just over 303 million. Contemporary sources report that health officials could do nothing to control the spread of the epidemic, and that it affected all classes of society. However, the intensity of the epidemic differed greatly across regions with the most affected districts in the Central Provinces suffering greater than 10% mortality, while the least affected had no appreciable increase in deaths during 1918. This paper draws on these within-India differences in the extent to which districts were exposed to the influenza epidemic in order to study district-level responses.

2 Neo-Malthusian and Unified Growth Theories

2.1 Contemporary Malthusian views of Indian Demography and Growth

Malthus' theory was partially inspired by the example of India, and his writings in turn inspired the writings and policies of the British government in India (Caldwell 1998). Indian Census reports consistently bemoan the lack of the "preventative" check of fertility control and the dominance of the "positive check" of famine and epidemics. The 1931 Census report contains a particularly long discussion of Malthusian dynamics, stating that although "a mere superfluity of food supply is not enough, as it only enables the possessor to breed up to the subsistence level again" there are now signs that birth control may be on the horizon with the creation of a "Neo-Malthusian League with two Maharajas, [and] three High Court Judges" in Madras. The census report authors found robust support for their emphasis on the positive check from the life tables calculated by actuaries after each census took place: life expectancy was 25 years in 1891, 24 in 1901, 23 in 1911 and had fallen to 20 by 1921.

Birth rates in this period are more difficult to estimate since contemporary censuses did not record the number of births per woman. Visaria and Visaria (1983) report a tight range of estimates of the birth rate by different scholars of between 45 to 50 births per 1000 people. These rates are remarkably stable from 1881 to 1941, and are essentially the same as those reported by Rele (1987) using a similar methodology for the years 1951-1966 when the population was growing at its fastest. Even as economic growth increased after independence and (total) GDP grew substantially, birth rates appear to have not increased beyond levels observed in the 1880's (although these results are subject to any changes in relative Census coverage of adults and young children). This pattern contrasts sharply with that of England, which saw increasing fertility from the mid-17th to mid-19th century as death rates dropped, a pattern that had been influential in informing thinking on long-term growth (Galor 2005).

This stability, combined with the lack of variation in birth rates across different religious and social groups, leads Visaria and Visaria (1983) to conclude that the only check on fertility was the prohibition on widow remarriage. There is, furthermore, essentially no evidence of effective traditional birth control methods. Nevertheless, fertility rates are below the biological maximum. Despite the relatively young age of marriage in India in the early 20th century, there remained a significant margin to increase fertility through earlier marriages. In 1921, 19% of women aged 15-20 were unmarried, and 60% of girls aged 10-15 were unmarried.

The (apparent) long-term stability of birth rates sheds some light on the nature of the Malthusian dynamics at play over time. If birth rates are stable, then changes in population occur largely through changes in child survival rates, which partially depend upon parental investment in children. Thus the sharp distinction between quality and quantity is somewhat blurred: higher quality children are both more likely to survive and (if they do survive) may be stronger and better educated.

2.2 Modern Malthusian and Unified Growth

A common prediction of a broad class of unified growth models is that, in the Malthusian steady state, changes in income should have no effect on child quality or education. In Galor and Weil (2000) parental education choices respond only to expectations of future growth, which is in turn determined by the current population size. Interpreted literally, this model suggests that the short-term effects of the influenza epidemic would then be to reduce growth and hence education. Any increases in wealth that households might receive after the influenza would be expended purely on fertility and (outside the Malthusian equilibrium) on consumption. However it may be that models emphasizing endogenous productivity growth as the driver of demographic change are less well suited to the Indian context, where most productivity growth was due to the introduction of technologies developed abroad rather than domestic innovation.

A branch of unified growth theory explicitly examines the effects of large mortality events (such as an epidemic) on growth. Lagerlöf (2003) argues that a period of low mortality (due to a lack of epidemic disease) can increase population density, facilitating the spread of ideas and increasing the returns to investment in human capital. In his model, the influenza epidemic would decrease investment in human capital in both the Malthusian and post-Malthusian eras. Soares (2005) argues that it is not epidemic mortality but rather baseline life expectancy and infant mortality that affects fertility and human capital choices. Soares considers a "Malthusian" state with minimal investment in education or technology, in which moderate increases in income would once again lead to no effect on human capital and an increase in fertility. In contrast to the other models considered here, outside the Malthusian equilibrium increases in wealth caused by the influenza epidemic might lead to both greater numbers and quantity of children since parents value both (as in Becker and Lewis (1973)). The Indian time series relationships of life expectancy, infant mortality, and fertility do not, however, correspond to the prediction of Soares' model in the post-Malthusian era. While both life expectancy and infant mortality fell from 1901 to 1951, there was no significant decrease in fertility as Soares' model would suggest (Visaria and Visaria 1983; Dyson 1989).

Though the theories we cite here are just a glimpse of the large body of work on unified growth (see Galor (2005) for more details) we argue that their common result that increases in resources are devoted entirely to fertility and consumption (not to child quality) is fundamental to the view that the world once had a discrete "Malthusian" era. If this were not the case, then a tremendous increase in land per capita, such as the Black Death in Europe, would have resulted in a large increase in the next generation's human capital. This, in turn, would have increased that generation's wealth (or health, as in Lagerlöf (2003)) which would have been further transmitted to their children, potentially implying the industrial revolution would have occurred in the 14th rather than than the 18th. Since this is clearly counterfactual, growth theorists have formulated models in which positive shocks have no child quality effects at low levels of income.

3 India and the 1918 Influenza Pandemic

The 1918 influenza epidemic caused approximately 50 million deaths worldwide, making it (after HIV) the second most deadly pandemic of the modern era. Among those countries affected by the influenza India stands out because of its extraordinarily high death rate of 11 to 14 million people, and because the government was at least attempting to collect comprehensive data on vital statistics. In strong contrast with the HIV epidemic, the influenza struck very quickly and all attempts to control it in India were acknowledged as failures. Finally, we may hope that medical technology has advanced to the point that the 1918 influenza was the last pandemic of its scale; it thus represents a unique opportunity to study a massive and exogenous population shift.

The patterns of mortality caused by the influenza differed significantly from baseline deaths and the patterns of deaths caused by most epidemics. Under normal circumstances the highest death rates were among infants and the elderly. In contrast, the influenza primarily killed adults between the ages 20 and 45, and particularly affected women. Many researchers now conclude that this unusual pattern is due to the disease inducing a cytokine storm—a massive and deadly overreaction of the body's immune system—that may have been the cause of death for many influenza victims (Kobasa et al. 2007). Prime age adults, having stronger immune systems, would then be most vulnerable.

However, the inter-group and inter-regional patterns of mortality in India suggest that weaker populations may have experienced higher mortality. Mills (1986) cites data showing that in Bombay City mortality among low caste (and presumable poorer) Hindus was 3 times that of other Hindus, and more than 8 times that of Europeans. Similarly, Appendix Figure 8 shows that mortality in 1917 is highly correlated with mortality in 1918, suggesting that areas in which negative shocks had recently struck were most vulnerable to the epidemic.

Several new studies have re-estimated the impact of the 1918 influenza both in India and elsewhere. Chandra, Kuljanin, and Wray (2012) and Hill (2011) re-examine the overall mortality from the influenza and find it to be between 11 and 13.88 million, somewhat lower than influential earlier estimates. Several papers have examined the effects of the influenza on those in utero when their mothers were sickened by it (Almond 2006; Lin and Liu 2014). Most closely related to this study are the three works that examine the subsequent economic outcomes of areas differentially affected by the influenza. In the context of the United States, Brainerd and Siegler (2003) find that states with higher mortality in 1918 grew faster during the 1920s, and Garrett (2009) finds a positive effect on wages. In contrast, Karlsson, Nilsson, and Pichler (2014) see no effect on wages in Sweden, and a decrease in capital earnings and an increase in poorhouse rates. Finally, we note that the scale of mortality studied in the United States and Sweden is an order of magnitude less than in India: in the most affected areas of these countries mortality was around 1%–in India it was greater than 10%.

Figure 2 shows the geographical incidence of the influenza across British India, proxied by number of deaths from September 1918-January 1919 as a fraction of the 1911 population. Dotted areas on the map denote the many Princely States where vital statistics data is unfortunately not available. There is substantial variation across the British-ruled parts of the country: eastern and central India suffered worst from the epidemic, with many districts experiencing greater than 10% mortality. Northern India was also heavily affected, although death rates here seem more variable. In contrast, however, the southern and eastern portions of the country show relatively low mortality–roughly comparable to normal seasonal rates.

There is substantial uncertainty over why certain areas were more affected by the influenza than others. A natural hypothesis might be that districts with higher population density might have greater death rates, since transmission of the influenza would be easier (Chandra et al. 2013). However, this does not appear to be true in the Indian data. Figure 9 shows that that the non-parametric relationship between deaths and population density is actually negative. These results are partially driven by the fact that Bengal, one of the most densely populated provinces, was least affected by the influenza epidemic. Indeed, regressions of death rate on density show no relationship once province fixed effects are included.[1]

Mills (1986) cites earlier sources arguing for a role of the diurnal temperature variation, and finds suggestive evidence for this in the Bombay Presidency. More recently, laboratory research on guinea pigs has determined that low relative or absolute humidity and low temperatures are conducive to the spread of influenza (Lowen et al. 2007; Shaman and Kohn 2009).[2] We present evidence below that the spread of influenza across India districts was consistent with these laboratory results.

4 Data

Our empirical analysis is based upon a unique dataset covering Indian economic and demographic outcomes from 1891 to 1931. Our basic units of analysis are the districts of British India—a set of just over 200 geographical units that constitute the smallest area for which consistent data are available. The boundaries of many districts changed substantially over the 40-year period of our analysis. Whenever possible we re-combine district information to generate a consistent panel based on the 1891 district boundaries. However, when it is not possible to perfectly reconstitute the 1891 districts (for example when a new district is created from parts of several others) we treat the new districts as independent of the original district from which they were assembled or extracted. Thus our data do not comprise a balanced panel in that some districts appear and disappear partway through the time period in consideration. However it is balanced in the sense that the geographical area under consideration remains constant throughout, despite being divided differently in different periods.

[1] Chandra, Kassens-Noor, Kuljanin, and Vertalka (2013) find that more densely populated districts grew more slowly after 1918, and interpret this as the effect of the influenza. However, it is difficult to rule out that catch-up effects are not biasing their results upwards.

[2] The medical literature suggests different physical mechanisms for the humidity and temperature effects. At high humidities exhaled respiratory droplets settle too rapidly to contribute to influenza virus spread. Researchers found that at high temperatures study animals exhibited less viral shedding, although the reasons for this are less clear.

We use GIS maps of historical Indian districts to link climate data (measured at discrete weather stations) and anthropometric data (measured in the districts of post-independence India) to the 1918 districts. Unfortunately we only have detailed maps corresponding to 1891-era district boundaries. For these outcomes we link each district created post-1891 to the 1891 district with which it shares the largest proportion of land area.

We drop data from the major cities of Delhi, Madras, Bombay and Calcutta. These cities had large populations of immigrants and non-Indians, and had economic environments very different from the predominantly agrarian societies that prevailed in the rest of India.

Population, Education, and Migration We collect population, education, and migration data from the Census of India 1891, 1901, 1911, 1921, and 1931, limiting our analysis of the education and migration data to the last three rounds. The Census of India data provide a consistent and highly detailed dataset on Indian society and demographics. While measuring its accuracy is difficult, the post enumeration checks conducted in 1951 found that the census missed only 1.1% of the population. Mukerji (1982) revises these estimates and extends them to earlier years, concluding that net omission rates between 1931 and 1901 were always below 2.3%, and below 1% in 1911 and 1901.

The census data report population by age range, although these data are notoriously noisy due to age misstatement. To attenuate these problems we examine only a single population subgroup: women of reproductive age, defined as ages 15 to 40. This is both the group that suffered most from the influenza, as well as a particularly relevant population base when measuring the changes in fertility induced by the influenza.

The census collected literacy data as well, cross-tabulated by broad age groups and districts. We limit our analysis of these data to 1911 and later, when the definition of literacy was constant across provinces. Census enumerators defined literate individuals as those who could "write a letter to a friend and read the answer to it", and enumerators were tested respondents to verify their competence (Chaudhary 2010). We focus our analysis on literacy in native languages.

We also collect the Census data on birthplace as a measure of migration, again from 1911-1931. In most province-years the Census recorded the place of birth of each individual, specifying the exact district of birth if this district lay within the same province as the district of residence, and the province (if born in India) or nation of birth otherwise. In this case we can identify both the number of immigrants residing in each district, as well as the number of emigrants from each district who have moved to other districts in the same province. Unfortunately the 1931 Census frequently reports only the number of inhabitants of each district born outside the province of residence, substantially limiting our ability to track migration of individuals in response to the 1918 epidemic that took place after 1921.

Vital Statistics The provincial Sanitary Commissions collected and published extraordinarily detailed information on the vital statistics of colonial India. Each year's Sanitary Report contains data on deaths and births, as well as death data classified by cause, age, month, and religious affiliation (and frequently includes cross-tabulations of these categories). All data are reported at the district level, and occasionally at even finer administrative units.

The major concern about the vital statistics data are their accuracy. Accounts of the quality of vital statistics data from this period are frequently highly critical; for example, the 54th Annual Report of the Director of Public Health, Bengal writes "in most districts the verification of births and deaths is a mere farce" (cited in Meikle (1926)). However, contemporary analyses occasionally took a more positive tone. The 1931 Census report suggests that deaths registration in Punjab is near-complete, and the 1921 report makes a similar claim for Madras. More modern authors have also reached positive conclusions, Dyson (1989) for the province of Berar, and Hill (2011) for all of India.

Regardless of the average coverage level, the necessary condition for the unbiasedness of our results is that collection of vital statistics was not differentially affected by the influenza. A particular concern is that if the village record keepers themselves were sickened or killed, the ex-post reconstructed totals might be especially inaccurate in areas with high influenza mortality (Census of India 1921, Vol. I). We can test for deaths underreporting by examining whether causes of death that seem *a priori* unrelated to the influenza also respond to influenza mortality. Table 11 shows the differential death rates by cause in districts most affected by influenza in 1918 and 1919. Reassuringly, the three categories least likely to be affected by the influenza—injuries, accidents and animal attacks—as well as suicides show no significant decrease in 1918 and 1919 and indeed show a small increase.

The other death causes show interesting patters as well. While 'fever' displays a predictably large increase, reported deaths due to digestive and respiratory ailments both decrease significantly, and smallpox shows a large but insignificant decrease. Mills (1986) speculates that this decrease may either be due to influenza disproportionately causing deaths in those who might otherwise have soon died of other diseases, or a general tendency to mis-diagnose other diseases as influenza during the peak of the outbreak.

Wages The British administration collected detailed data on wages in India, although unlike the population and vital statistics data the wage data were collected inconsistently over time and across provinces. Until 1906 most provinces reported annual wages in very broad categories of skilled and unskilled labor. After criticism for inaccuracy, wage data switched to being collected on a quinquennial basis, and the 1911 and 1916 quinquennial wage censuses collected substantially more detailed information from all provinces at roughly the same time. After this, however, different provinces diverged in their frequency and depth of data collection, with some maintaining the quinquennial system and others producing only a few reports between 1916 and Indian independence in 1947.

We have assembled as much of these data sources as possible, with a summary reported in Appendix Table 8. Virtually all wage censuses collected data on daily wages for unskilled agricultural labor, which therefore constitutes the most consistent time series. In addition, we focus on three skilled non-agricultural occupations: carpenters, blacksmiths and masons. These are the most commonly reported occupations, and are hence the least subject to problems based on differential reporting across years. All the wage censuses take pains to explain the complexity of Indian labor markets during this period. Many laborers worked under long-term contracts for landlords, and were paid largely in kind for their work.

Anthropometrics Perhaps the most relevant dimension of investments in child quality in early 20th century India is nutrition. Since we have little contemporaneous evidence of resources expended on feeding children, we examine the heights of male adults born in the years before, during, and after the influenza. These data come from a remarkable series of anthropometric surveys conducted in India during the 1940s to early 1970s. The first two of these, Mahalanobis et al. (1949) and Majumdar et al. (1958), were state-level studies carried covering only the United Provinces (1941) and Bengal (1945).[3] These were extended by much more data collected by the All India Anthropological Survey (AIAS) in the 1960s and published in Rakshit et al. (1988) and Basu et al. (1994), which covered almost the entire country. In total we have height measurements of 45,288 men, ranging in age from 18 to 63, along with the district in which the individual resided at the time of measurement.

As with many anthropometric surveys, the selection into the sample was influenced by the purposes for which the data were originally collected. The main purpose of all these surveys was to document the variation in physical characteristics across the many different ethnic groups of India. Thus the tribal groups are heavily oversampled relative to their share of the population. Within each group-district pair, it seems that the sur-

[3] Interestingly, these studies contain some of the first applications of the Mahalanobis distance measure, used to estimate the similarities between different Indian castes and tribes based upon detailed measures of facial features.

veyors attempted to measure a representative sample of individuals, choosing roughly 50 from each group in each district. Individuals were not selected based on observable characteristics, and the authors claim that "the sample was free from any selection bias" (for a further discussion of sampling in the AIAS see Guntupalli and Baten (2006)). For the purposes of this study perfectly random sampling is not necessary—we require only that the relative selection of individuals born before and after the influenza epidemic is not different across districts that were more or less affected by it.

Agriculture The main channel through which the Influenza might have led to greater wealth of survivors is through an increase in the amount of available agricultural land per capita. The impact of the 1918 influenza on land farmed has been the subject of a well-known controversy in Development Economics regarding the implications of declines in total area cropped for theories of surplus labor (Schultz 1964; Sen 1967a; Schultz 1967; Sen 1967b). The main source for agricultural data on colonial India is the Agricultural Statistics of India (ASI) series, from which we draw the "net area cropped" variable. However, from 1921-1928 the ASI did not publish district-level data on area cropped. For this period we assemble the equivalent data from the state-level Season and Crop Reports.

The area sown is, of course, only part of the determinants of agricultural output—thus we also collect data on the measured yields of the crops grown in British India at this time. The colonial administration compiled "standard yield" data at 5 year intervals. These statistics were intended to "represent the average outturn on average soil in a year of average character" (ASI, 1931-2) and form the basis upon which all other calculations of annual yields and harvest size were calculated. Ideally these yields should have been measured by government experts carrying out crop-cutting experiments, although in some major provinces (Bombay and Madras) it appears that yield estimates were based upon the opinions of revenue officers who assessed yields for taxation purposes. We digitized these yield data from 1916-17 to 1931-32 for all districts and provinces of British

India.

Climate Data We collected climate data primarily from the Indian Monthly Weather Review (IMWR) from 1918 and 1919, a government publication listing the readings from weather stations across the country. The IMWR reports detailed data from 214 stations per month, of which 175 lie in British India. A important feature of the IMWR is that for most meteorological variables it reports both the monthly measurement, as well as the amount by which this observation differs from the normal value for that month.[4] We collect data on both the real value and departure from normal for temperature (monthly min, max and mean), humidity, and rainfall (total rainfall, number of rainy days, and heaviest rainfall). We also collect the observed value for daily temperature range, for which the historical normal is unfortunately not recorded.

WWI Mortality The influenza struck India just as WWI was coming to an end, raising the possibility that the effects of the influenza might be conflated with those of participation or casualties in the War. Indian soldiers participated extensively in the war, with 957,000 serving and 74,260 causalities. While the broad patterns of influenza and war impacts appear different—the influenza struck hardest in the Central Provinces and Bombay, while at least a third of the recruitment was from Punjab alone—it is still possible that within-province variation in the two factors might be correlated. To account for this we digitize the Commonwealth War Graves Commission (CWGC) list of Indian military personnel who died in WWI, expanding upon the work of Vanden Eynder (2011). In the majority of cases we are able to determine the soldier's district of birth, thereby giving us a measure of how strongly each part of India was affected by mortality from the War. Unfortunately recruitment data are not available for all of India, although in Punjab (where it is recorded by Leigh (1922)) recruitment and casualties are highly correlated at the district

[4]The 1918 IMWR reports the normal values are "derived from the data of the 33 years 1878-1910; in the case of some of the most recently started observatories the period is shorter, but it is never less than five years".

level.

5 Results

Our baseline specification is a difference in differences regression:

$$y_{it} = \beta_0 + \sum_{\phi \in \Phi} \beta_\phi DeathRate_{1918,i} \times \mathbb{I}\left[t = \phi\right] + \mu_i + \eta_t + \epsilon_{it} \tag{1}$$

in which i indexes districts, t years, and μ_i and η_t are district and year fixed-effects. If Φ defines the set of years in which we wish to estimate effects, then the β_ϕ coefficients represent the differential effects of greater district-level influenza mortality on outcome y_{it} in those years. The variable $DeathRate_{1918,i}$ is measured as the ratio of deaths from all causes in district i in 1918 divided by the population in that district in 1911.[5] In some specifications we allow Φ to be the set of all years between 1891 and 1931, then graph the estimated set of β_ϕ coefficients to show a year-by-year impact study of the influenza's effects.

While there is little evidence that influenza incidence was correlated with any pre-existing trends (indeed our results suggest that it was not), for additional robustness we also pursue an instrumental variable (IV) strategy using the weather conditions across Indian districts in late 1918/early 1919. This approach may also reduce attenuation bias from (classical) measurement error in the death rate information used to construct $DeathRate_{1918,i}$. The specific IVs we choose are the departures from the monthly normal temperature and humidity in the months of September 1918 through January 1919. As Figure 3 demonstrates, the Influenza mortality was almost entirely contained with in this period, and these meteorological variables best correspond to the environmental factors discussed in

[5]Many of these deaths were likely unrelated to the influenza. However our fundamental interest is in the effects of high mortality rather than influenza mortality per se, so including these other deaths should not affect the analysis unless there was another endogenous shock to mortality in 1918. We do not know of any other shock of this nature.

Section 3 that are most correlated with the spread of Influenza. Since, under this IV framework, there are multiple endogenous variables in equation (1) the first-stage regressions do not have a concise non-matrix representation. However, the single variable analogue can be represented as a first stage regression of the form:

$$DeathRate_{1918,i} \times \mathbb{I}\left[t=\phi\right] = \alpha_{\phi,0} + \sum_{m=1}^{M} \left(\gamma_{\phi,m} \Delta Temp_{i,m} + \lambda_{\phi,m} \Delta Humidity_{i,m}\right) \times \mathbb{I}\left[t=\phi\right] + \zeta_{\phi,i} \tag{2}$$

where as before ϕ indexes the future time period in which the effects of the influenza are being estimated, i indexes the district, and m indexes the months whose weather conditions are used as instruments.

Appendix Table 10 shows the relationship between weather conditions and influenza mortality in the main months of the epidemic. The first three columns use the observed levels of temperature and humidity as dependent variables, and are presented to show the main effects. The last 3 columns present results using the deviations from the historical monthly normal temperatures, which are the instruments actually employed in the estimation. Climatic variables, in particular humidity are highly predictive of influenza mortality, with an R^2 of .42 in November, the most deadly month. The departures from normal are highly significant, although as might be expected they are less powerful predictors than the absolute levels. Interestingly the effect of temperature is the opposite of what is found in laboratory studies, which may be due to the fact that the ultimate cause of death was often bacterial pneumonia, which may be positively correlated with temperature (Mills 1986). These results show that the climate deviations from normal have significant power as instruments month-by-month, and more so when used all together as in the IV specifications in equation (2).

We begin our presentation of the results with a test of the main implication of Malthusian theory: the test of the effects of the influenza on fertility. Table 1 shows the impact on the number of births reported per district from the vital statistics data. The first column examines only the log of the total number of births, and finds a negative effect. This seemingly paradoxical result is due to the fact that the influenza substantially lowered the number of women of reproductive age, who were the most heavily affected. When births are renormalized by the number of women aged 15 to 40 we find a highly significant positive effect on birth rates in the next decade. Since births per reproductive-age woman were roughly 0.2 per year in the years before the influenza, a district with an influenza death rate of 10% (the 84th percentile) would have experienced a roughly 15% increase in fertility according to our estimate in column 2, or a 19% increase according to the log specification in column 3. Columns 4 and 5 repeat the estimates using the temperature and humidity instruments. Point estimates increase substantially, with the fertility effect of 10% mortality now estimated at 22.5-26% (as compared with 15-19% in columns 2 and 3).

How long would it take for this district with 10% mortality to return to its previous population size? A simple back-of-the-envelope calculation puts these numbers into perspective relative to population numbers. Assuming that the population was stable prior to the influenza and using the 1921 population fraction of reproductive age women (19%), then it would take 18.6 years for the population to return to its pre-influenza level using the estimates in column 2, and 13.0 years using those in column 4. Thus we find strong support for the hypothesis of Malthusian re-convergence to stable (relative) populations in the medium term.

Scaling the number of deaths by the number of women of reproductive age likely substantially understates the effects on fertility because many women were widowed by the influenza. Given the traditional Hindu prohibition on widow remarriage, women who lost their husbands during the influenza were very unlikely to have had children following the epidemic.[6] This prohibition seems to have partially extended to Muslims as wells,

[6] Data on marriage status exist and we are in the process of entering it.

as the fraction of Muslim women who were widows in 1911 (14.7%) was not substantially lower than the Hindu fraction (18.7%). Mill's (1986) estimate that 18-19% of marriages were dissolved by influenza in the Bombay Presidency puts the scale of adjustment into some perspective.

Our next results examine the effects on child quality. Perhaps the most natural measure of investment in child quality during this era is that of the sorts of nutritional investments that lead to larger height in adulthood. We carry out the analysis of this outcome using a modified version of equation 1, where our unit of observation is at the individual level, rather than the district. Individuals measured in districts that do not correspond cleanly to 1891 districts are replicated in the data with each new observation weighted by the fraction of their district of measurement that corresponds to the 1891 district in which influenza mortality is measured. This approach allows us to include controls for the caste or tribe of individuals, the groups that constituted the sampling frame of the anthropological surveys from which the data are gathered.

Table 2 shows the impact of the influenza on (the log of) height. Consistent with other studies of the effects of influenza (Lin and Liu 2014) we find a negative and significant effect of the influenza on the heights of children in utero during the epidemic. However, we find no effects on heights of children born in the medium and long-term after the epidemic. Coefficients in the decade from 1921-1930 are positive and have p-values under .15 in the OLS and IV specifications in columns 1 and 4, but results in column 2 controlling for the individual's ethnic group are small and precisely measured. Figure 5 shows these same results graphically, demonstrating a fairly consistent lack of impact in the years after the influenza. Results in columns 5-7 examine, using a quantile regression, the effects on the 20th, 50th, and 80th percentiles of the height distribution, conditional on district and time fixed effects. We find some suggestive evidence that effects (if they exist) are among the lower end of the height distribution, but we cannot test these for statistical significance due to the limitations of statistical software. While measurement error (in particular, due to potentially noisy age data) may play a part in explaining the lack of impact, it is notable that the age data are sufficiently detailed to estimate a negative and significant effect of the influenza itself.

Our second measure of child quality is the education, specifically the literacy rate, of adults and children. We present these results in Table 3, with the first panel displaying the OLS specification and the second the IV specification. The most striking result is the very large and significant impact on overall, and especially male, literacy by 1931. Baseline (overall) literacy was 7.2% so the one percentage point increase in 1931 literacy implied by the IV results suggests that a 10% influenza death rate would raise literacy by 13.8%. The results derive entirely from changes in male literacy, and are of similar magnitudes in all age groups. Curiously the influenza appears to lower female literacy rates, suggesting that resources may be transferred from male to female children. However, we interpret these results with caution, first because baseline female literacy was extremely low during this period (less than 1%) so small absolute changes may appear to be large proportional changes, and second because the negative effects on female literacy already appear, albeit of a smaller magnitude, in 1921. It is possible that educated women may have suffered disproportionate mortality during the influenza, lowering literacy rates.

A potential concern underlying all the results displayed so far is that the effects of migration may be diluting (or enhancing) the impact of the influenza. Fortunately the Census collected detailed data on individuals' birthplaces in 1921, and broad data in 1931. Table 4 displays results concerning the migration impact of the influenza epidemic based on these data. The first column looks only at 1921, when district-level birthplace data are available from all provinces and the short-term effects of influenza migration are likely to be strongest. The relationship between influenza death and the number of individuals reporting birthplaces outside of the district is insignificant and slightly negative. Column 2 expands the analysis to 1931, creating a balanced panel of districts in which district-level birthplace information is reported in all census years. Again, we find no significant migration effects. Finally, column 3 looks at cross-province migration—migration events that crossed province boundaries, on which we have complete data in all periods–and still finds no statistically significant impact on migration. Thus our results confirm the low levels of internal migration in colonial India, even in response to a shock as major as the influenza.

The positive effect of population loss on agricultural output (and hence wealth) per capita is an important link in the chain of reasoning that argues for a Malthusian response to the influenza. If the agricultural area cropped decreased in response to the population loss, or if yields per acre decreased, then we should not expect a major impact. We therefore examine the impact on reported area cropped in Table 6. While the influenza does decrease farmed area during its prevalence in 1918-1919, we find no long-term impacts. Figure 6 tells the same story: a short decrease, then a rapid return to normal.[7] Thus the evidence points to a substantial increase in area cropped per capita.

The area of land farmed is only one determinant of agricultural output; post-influenza agricultural incomes could still have decreased if yields went down. To examine this outcome we estimate a pooled version of equation (1), in which each observation is at the crop-district-year. We also include controls for irrigation, the season of the crop, and other pertinent crop-specific variables (such as whether cotton is cleaned). The results in Table 7 are perhaps surprising: column 1 shows that average yields actually increased both in the very short term (1921) and in the medium term (1926). The further columns suggest that this result is largely driven by yields on rice (the most commonly reported crop in the

[7]The 1918-19 Season and Crop for the Central Provinces and Berar report remarks specifically upon the lack of impact on area cropped: "Such a mortality as that caused by the influenza epidemic of 1918-19 might well have been expected to cause at least a temporary set back in the agriculture of the province, but of this there is, as yet, little or no evidence. Nothing has impressed me more of late than the stoical indifference, with which the cultivating classes settled down, immediately after the epidemic, into their normal groove. In thousands of villages of the plateau the decrease in the number of able-bodied males was appalling, but "heirs" were at once forthcoming, often summoned from great distances, and the decrease in the area occupied for cultivation...was insignificant. As for the concomitant crop failure, severe and widespread though it was, the province took it in its stride. The decrease on the normal in the areas under wheat and juar, due mainly to scarcity of seed, may be confidently expected to disappear after a single good year." (pg. 8)

agricultural data), but also appears in non-cereals in 1926 (albeit at a smaller level). What could be driving this effect? One potential hypothesis is that yields may be increasing in farmer income, perhaps because the now wealthier farmers were less credit-constrained and could thus invest in more input use. However, the Malthusian dynamics of early 20th century India substantially complicate the question. Districts with higher yields also had higher population density, implying that the influenza (which mostly struck areas of lower density), was correlated with lower initial yields. Thus the positive effect on yields may simply be these low-yield areas catching up.[8]

We conclude our analysis with an inquiry into the impact on wages. Our analysis here parallels that used for the agricultural variables described above: we regress log wages, pooled over all occupations, on the measures of influenza intensity in equation (1) as well as indicators for the specific occupation that the wage is quoted for, urban and rural dummies, and dummies for whether the wage is paid in kind or cash.[9] Perhaps surprisingly, we find no overall significant effect of the mortality on wages, in column 1, although coefficients are positive as we would expect from a reduction in labor supply. However, the overall null effect disguises offsetting effects in the skilled an non-skilled sectors. The wage rate for unskilled agricultural workers, shown in column 2, does not change significantly, even appearing to decrease. In contrast, skilled non-agricultural wages go up when using pooled occupational wages, in column 3. And the results in columns 4 and 5 imply that his effect was true for all skilled occupations for which we have data, other than masons. While it is not surprising that skilled wages would react more than unskilled labor (perhaps due to greater labor market frictions in that sector), the degree of difference is large, suggesting that there are features of the market for unskilled agricultural labor that are conflating these results.

[8]To address this issue we are collecting several additional earlier years of yields data that will permit us to measure pre-existing trends in yields.

[9]A more representative approach would be to weight each occupation by its population share. We are currently digitizing the occupational data that will allow us to estimate this specification.

6 Conclusion

[To be completed.]

References

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

Basu, A., D. Mukherjee, P. Dutta, D. Bose, M. Basu, G. Ghosh, G. Kumar, and F. Huq (1994). *All India Anthropometric Survey: North Zone Basic Anthropometric Data, Vol. I.* Anthropological Survey of India, Madhya Pradesh, Calcutta.

Becker, G. S. and H. G. Lewis (1973). On the Interaction between the Quantity and Quality of Children. *The Journal of Political Economy*, S279–S288.

Brainerd, E. and M. V. Siegler (2003). The Economic Effects of the 1918 Influenza Epidemic. Technical report, CEPR Discussion Papers (3791).

Caldwell, J. C. (1998). Malthus and the Less Developed World: The Pivotal Role of India. *Population and Development Review 24*(4), pp. 675–696.

Chandra, S., E. Kassens-Noor, G. Kuljanin, and J. Vertalka (2013). A geographic analysis of population density thresholds in the influenza pandemic of 1918–19. *International journal of health geographics 12*(1).

Chandra, S., G. Kuljanin, and J. Wray (2012). Mortality from the Influenza Pandemic of 1918-1919: The Case of India. *Demography 49*, 857–865.

Chaudhary, L. (2010). Taxation and educational development: Evidence from British India. *Explorations in Economic History 47*(3), 279–293.

Dyson, T. (1989). The population history of Berar since 1881 and its potential wider significance. *The Indian Economic and Social History Review 26*(2), 167–201.

Galor, O. (2005). Unified Growth Theory. In P. Aghion and S. Durlauf (Eds.), *The Handbook of Economic Growth*. North Holland.

Galor, O. and D. N. Weil (2000). Population, technology, and growth: From Malthusian stagnation to the demographic transition and beyond. *American economic review*, 806–828.

Garrett, T. A. (2009). War and pestilence as labor market shocks: US manufacturing wage growth 1914–1919. *Economic Inquiry 47*(4), 711–725.

Guntupalli, A. M. and J. Baten (2006). The development and inequality of heights in North, West, and East India 1915–1944. *Explorations in Economic History 43*(4), 578–608.

Hill, K. (2011). Influenza in India 1918: excess mortality reassessed. *Genus 67*(2).

Karlsson, M., T. Nilsson, and S. Pichler (2014, March). 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.

Kingsley, D. (1951). The Population of India and Pakistan. *Princeton University*.

Kobasa, D., S. M. Jones, K. Shinya, J. C. Kash, J. Copps, H. Ebihara, Y. Hatta, J. Hyun Kim, P. Halfmann, M. Hatta, F. Feldmann, J. B. Alimonti, L. Fernando, Y. Li, M. G. Katze, H. Feldmann, and Y. Kawaoka (2007, January). Aberrant innate immune response in lethal infection of macaques with the 1918 influenza virus. *Nature 445*, 319–323.

Lagerlöf, N.-P. (2003). From Malthus to Modern Growth: Can Epidemics Explain the Three Regimes?*. *International Economic Review 44*(2), 755–777.

Leigh, M. S. (1922). *The Punjab and the War*. Superintendent, Government Printing, Punjab.

Lin, M.-J. and E. M. Liu (2014, June). Does in utero exposure to Illness matter? The 1918 influenza epidemic in Taiwan as a natural experiment. *Journal of Health Economics 37*, 152–163.

Lowen, A. C., S. Mubareka, J. Steel, and P. Palese (2007). Influenza Virus Transmission Is Dependent on Relative Humidity and Temperature. *PLoS Pathogens 3*(10), 1470–1476.

Mahalanobis, P. C., D. Majumdar, M. Yeatts, and C. R. Rao (1949). Anthropometric survey of the United Provinces, 1941: a statistical study. *Sankhyā: The Indian Journal of Statistics*, 89–324.

Majumdar, D. N., C. R. Rao, and P. Mahalanobis (1958). Bengal anthropometric survey, 1945: A statistical study. *Sankhyā: The Indian Journal of Statistics*, 201–408.

Meikle, H. (1926). Age Distribution and Rates of Mortality deduced from the Indian Census Returns of 1921 and previous enumerations. Technical report, Calcutta: Government of India Central Publication Branch.

Mills, I. (1986). The 1918-1919 Influenza Pandemic–The Indian Experience. *The Indian Economic and Social History Review 23*(1), 1–40.

Mukerji, S. (1982, Nov.). Under-Enumeration in India Censuses: Impact on Inter-Censal Population Growth: 1901-81. *Economic and Political Weekly 17*(46/47), 1870–1875.

Rakshit, H. K., R. Singh, S. Ahmad, D. Sastry, S. Ali, and R. Bhale (1988). *All India Anthropometric Survey, South Zone: Basic Data*. Anthropological Survey of India, Ministry of Human Resource Development, Department of Culture, Government of India.

Rele, J. R. (1987). Fertility Levels and Trends in India, 1951-81. *Population and Development Review 13*(3), pp. 513–530.

Schultz, T. W. (1964). *Transforming traditional agriculture*. New Haven: Yale University Press.

Schultz, T. W. (1967). Significance of India's 1918-19 Losses of Agricultural Labour-A Reply. *The Economic Journal*, 161–163.

Sen, A. K. (1967a). Surplus labour in India: A critique of Schultz's statistical test. *The Economic Journal*, 154–161.

Sen, A. K. (1967b). Surplus labour in India: A rejoinder. *The Economic Journal*, 163–165.

Shaman, J. and M. Kohn (2009, March). Absolute humidity modulates influenza survival, transmission, and seasonality. *Proceedings of the National Academy of Sciences of the United States of America 106*(9), 3243–3248.

Soares, R. R. (2005, June). Mortality Reductions, Educational Attainment, and Fertility Choice. *The American Economic Review 95*(3), 580–601.

Vanden Eynder, O. (2011). Military Service and Human Capital Accumulation: Evidence from Colonial Punjab. Technical report, Working Paper, Paris School of Economics.

Visaria, L. and P. Visaria (1983). Population (1757-1947). In D. Kumar and M. Desai (Eds.), *The Cambridge Economic History of India*, Volume 2, pp. 463–532. Cambridge University Press. Cambridge Histories Online.

Young, A. (2005). The Gift of the Dying: The Tragedy of Aids and the Welfare of Future African Generations. *The Quarterly Journal of Economics 120*(2), 423–466.

Tables

Table 1: Fertility Impacts

	(1)	(2)	(3)	(4)	(5)
	Log births	Births per woman aged 15-40	Log births per woman aged 15-40	Births per woman aged 15-40	Log births per woman aged 15-40
	OLS-FE	OLS-FE	OLS-FE	IV-FE	IV-FE
1918 death rate ×	-0.597*	-0.0727	0.128	-0.157	-0.400
years 1918-1920	(0.314)	(0.0516)	(0.310)	(0.0695	(0.446)
1918 death rate ×	-0.660	0.314***	1.909***	0.451***	2.600***
years 1921-1930	(0.431)	(0.0694)	(0.408)	(0.094)	(0.539)
Observations	7098	6923	6923	6906	6906
District FE	Yes	Yes	Yes	Yes	Yes

Instruments are the monthly deviations of mean temperature and humidity from historical averages from Sept. 1918-Jan. 1919. Standard errors in parentheses clustered at the district level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 2: Height Impacts

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS-FE	OLS-FE	IV-FE	20th %-tile	50th %-tile	80th %-tile
1918 death rate ×	-0.114**	-0.139***	-.115*	-0.0964	-0.136	-0.162
years 1918-1920	(0.0534)	(0.0484)	(0.0610)			
1918 death rate ×	0.0463	0.0249	0.0609	0.0748	0.0245	-0.0139
years 1921-1930	(0.0362)	(0.0283)	0.0429			
1918 death rate ×	0.0329	0.0151	0.0397	0.0752	0.00840	-0.0354
years 1931-1940	(0.0362)	(0.0264)	(0.0417)			
District, time FE	Yes	Yes	Yes	Yes	Yes	Yes
Group FE	No	Yes	No	No	No	No
Observations	48503	48503	48503	48503	48503	48503
R-sq	0.072	0.125	0.074			

Instruments are the monthly deviations of mean temperature and humidity from historical averages from Sept. 1918-Jan. 1919. All regressions weighted by the fraction of the 1951 or 1961 district in which the individual was measured lies in each 1891 district. Standard errors in parentheses clustered at the district level (not presently available for the quantile regression in columns 4-6). * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 3: Literacy Impacts

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Whole Population			Ages 10-15			Ages 15-20			Ages 20 and older		
	All	Male	Female	All	Male	Female	All	Male	Female	All	Male	Female
						Panel A. OLS						
1918 death rate ×	-0.00172	0.0177	-0.0268***	-0.0105	0.0281	-0.0524***	0.0660*	0.119**	-0.0159	-0.0110	0.00666	-0.0367***
year = 1921	(0.0201)	(0.0331)	(0.00920)	(0.0247)	(0.0395)	(0.0184)	(0.0365)	(0.0593)	(0.0176)	(0.0233)	(0.0413)	(0.0109)
1918 death rate ×	0.0586*	0.165***	-0.0482**	0.0920**	0.210***	-0.0457*	0.0528	0.167*	-0.0596*	0.0832**	0.230***	-0.0666***
year = 1931	(0.0315)	(0.0521)	(0.0195)	(0.0438)	(0.0701)	(0.0271)	(0.0526)	(0.0916)	(0.0322)	(0.0394)	(0.0656)	(0.0204)
						Panel B. Instrumental Variables						
1918 death rate ×	0.00136	0.0205	-0.0170	0.0180	0.0654	-0.0232	0.0817	0.156*	-0.0124	-0.0268	-0.0144	-0.0289**
year = 1921	(0.0269)	(0.0452)	(0.0126)	(0.0358)	(0.0554)	(0.0234)	(0.0499)	(0.0829)	(0.0291)	(0.0338)	(0.0611)	(0.0133)
1918 death rate ×	0.100**	0.261***	-0.0684***	0.118**	0.312***	-0.0832**	0.116	0.310**	-0.0984**	0.120**	0.315***	-0.0842***
year = 1931	(0.0430)	(0.0693)	(0.0238)	(0.0568)	(0.0901)	(0.0338)	(0.0773)	(0.129)	(0.0416)	(0.0547)	(0.0901)	(0.0251)
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	598	598	598	596	596	596	596	596	596	596	596	596
Mean of dep. var.	0.0600	0.104	0.0128	0.0601	0.0939	0.0195	0.0887	0.150	0.0240	0.0809	0.144	0.0139

Instruments are the monthly deviations of mean temperature and humidity from historical averages from Sept. 1918–Jan. 1919. Standard errors in parentheses clustered at the district level. * p<0.10, ** p<0.05, *** p<0.01.

Table 4: Migration

	(1) Different district	(2) Different district	(3) Different province
1918 death rate ×	-0.0402	0.268	-0.0190
year = 1921	(0.0707)	(0.183)	(0.0434)
1918 death rate ×		0.0232	-0.0316
year = 1931		(0.113)	(0.0487)
District FE	Yes	Yes	Yes
Observations	392	227	587
R-squared	0.042	0.090	0.004
Mean of dep. var.	0.0990	0.118	0.0378

The sample in columns 1 is limited to the years 1911 and 1921, and includes all provinces. The sample in columns 2 is restricted to only those provinces that report data on the number of individuals living in their district of birth in all three census years of 1911, 1921, and 1931. Column 3 uses all years and all provinces. Standard errors in parentheses clustered at the district level. * p<0.10, ** p<0.05, *** p<0.01.

Table 5: Effects on Log Wages

	(1) All occupations	(2) Unskilled ag. labor	(3) Skilled non-ag. labor	(4) Carpenters	(5) Blacksmiths	(6) Masons
1918 death rate ×	0.296	0.526	0.740	2.003**	1.517	1.396
years 1918-1920	(1.027)	(1.113)	(1.352)	(1.002)	(1.265)	(1.319)
1918 death rate ×	0.0885	-0.651	2.008***	2.356***	1.952*	0.140
years 1921-1930	(0.705)	(0.910)	(0.751)	(0.634)	(0.992)	(1.332)
1918 death rate ×	0.782	-0.295	1.962***	1.550**	2.145**	-0.0656
years 1931+	(0.677)	(0.890)	(0.617)	(0.756)	(0.903)	(1.365)
District, Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12484	3548	5356	1536	1062	837
R-squared	0.664	0.700	0.731	0.806	0.803	0.642

Columns 1-3 contain controls for the specific occupation for which the wage was collected. All regressions include controls for urban or rural wages, and whether the wage was paid in cash or in kind. Standard errors in parentheses clustered at the district level. * p<0.10, ** p<0.05, *** p<0.01.

Table 6: Impact on Area Cropped

	(1) Log of net area cropped	(2) Log of net area cropped
1918 death rate ×	-0.980*	-0.841*
years 1918-1920	(0.526)	(0.500)
1918 death rate ×	0.629	0.644
years 1921-1930	(0.575)	(0.577)
District FE	Yes	Yes
Controls for rainfall	No	Yes
Observations	6654	6626
R-squared	0.094	0.102

Net area cropped is equal to total area cropped net of area double cropped. Weather controls include quadratics in current and lagged rainfall, and contemporaneous values of the number of days rain has stopped since monsoon onset, an indicator for monsoon failure, and the date of first rains. Standard errors in parentheses clustered at the district level. * p<0.10, ** p<0.05, *** p<0.01.

Table 7: Impact on Yields

	(1)	(2)	(3)	(4)
	Log yields			
	All crops	Rice	Wheat	Non-cereals
1918 death rate ×	0.733**	1.163**	-0.183	0.360
year = 1921	(0.307)	(0.541)	(0.409)	(0.374)
1918 death rate ×	0.866**	1.800***	-0.127	0.687*
year = 1926	(0.436)	(0.605)	(0.425)	(0.380)
District, year FE	Yes	Yes	Yes	Yes
Crop type, season, farming technique, processing controls	Yes	Yes	Yes	Yes
Observations	7710	1329	752	3507
R-squared	0.853	0.262	0.665	0.933

Table reports results at the crop-district-year level on pooled yields data for all crops. All regressions include controls for type of crop, season of farming, whether irrigated, how crop was sown, whether output is husked, and whether cotton is American or indigenous. Standard errors in parentheses clustered at the district level. * p<0.10, ** p<0.05, *** p<0.01.

Figures

Figure 1: Deaths from Influenza

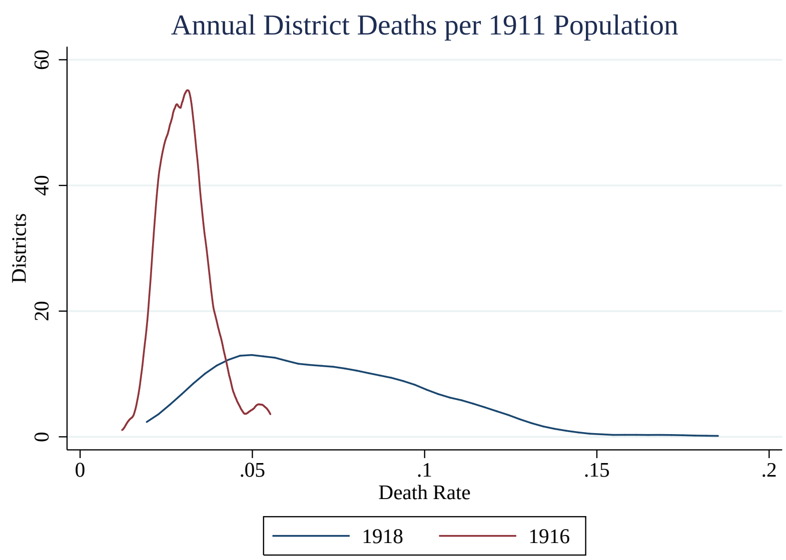


Figure 2: Geographical Incidence of Influenza Deaths

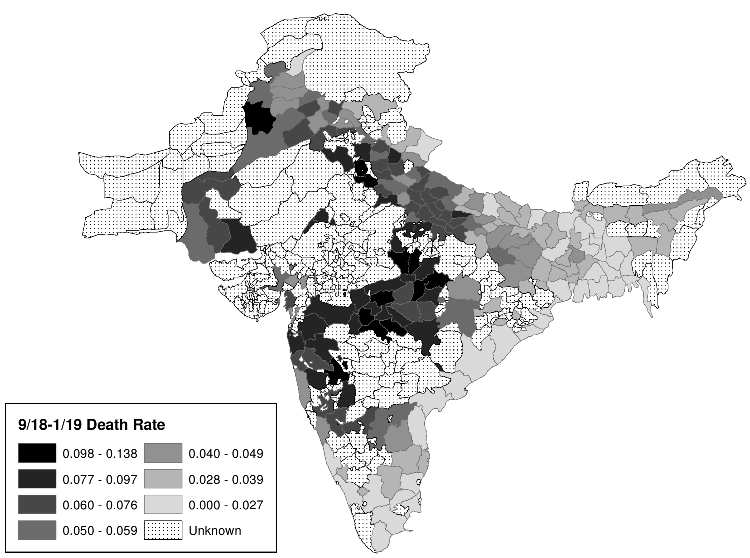


Figure 3: Timing of Influenza Deaths

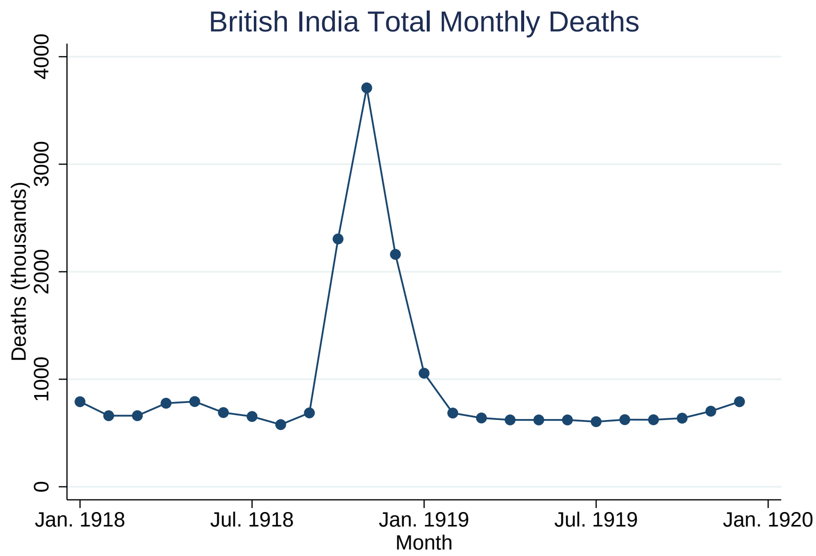


Figure 4: Deaths by Age

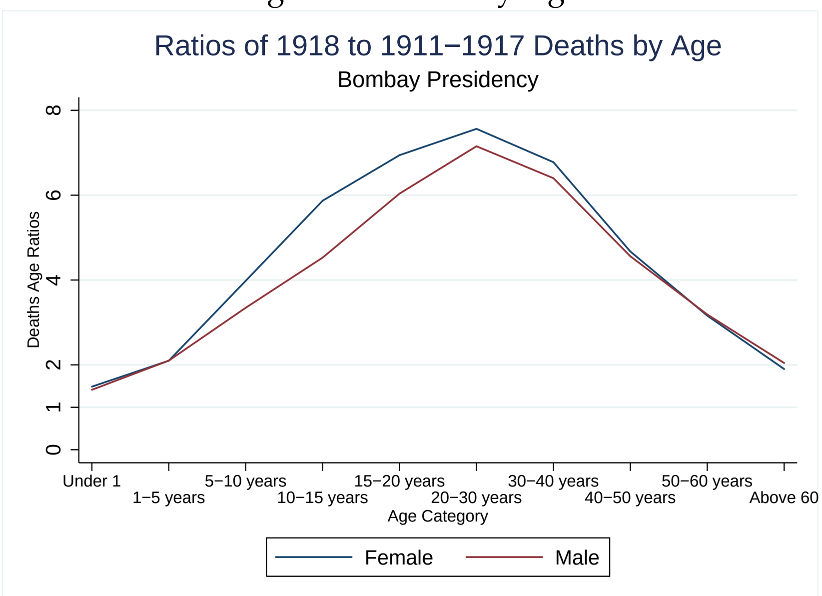


Figure 5: Heights

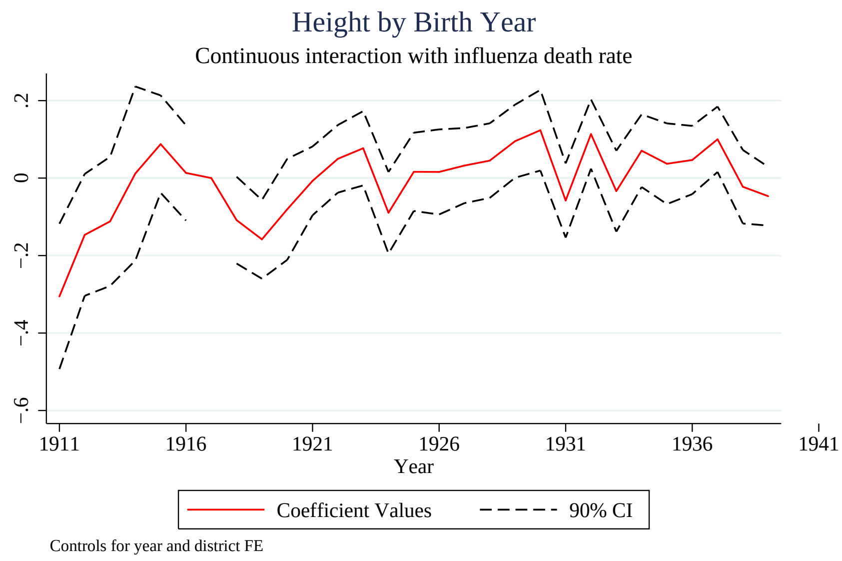
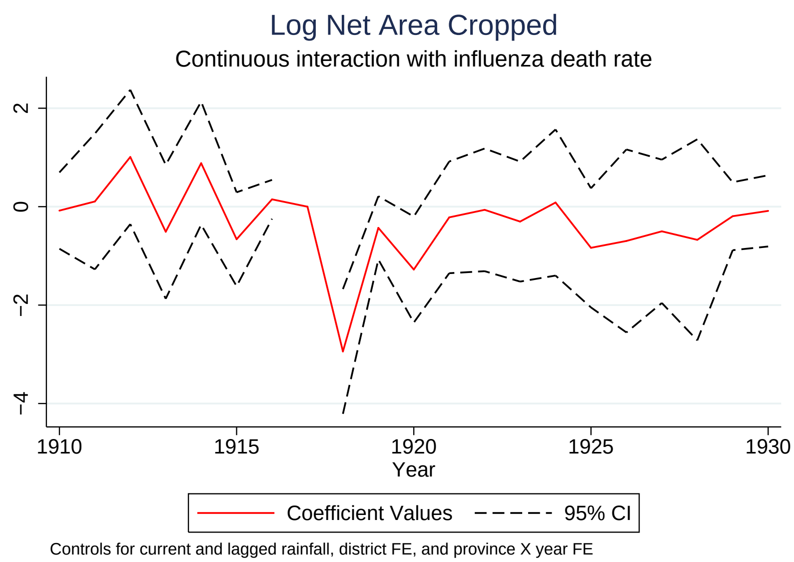


Figure 6: Impact on Agricultural Area Cropped



Appendix

Appendix Tables

Table 8: Availability of Wage Data

Province	Years of data available
Ajmer-Merwara	1891-1906, 1911, 1916,
Bengal	1891-1909, 1911, 1916, 1925
Bombay	1891-1922
Central Provinces & Berar	1891-1907, 1910-1920, 1922, 1923
Coorg	1891-1907, 1911, 1913, 1916, 1918, 1923, 1926, 1932, 1935, 1941
Madras	1891-1906, 1911, 1916
Northwestern Frontier	1891-1906, 1911, 1916, 1922, 1928
Punjab	1891-1908, 1911, 1916, 1917, 1922, 1927, 1932, 1937, 1943
United Provinces	1891-1906, 1911, 1916, 1928, 1934, 1939, 1944

Table 9: District-level Influenza Death Rates by Province

Province	Mean	SD	10th Percentile	90th Percentile
Assam	0.036	0.005	0.030	0.044
Bengal	0.029	0.009	0.019	0.043
Bombay	0.068	0.018	0.046	0.092
CP and Berar	0.087	0.024	0.057	0.114
Madras	0.030	0.014	0.018	0.055
Punjab	0.059	0.022	0.040	0.081
United Provinces	0.057	0.016	0.034	0.082

Influenza death rates are measured as the sum of district-level deaths between September 1918 and January 1919, divided by the population of the district in 1911.

Table 10: First Stage Regressions

	Actual Readings			Departure from Normal		
	Oct.	Nov.	Dec.	Oct.	Nov.	Dec.
Humidity	-0.001***	-0.001***	-0.002***	-0.003***	-0.002***	-0.000
	[-15.216]	[-12.914]	[-9.134]	[-7.509]	[-10.139]	[-0.607]
Mean Temperature	0.001***	0.001***	0.000	0.003	0.006***	0.002
	[3.863]	[3.285]	[0.379]	[1.352]	[5.182]	[1.447]
R-squared	0.553	0.512	0.322	0.298	0.317	-0.000
N	203	203	203	203	203	203

t-statistics in brackets. * p<0.10, ** p<0.05, *** p<0.01.

Table 11: Deaths by Cause

	(1) Digestive ailments	(2) Fever	(3) Physical injuries & animal attacks	(4) Respiratory ailments	(5) Smallpox	(6) Cholera
1918 death rate $\times$ year $\in \{1918, 1919\}$	-1.566*	5.968***	0.532	-4.455***	-4.305	-1.681
	(0.814)	(0.527)	(0.500)	(1.158)	(2.857)	(2.699)
District, Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1992	1993	1992	1977	1929	1847
R-squared	0.027	0.675	0.061	0.205	0.114	0.145

Standard errors in parentheses clustered at the district level. * p<0.10, ** p<0.05, *** p<0.01.

Appendix Figures

Figure 7: Humidity in November, 1918

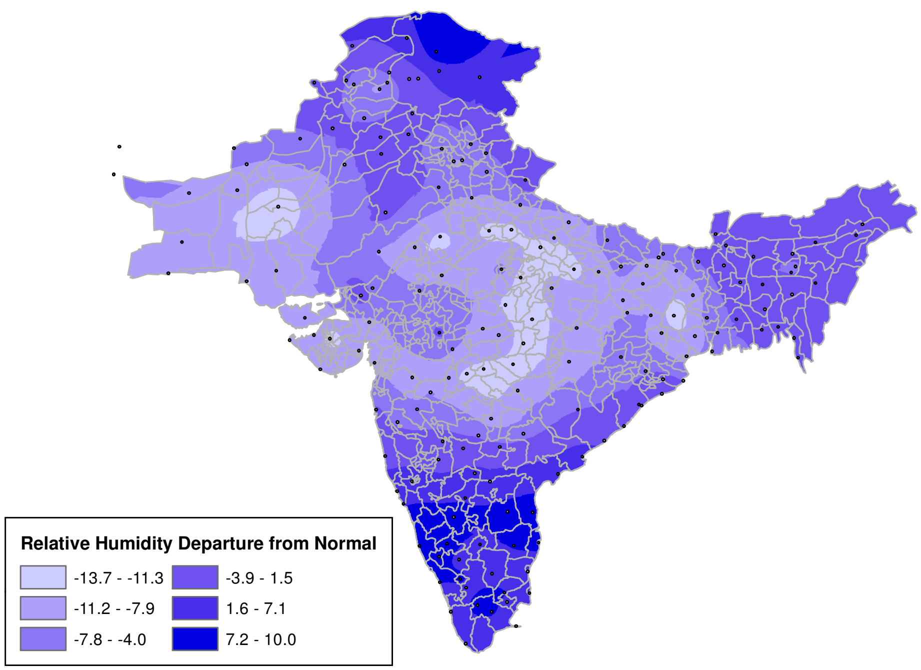


Figure 8: Pre-trends in Deaths

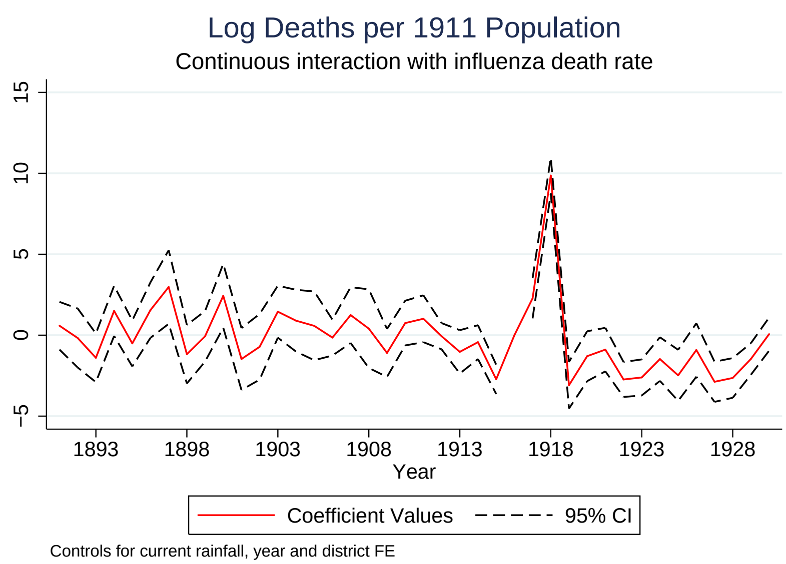


Figure 9: Relationship with Population Density