Dynamics of a Malthusian Economy: India in the Aftermath of the 1918 Influenza*

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. Consistent with most Malthusian unified growth theories, we find that individuals in heavily affected districts had more children in the aftermath of the influenza. Contrary to these theories, we also find that these children were taller and better educated. Our results suggest that the preference for child quality existed even in societies that appeared Malthusian both to contemporary observers and modern historians.

*This version: February 2016. We thank Sunny Lee, Andrew Mollerus, Fatima Aqeel, Ting Wang, and Naureen Rashid for excellent research assistance. We are very grateful to Shameel Ahmad, Prashant Bharadwaj, Guilhem Cassan, Latika Chaudhury, Arvinda Guntupalli, and Takashi Kurosaki for their generosity in sharing data.

1 Introduction

Households' fertility and human capital 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 spatial 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 per-capita income, that allows us to explore Malthusian dynamics.

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.

In our first set of results, we investigate the spatial distribution of population in the cross-

section just before the Influenza struck. In line with the predictions of a Malthusian model in long-run equilibrium, regions with favorable agro-ecological endowments (such as soil quality, groundwater availability, or terrain) have substantially higher output and populations but no higher output per capita (or proxies for human capital). We also find that agricultural wages are lower in high-productivity regions, consistent with a strong population response to agricultural income. In short, our cross-sectional findings imply that, in line with the views of contemporary observers, India on the eve of the Influenza was a plausibly Malthusian environment.

The Malthusian model makes strong predictions about how output, wages and fertility will respond to a sudden change in the population level as was brought about by the 1918 Influenza. Our next set of results compares these predictions to what actually happened in India in the aftermath of the Influenza. Broadly speaking, our findings are consistent with Malthusian mechanisms, suitably amended. Our most surprising result is that the total level of agricultural output (or equally, the amount of land cultivated, an alternative proxy for agricultural output that is better measured than yields) saw very little reduction (outside of the Influenza year itself) in the face of the enormous reduction in population due to the Influenza. Agricultural wages, too, show no statistically significant increase. This is all the more surprising given that Influenza affected working-age adults with the greatest ferocity. Because we see only relatively minor migration responses, these findings suggest that survivors' labor supply increased substantially—largely filling the void left by deceased workers lost to Influenza. Consistent with this, we see very slow fertility (and weak human capital investment, at least as proxied by literacy measures) responses to the Influenza.

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 features an algebraic model of Malthusian dynamics recognizable to modern economists, demonstrating that "a mere superfluity of food supply is not enough, as it only enables the possessor to breed up to the subsistence level again". 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 in 1947 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 adoption 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).

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 in Table 12 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 9 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. These patterns imply that an important component of our analysis will be devoted to disentangling the behavioral changes due to the influenza with changes in the population composition that it may have caused.[1]

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. 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

[1]Even if our results were entirely driven by selection, they would still be informative on the quality-quantity trade-off as long as this selection is primarily along the wealth dimension. To bias our results, influenza mortality must be correlated with innate preferences for fertility and human capital of children, holding income constant.

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 10 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.[2]

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).[3] We present evidence below that the spread of influenza across India districts was

[2]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.

[3]The medical literature suggests different physical mechanisms for the humidity and temperature effects.

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.

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.

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.

Population, Education, and Migration We collect population, education, occupation, and migration data from the Census of India 1891, 1901, 1911, 1921, and 1931, limiting our analysis of the education, occupation and migration data to the last three rounds. The Census of India data for this period 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 collected literacy data as well, cross-tabulated by broad age groups and districts. Our primary specifications 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 male literacy in native languages.

Our analysis of the impacts on the occupation distribution focuses on the broad categories of industrial and agricultural workers, whose definitions remain constant from 1911-31. Unfortunately data on labor force participation is not as consistent since the definition of a 'worker' in 1931 is more restrictive than that used in earlier censuses. While this change lowers observed labor force participation, we have no reason to believe that this would interact with influenza mortality.

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). This study contains the first collection of these data spanning a 40-year period and encompassing the entire South Asian subcontinent.

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. Despite these more optimistic evaluations, measurement error is surely a major issue in the measured influenza death rates, and suggests serious concerns due to attenuation bias. We address these concerns through an instrumental variables strategy discussed below.

In addition to classical measurement error, 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). Fortunately the data on the cause of death allows us to 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 tests whether death rates from other diseases in 1918 varied significantly with

the incidence of influenza, relative to the rest of the 1911-21 decade. Reassuringly, the category least likely to be affected by the influenza—injuries, accidents and animal attacks–shows no significant decrease in 1918 and indeed a very small increase. While 'fever' displays a predictably large increase, no other source of disease varies significantly with influenza. Reported deaths due to smallpox and cholera both decrease, and respiratory ailments increase.[4] Finally, we note a significant increase in female suicides, perhaps a consequence of the large number of widows created by the influenza.

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).[5] 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 (Bodenhorn, Guinnane, and Mroz 2013). The main purpose of all these surveys was to document the variation in physical characteristics across different ethnic groups of India. Thus the tribal groups are heavily oversampled relative to their share of the population. Within each group-district pair, sur-

[4]Mills (1986) speculates that decreases 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. The increases in respiratory deaths may be due to opportunistic infections, particularly pneumonia.

[5]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. From 1921-1928 the ASI did not publish district-level data on area cropped so we assemble the equivalent data from the state-level Season and Crop Reports.[6]

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

[6]The province of Assam did not publish state-level agricultural for the 1921-1928 period. Rather than unbalance the sample of districts during this crucial period we drop Assam from all analyses of agricultural land cropped.

1931-32 for all districts and provinces of British India.

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, 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. All wage regressions include controls for in-kind payment, and any additional contract details.

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.[7] We collect data on both

[7]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

the actual value and the normal value for temperature (monthly min, max and mean), absolute humidity, and rainfall (total rainfall, number of rainy days, and heaviest rainfall). These weather station-level values are then interpolated across India through krigging, with each district assigned the spatial mean value of the climate variable within the district boundaries.

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 level.

5 Results

Our baseline specification is a difference in differences regression:

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

in which i indexes districts, p provinces, t years, and μ_i and η_{pt} are district and province×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

years".

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 September 1918 through January 1919 divided by the population in that district in 1911.[8] 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. The province×year fixed effects absorb any correlations between the broad geographical incidence of the influenza and the secular evolution of Indian regions.

While there is little evidence that influenza incidence was correlated with any pre-existing trends (indeed our results suggest that it was not), the measurement error in mortality suggested by the vital statistics reports may cause substantial attenuation bias. We thus supplement our OLS analysis with an instrumental variables approach. The specific IV we choose is the value of absolute humidity October 1918. As Figure 3 demonstrates, the Influenza mortality was surging during this period, and this meteorological variables corresponds to the environmental factors discussed in Section 3 that were shown to facility the spread of influenza. However, since the measured October, 1918 humidity of a district is potentially correlated with other long-term trends (for example, coastal districts are both more humid and more exposed to trade shocks, we also include controls for the normal October absolute humidity of the district.

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:

$$DeathRate_{1918,i} \times \mathbb{I}\left[t = \phi\right] = \alpha_{\phi,0} + \left(\gamma_\phi Humidity_{i,1918} + \lambda_\phi Humidity_{i,Norm}\right) \times \mathbb{I}\left[t = \phi\right] + \zeta_{\phi,i} \tag{2}$$

[8]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 late 1918/early 1919. We do not know of any other shock of this nature.

where as before ϕ indexes the future time period in which the effects of the influenza are being estimated, i indexes the district, and $\{1918, Norm\}$ indicates either the realized October 1918 humidity, or the normal October humidity. In addition its advantages in alleviating attenuation bias, the IV estimates also provide a robustness check for the OLS results; while influenza incidence might be correlated with some unobserved characteristics or districts, it is difficult to see how a single month's humidity would be.

Appendix Table 10 shows the relationship between weather conditions and influenza mortality in the main months of the epidemic. The first two columns show the first stage regression of the influenza death rate on October 1918 absolute humidity both without province fixed-effects (column 1) and with them (column 2). In both cases the first stage is very strong, with F-statistics of over 70. Columns 3 and 4 introduce the additional potentially promising instruments of November and December humidity. While these have some explanatory power, they ultimately weaken the first stage; we omit them from future specifications.

Fertility 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 two main measures of fertility: the number of births (from the vital statistics data) and the ratio of children to women (from the census data). We renormalize the birth rate by the number of women aged 15 to 40 to adjust for the impact of the influenza on the number of women of reproductive age (a group that was among the most heavily affected). While OLS effects on births are small (column 1) instrumenting births with Oct. 1918 humidity reveals a much larger and significant effect (column 2). The coefficient implies that a 10% influenza death rate–roughly the 85th percentile–increases birth rates by over 18% in the years 1921-31.

A second measure of the fertility effect of the influenza comes from the ratio of children to adult women in the 1921 and 1931 Census population data. This measure is largely independent of the vital statistics data, since it was collected by a separate government agency, and has the advantage that it provides a better measure of surviving children at

the cost of some annual granularity. Reassuringly, this evidence is highly consistent with the births data. Child-women ratios rise in 1921 due to excess female influenza mortality, but remain significantly greater than pre-influenza levels for the cohort born in 1921-1931. The instrumental variable estimates in column 4 suggest that districts with a 10% greater influenza mortality had more than 25% more children under the age of 10 by 1931 relative to unaffected districts.

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 17 years for the population to return to its pre-influenza level using the estimates in column 2. However, it is not clear whether the birth rate indeed remained high for this duration after the influenza. The event study graph in Figure 5, showing the yearly plot of the regression coefficients, seems to indicate that birth rates were generally falling in the influenza-affected areas by the late 1920s.

Quality 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. We find strong evidence of an increase in heights of children born 3-13 years after the epidemic in both OLS (column

1) and IV (column 3) specifications. A primary concern is that these results may be driven by selection, as the influenza may have differentially killed shorter individuals. Two pieces of evidence suggest this is unlikely: First, we see no difference in heights of adults born prior to the influenza in districts with higher death rates. Second, controlling for whether an individual belongs to a scheduled ("untouchable") caste or tribal ethnic group (columns 2 and 4) attenuates the results but they retain significance. Unlike other studies of the effects of the influenza epidemic (Lin and Liu 2014) we very weak evidence for a negative effect of the influenza on the heights of children in utero during the epidemic.

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. Both show some evidence of selection, as the fraction of literate men over age 20 increases immediately after the influenza in 1921. However, the instrumental variables results in Panel B also show a significant increase in literacy for boys aged 10-15 in 1931. This age group is precisely where we would expect the largest impact: the set of children learning to read in the years after the influenza. If we view the 1921 coefficient as a proxy for selection, then the fact that the 1931 impact is more than twice as large suggests that the effect is due to behavioral change. However, the IV standard errors are such that we cannot reject equality between the 1921 and 1931 coefficients ($p = .136$), so we must treat the result with caution.

Our third measure of child quality is the fraction of children aged 10-15 who are married or widowed. Child marriage was common in early-20th century India, with 27.8% of individuals in this age range married. There is extensive evidence from modern research (e.g. Sekhri and Debnath (2014)) on the negative impacts of child marriage on human capital accumulation, and such views were also common in 1918 India with well known figures such as Gandhi denouncing child marriage as an impediment to schooling. Table 4 tests whether the influenza induced parents to postpone marriages, and finds strong evidence that this is the case. IV results suggest that a 10% influenza death rate causes an almost 10% decrease in the fraction

of married 10-15 year-olds, a huge effect relative to the baseline level of 27.8%. Some of this impact may be due to selection, reflected in the negative (but insignificant) impact on the child marriage rate in 1921, but the difference between 1921 and 1931 effects remains large and significant. Interestingly, results do not appear significantly differentiated by gender.

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 we can quantify these effects using detailed census data on individuals' birthplaces in 1921, and broad data in 1931. Table 5 displays these results. The first and third columns 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. We find no significant effects, although the IV coefficient is large. Columns 2 and 4 examine the fraction of inhabitants of each district from a different province. Here, at least in the IV specifications, we do find a significant migration effect: higher mortality districts have more inhabitants born in other provinces. The magnitude of the impact implies that a district with 10% mortality would have a roughly 3.5% increase in individuals from another province. While this is small in absolute terms, it represents a doubling of baseline number.

These results suggest that migration may have played a role in re-populating the districts affected by the influenza. However, several pieces of evidence suggest that migration is unlikely to be driving the main impact on child quality and quantity. First, we find the greatest results in 1931, when immigration is leveling off. The fraction of out-province immigrants increases by roughly 51% between 1921 and 1931, while the changes in fertility, height, education, and marriage age increase by much larger factors, suggesting they are not driven (solely) by migration. Second, in order for migration to be driving our observed results, it would be necessary for immigrants to have higher birthrates, be taller and better educated, and marry later than the native population. While this type of selection (which we find implausible) would require a re-framing of the Malthusian model, if individuals' choice to relocate to areas of (relatively) lower density is associated with differential fertility and

human capital investment choices, this confirms the fundamental relationships behind the Malthusian model.

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. Panel A demonstrates that while the influenza does decrease farmed area during its prevalence in 1918-1920, the longer-term impact (i.e. that in 1921-1930) is smaller and not precisely estimated. Figure 7 tells a similar story: a substantial short-run decrease, and then a rapid return to normal.[9] Similarly, Panel B shows no effect of the influenza on crop yields. These null results, combined with the large population decreases due the influenza deaths confirm that there was a substantial increase in agricultural output per capita.

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.[10] Perhaps surprisingly, we find no overall significant effect of the mortality on wages, in column 1. However, the overall null effect disguises offsetting effects in the skilled and non-skilled sectors. The wage rate for unskilled agricultural workers, shown in column 2, does not change significantly, even

[9]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)

[10]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.

appearing to decrease. In contrast, skilled non-agricultural wages go up when using pooled occupational wages, as in column 3. 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.

6 Conclusion

The impact of income shocks on the fertility and educational decisions of very poor households is both a important historical debate and a salient modern policy question. Malthusian theory and its unified growth theory successors posit that unless society has begun the growth process, for instance by increasing returns to education or by improving public health, increased income will be devoted exclusively to consumption and fertility and thus any welfare improvements quickly eroded by a larger population. In contrast, the literature on the quality-quantity trade-off suggests that increases in income translate into greater human capital and, ultimately, higher productivity and future growth.

This paper presents evidence that, even in an economic environment with no (positive) cross sectional relationship between agricultural productivity and human capital, an income shock due to the 1918 Influenza does induce increases in height, education, and decreased age of marriage.

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.

Bodenhorn, H., T. W. Guinnane, and T. A. Mroz (2013, May). Sample-selection bias in the historical heights literature. Yale University Economic Growth Center Discussion Paper No. 1023.

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.

Sekhri, S. and S. Debnath (2014). Intergenerational Consequences of Early Age Marriages of Girls: Effect on Children's Human Capital. *The Journal of Development Studies 50*(12), 1670–1686.

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)
	Log births per woman aged 15-40		Children 0-10 per woman aged 15-40	
	OLS	IV	OLS	IV
1918 death rate ×	-1.673***	-4.388***	0.577*	2.417***
years 1918-1920 or 1921 census	(0.338)	(0.800)	(0.336)	(0.579)
1918 death rate ×	0.501	1.831**	0.656**	2.596***
years 1921-1930 or 1931 census	(0.490)	(0.844)	(0.328)	(0.587)
R-squared	0.354	0.309	0.392	0.354
Observations	4123	4123	624	621
Province×period FE	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes

Data from columns 1,2 drawn from annual vital statistics on births for the 1911-1930 period, normalized by the population of women aged 15-40 reported in the previous decennial census. Data in columns 3-4 are drawn from the 1911,21 and 31 decennial census population tables. Independent variables are the interaction of the influenza mortality rate with indicators for the 1918-1920 and 1921-1930 time periods (cols. 1,2), or the 1921 and 1931 decennial censuses (cols. 3-4). Instruments are the October 1918 absolute humidity value interacted with dummy variables for time periods (columns 1,2) or subsequent census years (columns 3,4). All specifications with instrumental variables also include controls for the normal October absolute humidity of the district, interacted with the relevant time indicators. Standard errors in parentheses clustered at the district level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 2: Height Impacts

	(1) OLS	(2) OLS	(3) IV	(4) IV
1918 death rate ×	-0.000481	-0.00423	0.0498	0.0403
years 1918-1920	(0.0872)	(0.0822)	(0.171)	(0.173)
1918 death rate ×	0.122**	0.125**	0.280**	0.235**
years 1921-1930	(0.0547)	(0.0561)	(0.119)	(0.116)
1918 death rate ×	0.0804*	0.0851*	0.157	0.117
years 1931-1940	(0.0485)	(0.0483)	(0.124)	(0.114)
Province×period FE	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
Caste/Tribe FE	No	Yes	No	Yes
Observations	49124	49124	49124	49124
R-squared	0.075	0.147	0.075	0.147

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. * p<0.10, ** p<0.05, *** p<0.01.

Table 3: Male Literacy Impacts

	(1) All Ages	(2) Ages 10-15	(3) Ages 15-20	(4) Ages 20 and older
		Panel A. OLS		
1918 death rate ×	0.0918*	0.0438	0.114	0.155**
year = 1921	(0.0522)	(0.0662)	(0.102)	(0.0683)
1918 death rate ×	-0.0434	0.0458	-0.0820	-0.0180
year = 1931	(0.0677)	(0.109)	(0.122)	(0.0883)
		Panel B. Instrumental Variables		
1918 death rate ×	0.193*	0.186	0.144	0.359**
year = 1921	(0.111)	(0.137)	(0.182)	(0.157)
1918 death rate ×	0.243	0.459**	0.318	0.339
year = 1931	(0.175)	(0.227)	(0.293)	(0.234)
Province×year FE	Yes	Yes	Yes	Yes
District, year FE	Yes	Yes	Yes	Yes
Observations	621	621	621	621
Mean of dep. var.	0.103	0.0924	0.148	0.142

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: Impact on Age of Marriage

	(1)	(2)
	Fraction of 10-15 year-olds married or widowed	
	OLS-FE	IV-FE
1918 death rate ×	-0.0526	-0.195
1921 census	(0.143)	(0.201)
1918 death rate ×	-0.681**	-0.958**
1931 census	(0.272)	(0.416)
Province×period FE	Yes	Yes
District FE	Yes	Yes
Observations	621	621
R-squared	0.286	0.296
Mean of dep. var.	0.278	0.278

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

Table 5: Migration

	(1) Different district	(2) Different province	(3) Different district	(4) Different province
	OLS-FE	OLS-FE	IV-FE	IV-FE
1918 death rate ×	0.0371	0.0985	0.676	0.229*
year = 1921	(0.246)	(0.0804)	(0.431)	(0.131)
1918 death rate ×		-0.0319		0.345**
year = 1931		(0.0700)		(0.151)
District FE	Yes	Yes	Yes	Yes
Province×year FE	Yes	Yes	Yes	Yes
Observations	410	619	410	619
R-squared	0.025	0.144	0.019	0.103
Mean of dep. var.	0.103	0.0366	0.103	0.0366

The sample in columns 1 and 3 is limited to the years 1911 and 1921, Columns 2 and 4 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 6: Impacts on Agricultural Output

	(1)	(2)
Panel A: Log of net area cropped		
	OLS-FE	IV-FE
1918 death rate ×	-1.692***	-2.595***
years 1918-1920	(0.485)	(0.906)
1918 death rate ×	-0.195	0.360
years 1921-1930	(0.355)	(1.095)
Panel B: Log of crop yields		
	OLS-FE	IV-FE
1918 death rate ×	-0.322	0.478
years 1921	(0.524)	(1.691)
1918 death rate ×	-0.104	1.467
years 1926	(0.530)	(1.582)
1918 death rate ×	-0.597	-1.329
years 1931	(0.485)	(1.719)
Province×period FE	Yes	Yes
District FE	Yes	Yes
Observations	3837	3835
R-squared	0.203	0.200

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: Effects on Log Wages

	(1) All occupations	(2) Unskilled ag. labor	(3) Skilled non-ag. labor
Panel A: OLS-FE			
1918 death rate ×	-0.320	-0.948	1.161
years 1918-1920	(1.125)	(1.579)	(1.295)
1918 death rate ×	-1.255	-1.764	2.204**
years 1921-1930	(1.087)	(1.328)	(1.026)
1918 death rate ×	-0.736	-2.215	1.297
years 1931+	(0.854)	(1.401)	(0.812)
Panel B: IV-FE			
1918 death rate ×	-0.382	0.0916	1.252
years 1918-1920	(2.164)	(2.677)	(2.772)
1918 death rate ×	-2.249	-1.911	3.859**
years 1921-1930	(1.551)	(1.732)	(1.801)
1918 death rate ×	-1.390	-1.233	2.714*
years 1931+	(1.318)	(1.938)	(1.613)
Province×period FE	Yes	Yes	Yes
District, Time FE	Yes	Yes	Yes
Observations	13212	3832	5600

Columns 1 and 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.

Figures

Figure 1: Deaths from Influenza

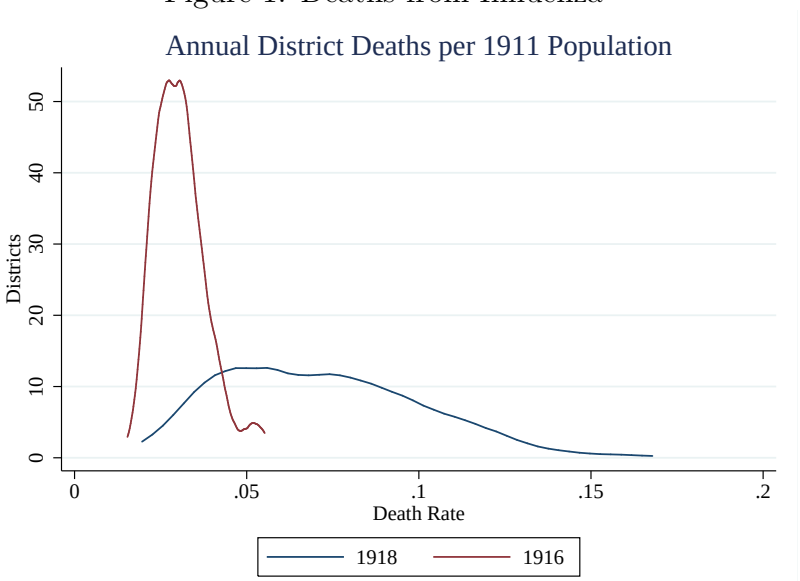


Figure 2: Geographical Incidence of Influenza Deaths

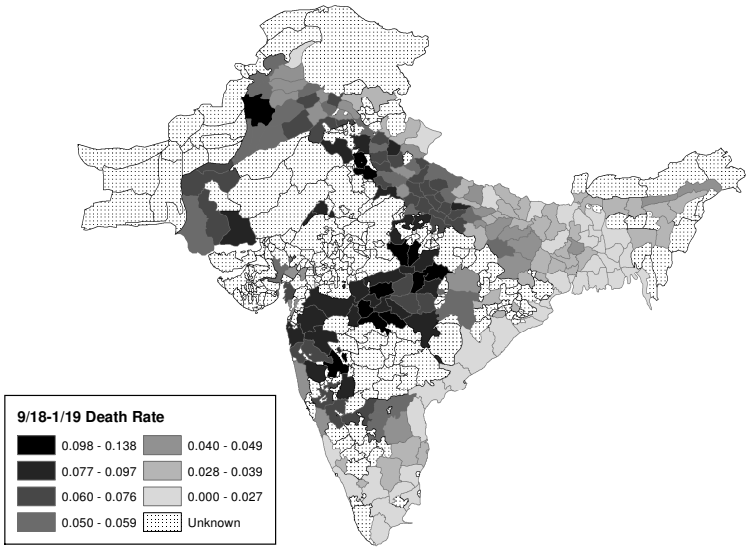


Figure 3: Timing of Influenza Deaths

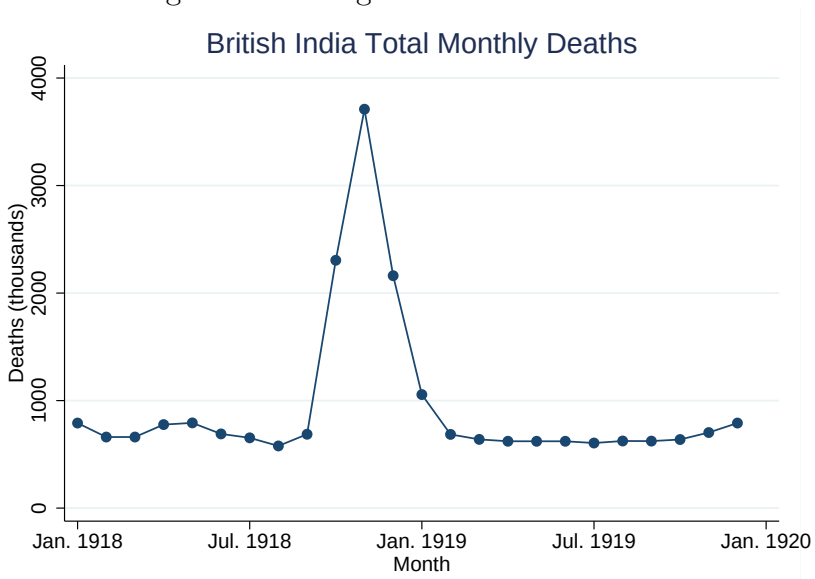


Figure 4: Deaths by Age

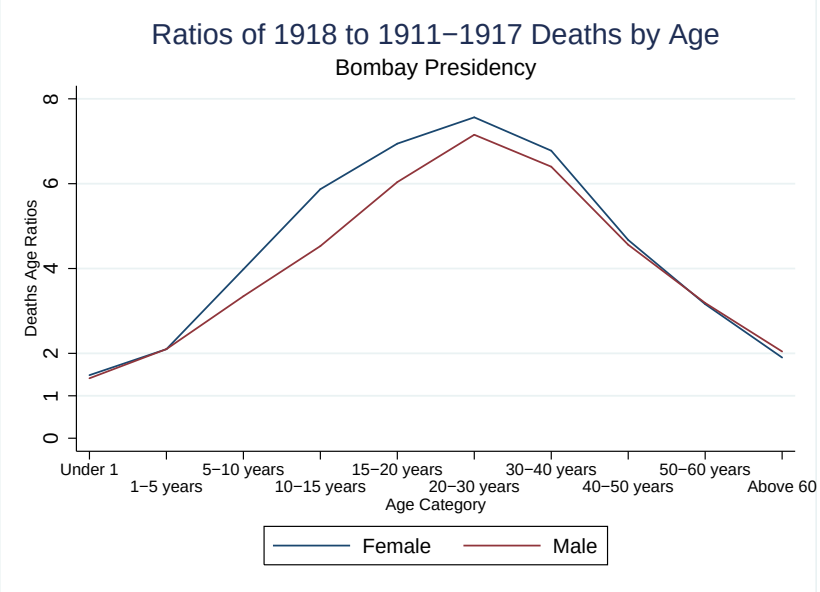


Figure 5: Log. Births per Woman Age 15-40

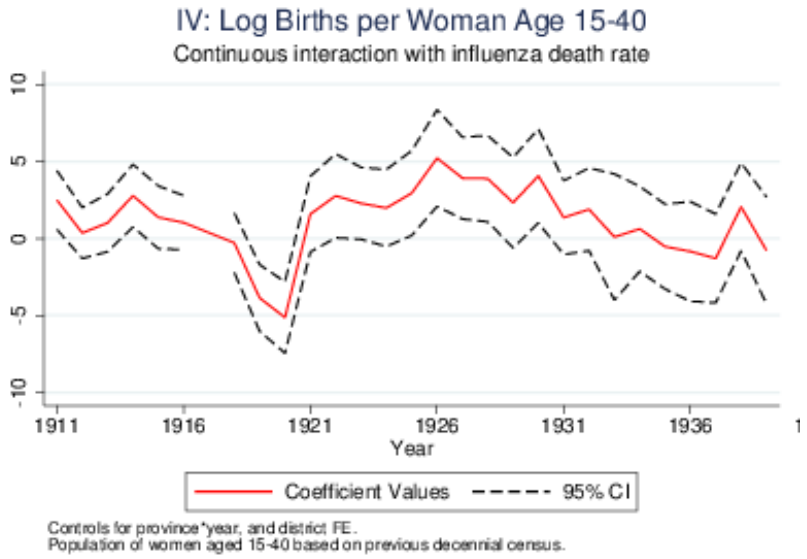


Figure 6: Heights

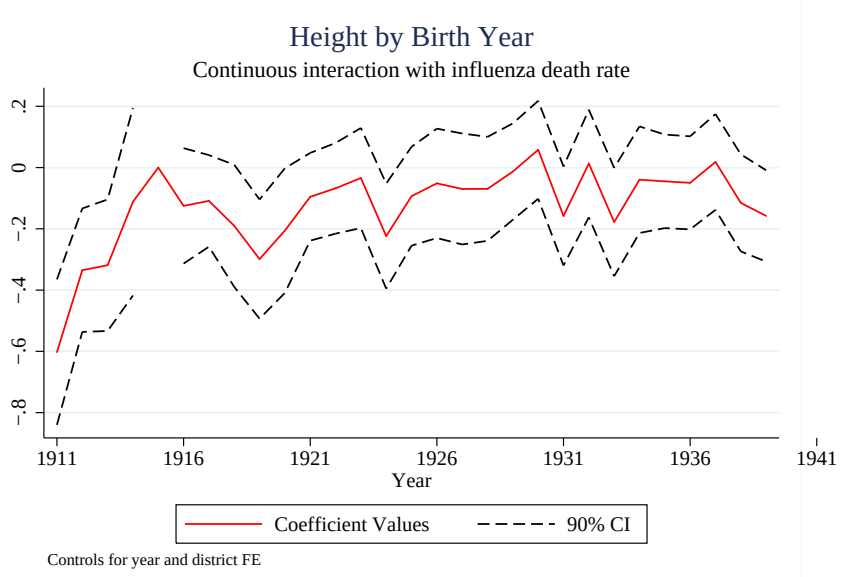
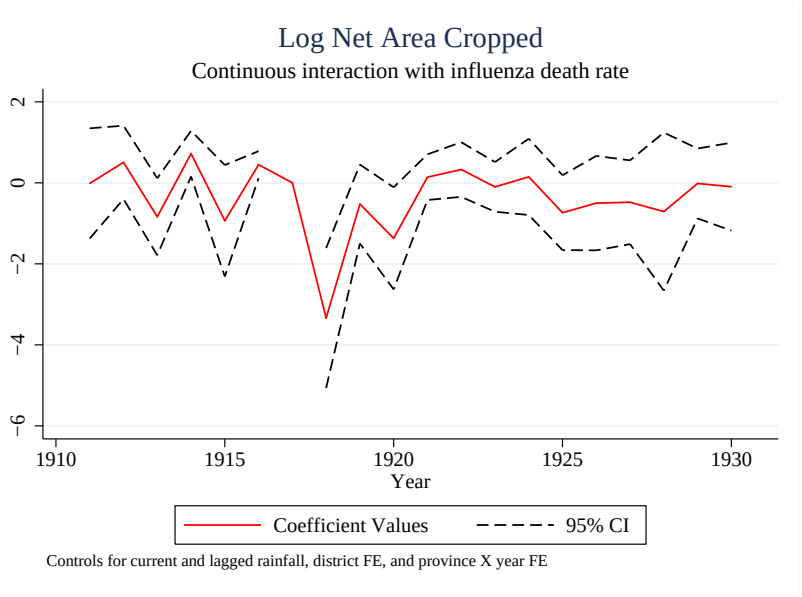


Figure 7: Impact on Agricultural Area Cropped



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
North-Western 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

	(1)	(2)	(3)	(4)
	Log. Influenza death rate			
Oct. Humidity	-0.278***	-0.255***	-0.282***	-0.252***
	[-9.418]	[-8.483]	[-10.084]	[-7.478]
Nov. Humidity			-0.311***	-0.064
			[-5.645]	[-0.899]
Dec. Humidity			0.314***	0.031
			[6.019]	[0.360]
Province FE	No	Yes	No	Yes
R-squared	0.510	0.686	0.614	0.691
N	208	208	208	208
F-statistic on instrument(s)	88.70	71.96	59.22	22.78

Robust t-statistics in brackets. All regressions include controls for the historical district-level normal humidity value for the relevant months included in the set of instruments. * p<0.10, ** p<0.05, *** p<0.01.

Table 11: Deaths by Cause

	(1) Digestive ailments	(2) Fever	(3) Injuries & animal attacks	(4) Respiratory ailments	(5) Smallpox	(6) Cholera	M
Influenza death rate × year = 1918	-0.406	15.514***	0.288	1.986	-1.106	-3.564	
	(1.484)	(1.363)	(1.131)	(1.459)	(5.535)	(7.554)	
District FE	Yes	Yes	Yes	Yes	Yes	Yes	
Province×year FE	Yes	Yes	Yes	Yes	Yes	Yes	
Observations	2294	2294	2294	2294	2294	2294	
R-squared	0.159	0.798	0.091	0.657	0.374	0.392	

Outcome variable is the log of deaths by cause + 1 to ensure uniform sample size. The influenza death rate is defined as the total of Sept. 1918-Jan. 1919 deaths divided by 1911 district population. The dataset consists of all districts of British India from 1911 to 1920. Standard errors in parentheses clustered at the district level. * p<0.10, ** p<0.05, *** p<0.01.

Table 12: Mortality by Race/Caste: Influenza and All Causes, Bombay City, 1918 (Mortality/1,000 population)

	Influenza	All Causes	Influenza Fraction
Europeans	8.3	29.1	0.29
Parsees	9.0	29.5	0.31
Eurasians	11.9	42.0	0.28
Jews	14.8	40.2	0.37
Indian Christians	18.4	53.7	0.34
Caste Hindus	18.9	53.3	0.35
Mohammedans	19.2	61.2	0.31
Low Caste Hindus	61.6	162.7	0.38

Table taken from **?**) with influenza fraction calculated. See sources cited in **?**).

Appendix Figures

Figure 8: Humidity in November, 1918

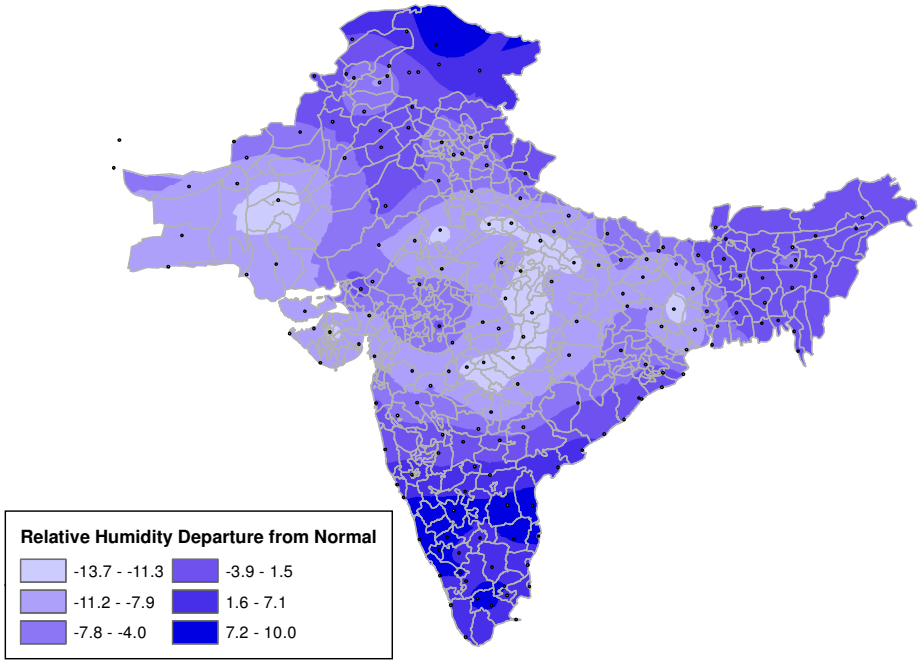


Figure 9: Pre-trends in Deaths

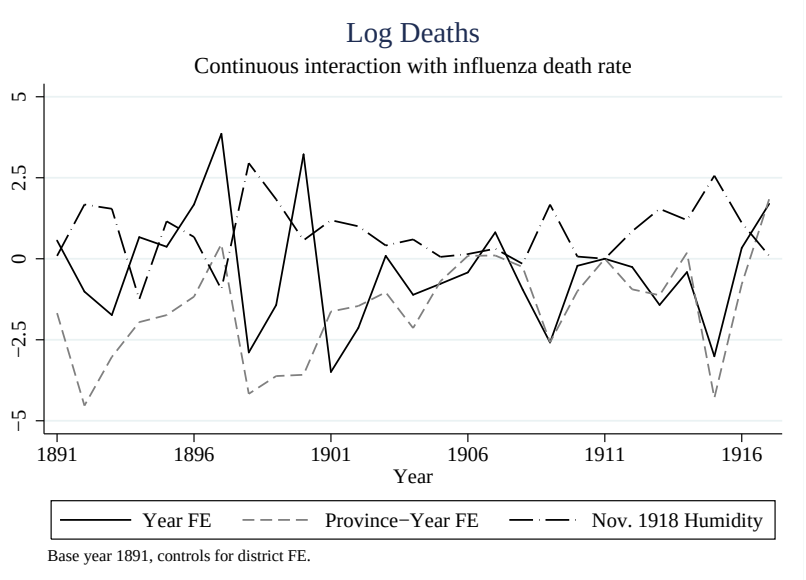


Figure 10: Relationship with Population Density