The Great Revolt and its Legacy: Understanding Vaccine Hesitancy in Colonial India*

Malik Altaf Hussain† Noel D. Johnson‡

June 3, 2026

Abstract

This paper examines whether exposure to colonial violence shaped later engagement with public health initiatives in India. Using a new district-level panel on smallpox vaccination from 1870 to 1919, we study the legacy of the 1857 revolt and the British repression that followed. Districts located closer to Lucknow, where retribution was most extensive, consistently show lower vaccination uptake in the decades after the revolt. The decline is strongest in the immediate aftermath and fades across cohorts. The results suggest that historical experiences of state violence can produce lasting mistrust, reducing participation in programs that are otherwise materially beneficial.

JEL Codes: N35, I12, O15, F54, N45.

Keywords: Disease, Smallpox, Colonialism, India, Vaccine Hesitancy, Trust, State Capacity

*We'd like to thank the Center for the Study of Public Choice and the Mercatus Center for their generous financial support. We also benefited from comments from Alex Tabarrok, Karen Clay, Eduardo Montero, Matthew Jackson, Mark Koyama, Desiree Desierto, Romain Wacziarg, Walker Hanlon, Daniel Bogart, James Fenske, Shruti Rajagopalan, Bishnupriya Gupta, Tirthankar Roy, Dave Donaldson, Alexander Lee, Natalya Naumenko, Vincent Geloso, Jonathan Schultz, and seminar audiences at Virginia Tech, the University of Connecticut, the Markets and Society Conference, and George Mason University.

†Department of Economics and the Mercatus Center, George Mason University, mhussa5@gmu.edu

‡Department of Economics and the Mercatus Center, George Mason University, njohnsoL@gmu.edu.

1 Introduction

Between the mid-nineteenth and early twentieth centuries the British colonial state in India undertook extensive vaccination campaigns aimed at controlling smallpox. This disease was a major public health issue at the time that caused repeated epidemics with case fatality rates of 20-30% and left survivors with permanent disfigurement, scarring, and visual impairment (Banthia and Dyson, 1999; Arnold et al., 1993). Beginning with the introduction of vaccination in 1802, and intensifying after the 1860s, colonial authorities sent vaccinators across urban areas and into rural districts, offering vaccinations free of cost (Bhattacharya et al., 2005). These campaigns represented some of the earliest and most ambitious colonial investments in public health and for many households they constituted the first direct experience with a state-mediated medical intervention (Bhattacharya, 1998; Arnold et al., 1993).

Adoption of vaccination however, was highly uneven across regions, with the result that many districts continued to experience high mortality and morbidity from smallpox despite the availability of free vaccination (Banthia and Dyson, 1999). This suggests that adoption decisions were not solely determined by the perceived benefits of vaccination but were also shaped by the political and institutional environment in which households interacted with the state. In particular, it raises the question whether in colonial India prior experiences of state violence may have discouraged engagement with even welfare-enhancing interventions like vaccination (Lowes and Montero, 2021; Alsan and Wanamaker, 2018).

This paper examines these dynamics through a specific historical episode, namely the Great Revolt of 1857 and its violent suppression by British colonial forces. The revolt emerged from accumulated grievances over land dispossession, disruptions to traditional agrarian structures, and perceived threats to cultural and religious practices, which combined with immediate military triggers to generate widespread resistance across North India (Mukherjee, 2002; Stokes, 1978; Metcalf, 2015). The British response was characterized by systematic violence, including mass executions, public punishments, property destruction, and large-scale land confiscations intended to reassert authority and dismantle the social bases of resistance (Kolsky, 2010; Mukherjee, 2002). Contemporary records indicate near-universal participation of adult males in the rebellion in Oudh (present-day Awadh), leading to widespread collective punishment that touched entire communities (Mukherjee, 2002). While the most detailed records of British violence in this region center on Lucknow, the retribution and suppression of revolt extended into the surrounding countryside.

These experiences lived on in local memory through oral traditions, family narratives, and community practices, shaping how people viewed the colonial state for generations afterward (Pandey et al., 2001). What makes this particularly relevant for understanding vaccination adoption is that the same administrative machinery that had enforced these punitive measures later became responsible for implementing public health campaigns. Households were now asked to accept vacci-

nations from the very agents of a state they had experienced as violent and coercive (Arnold et al., 1993; Hardiman, 2006). This transition from violent suppression to public health outreach offers a unique empirical context to test whether exposure to state violence creates persistent barriers to adopting welfare-improving technologies.

We construct a novel district-level panel dataset on smallpox vaccination rates from 1870 to 1919 across British India, digitized from annual colonial vaccination reports. We examine whether districts geographically and linguistically closer to Lucknow, a major center of resistance and reprisals during the 1857 revolt, show persistently lower vaccination rates in subsequent decades. To measure geographic proximity we calculate least cost travel costs from each district to Lucknow using historical transportation networks, accounting for changes in railway construction over time. We also construct a measure of linguistic distance from Lucknow using data from *Ethnologue*. For our outcome, we focus on vaccination rates for infants under one year of age for two key reasons. First, this measure avoids the mechanical effects present in older age groups, where districts with higher initial vaccination rates naturally have fewer unvaccinated individuals remaining in later years, creating spurious patterns of decline. Second, infant vaccination captures parental decisions that required direct interaction with colonial vaccinators, allowing us to test how prior experiences with state violence shaped willingness to engage with colonial authorities. By examining decade-by-decade patterns in infant vaccination rates across districts with varying proximity to Lucknow we trace how the memory and impact of the 1857 revolt evolved over time.

We find that districts closer to Lucknow, both geographically and linguistically, had significantly lower infant vaccination rates in the decades following the 1857 revolt. A one standard deviation improvement in geographic access to Lucknow was associated with approximately 14 to 17 fewer infant vaccinations per 1,000 inhabitants annually during the 1880s, equivalent to a 1.4 to 1.7 standard deviation decrease. Effects peak in the 1880s and 1890s and fade by the 1910s, consistent with the survival of the cohort aged 10–25 in 1857. The linguistic-distance effect shows the same peak-and-fade pattern under a close-kin parameterization and a steady-to-growing pattern under a broader-cultural parameterization, suggesting two complementary channels of transmission. Placebo tests using proximity to Delhi, Calcutta, or Bombay yield no comparable patterns, indicating our findings reflect the specific legacy of violence around Lucknow rather than general proximity to colonial centers.

This paper contributes to several interrelated literatures. First, there is an extensive literature examining public health initiatives in colonial contexts—notably vaccination campaigns aimed at reducing infectious diseases such as smallpox. Banthia and Dyson (1999) provide foundational evidence on the burden of smallpox in colonial India, highlighting its devastating mortality and morbidity impacts. Arnold et al. (1993) and Bhattacharya et al. (2005) explore British public health policy in India, emphasizing the intersection of medicine, governance, and colonial ideology.

More recent contributions by Cutler and Miller (2005), Alsan and Goldin (2019), and Dupas (2011) reinforce the significance of historical public health campaigns in influencing long-term health outcomes.

We also contribute to the literature on the economic and political impacts of colonialism. Colonial rule has had profound economic, social, and political legacies, as extensively documented by Acemoglu et al. (2001), Acemoglu et al. (2002), Nunn (2008), Dell (2010), and Banerjee and Iyer (2005). These studies argue that the type and nature of colonial institutions critically influenced post-colonial development trajectories. More specifically relevant to the Indian context, Bose and Jalal (2022) and Roy (2020) emphasize the role of British colonial policies in shaping economic and institutional frameworks within India. Recent research, such as Dell and Olken (2020) and Lowes and Montero (2021), connects historical colonial events and policies directly to contemporary trust and social cohesion, providing a theoretical and empirical basis for understanding how colonial violence and repression may have lasting effects on governance and public trust.

Lastly, there is an important and growing literature on the determinants of vaccine hesitancy and trust. Trust has been widely recognized as a critical determinant of vaccine acceptance in public health literature (Larson et al., 2018; Dubé et al., 2015). Recent literature emphasizes that trust in institutions significantly influences vaccine acceptance, particularly when those institutions have histories marked by violence or coercion (Alsan and Wanamaker, 2018; Lowes and Montero, 2021).

Our central marginal contribution is to demonstrate that *non-medical* state violence, a coercive episode unrelated to public health, can produce persistent under-engagement with later medical campaigns. The closest comparators in this literature focus on medical episodes: Lowes and Montero (2021) shows that French colonial sleeping-sickness campaigns shaped contemporary medical mistrust in Central Africa, and Alsan and Wanamaker (2018) document the long shadow of the Tuskegee syphilis study on Black Americans' health-care use. Both identify a medical injury as the source of medical mistrust. We show that an event with no direct medical content, the suppression of an armed political rebellion, generated mistrust that depressed engagement with a subsequent and unrelated public-health technology more than a generation later. The mechanism is therefore not specific to medical experiences but reflects a more general transmission of mistrust across domains. Our findings also contribute to research on the long-run effects of historical violence (Nunn, 2008; Dell, 2010) by demonstrating a specific channel, vaccination uptake, and to the literature on state capacity and legitimacy (Acemoglu et al., 2005; Besley and Persson, 2011) by showing that the same coercive capacity used to establish authority can simultaneously undermine the legitimacy required to deliver public goods. We document the channel through three complementary pieces of evidence: a distance-to-Lucknow gradient that we decompose into a discrete Awadh-region effect (the area uniformly affected by the 1858 Oudh Settlement) and a continuous within-non-Awadh gradient radiating outward from Lucknow; an independent linguistic-distance

gradient that tracks the same direction with a similar temporal pattern under close-kin parameterizations and a longer-lived cultural-transmission channel under broad-cultural parameterizations; and suggestive mortality evidence in the 1890s consistent with the implied welfare loss.

The remainder of the paper is organized as follows. Section 2 provides historical background on smallpox in colonial India. Section 3 introduces the Great Revolt of 1857 and the British retributive response. Section 4 develops our theoretical framework for vaccine diffusion and the cohort-based preference-formation channel. Section 5 describes the data. Section 6 presents our empirical analysis and results. Section 7 concludes. An appendix provides additional tables, figures, and diagnostic results for online publication.

2 Smallpox in India

2.1 The Science of Smallpox

Smallpox was caused by the *variola* virus, transmitted via respiratory droplets and contaminated objects (Fenner et al., 1988). The dominant Indian strain, *variola major*, had a case fatality rate of 20–30% overall and reached 80% among children under five (Geddes, 2006; Dixon, 1962). The disease was highly transmissible—each infected person transmitted to roughly five to seven others on average (James, 1909; Fenner et al., 1988; Rogers, 1945)—and had been endemic to South Asia for centuries before British rule (Hopkins and Lythcott, 1983; Habib et al., 1982). Between 1871 and 1900, registered smallpox deaths averaged over 120,000 annually in colonial India, with actual mortality likely much higher given pervasive underregistration of infant deaths (Arnold et al., 1993; Banthia and Dyson, 1999). The British introduced vaccination using the vaccinia virus, which conferred immunity by exposing patients to a milder related virus. Despite its medical effectiveness, vaccination uptake faced significant obstacles in colonial India from logistical difficulties, religious objections, and the trust deficit that emerged after the 1857 rebellion (Bhattacharya et al., 2005).

2.1.1 The British Smallpox Vaccination Campaign

The British introduced smallpox vaccination to India in 1802, making it one of the first organized public health efforts in the colony (Bennett, 2007; Lahariya, 2014). The first successful vaccination took place in Bombay on June 14, 1802, using lymph that had traveled from Europe via Vienna, Baghdad, and Basra (Bennett, 2007, pp. 204–205). This initial success allowed colonial authorities to create a local supply chain and expand vaccination to Madras, Hyderabad, and Calcutta within the same year. However, the campaign encountered immediate hurdles including logistical difficulties in transporting and preserving the vaccine, concerns about ritual pollution from the procedure, and widespread suspicion of colonial medical interventions among the Indian population (Bennett, 2007; Bhattacharya et al., 2005, pp. 209–212).

Over the following decades, the colonial government developed formal administrative structures for vaccination. By 1850, authorities recognized vaccination as the most effective method to control smallpox and began organizing vaccination services under the Sanitary Departments (James, 1909; Bhattacharya et al., 2005). The new system differed fundamentally from traditional variolation practiced by *tikadars*, who operated without formal qualifications or government oversight. The colonial vaccination system employed two categories of workers: paid vaccinators who received government salaries, and licensed vaccinators who could charge small fees for each operation (Bhattacharya et al., 2005). Two distinct administrative approaches emerged: The Bombay System (established 1827) sent teams of traveling vaccinators directly to people's homes, supervised by European Superintendents who toured districts to verify records. The Dispensary System (introduced in Calcutta in 1837) operated through fixed locations where Civil Surgeons offered free vaccinations, supplemented by traveling vaccinators for rural areas (James, 1909; Bhattacharya et al., 2005).

The 1857 revolt altered Indian attitudes toward the British administration. After the British suppressed the rebellion with extreme violence, trust in colonial governance declined sharply (Malleson, 1891). Rumors spread that the British were using vaccination to spread plague as a method to weaken the population and prevent future uprisings (Naraindas, 2003). Opposition to vaccination intensified and came from multiple sources. Many Indians questioned whether the vaccine actually worked, viewing it as an unproven foreign medical practice (Arnold et al., 1993; Harrison, 1994; Bhattacharya et al., 2005). Religious objections were particularly strong in regions where people worshipped *Sitala*, the goddess associated with smallpox. Caste concerns created additional barriers, as upper-caste Hindus objected to receiving lymph collected from lower-caste donors. The use of cows in lymph production generated further resistance, given the sacred status of cattle in Hinduism (Bhattacharya et al., 2005; Arnold et al., 1993; Banthia and Dyson, 1999). The vaccination procedure itself could be painful and occasionally led to post-vaccinal deaths, serious complications, and unsuccessful vaccinations, which reinforced existing suspicions (Banthia and Dyson, 1999). Traditional variolators (*tikadars*) also opposed the new system, fearing loss of their livelihood (Lahariya, 2014; Bhattacharya et al., 2005).

In response to continued high smallpox mortality, the British attempted to increase vaccination coverage through legal compulsion in the late nineteenth century, but these efforts reached only a small fraction of the population. Authorities introduced compulsory vaccination in Bombay in 1877 and Karachi in 1879 after observing that smallpox deaths remained high in areas where traditional variolation continued. The Government of India Act of 1880 banned inoculation and mandated childhood vaccination in select municipalities and military cantonments. However, rural areas remained largely exempt from these requirements, except for some administrative circles in Bengal. According to James (1909), areas with mandatory vaccination contained less than 7 percent of the total population covered by vaccination efforts. This limited reach meant that

compulsion played a minimal role in expanding vaccination coverage, and the colonial government continued to rely primarily on voluntary acceptance of the vaccine.

Despite these obstacles, vaccination campaigns achieved substantial reductions in smallpox mortality across India. James (1909) documented sharp declines in death rates between 1867 and 1907. In Bombay, deaths fell from 537 to 240 per million. The United Provinces experienced an even greater decline, from 1,587 to 168 per million. Berar's rate dropped from 1,083 to 183 per million. Other regions including the Central Provinces, Punjab, the North-West Frontier, and Madras recorded approximately 50 percent reductions in mortality. The frequency of major epidemics also declined markedly. Before 1886, smallpox outbreaks occurred regularly across India, with repeated epidemics affecting Bombay, the United Provinces, Punjab, Madras, and Berar. After vaccination coverage expanded in these provinces, they experienced no further major epidemics (James, 1909). These mortality reductions occurred even though vaccination remained voluntary for most of the population, suggesting that sufficient numbers of Indians accepted vaccination despite their reservations about colonial medicine. The success varied by region and administrative system, but the overall pattern showed that vaccination could reduce smallpox deaths when enough people participated, even without widespread compulsion (Banthia and Dyson, 1999).

3 Mistrust and The Great Revolt of 1857

3.1 Causes of the Revolt

The Great Revolt of 1857 emerged from widespread economic and social disruption caused by British administrative changes. When the British annexed Awadh in 1856, they dismantled the *taluqdari* system that had governed land ownership and local administration for generations. Mukherjee (2002) documents that this change removed approximately 21,000 *taluqdars* from their positions, while thousands of their dependents and peasants lost both protection and access to traditional support networks. The colonial administration introduced a new land revenue system that increased tax demands and required farmers to navigate unfamiliar bureaucratic procedures. Many villages could not meet these new revenue requirements, leading to increasing debt and land forfeiture. British revenue records and petitions analyzed by Mukherjee (2002) show that both *taluqdars* and peasants experienced what he describes as "the collapse of all customary protection." By early 1857, government documents indicate that "distrust and fear of the government was nearly universal in the Awadh countryside" (Mukherjee, 2002).

The revolt gained momentum through the connection between military and civilian grievances. The Bengal Army was comprised of soldiers recruited mostly from Awadh, and these sepoys maintained strong ties to their home villages. When soldiers returned home on leave, they found their families struggling with land loss and new forms of taxation (Mukherjee, 2002; Metcalf, 2015). This overlap

between military discontent and rural distress transformed what started as an army mutiny into a popular uprising. While the introduction of the Pattern 1853 Enfield rifle cartridge, rumored to contain cow and pig fat, served as the immediate trigger, Mukherjee (2002) argues that this incident merely crystallized existing tensions. The cartridge controversy united religious concerns, social grievances, and economic hardships into a single issue that enabled rapid mobilization across Awadh's towns and villages (Metcalf, 2015; Chandra et al., 1987).

3.2 Course of the Revolt

Lucknow became the focal point of resistance in Awadh, serving as the region's political, administrative, and symbolic center. According to Mukherjee (2002) and Fremont-Barnes (2014), approximately 7,000 rebels entered Lucknow by late June 1857, with *taluqdars* and their armed followers forming the initial core of resistance. These numbers grew rapidly as British forces retreated and news of early rebel successes spread throughout the region. At the peak of fighting, rebel forces in and around Lucknow numbered between 30,000 and 40,000 fighters (Mukherjee, 2002). British sources later claimed the presence of up to 100,000 defenders, though Mukherjee (2002) and Fremont-Barnes (2014) suggest this figure likely includes irregular fighters, camp followers, and civilian supporters rather than active combatants. The British Residency housed approximately 3,000 people during the siege from June to November 1857, including British and Indian troops, civilians, and those who remained loyal to the colonial government (Mukherjee, 2002; Sen, 1957).

The rebels organized an effective siege that lasted several months and drew support from the surrounding countryside. Mukherjee (2002) and Fremont-Barnes (2014) provide a detailed account of rebel tactics—they established artillery positions around the Residency, dug defensive trenches, blocked supply routes, and maintained a tight blockade of the compound. Rural networks sustained the siege through patron-client relationships, with *taluqdars* providing food, recruits, and intelligence to rebel forces. Peasant participation reflected both economic desperation and belief in the legitimacy of resistance against colonial rule (Mukherjee, 2002; Stokes, 1978). The prolonged fighting had a major impact on Lucknow's economy as markets closed, agricultural production stopped, and food became scarce during the crucial kharif harvest season. British relief efforts failed repeatedly until Sir Colin Campbell finally broke through in March 1858 with a larger force, but by then much of Lucknow lay in ruins and rebel resistance had spread throughout the countryside (Mukherjee, 2002; Metcalf, 2015).

3.3 Retribution and Social Memory

The British response combined mass executions with systematic property confiscation that affected nearly every family in Awadh. Reports spread throughout the region that "the British Officers

enticed 1,000 sepoys to come in on promises of forgiveness and having secured them, put the whole body to death in one night" (Enclosure to Forsyth to Edmonstone, 14 August 1858, as cited in Mukherjee, 2002, p. 168). Whether accurate or not, these accounts were "reported far and wide in Awadh" and shaped local understanding of British methods. The colonial forces executed rebels through public spectacles by blowing them from cannons in village squares and hanging them along main roads specifically to terrorize the population (Mukherjee, 2002, pp. 255–260). They destroyed at least 1,500 houses in Lucknow, confiscated over half a million acres of land in Oudh, and demolished religious sites including the Qadam Rasul shrine (Mukherjee, 2002; Metcalf, 2015; Oldenburg, 2014, pp. 260–264, 286–287). Even British officials acknowledged the legitimacy of local grievances, with one admitting that "under these circumstances, the hostilities which have been carried on in Oude, have rather the character of legitimate war than that of rebellion" (as cited in Mukherjee, 2002, p. 170). Families who lost land, witnessed executions, or fled their villages transmitted these experiences through folk songs, personal accounts, and warnings about British treachery that persisted for generations (Llewellyn-Jones, 2014). When colonial vaccinators arrived in these villages in subsequent decades, they encountered families who remembered the atrocities committed by the British, either directly or through folklore, memories that likely contributed to a trust deficit that made cooperation with colonial programs, such as vaccination, difficult to accept (Llewellyn-Jones, 2014; Bhattacharya et al., 2005).

If this collective memory of violence shaped household decisions about engaging with colonial authorities, we would expect districts closer to Lucknow to have lower vaccination rates in subsequent decades. Testing this hypothesis, however, requires addressing two measurement issues. First, yearly vaccination data reflect both new adoptions and the shrinking pool of unvaccinated individuals, making it difficult to separate changing attitudes from mechanical saturation effects. Second, any effect of the 1857 violence should weaken over time as the directly affected generation ages, but we need to specify when this cohort would have influenced vaccination decisions. The next section develops theoretical frameworks to address both issues, allowing us to distinguish the legacy of colonial violence from other barriers to vaccination adoption.

4 Theory: Vaccine Diffusion and Preference Formation

We face two substantial theoretical questions. First, how does a vaccine diffuse through a population? Second, how do traumatic events translate into preferences over practices such as the willingness to allow a colonizing government to vaccinate one's infant? We tackle each of these in turn.

4.1 A Model of Vaccine Diffusion

Understanding vaccine diffusion is critical when analyzing the success or failure of public health campaigns. The Bass Model provides a powerful theoretical framework to describe the process by which a new practice—such as vaccination—spreads through a population (Bass, 1969; Jackson, 2008). This model conceptualizes diffusion as initially driven by a few early adopters (innovators) and subsequently by followers (imitators), influenced by the number of individuals who have already adopted the innovation.

The Bass Model characterizes adoption through two parameters: the coefficient of innovation, p, which measures the propensity to adopt independently of the actions of others, and the coefficient of imitation, q, which captures the tendency to adopt influenced by existing adopters. The model can be expressed by the following differential equation:

$$\frac{dF(t)}{dt} = (p + qF(t))(1 - F(t)) \tag{1}$$

Where $F(t)$ is the cumulative fraction of adopters at time t, p is the innovation coefficient, and q is the imitation coefficient.

Integrating this differential equation results in the well-known S-shaped diffusion curve, capturing the acceleration of adoption at intermediate levels and the subsequent deceleration as saturation is approached. This is illustrated in Figure 1.[1]

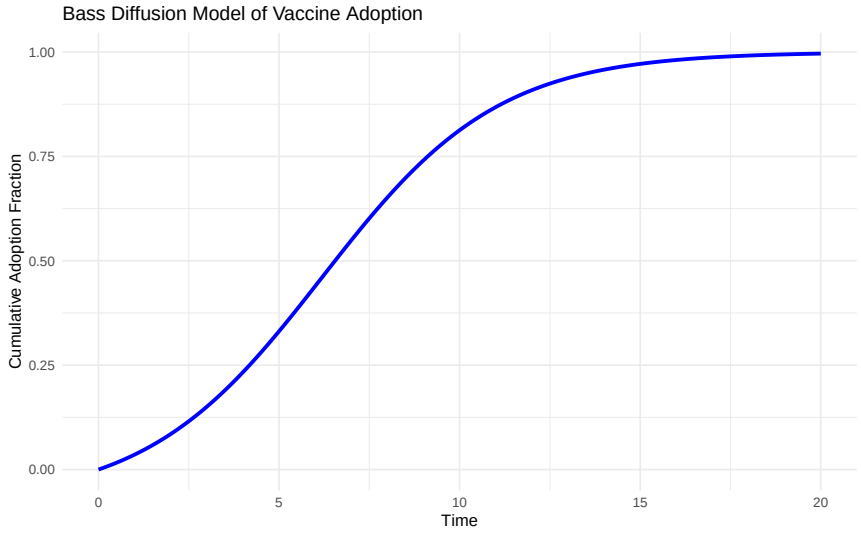


Figure 1: Bass Diffusion Curve

Initially, when adoption is low ($F(t) \approx 0$), the adoption rate is primarily driven by p, the innovators. As $F(t)$ grows, the influence of imitators ($qF(t)$) becomes more significant. Eventually, as

[1]For illustrative purposes we assume $p = 0.03$, $q = 0.38$, and that $F(0) = 0$.

cumulative adoption approaches saturation ($F(t) \approx 1$), the adoption rate slows down due to the scarcity of non-adopters. The medical authorities, in effect, run out of people to vaccinate.

In the context of our study, the Bass Model highlights a critical empirical challenge. Our data measure yearly vaccination rates at the district level rather than the cumulative proportion vaccinated (i.e. we have data on the slope of the Diffusion Curve). According to the Bass framework, as the cumulative fraction vaccinated ($F(t)$) increases, the yearly rate of new vaccinations must decline simply due to a decreasing pool of unvaccinated individuals, independent of other influences like trust or institutional effectiveness. Thus, directly analyzing yearly district-level vaccination rates may confound the genuine effects of factors such as mistrust toward vaccination campaigns with this mechanical decline induced by rising cumulative vaccination levels. This may be particularly problematic for us given that we would expect any mistrust towards colonial institutions generated by the 1857 massacres to attenuate over time as the memory of the event fades from the population. If we were to naively use the data on adult yearly vaccination rates, this attenuation due to fading memory of 1857 would be confounded by the mechanical decline in vaccination rates due to the rising level of vaccination rates in the population.

We can verify directly from our data that this saturation pressure is operating, without recourse to parametric estimation. Figure 5 (presented in Section 5.1 below) shows that infant vaccination rates rose monotonically across our sample period, with no sign of leveling off. By contrast, Figure 19 in the Appendix shows that the all-ages vaccination rate climbed sharply through the 1870s but plateaued at approximately 3.5% per year by the 1880s and remained roughly flat through the 1910s. The divergence (rising infant rates against a flat all-ages rate) is the direct empirical signature of the Bass saturation force: the existing population was approaching the bound $F(t) \to 1$, while the renewable infant cohort remained outside that constraint. This pattern alone motivates our choice to use infant vaccination as the dependent variable in our main analysis.[2]

To circumvent this potential source of bias we adopt an alternative strategy. We use the yearly vaccination rates for children under one year old as our dependent variable. This population is renewed annually providing a consistent and unaffected base each year free from the mechanical reduction associated with rising cumulative vaccination. Consequently, variations observed in this rate should more reliably reflect external influences, including trust or mistrust, institutional effectiveness, and the legacy of historical events such as the 1857 Revolt.

The next question we must address then is, "during which years do we think the population would

[2]We complement the graphical argument with a parametric Bass fit on the only district in our panel with sufficient pre-1870 data to anchor an initial-level estimate (Kolaba, part of "Old Bombay"; Appendix Figure 16), combined with the 1872 Bombay City census reading of 29.6% adult vaccination (Banthia and Dyson, 1999; *Census of Bombay City*, 1872, p. 656). The fit yields $\hat{p} = 0.006$ (p-value 0.00127) and $\hat{q} = 0.096$ (p-value $<$ 2e-16) and reproduces the flattening pattern: the fitted decadal increment falls from 13% in the 1880s to about 3% in the 1910s. The imputation of Kolaba's rates for 1874–1880 and 1881–1888 reflects this data limitation rather than a methodological choice. We treat the parametric estimate as supporting evidence rather than primary motivation.

be composed of people most likely to have been both affected by the 1857 massacres and alive to exert influence over family decisions on whether or not to allow infants to be vaccinated?"

4.2 Preference Formation

Research in developmental psychology and economics consistently suggests that the most influential window for forming attitudes toward institutions and societal trust is adolescence and early adulthood (ages roughly 10–25 years). Erikson (1968) emphasizes adolescence as a critical period for identity formation, including orientations toward authority figures and institutions. Giuliano and Spilimbergo (2025) review the literature in social psychology and identify two theories that support the idea of a "Formative Age Hypothesis". These include the *Impressionable Years Hypothesis* which argues that core beliefs (e.g. trust in institutions) are shaped during impressionable years and then persist into adulthood (Krosnick and Alwin, 1989). Similarly the *Increasing Persistence Hypothesis* argues that individuals form their beliefs during their impressionable years and then their openness to new beliefs declines as they age and experience "... a decrease in interest in events distant from one's immediate life" (Glenn, 1980). Giuliano and Spilimbergo (2025) also point to the literature on neurological development which provides evidence for the "synaptic tagging and capture" hypothesis which argues that the volume of grey matter in the brain grows through adolescence and then gradually declines in adulthood (Frey and Morris, 1997). This has been argued to support the idea that influences experienced during adolescence are more likely to generate the synaptic connections which form the basis for beliefs into adulthood.

We take guidance from these literatures and assume that the cohort of individuals who would be most affected by the 1857 massacres would be those aged 10 to 25 years when they occurred. This means people born between 1832–47 who were "close" to Lucknow were the most likely to have their attitudes towards British institutions negatively molded by the events of 1857.[3]

During what years should we expect the influence of the 1832–47 cohort to be affecting family decisions? We approach this problem by constructing a "treatment exposure curve" which shows the probability of survival for someone born in our treatment cohort for each of the years in our data set, *conditional on them being alive in 1857*. We can then investigate whether the treatment effect of Lucknow exposure in 1857 weakens in a reasonable way as the 1832–47 cohort dies out.

We take data from Visaria and Visaria (1982) on conditional mortality in India during the second half of the nineteenth century and use them to build a simple mortality model. The standard exponential mortality function is $S(t) = e^{-\lambda t}$, where $S(t)$ is the probability of surviving until period t, λ is the constant mortality rate over the time period, and t is the number of years elapsed.

[3]While it is possible that preferences may have been vertically transmitted between the 1832–47 cohort and their children, the literature generally suggests that there is *intergenerational attenuation* and we should expect to see the strongest effect in the cohort of 1832–47 rather than their offspring (Bisin and Verdier, 2001).

We can approximate the hazard rate for a given life expectancy at age x which we will call e_x. This hazard rate is given as $\lambda_x \approx \frac{1}{e_x}$. So, using example data from Visaria and Visaria (1982), if conditional life expectancy at age 20 was 38 years, then $\lambda_{20} \approx 0.026$. We apply this reasoning to multiple initial ages to approximate the conditional survival probability for the 1832–47 cohort in each year of our sample. Appendix Table 1 shows our calculations.[4]

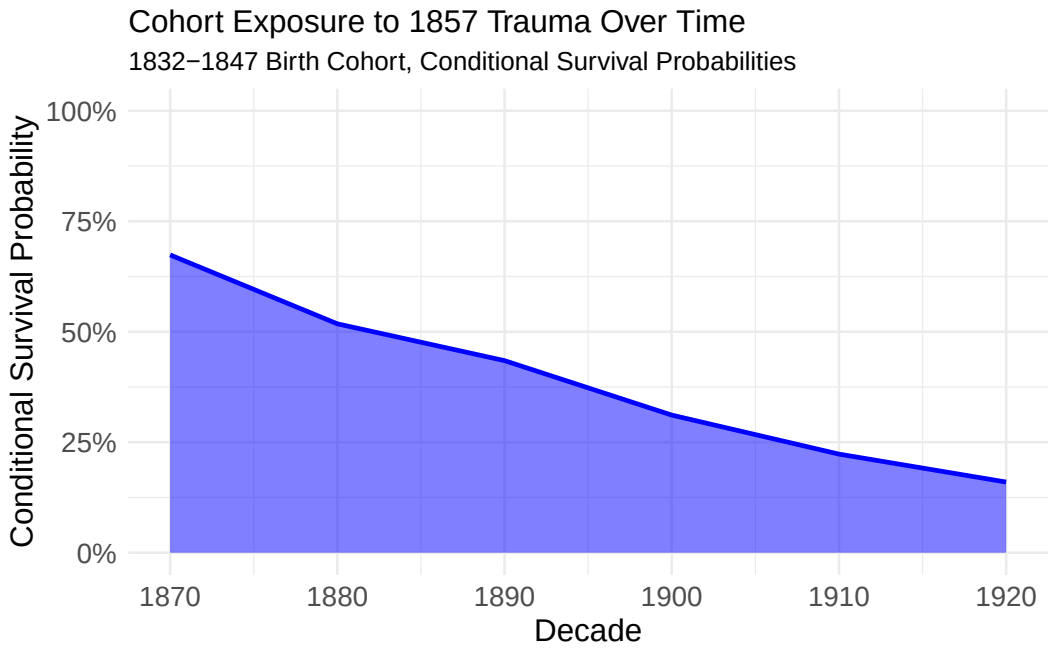


Figure 2: 1857 Treated Cohort Survival Over Time

As can be seen in Figure 2, conditional on having been alive in 1857, the probability of survival of an individual in our treated cohort is around 70% at the beginning of the time period. This declines to around 27% by 1900 and around 13% by 1920. While these numbers are necessarily approximate, they give a reasonable estimate of what to expect in terms of observing the impact of 1857 on vaccination rates in later periods. The cohort most likely to have been affected by 1857 will have been severely attenuated after 1900 or so. As such, in the absence of strong vertical transmission of beliefs, we also expect to see less effect of proximity to Lucknow after 1900.

The cohort-attenuation channel above is not the only mechanism through which exposure to 1857 violence might affect later vaccination uptake. A complementary channel operates through persistent inter-group cultural distance: institutionalized practices, religious networks, and language-mediated transmission can preserve and even widen attitudes toward colonial institutions through mechanisms that do not require the original cohort to remain alive. Such horizontal-transmission channels are emphasized in Bisin and Verdier (2001) and Giuliano and Spilimbergo (2025). Em-

[4]Use age 20 as our reference point until 50 (since cohort members would have passed age 50 around 1882–1897), then switch to age 30 for more accurate older age estimates.

pirically, the two channels imply different temporal patterns: cohort transmission predicts a peak-and-fade trajectory tied to the 1832–47 cohort's survival, while broader cultural transmission can predict a steady or growing effect. We will return to this distinction in Section 6.5.2, where the two parameterizations of our linguistic-distance measure ($\delta = 0.5$ and $\delta = 0.05$) will let us distinguish them empirically.

5 Data

5.1 Colonial Vaccination Reports and Vaccination Rates

Between 1869 and 1928, the British colonial administration in India produced annual vaccination reports documenting smallpox vaccination efforts across the subcontinent. These reports constitute one of the earliest systematic public health data collection efforts in a colonial context, recording district-level vaccination counts, population estimates, and, crucially for our analysis, age-specific data that distinguished vaccinations among infants under one year from those administered to older individuals (Bhattacharya et al., 2005).

We collect these archival data from annual and triennial vaccination reports produced by provincial sanitary departments across British India, digitized by the National Library of Scotland.[5] Given that district boundaries were altered in significant ways by the British over the course of our study period, we choose to use the largest district city to define the coordinates of the data contained in the district level reports. This allows us to avoid having to harmonize specific borders of the districts—a possibly impossible task.

For this study, we focus on the years 1870 to 1919. We have several reasons for restricting our focus to these years. First, earlier years, especially 1869, contain very few observations. Second, our analysis below is based on studying the evolution of vax rates by decade and the decade 1920-1928 contains incomplete coverage. Third, our treated cohort has aged out by the 1920s. Finally, and most importantly, the relationship between public health and the colonial authorities was breaking down by the 1920s for reasons unrelated to the hypothesis of this paper. Namely, in April of 1919 the Amritsar Massacre took place, which severely undermined Indian trust in the Raj. In 1920 Gandhi began his, ultimately successful, campaign of non-cooperation against the British.

[5] Collections cover Assam (1875–1927), Bengal (1869–1929), Bombay and Sindh (1870–1923), Central Provinces and Berar (1868–1920), Eastern Bengal and Assam (1905–1912), Hyderabad Assigned Districts (1872–1903), Madras (1894–1920), Mysore (1878–1880), North-West Frontier Province (1904–1920), North-Western Provinces and United Provinces (1866–1922), Oudh (1871–1873), and Punjab (1867–1920). See Appendix Table 3 for detailed URLs.

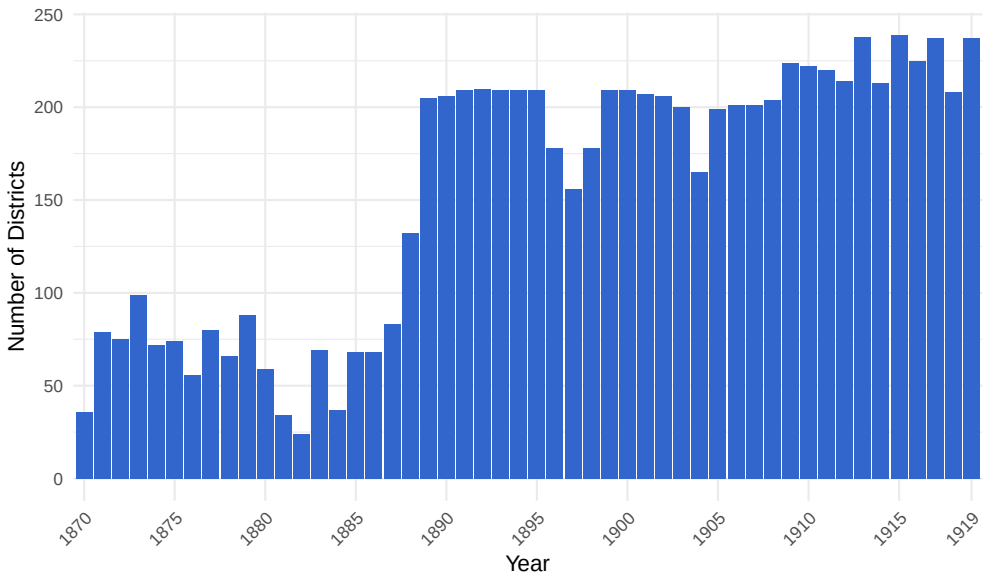


Figure 3: Number of districts with reported vaccination data, by year, 1870–1919. The panel is unbalanced because not all provinces appear in every year of the colonial vaccination reports and because some districts entered the digitised sample only after the 1880s expansion of provincial reporting. Source: digitised colonial vaccination reports (see Section 5.1 and Appendix Table 3).

Our base dataset contains 7,746 district-year observations. As illustrated in Figure 3 coverage expanded substantially in the 1880s as colonial administrative capacity strengthened and provincial sanitary departments adopted standardized reporting formats (Bhattacharya et al., 2005). By 1890, the dataset captures vaccination data for 211 districts, which grows to over 240 districts annually by the 1910s.

Figure 4 shows the location of our district cities. It should be emphasized that, since we have an unbalanced panel, not all of these districts will be in all of the regressions.[6] To get a better sense of this, the Figure shows a grey 1,000 km buffer around Lucknow. This represents our main sample. Since our identification strategy depends on "distance" to Lucknow, we expect that, at a certain distance, there should be no meaningful relationship between proximity to Lucknow and infant vaccination rates. In our baseline regressions we choose 1,000 km as the cut-off. We investigate the robustness of this assumption in Section 6.7.

[6]Our baseline regressions will contain 79 districts for 1870-79, 78 districts for 1880-89, 149 districts for 1890-99, 203 districts for 1900-09, and 206 districts for 1910-19.

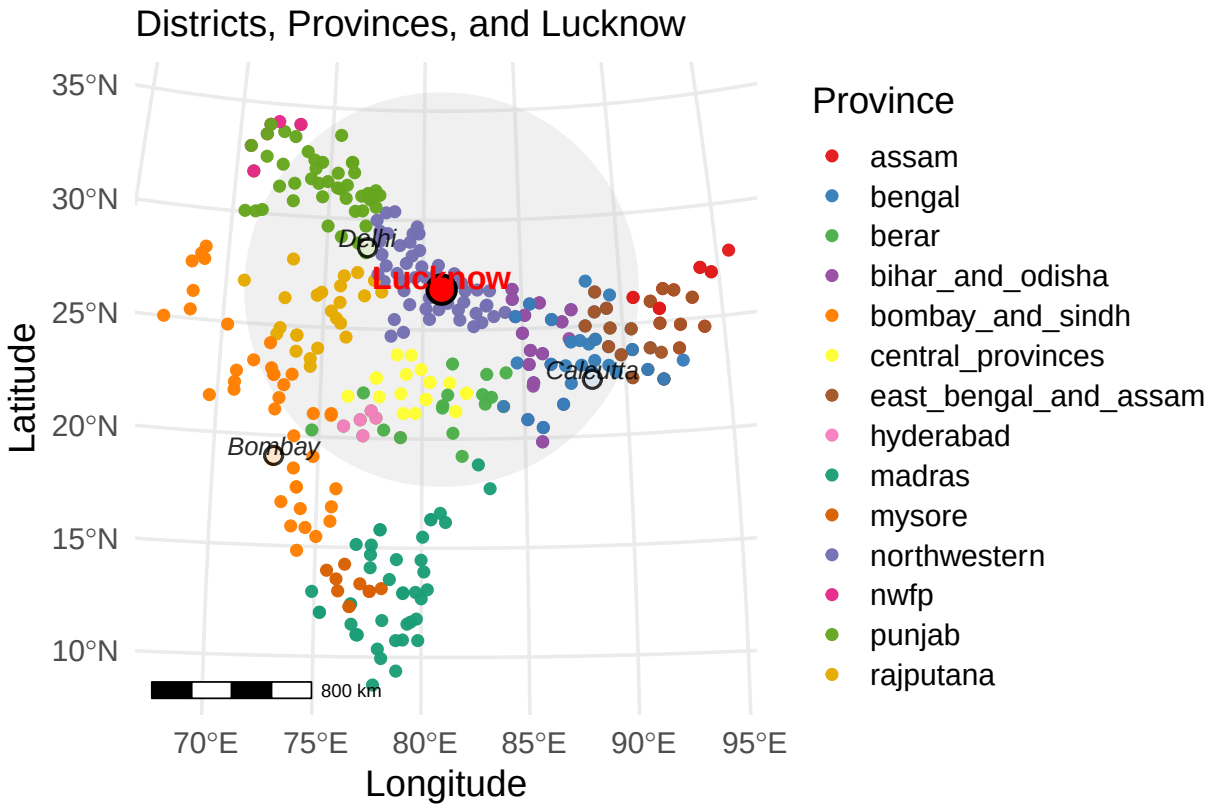


Figure 4: Districts, Provinces, and Lucknow

Figure 4 also shows the province in which each district is located. This is important since in our preferred specifications we will include province fixed effects. As such, our main results will be based on within province variation in distance to Lucknow and vaccination rates. The Figure also indicates the location of Delhi, Calcutta, and Bombay. We will be exploiting these cities as placebos in Section 6.7.

The absolute scale of vaccination efforts was considerable. Annual vaccinations increased from under 40,000 in 1869 to over 9 million by the 1920s, demonstrating the colonial state's expanding capacity to conduct public health campaigns. Considered relative to district populations, these figures reveal seemingly modest levels of coverage. We measure vaccination intensity as annual vaccinations per capita at the district level. These rates averaged 3.6% annually.[7] While this may seem like a low annual rate of vaccination two considerations should be kept in mind. First, there was quite a bit of variation in rates across districts. Figure 5 indicates that in the 1890s the interquartile range of vaccination rates was about 1.7%. Second, according to our simulation exercise using data from Bombay in Section 4.1 one or two percentage points makes a significant

[7]Specifically, we construct Vaccination Rate$_{dt}$ = Total Vaccinations$_{dt}$/Total Population$_{dt}$, where Total Vaccinations$_{dt}$ represents all vaccinations in district d during year t. This is similar to how colonial administrators evaluated campaign reach and provides a consistent measure across our full geographic and temporal span.

difference in population vaccination rates over a few decades.

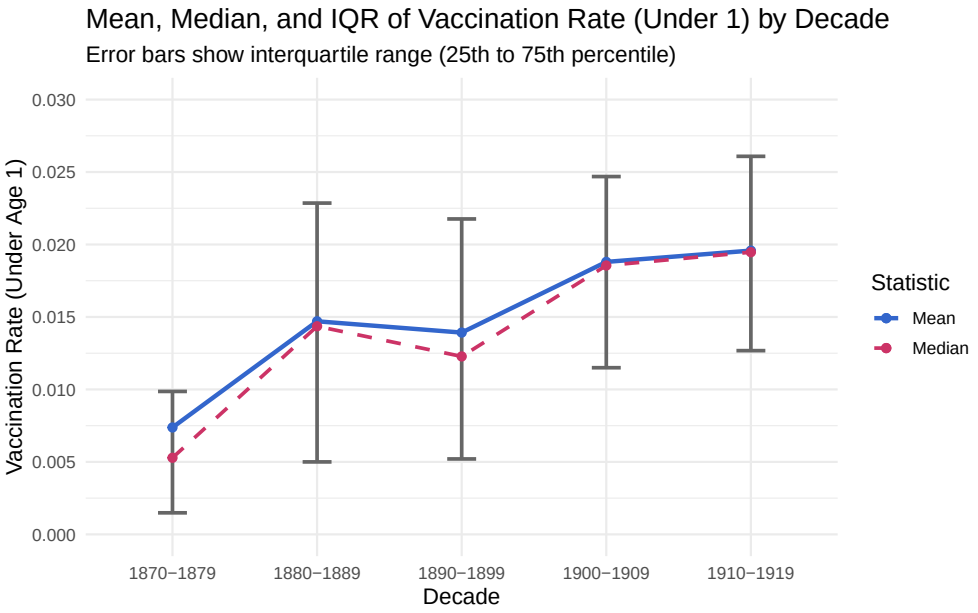


Figure 5: Mean infant vaccination rate (vaccinations administered to children under one year of age, divided by total district population) by decade across the 1870–1919 study period. The plot shows the mean, median, and interquartile range across districts in our panel for each decade. Source: digitised colonial vaccination reports (see Section 5.1).

For most of our analyses we will collapse the 1870-1919 yearly panel into decades (i.e. 1870-79, 1880-89, 1890-99, 1900-09, and 1910-19). We do this for several reasons. First, given that both trust in the colonial health system as well as the least cost path to Lucknow evolved relatively slowly over time, the identifying variation we're using is more realistically at the decade level rather than the yearly level. Second, given the unbalanced nature of the panel, districts can sometimes be present for a few years, then drop out of the records for a few years, and then return. Collapsing the panel at the decade level allows us to avoid having to interpolate values for these missing districts. Finally, there can be quite a bit of yearly variation in vaccination rates that is almost certainly not due to variation in distance to Lucknow. Using decade averages allows us to smooth over these fluctuations and focus on the variation that most interests us.

5.2 Distance to Lucknow

We investigate two different measures of distance to Lucknow. The first will be based on geographic distance and will be constructed as the cost of taking the least cost travel path through the (time-varying) transportation network to Lucknow. Our second measure takes inspiration from the recent literatures on market formation and cultural transmission and focuses on the linguistic distance of districts from Lucknow.

5.2.1 Geographic Distance

One of our main variables of interest is the degree of geographic and economic accessibility to Lucknow—the epicenter of the 1857 events—for all major cities and districts across colonial India. Our hypothesis is that cities with lower travel costs to Lucknow were more exposed, both directly and indirectly, to the spread of violence, news, and later to government interventions such as vaccination campaigns. We claim (and will provide evidence in Section 6.7) that the impact of proximity to Lucknow depended not on just linear distance but rather on the evolving structure of the transport network, which changed substantially over the late nineteenth century.

There are several possible approaches to quantifying access to Lucknow. Great-circle (Euclidean) distance would be the simplest method and involves calculating the straight-line geographic distances between cities. This is easy to calculate but fails to capture historical realities of travel such as the growing rail network and the presence of navigable waterways. Alternatively, we can calculate the least-cost travel cost between each city and Lucknow. Here, the "distance" between two locations is defined not by kilometers but by the cumulative cost (relative to the normalized cost of using rail) of traversing the lowest-cost route, given the transport modalities available at the time.

To calculate least-cost travel costs for each city and year, we do the following: We assemble georeferenced layers for the main transport modalities available in colonial India—navigable rivers, seas, railways, and "portage" (land travel not facilitated by modern transport infrastructure). See Section 5.3 for data description and sources. Each modality is associated with a travel cost per unit distance based on historical estimates from Donaldson (2018).[8] We overlay a regular lattice of 25 km $\times$ 25 km grid cells across the Indian subcontinent. For each grid cell, we extract the lowest-cost available transport technology for each period in the sample. We then apply Dijkstra's algorithm to compute the path of least cumulative cost from Lucknow to each district city. Finally, for each city and period, we record both the optimal path and the least-cost travel cost associated with the optimal path.

[8]Donaldson's estimates are: Rail=1.0, Coast=2.25, River=3.0, Road=4.5. As such, our measure of least cost travel costs are going to be in terms of normalized railroad-equivalent units.

(a) 1860 (b) 1880

Figure 6: Least Costs Paths Between Lucknow and Bombay

For example, Figures 6a and 6b show both the cost rasters and the evolving least cost travel paths between Lucknow and Bombay in 1860 and 1880 respectively. The cost rasters show the rivers, coast line (sea), and rail lines in light green, brown, and white respectively. The least cost paths are in red. The Figures make clear that the evolution of the rail system mattered for the "distance" to Lucknow over time. In 1860 the cumulative least-cost travel cost between Lucknow and Bombay was 126. In 1880 this had fallen to 45.

Figure 7 summarizes how the average cost to reach Lucknow from all cities fell substantially over the period, reflecting the rapid expansion of the rail network. The plot shows the mean, median, and interquartile range of the least-cost travel cost between each city in the data set and Lucknow by decade. In 1860, the mean cost of travel was about 80 with a IQR of about 45 to 115. By 1900 the mean had dropped to 30 with an IQR of 15 to 45. As we will show in Section 6.7, this variation in travel costs generated by the rail network matters a lot for our hypothesis. In this case measurement is important—it was access to Lucknow through the transport network that mattered for the diffusion of beliefs about the colonial health authorities, not simple straight line distance.

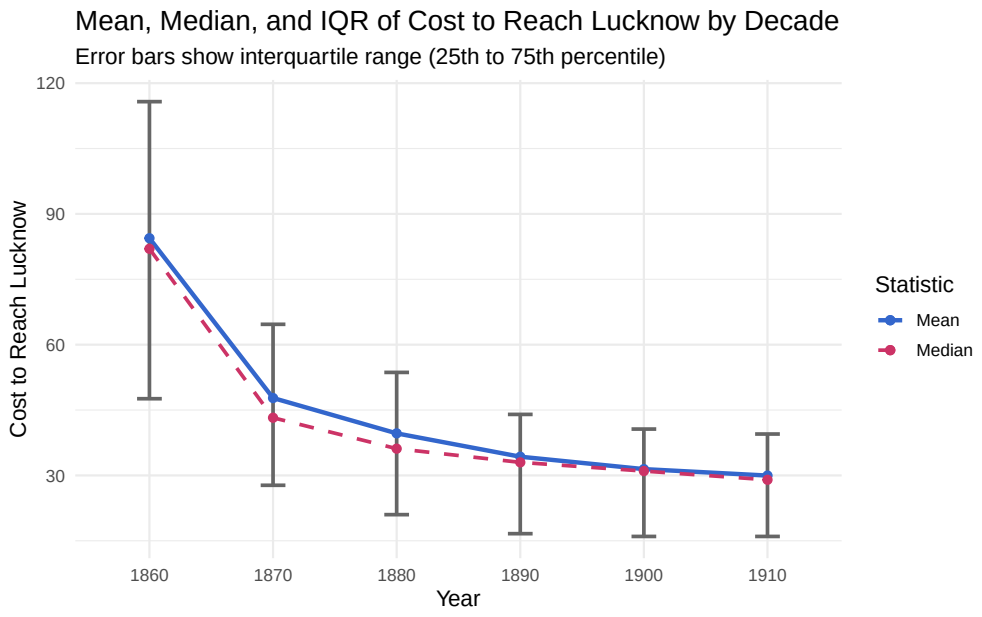


Figure 7: Least Cost Travel Cost to Lucknow by Decade

A remaining question concerning geographic distance to Lucknow is how to incorporate the evolving rail network, and its impact on lowering travel costs, into our primary measure. One simple approach would be to construct our treatment variable as access to Lucknow in 1860 (or as close to 1857—the year of the massacres—as possible). A related option would be to assume that in every decade we analyze the relevant measure of cost to Lucknow is just whatever it was at the beginning of that period. We find neither of these approaches to be attractive since they fail to incorporate the cumulative exposure of a district to Lucknow over time. For some cities, like Bombay, the expansion of the rail network made a very real difference, in a relatively short span of time, on the geographic access to the location of the 1857 violence.

An alternative approach is for us to incorporate information on the changing access to Lucknow over time in our measure of Lucknow access for district i in period t. There are several ways we could do this. The most flexible would be for us to create a discounted time-weighted average of access cost for each period. However, we have been unable to find any convincing guidance on what the appropriate discount rate for this particular circumstance would be. As such, we choose to construct a more transparent measure which is to take the simple average of Lucknow access for every district up to the current period. That is, we can calculate the mean access cost between 1857 and a given period t as:

$$\text{AccessLucknow}_{i,t} = \frac{1}{T} \sum_{s=1860}^{t} \text{PeriodCost}_{i,s} \tag{2}$$

Where $AccessLucknow_{i,t}$ is the average least cost travel cost for district i to Lucknow in period t. T is the number of periods before the current period (inclusive of the current period). $PeriodCost_{i,s}$ is the least cost travel cost for district i to Lucknow in the historical period/decade s. So, for example, if we have two districts (Bombay and Lucknow) and two periods (1860 and 1880), then $AccessLucknow_{Bombay,1880}$ would be equal to $85.5 = \frac{126+45}{2}$.

Given the historical and geographic complexity of colonial India, this approach captures both the physical and economic realities of travel, accounting for different modalities of transport. It also provides a flexible framework for assessing how the evolution of the transport network (such as the expansion of rail lines in the late nineteenth century) affected access and, by extension, the diffusion of beliefs or information originating in Lucknow.

5.2.2 Linguistic Distance

Recent research in economics increasingly links linguistic distance to patterns of market development, trade, and cultural transmission. Linguistic similarity is associated with greater bilateral trade flows, facilitating communication, trust, and the efficient exchange of market-relevant information between trading partners (Fenske and Kala, 2021; Anderson and Van Wincoop, 2004; Egger and Lassmann, 2012). More broadly, linguistic and cultural differences are correlated with ancestral distance, shaping a variety of economic outcomes including technology diffusion, institutional quality, and economic integration (Spolaore and Wacziarg, 2018).

Within Indian economic history specifically, linguistic and cultural divisions—alongside caste and religious differences—have played prominent roles. Linguistic boundaries influenced the flow of commercial information, affecting patterns of industrial segregation and specialization (Gupta, 2014). Furthermore, linguistic and ethnocultural fragmentation impeded collective action and investment in public goods, such as education. For example, caste, religious, and linguistic divisions contributed to systematically lower schooling investments, with persistent economic consequences extending well into the late 20th century (Chaudhary, 2009; Chaudhary and Garg, 2015).

South Asia has remarkable linguistic diversity, containing numerous distinct languages from four major families: Indo-European, Dravidian, Austro-Asiatic, and Sino-Tibetan (Asher, 2008). The Indo-European languages, including Hindi and Bengali, arrived roughly 3,500 years ago and largely replaced Dravidian languages across northern regions (Haak et al., 2015). Dravidian languages—Tamil, Telugu, Kannada, and Malayalam—dominate the southern parts of India and have ancient literary traditions dating back to at least the second century BCE (de la Fuente, 2005). Austro-Asiatic and Sino-Tibetan languages, present since prehistoric times, persist in smaller linguistic communities across eastern and northeastern regions (Asher, 2008).

South Asian languages, though sometimes sharing vocabulary due to historical contact and bilin-

gualism, remain distinct in grammar, phonology, and origin (de la Fuente, 2005; Montaut, 2005). This linguistic complexity reflects long histories of population migration, isolation, and cultural persistence, with language boundaries remaining notably durable despite sustained interactions among different linguistic groups (Montaut, 2005).

We use data contained in the Ethnologue database to construct measures of the distance between Indian language groups (Eberhard et al., 2025). This database categorizes every language in the world to build a language tree. This tree has a maximum of 15 branches. For example, Appendix Figure 27 shows an example of a language tree for major Indian languages. These classifications are primarily taken from Frawley (2003). A standard way to measure the similarity between two language groups is to count how many branches on the language tree they share (Fearon and Laitin, 2003; Esteban et al., 2012). Thus, in our example Figure Punjabi shares two branches with Gujarati. As such, those languages are more similar than Punjabi and Marathi which only share a single branch. More generally, we calculate,

$$d_{L,i} = 1 - \frac{SharedBranches^{\delta}}{15} \tag{3}$$

Where $d_{L,i}$ is the linguistic distance between Lucknow and district i and δ is a parameter indicating the sensitivity of the measure to differences in language groups. Fearon and Laitin (2003) originally takes δ to be 0.5 whereas other, more recent, papers have used a smaller measure (usually $\delta = 0.05$) (Esteban et al., 2012; Fenske and Kala, 2021). The larger the δ used the more sensitive the measure is to differences in the number of shared branches between two districts. Thus, in our simple example, $d_{Punjabi,Gujarati} = 0.64$ using $\delta = 0.5$. Using $\delta = 0.05$ it is 0.1.

We report results below under both parameterizations. The two values capture different aspects of linguistic distance: $\delta = 0.5$ down-weights distant relationships in the language tree so that only districts sharing many branches register as "close," rendering the measure sensitive primarily to *close* linguistic kin (Awadhi–Hindi–Bhojpuri in our setting). $\delta = 0.05$ spreads the distance more uniformly across the tree, making the measure sensitive to *broad* cultural distance across language families. As we show in Section 6.7, the two parameterizations yield distinct temporal patterns: $\delta = 0.5$ produces a peak-and-fade trajectory consistent with the cohort attenuation we model in Section 4.2, while $\delta = 0.05$ produces a steady-to-growing effect over the full sample period—suggestive of a complementary, longer-lived cultural-transmission channel that does not depend on the survival of any specific birth cohort.

5.3 Controls

Some of our controls come directly from the vaccination reports. For example, the total population of the district as well as the spending on smallpox vaccination in the district are usually reported.

Using these variables we construct a measure of vaccination spending per capita. Controlling for this last variable is important as it helps us distinguish between supply driven explanations for vaccine uptake versus demand. The reports also always list the province in which the district is located, which we will include as fixed effects in our preferred specifications (see Figure 4).[9]

Our shape files describing the growth of the rail network are from Donaldson (2018). Appendix Figure 23 shows how the network expanded from 1860 to 1910. We use these data in both the construction of control variables for distance of a district to the nearest rail line in each decade, as well as in our creation of the access to Lucknow variables.

Our measure of terrain ruggedness comes from Nunn and Puga (2012).[10] This measure is important as it is possible that harder to access districts may have received fewer resources from the British vaccinators. Alternatively, more remote districts may have been more hesitant to take up the vaccine for cultural reasons. Appendix Figure 21 illustrates our measure.

We take our measure of caloric suitability from Galor and Özak (2016).[11] One reason we include it as a control is that it is possible, for Malthusian reasons, that places with greater agricultural suitability were also more developed in the late 19th century. This, in turn may have impacted both the willingness to accept the vaccine or the ability of the local government to distribute the vaccine. Appendix Figure 20 shows the variation in caloric suitability over the Indian subcontinent.

Appendix Figure 22 shows major navigable rivers. We take these data from the Food and Agriculture Organization of the United Nations.[12] We follow the literature in hydrology and restrict our attention to rivers with a Strahler number greater than 2.[13]

6 Analysis

6.1 Empirical Strategy

Our identification strategy exploits spatial variation in exposure to British retributive violence following the 1857 revolt. Districts closer to Lucknow, the epicenter of both the rebellion and its violent suppression, experienced more intense collective punishment, including mass executions, property confiscations, and the destruction of entire villages. Contemporary British reports documented that "three fourths of the adult male population of Oudh, had been in rebellion," leading to widespread retaliation that affected entire communities rather than select individuals (Mukher-

[9] All reports can be downloaded using the links in Appendix Table 3.

[10] `https://diegopuga.org/data/rugged/`.

[11] `https://ozak.github.io/Caloric-Suitability-Index/`.

[12] `https://data.apps.fao.org/catalog/dataset/76e4aacc-b89e-4091-831f-63986fe029f9/resource/e8e609ed-b22f-4a0c-a482-8843618ce040`.

[13] The Strahler number measures the number of branches coming "before" the current branch.

jee, 2002). By examining infant vaccination rates from the 1870s through 1910s, we test whether this violence generated lasting mistrust that reduced vaccination adoption decades after the revolt ended.

Our primary estimating equation is:

$$vaxrate_{id} = \alpha + \beta \cdot Prox_{id} + X'_{id}\gamma + \mu_p + \varepsilon_{id}, \tag{4}$$

where $vaxrate_{id}$ denotes the infant vaccination rate in district i during decade d, measured as vaccinations per capita for children under one year of age. The coefficient of interest, β, captures how proximity to Lucknow relates to vaccination rates. We focus on infants because this population renews annually through births, eliminating the mechanical relationship between past vaccination coverage and current vaccination opportunities that affects older age groups.

We employ two complementary proximity measures for $Prox_{id}$, estimated in separate regressions and constructed as discussed in Section 5.2. First, we use travel costs to Lucknow based on historical transport networks, which capture actual accessibility better than linear distance. British punitive expeditions followed existing railways, rivers, and roads when spreading outward from Lucknow, making districts along these routes more exposed to retributive violence regardless of their straight-line distance. Second, we use linguistic distance to capture cultural proximity to Lucknow (Fenske and Kala, 2021). Districts sharing linguistic ties maintained kinship networks and communication channels that facilitated both participation in the revolt and subsequent transmission of collective memories about British atrocities.

The vector X_{id} includes controls for terrain ruggedness, caloric suitability, minimum distance to the sea and navigable rivers, total population, per capita vaccination spending, and railway access in the corresponding decade. We standardize all variables before estimation, allowing coefficients to be interpreted as the change in vaccination rates (in standard deviations) associated with a one standard deviation change in the explanatory variable. Since our proximity measures increase with distance from Lucknow—higher travel costs and greater linguistic distance for more remote districts—a positive β indicates that districts farther from Lucknow have higher vaccination rates, consistent with persistent mistrust in areas that experienced British retribution.

We estimate Equation (4) separately for each decade from the 1870s through 1910s using OLS with Conley standard errors to account for spatial correlation. Estimating decade-by-decade allows us to trace how the relationship between proximity to Lucknow and vaccination rates evolves as the cohort that experienced 1857 ages and dies. According to our survival estimates in Section 4.2, approximately 70% of the treated cohort (those aged 10–25 in 1857) remained alive in the 1870s, falling to 27% by 1900 and 13% by 1920. This temporal pattern provides a test of our mechanism:

if proximity to Lucknow affects vaccination through memories of 1857, the effect should weaken as this cohort's influence on household decisions diminishes.

Our identification assumes that, conditional on controls and province fixed effects, proximity to Lucknow affects vaccination rates only through exposure to 1857-related violence and its lasting impact on trust. This assumption is plausible for three reasons. First, after the revolt's suppression, the British moved the regional capital from Lucknow to Allahabad in 1858, reducing Lucknow's administrative importance for most districts (Metcalf, 2015). Second, Lucknow was not a major commercial hub for districts outside Awadh. Calcutta, Bombay, and Delhi dominated inter-regional trade networks (Yang, 1999). Third, our placebo tests using these alternative cities find no similar patterns, suggesting that proximity to colonial centers per se does not drive vaccination rates. Province fixed effects, μ_p, further strengthen our argument by absorbing time-invariant differences in administrative capacity, disease environment, and cultural attitudes toward both colonial authority and medical interventions.

To verify that our results reflect the specific legacy of 1857 violence rather than general spatial patterns, we conduct several robustness checks including placebo tests using travel costs to other major colonial centers (Delhi, Calcutta, and Bombay), sensitivity analysis using different distance thresholds, and alternative specifications. We also decompose the distance-to-Lucknow gradient into two components: a discrete Awadh-vs-rest contrast (the area uniformly affected by the 1858 Oudh Settlement) and a continuous within-non-Awadh gradient radiating outward from Lucknow. As we show in Section 6.7, the two components contribute roughly equal shares of the headline effect, with the within-Awadh access gradient being approximately flat—consistent with the Oudh Settlement disrupting the entire Awadh region uniformly rather than smoothly with distance. We discuss these tests in detail in our results section below.

A note on the time variation in our access measure. Because $Access_{id}$ is constructed as a cumulative mean of decadal least-cost travel costs to Lucknow, inter-decade correlations are very high—adjacent-decade correlations from the 1870s onward all exceed 0.99 (see Table 7). These high correlations reflect rank preservation across decades, not equivalence of absolute accessibility: as the rail network expands and other transport links develop, absolute travel costs change substantially even where district rankings do not. That the absolute changes matter for identification is documented in Section 6.7: replacing least-cost travel cost with linear (Euclidean) distance, a measure that preserves the rank ordering but discards time-varying transport content, yields markedly weaker coefficients in every decade. We therefore interpret $Access_{id}$ as a decade-specific *exposure* measure—the average accessibility a district faced through its relevant cohort's formative years—with both the cross-sectional ordering and the time variation in absolute accessibility contributing to identification.

6.2 Geographic Distance

Table 4 reports estimates of equation 4 without province fixed effects. The dependent variable is the standardized infant vaccination rate, calculated as vaccinations per capita for children under one year of age. The main explanatory variable is the standardized measure of travel costs to Lucknow, constructed as the time-averaged least-cost path from each district to Lucknow using historical transportation networks. Higher values indicate greater travel costs and thus worse access to Lucknow.

The coefficient on access to Lucknow is positive and statistically significant in the 1870s (0.711, $p < 0.05$), 1880s (1.412, $p < 0.01$), and 1890s (0.669, $p < 0.001$). The coefficient remains positive but smaller in magnitude in the 1900s (0.279) and the 1910s (0.192), and is not statistically distinguishable from zero in either of those decades under the no-FE specification. Because our access measure increases with travel costs, these positive coefficients indicate that districts with better access to Lucknow (those with lower travel costs) had systematically lower infant vaccination rates.[14] This inverse relationship between proximity to Lucknow and vaccination uptake is consistent with our hypothesis that exposure to British retributive violence following the 1857 revolt generated persistent mistrust in colonial institutions, including public health programs.

The effects are economically large. In the 1880s, a one standard deviation decrease in travel costs to Lucknow is associated with a 1.412 standard deviation decline in infant vaccination rates. With a mean infant vaccination rate of approximately 2% and standard deviation of 1%, this represents a 1.4 percentage point reduction, or about 70% relative to the mean. To put this in context, if a district gained rail access to Lucknow, our estimates imply that vaccination rates would drop from 2 per 100 infants to 0.6 per 100.[15] Given that smallpox caused over 100,000 deaths annually in India during the late nineteenth and early twentieth centuries, with case fatality rates of 20–30%, lower vaccination coverage in the 1870s through 1910s would have led to substantial excess mortality in districts closer to Lucknow (Arnold et al., 1993).

The coefficients follow a clear temporal pattern. The effect nearly doubles from 0.711 in the 1870s to 1.412 in the 1880s. This timing corresponds to when the cohort aged 10–25 during the revolt would have been making vaccination decisions for their own infants. Between the 1880s and 1900s, the coefficient falls by 80%, from 1.412 to 0.279. By the 1910s, it drops to 0.192 and becomes

[14] An outcome-construction robustness check that replaces the vaccination rate with the standardized log count of under-one vaccinations ($\log(1 + \text{under-one vaccinations})$) and controls directly for log total population produces the same temporal pattern. The 1870s, 1880s, and 1890s coefficients remain positive and statistically significant, and the 1900s and 1910s coefficients are small and not significant. Appendix Table 18 reports the regression.

[15] A new rail connection typically reduced travel costs by approximately one standard deviation in our data. As discussed in Section 5.2, we calculate least-cost travel paths using historical transport networks with rail normalized to 1.0 cost unit as per Donaldson (2018). The Lucknow-Bombay route, for example, saw costs fall from 126 to 45 units between 1860 and 1880 (Figures 6a–6b). With mean travel costs around 80 and an interquartile range of 45–115 in the 1860s (implying a standard deviation of approximately 52), this 81-unit reduction represents roughly 1.5 standard deviations (Figure 7).

statistically insignificant. As shown in Section 4.2, the survival probability of the directly affected cohort fell from approximately 70% in the 1870s to 27% by 1900. The declining coefficients track this demographic shift. These patterns are consistent with the hypothesis that proximity to sites of colonial violence affected vaccination behavior primarily through the generation that directly experienced the 1857 events. The weakening and eventual disappearance of the effect as this cohort aged out of child-rearing decisions suggests limited persistence of these effects across generations.

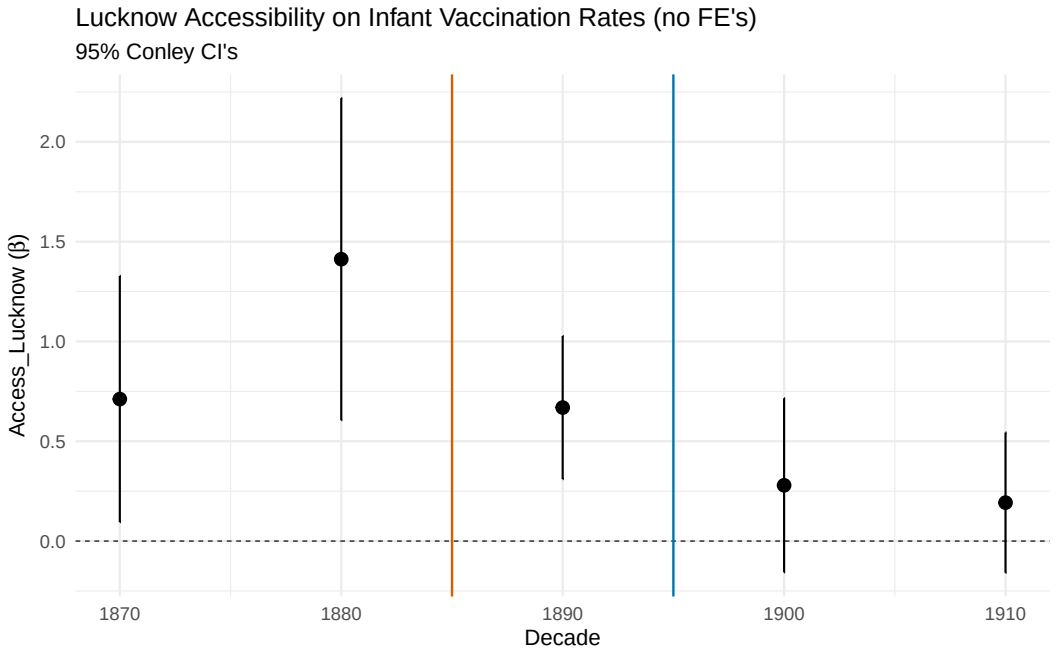


Figure 8: Effect of access to Lucknow on the infant vaccination rate, by decade, without province fixed effects. Each point is the coefficient on standardized cumulative least-cost travel cost to Lucknow from a separate decade-specific regression of the standardized infant vaccination rate on access and the headline control vector (no province fixed effects). Vertical bars are 95 percent Conley spatial HAC confidence intervals. A positive coefficient means districts with higher travel cost to Lucknow had higher infant vaccination. The sample is restricted to districts within 1,000 km of Lucknow.

6.3 Fixed Effects Results

Table 5 reports estimates of equation 4 with province fixed effects. Including these fixed effects accounts for time-invariant provincial characteristics that could influence both travel costs and vaccination rates, such as differences in administrative practices, disease environments, or prevailing attitudes toward colonial medicine. Figure 9 plots these coefficients with 95% Conley confidence intervals.

The coefficient on access to Lucknow remains positive across all decades. The estimates are 1.1 in the 1870s ($p < 0.001$), 1.7 in the 1880s ($p < 0.01$), and 0.8 in the 1890s ($p < 0.001$). The coefficients decline in the 1900s (0.6, $p < 0.05$) and 1910s (0.4, $p < 0.10$). As before, these positive

coefficients indicate that districts with better access to Lucknow (those with lower travel costs) had lower infant vaccination rates.

The estimated coefficients under fixed effects are broadly consistent with the baseline results, with magnitudes of similar or slightly larger size in the key decades. For example, the 1880s estimate under fixed effects remains large at 1.7, indicating a strong relationship within provinces between proximity to Lucknow and lower vaccination rates. The estimates are highest in the 1880s and decline in the following decades, which aligns with the proposed cohort mechanism. Including province fixed effects helps address the possibility that differences in administration or disease environments across provinces account for the observed relationship.[16]

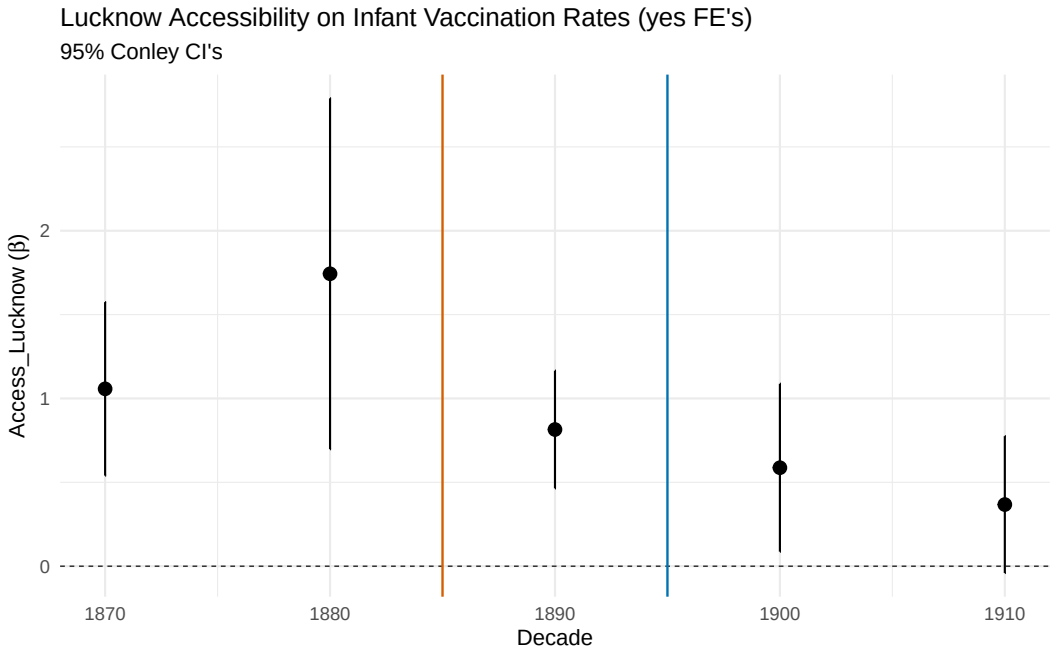


Figure 9: Effect of access to Lucknow on the infant vaccination rate, by decade, with province fixed effects. Each point is the coefficient on standardized cumulative least-cost travel cost to Lucknow from a separate decade-specific regression of the standardized infant vaccination rate on access, the headline control vector, and province fixed effects. Vertical bars are 95 percent Conley spatial HAC confidence intervals. A positive coefficient means districts with higher travel cost to Lucknow had higher infant vaccination. The sample is restricted to districts within 1,000 km of Lucknow.

[16]Lucknow district itself is in our 1,000 km baseline sample at zero distance from itself, so one might worry that the gradient is mechanically driven by including the focal observation. Re-estimating the yes-FE specification on the subsample excluding Lucknow district yields essentially identical coefficients: the largest absolute change is 0.007 in the 1910s (from 0.367 to 0.361); the headline 1880s estimate is unchanged to two decimal places (1.7434 vs. 1.7406). The mechanism we identify reflects spatial spillover among districts exposed to retribution at Lucknow, not a single-observation effect of Lucknow itself. Appendix Figure 17 extends this leave-one-out check to every district in the 1880s sample. The 1880s access coefficient remains positive across all 78 leave-one-out samples, confirming that the result is not driven by any single district.

6.4 Within-District Identification

The province fixed effects in Section 6.3 absorb time-invariant *provincial* characteristics such as administrative practices, average disease environment, and prevailing colonial attitudes. They do not absorb characteristics that vary across districts within a province. Permanent geographic features, fixed cultural composition, the entire Awadh-vs-non-Awadh contrast embedded within the Northwestern Provinces, and time-invariant institutional differences across districts remain potential confounders.

To address these district-level confounders we stack the district-decade observations and estimate a single panel regression with district and decade fixed effects. The specification is

$$vaxrate_{id} = \sum_{d' \in \{1880,1890,1900,1910\}} \beta_{d'} \left(Access_{id} \cdot \mathbf{1}\{d = d'\} \right) + X'_{id}\gamma + \eta_i + \tau_d + \varepsilon_{id}, \tag{5}$$

where $vaxrate_{id}$ is the standardized infant vaccination rate in district i in decade d, $Access_{id}$ is the standardized cumulative least-cost travel cost from district i to Lucknow as of decade d, X_{id} is the control vector (distance to the rail network, total population, and vaccination spending per capita, all standardized), η_i is a district fixed effect, τ_d is a decade fixed effect, and ε_{id} is the error term. The 1870s is the omitted decade. We cluster standard errors by district.

District fixed effects absorb every time-invariant district characteristic, and decade fixed effects absorb shocks common to all districts. The decade-specific access coefficients $\beta_{d'}$ are identified from variation in access to Lucknow across decades within the same district, driven primarily by the expansion of the rail network and the consequent evolution of cumulative least-cost travel cost.[17]

Figure 10 plots the estimated $\beta_{d'}$ with 95% confidence intervals. The cohort-fade pattern survives this more demanding identification. The 1880s coefficient remains positive and statistically significant ($\hat{\beta}_{1880} = 0.259$, $p < 0.001$). The 1890s coefficient is roughly half its 1880s value ($\hat{\beta}_{1890} = 0.144$, $p = 0.05$). The 1900s and 1910s coefficients are indistinguishable from zero. The within-district peak coincides with the cross-sectional peak from Section 6.3, and the fade trajectory mirrors the conditional survival probability of the 1832–47 cohort plotted in Figure 2. Appendix Table 15 reports the regression alongside a balanced-panel variant restricted to the 62 districts observed in every decade. The 1880s coefficient and the fade pattern are preserved in the balanced sample.

[17] To probe whether the access result is sensitive to endogenous railway placement we also instrument cumulative least-cost access with straight-line geographic distance to Lucknow in the cross-sectional decade-by-decade specification. Appendix Table 17 reports OLS and IV estimates side by side. First-stage F-statistics range from 281 to 483, well above conventional weak-instrument thresholds. The IV coefficients are at least as large as their OLS counterparts in every decade, which is consistent with the view that endogenous railway placement is not the source of the headline relationship.

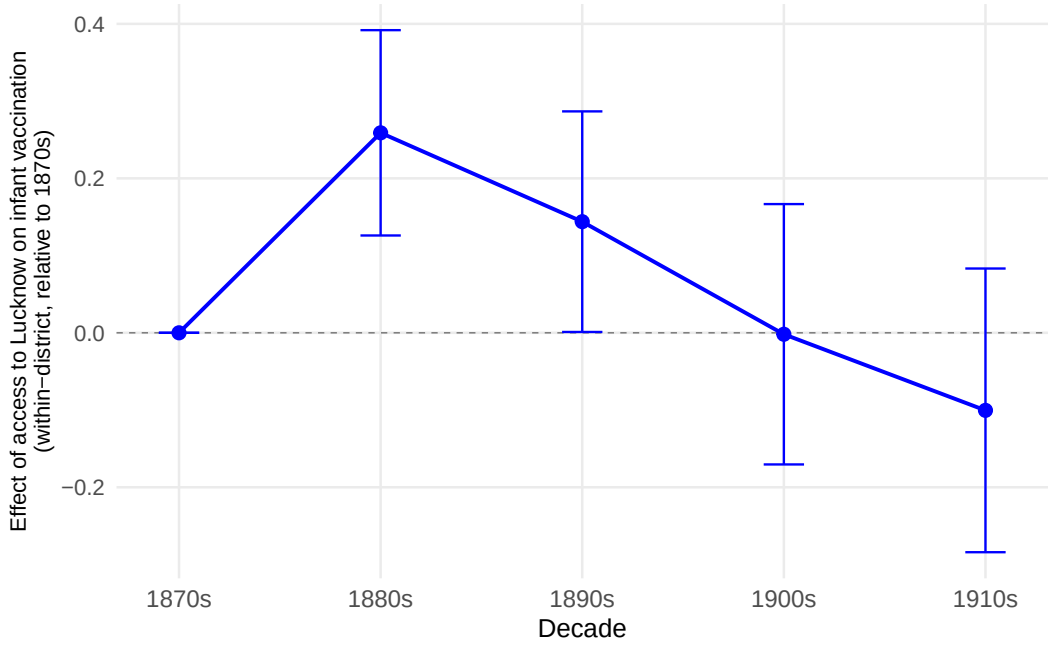


Figure 10: Within-district effect of access to Lucknow on infant vaccination by decade. Each point reports the coefficient on standardized cumulative least-cost travel cost to Lucknow interacted with a decade indicator from a single panel regression with district and decade fixed effects. The 1870s is the omitted baseline. The specification absorbs all time-invariant district characteristics. Vertical bars are 95% confidence intervals based on standard errors clustered by district. The sample is restricted to districts within 1,000 km of Lucknow. See Appendix Table 15 for the underlying regression and a balanced-panel variant.

6.5 Linguistic Distance

Table 6 reports results using linguistic distance from Lucknow, constructed from the Ethnologue database (see Section 5.2.2). Higher values indicate that the primary languages spoken in two districts share fewer linguistic branches and are therefore more linguistically distant from each other. This measure captures cultural connections that operated independently of transport routes—marriage networks, religious pilgrimages, and labor migration typically followed language boundaries rather than railways or rivers (Yang, 1999; Reed, 2022, Chapter 3). As anticipated in Section 4.2, the two parameterizations of the branch-decay parameter δ proxy for two distinct mechanisms. The Fearon-Laitin parameterization ($\delta = 0.5$) is most sensitive to close-kin linguistic ties and proxies for the cohort-driven channel. The Fenske-Kala parameterization ($\delta = 0.05$) is most sensitive to broad cultural distance and proxies for a longer-lived horizontal-transmission channel.

6.5.1 Close-kin transmission ($\delta = 0.5$)

Under the close-kin parameterization, the coefficient on linguistic distance increases from 0.184 in the 1870s to 0.462 in the 1880s ($p < 0.05$), declines to 0.172 in the 1890s, and falls to 0.005 and 0.016 in the final two decades; only the 1880s coefficient is statistically distinguishable from zero at conventional levels. In the 1880s, a one standard deviation increase in close-kin linguistic distance corresponds to a 0.462 standard deviation increase in vaccination rates, approximately five additional infant vaccinations per 1,000 inhabitants annually. The peak-and-fade pattern matches the geographic-access result and is consistent with the cohort-attenuation channel: language networks among close kin transmitted the memory of 1857 as long as the cohort that lived through the revolt was making household decisions, and this transmission weakened as that cohort aged out.

Cohort-survival formalization. The decade-by-decade pattern above is suggestive but cross-sectional. A sharper test pools the panel and asks directly whether the close-kin linguistic effect varies with predicted cohort survival. We stack the district-decade observations and estimate

$$vaxrate_{id} = \beta_L \left(\mathrm{LD}_i^L \cdot S_d\right) + \beta_P \left(\mathrm{LD}_i^P \cdot S_d\right) + X_{id}'\gamma + \eta_i + \tau_d + \varepsilon_{id}, \tag{6}$$

where LD_i^L is district i's standardized linguistic distance to Lucknow, LD_i^P is the corresponding standardized distance to a comparison city P, and S_d is the predicted survival probability of the cohort aged 10–25 in 1857 in decade d. District fixed effects absorb the main effects of the two linguistic-distance variables, both of which are time-invariant within district. Decade fixed effects absorb the main effect of survival. Identification of β_L and β_P comes from the interaction between linguistic distance and cohort survival within the same district across decades. Controls X_{id} are distance to the rail network, total population, and vaccination spending per capita, all standardized. Standard errors are Conley spatial HAC.

Table 11 reports the estimates for six pairwise comparisons.[18] The Lucknow interaction $\hat{\beta}_L$ is positive across all six pairwise comparisons and statistically significant in five of them. Point estimates range from 0.629 ($p < 0.01$) against Dacca to 0.696 ($p < 0.05$) against Bombay. The Calcutta column is the exception, with $\hat{\beta}_L = 0.373$ ($p = 0.36$). The placebo interactions $\hat{\beta}_P$ are statistically indistinguishable from zero in five of the six comparisons. The one exception is Lahore, where $\hat{\beta}_P = -0.756$ ($p < 0.001$) and the placebo coefficient runs in the opposite direction from the Lucknow effect. The cohort-attenuation pattern is therefore specific to linguistic proximity

[18] Delhi is excluded from this placebo set because Awadhi and the Delhi dialects of Hindustani sit in the same Western-Hindi branch of the Fearon-Laitin tree. Linguistic distances to Lucknow and Delhi correlate at 0.99998 under $\delta = 0.5$ and at 1.0000 under $\delta = 0.05$, so Delhi is mechanically collinear with Lucknow in any linguistic-distance regression. Delhi remains a meaningful placebo in the geographic-access regressions reported in Section 6.7, where travel costs to Lucknow and Delhi are not collinear at the district level.

to Lucknow rather than reflecting a generic relationship between linguistic distance from major South Asian cities and infant vaccination.[19,20]

6.5.2 Broad cultural transmission ($\delta = 0.05$)

The picture under the broad-cultural parameterization is different. The linguistic-distance coefficient grows over time rather than fading: from 0.43 ($p < 0.05$) in the 1870s, through 0.51 in the 1880s and 0.44 ($p < 0.05$) in the 1890s, to 0.83 ($p < 0.01$) in the 1910s (Appendix Table 10). The two parameterizations appear to pick up two distinct channels operating in parallel: close-kin linguistic transmission, captured by $\delta = 0.5$, fades with the dying cohort, while broad cultural distance, captured by $\delta = 0.05$, reflects more durable inter-group differences in how colonial institutions are received—differences that need not depend on the survival of any specific birth cohort and that may widen if subsequent events (later epidemics, the 1919 Amritsar Massacre, the rising non-cooperation movement) reinforce the cultural cleavage. Memory alone, transmitted through close kinship, was not sufficient to sustain resistance to colonial vaccination beyond the immediate cohort; broader cultural distance can sustain the gap on its own.

[19]The analogous specification under the $\delta = 0.05$ parameterization also yields a positive Lucknow $\times$ survival interaction (point estimates 1.26 to 2.16 across the six pairs, significant at the ten-percent level or better in five of six). This indicates that the broad-cultural channel discussed in Section 6.5.2 carries a cohort-attenuating component once the secular trend is absorbed by the decade fixed effects.

[20]Restricting the sample to districts within 5,000 km of Lucknow rather than the 1,000 km baseline used here gives the same qualitative pattern with larger magnitudes and tighter confidence intervals. The Lucknow $\times$ survival coefficient rises to the range 0.74 to 1.01 across the six pairwise comparisons under the wider sample.

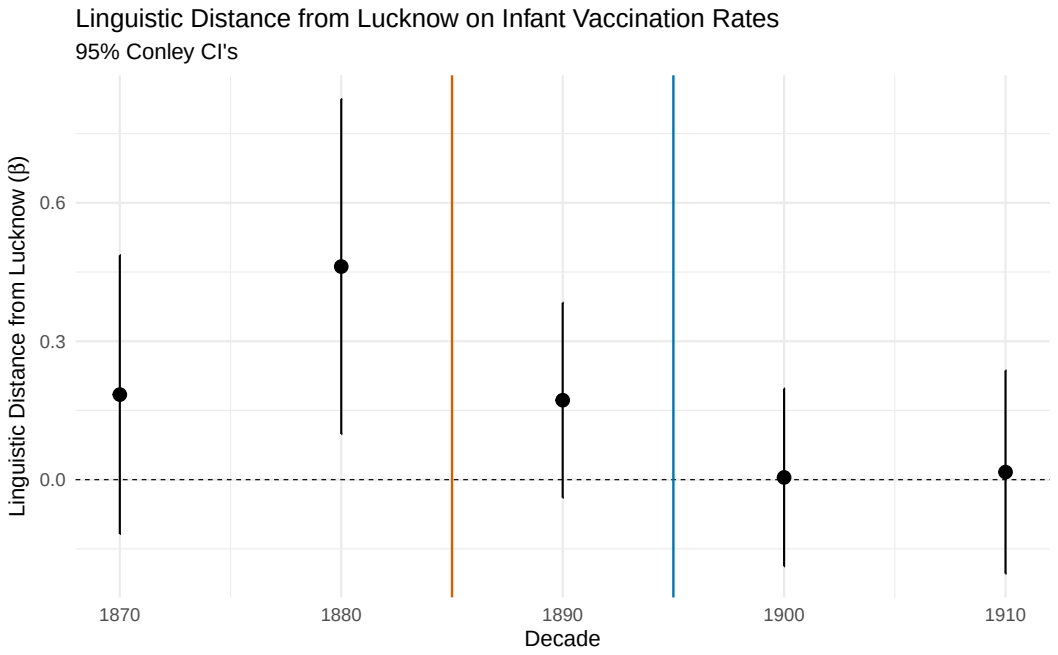


Figure 11: Effect of linguistic distance to Lucknow on the infant vaccination rate, by decade, without province fixed effects. Each point is the coefficient on standardized linguistic distance to Lucknow (Fearon-Laitin parameterization, $\delta = 0.5$) from a separate decade-specific regression of the standardized infant vaccination rate on linguistic distance and the headline control vector (no province fixed effects). Vertical bars are 95 percent Conley spatial HAC confidence intervals. A positive coefficient means districts more linguistically distant from Lucknow had higher infant vaccination. The sample is restricted to districts within 1,000 km of Lucknow.

6.6 Mechanism: Heterogeneity by Local Sepoy Activity

The cross-sectional and within-district results above show that proximity to Lucknow predicts lower infant vaccination, and the linguistic-distance results show that the pattern is specific to closeness in language as well as in travel. A natural next question is whether the effect is concentrated in districts that directly experienced the violence of 1857. We construct a district-level exposure indicator from documented sepoy events in 1857 and test whether the access-to-Lucknow gradient is steeper among exposed districts.

The exposure data come from David (2001), who lists 145 documented sepoy mutinies, disarmaments, and disbandments across the Bengal Army in 1857 (we exclude civilian jail-break events because the spatial concentration of jail-break locations reflects the placement of British administrative infrastructure rather than the geography of sepoy revolt).[21] Figure 12 plots the locations of

[21]The 145-event count excludes ten further events for which place names appear in David (2001) but could not be located. Place names were geocoded by hand using modern coordinates for the listed towns and cities, which David (2001) describes as relatively well-known. The raw event lists are in the replication archive under `Data/1857_additional_data/raw/`. The district-level exposure file is in the same archive under `Data/1857_additional_data/processed/`.

the events overlaid on the districts in our panel and the 1,000 km sample buffer around Lucknow. The events concentrate in the northern Gangetic plain, with a secondary cluster in Punjab and isolated incidents in central and western India.

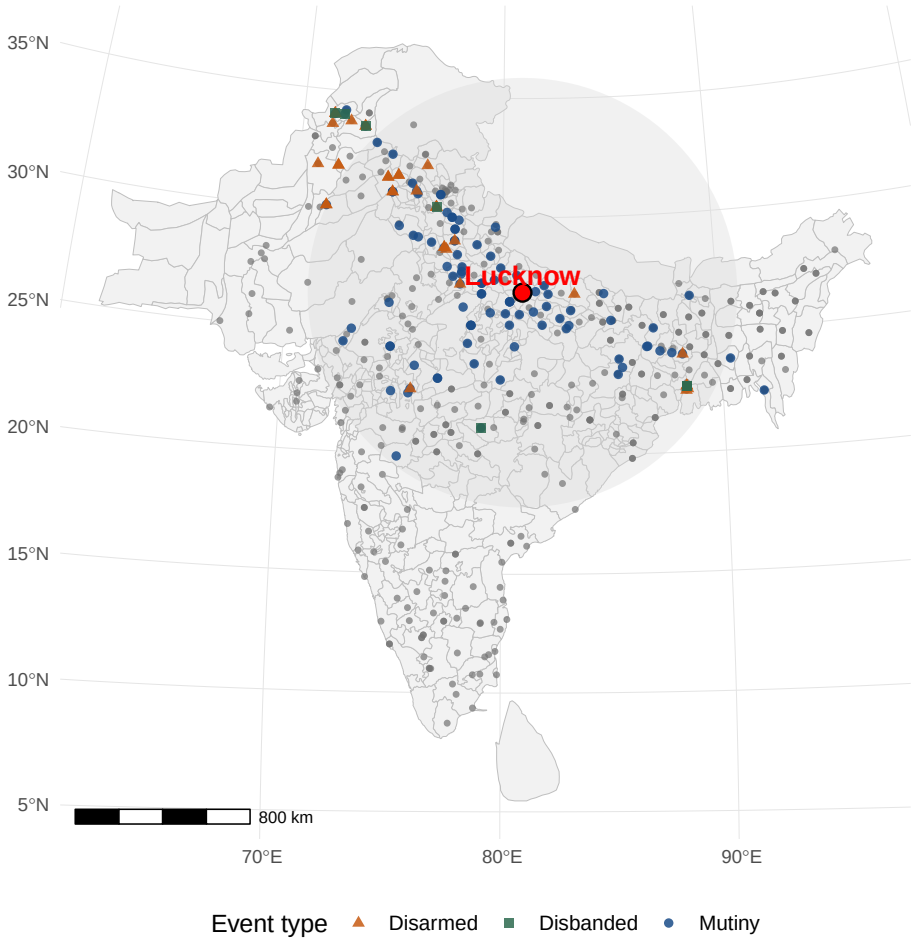


Figure 12: Location of documented sepoy events in 1857 (mutinies, disarmaments, and disbandments) overlaid on the districts in our panel. Lucknow is marked in red and the grey shaded region is the 1,000 km sample buffer. Events are drawn from David (2001). Jail-break events are excluded. Districts in our panel are shown as small grey points.

For each district we construct two binary exposure indicators, Sepoy50_i and Sepoy100_i, equal to one if the district lies within 50 km or 100 km respectively of at least one documented sepoy event. In the 1880s sample of 76 districts within 1,000 km of Lucknow, 46 districts satisfy the 50 km threshold and 60 satisfy the 100 km threshold. We estimate

$$vaxrate_{id} = \beta_1 Access_{id} + \beta_2 Sepoy_i + \beta_3 \left(Access_{id} \cdot Sepoy_i\right) + X'_{id}\gamma + \mu_p + \varepsilon_{id}, \tag{7}$$

separately by decade, where $Sepoy_i$ is one of the two exposure indicators, $Access_{id}$ is standardized

cumulative least-cost travel cost to Lucknow as in our baseline specification, X_{id} is the control vector from Equation (4), μ_p is a province fixed effect, and standard errors are Conley spatial HAC. The specification includes the main effects of $Access_{id}$ and $Sepoy_i$ separately, so the interaction β_3 is identified from the difference between the access-vaccination gradients in sepoy-exposed and non-exposed districts within province cells, not from the geographic overlap between sepoy events and proximity to Lucknow.

The coefficient of interest is β_3. Because $Access_{id}$ is positive when a district is farther from Lucknow and our headline finding is that vaccination rates rise with $Access_{id}$, a positive β_3 means the access-vaccination gradient is steeper in sepoy-exposed districts. Equivalently, closeness to Lucknow depresses vaccination more in districts that directly experienced documented sepoy activity in 1857.

Figure 13 plots $\hat{\beta}_3$ across the five decades, separately for the 50 km and 100 km exposure thresholds. The interaction is positive and statistically significant in the 1880s under both thresholds, with $\hat{\beta}_3 = 0.79$ ($p < 0.001$) at 50 km and $\hat{\beta}_3 = 1.01$ ($p = 0.012$) at 100 km. The implied magnitude is substantial. In the 1880s, the access-to-Lucknow gradient is approximately twice as large in districts within 50 km of documented sepoy activity as in districts without. The interaction is positive in all four remaining decades under the 50 km threshold and in three of four under the 100 km threshold, but none of these other decades reach conventional significance. The 1880s estimate aligns with the headline cohort-aging story. That decade is when the cohort aged 10 to 25 in 1857 was most relevant for infant vaccination decisions, and the result says the access-to-Lucknow effect in that decade is concentrated in districts that directly experienced sepoy activity. Appendix Table 12 reports the full decade-by-decade estimates.

Figure 13: Interaction coefficient $\hat{\beta}_3$ from Equation (7) by decade and exposure threshold. Vertical bars are 95% confidence intervals based on Conley spatial HAC standard errors. A positive coefficient means the access-to-Lucknow gradient is steeper in districts with documented sepoy activity within the stated radius than in districts without.

Placebo Regressions and Robustness

The geographic and linguistic results above are consistent with one another but each invites the question whether the underlying pattern reflects the specific legacy of 1857 violence rather than some general spatial gradient (proximity to any major colonial center, the inland–coastal axis, the Awadh-vs-non-Awadh contrast embedded in distance to Lucknow, or sample composition over the decades). We address each of these in turn through placebo regressions on alternative cities, sensitivity analyses on the construction of our distance measure, and a within-province decomposition that separates the Awadh-region effect from the continuous distance gradient. The results below collectively support our interpretation that proximity to Lucknow affects vaccination rates through the specific historical channel of mistrust generated by the 1857 revolt.[22]

[22] A complementary sample-composition check is reported in Appendix Table 16, which progressively restricts the estimating sample to the regions most directly connected to the events of 1857. The access-to-Lucknow coefficient remains positive across all decades when the sample is restricted to the North-Western Provinces and Oudh, the same region plus Punjab, or that region plus Punjab and Bengal. The result is therefore not driven by comparisons between districts near Lucknow and districts in southern or western India.

6.7 Placebo Tests

Figure 14 presents coefficients from regressions using travel costs to Delhi, Calcutta, and Bombay as alternative proximity measures. These cities served different functions in colonial India: Delhi was the former Mughal capital and site of intense but localized fighting in 1857 (Dalrymple, 2007); Calcutta was the administrative capital of British India; and Bombay was the primary commercial port on the western coast. If our results simply captured proximity to any major colonial center, we should observe similar patterns for these cities.

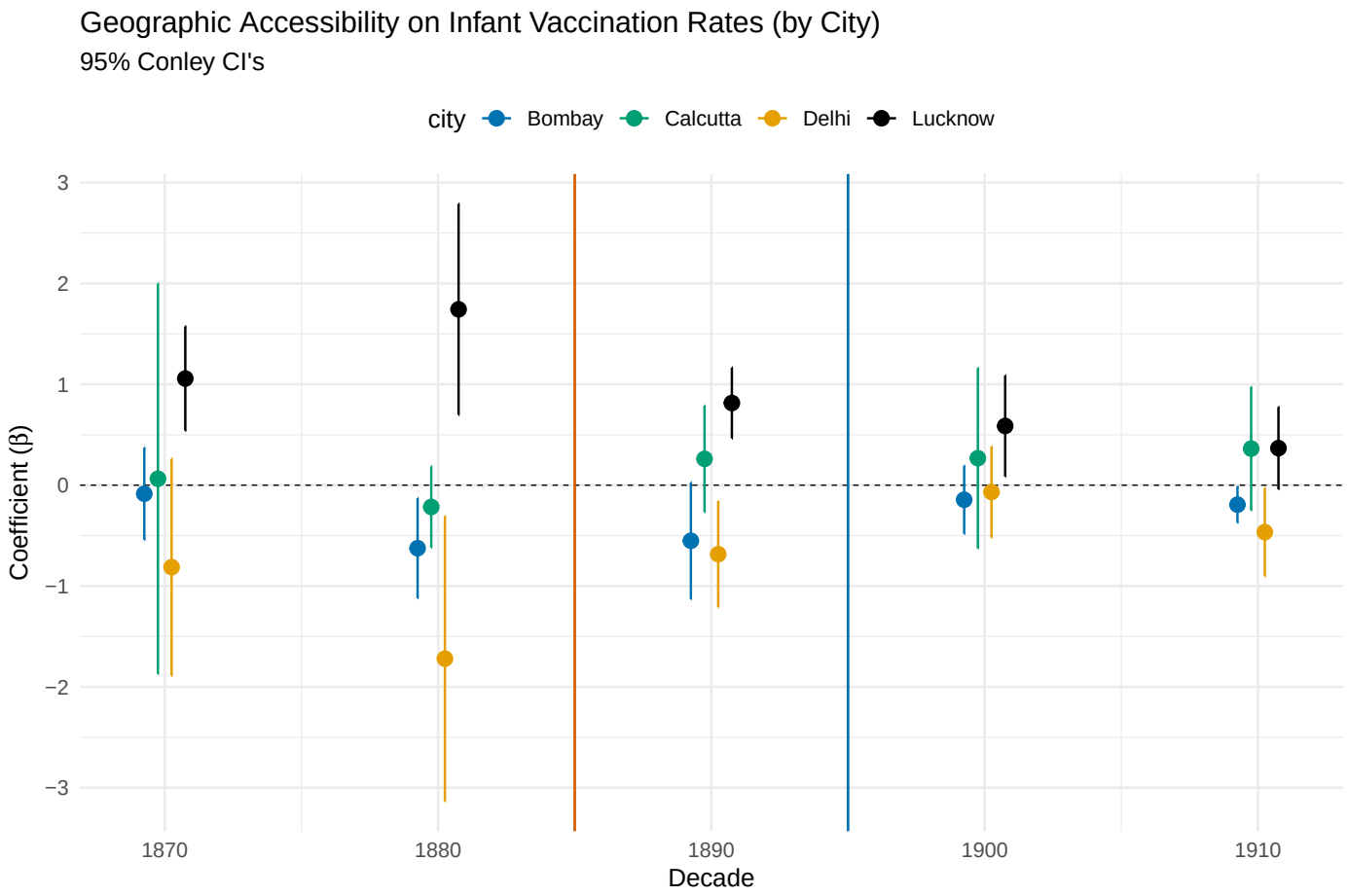


Figure 14: Coefficients on access to placebo cities (Delhi, Calcutta, Bombay), by decade, from province-fixed-effects regressions of the standardized infant vaccination rate on standardized cumulative least-cost travel cost to each placebo city and the headline control vector. Each subpanel reports a separate decade-by-decade specification with one placebo city in place of Lucknow. Vertical bars are 95 percent Conley spatial HAC confidence intervals. Unlike Lucknow (Figure 9), none of the placebo cities show consistent positive coefficients that peak in the 1880s. The sample is restricted to districts within 1,000 km of each placebo city respectively.

The placebo results show no consistent relationship between proximity to these alternative centers and vaccination rates. For Delhi, coefficients are negative but statistically insignificant across all decades, ranging from -0.778 in the 1870s to -0.018 in the 1900s. Calcutta shows marginally positive

coefficients in some decades but with wide confidence intervals and no clear temporal pattern. Bombay's coefficients fluctuate around zero without statistical significance. Unlike Lucknow, which shows consistent positive coefficients that peak in the 1880s and decline thereafter, none of these cities show the decade-specific pattern consistent with our cohort-based explanation. These null results, particularly in the case of Delhi where urban violence during 1857 was severe but did not generate rural institutional disruption as in Awadh, rule out the claim that exposure to 1857 violence alone explains long-run variation in vaccination uptake and instead support the view that the persistence of mistrust was rooted in the dismantling of rural institutional structures (Dalrymple, 2007; Mukherjee, 2002).

Delhi as a placebo: rural-vs-symbolic distinction. The Delhi placebo merits more discussion than the other two given how prominent Delhi was in the events of 1857. Delhi was the symbolic capital of the revolt, the seat of Bahadur Shah Zafar's restored Mughal court, the site of the British siege and storming of the city, and the location of the trial and execution of Mughal princes. Mass executions and demolitions in the Old City are extensively documented (Dalrymple, 2007). We treat Delhi as a placebo nonetheless because the post-1857 disruption in and around Delhi was concentrated in the city itself rather than diffused into the rural hinterland through institutional dismantling. The Awadh experience differed in kind, not only in degree: the Oudh Settlement of 1858–61 dispossessed approximately 21,000 *taluqdars* (Mukherjee, 2002), dissolving the patron–client relationships through which colonial authority had previously been mediated. Around Delhi, by contrast, the British re-established administrative institutions quickly and the rural revenue settlement was substantially less disruptive. Distance to Delhi therefore captures a different treatment than distance to Lucknow—one centered on urban siege violence, the other on the rural collapse of customary protections.

A horse-race specification including both access measures simultaneously formalizes this distinction. As shown in Appendix Table 9, the Lucknow coefficient remains positive and strongly significant ($\hat{\beta} = 1.45$, $p < 0.001$) when we condition on access to Delhi, while the Delhi coefficient is negative and statistically significant ($\hat{\beta} = -1.27$, $p < 0.001$). The signs are consistent with the interpretation above: districts closer to Lucknow had *lower* infant vaccination (the mistrust effect), whereas districts closer to Delhi had *higher* infant vaccination after conditioning on Lucknow access (consistent with Delhi as a center of post-1858 administrative re-establishment, where colonial vaccination programs were actively pursued). The relevant null—no Lucknow effect once we account for proximity to other major colonial centers—is decisively rejected.

Within-Awadh placebo: planned extension. A more demanding placebo design would compare districts *within* the post-1858 Awadh region, exploiting variation in the intensity of British retribution—specifically, the fraction of district land confiscated under the Oudh Settlement, or the

fraction of *taluqdars* dispossessed but not subsequently restored. We do not implement this design in the current draft because the requisite confiscation data are still being assembled by the second author, but it is the natural extension of the present analysis: it would replace the cross-Awadh-vs-non-Awadh comparison embedded in our distance-to-Lucknow measure with a finer within-Awadh contrast that holds the broader institutional and cultural context fixed. We discuss this extension as planned future work in the conclusion.

6.7.1 Robustness to Distance Thresholds and Sample Composition

Appendix Figure 25 examines the sensitivity of our results to the maximum distance threshold used to define our sample. Our baseline specifications restrict the sample to districts within 1,000 kilometers of Lucknow, but this choice is necessarily arbitrary. We therefore re-estimate our main specification using maximum distance cutoffs ranging from 500 to 1,500 kilometers. The results are stable in the 1870s and 1880s across all distance thresholds. At 500 kilometers, including only districts near Lucknow, the 1880s coefficient reaches 2.5, indicating stronger effects closer to the city. As we expand the radius to include more distant districts, coefficients become weaker but remain economically and statistically significant through 1,250 kilometers. Only at 1,500 kilometers, when we include districts as far as Tamil Nadu and Assam, do the effects weaken substantially. This pattern confirms that our results are driven by variation among districts plausibly connected to Lucknow rather than spurious correlations at extreme distances.

6.7.2 Alternative Distance Measures and Specifications

To assess whether our results depend on the construction of the proximity measure, we re-estimate the baseline regressions using linear (Euclidean) distance to Lucknow rather than least-cost travel time. Although the two measures are highly correlated at the district level ($r = 0.97$), the resulting estimates are vastly different. As shown in Appendix Figure 15, the coefficients using linear distance are smaller in magnitude, less precise, and statistically insignificant across all decades. For example, the 1880s coefficient declines from 1.7 under least-cost distance to 0.6 under linear distance, while standard errors rise by roughly 70 percent.

This divergence arises from differences in how the two measures capture exposure to British repression in 1857. Least-cost travel time incorporates terrain, rivers, and transport infrastructure, and therefore reflects actual accessibility to Lucknow. Linear distance, on the other hand, assumes uniform travel conditions across space, and fails to account for geographic and infrastructural barriers that shaped British troop movement. As a result, linear distance misclassifies districts that were formally close but effectively inaccessible due to natural obstacles, and vice versa.

The divergence between the two distance measures is most relevant in the region surrounding Lucknow, where the revolt and its suppression were concentrated. Although our regressions include

the full set of districts, the identifying variation comes largely from those close enough to have plausibly experienced direct exposure to colonial violence. Among these districts, even modest absolute differences between linear and least-cost distance translate into substantial differences in ranking, leading to shifts in how proximity—and by extension, exposure—is classified. This reordering matters for estimation: it affects which districts are treated as more or less exposed. For more distant districts, where British reprisals were unlikely and where the two measures align closely, the choice of proximity measure has little influence on the estimates.

Only the least-cost measure yields statistically significant results, which is consistent with the historical mechanism we aim to capture. If the estimated relationship were driven by broad spatial gradients or omitted variables correlated with distance, both measures—least-cost and linear—should produce similar effects. The fact that linear distance does not, despite its high correlation with least-cost distance, indicates that actual accessibility shaped exposure to repression. Districts that were geographically close but inaccessible due to terrain or lack of transport links experienced less violence, and the data reflect that difference. This strengthens the credibility of the identification strategy in our analysis.

6.8 Sample Composition and Alternative Specifications

Our unbalanced panel raises concerns that changing sample composition might drive the temporal patterns we observe. Districts entering the dataset in later decades might differ systematically from those present throughout, potentially confounding our cohort-based interpretation. To address this, we re-estimate our main specifications using only the 149 districts with vaccination data in the 1890s—our most complete decade—for all time periods. The results remain qualitatively unchanged: coefficients rise from the 1870s to peak in the 1880s before declining, with the temporal pattern remaining the same despite the restricted sample. The 1880s peak effect remains large at 1.7, though confidence intervals widen due to the reduced sample size.

Our main results are also robust to excluding province fixed effects, as shown in Tables 4 and 5. While the inclusion of fixed effects improves precision and accounts for provincial differences in administrative capacity, the temporal pattern of effects remains consistent in both specifications. These robustness checks further validate that our findings reflect the specific historical legacy of the 1857 revolt rather than mechanical features of our data or estimation strategy.

6.9 Awadh-vs-NWP Within-Province Falsification

The referee for an earlier version of this paper raised the concern that our distance-to-Lucknow measure cannot distinguish 1857-retribution mistrust from 1856-annexation effects, because the Oudh Settlement of 1858–61 dispossessed approximately 21,000 *taluqdars* (Mukherjee, 2002) and produced near-universal disruption across the Awadh region before subsequent retribution. To

probe this, we restrict our sample to the historical *northwestern* province (which encompassed both Awadh proper and the older Northwestern Provinces) and add an indicator for the twelve historical Awadh districts present in our panel.[23] We pool across decades, include decade fixed effects, and interact our access measure with the Awadh indicator.

The result (Appendix Table 14) is informative in a way we did not anticipate. Awadh districts have a substantially *lower* infant vaccination rate than other northwestern-province districts even after conditioning on access to Lucknow ($\hat{\beta}_{\text{Awadh}} = -0.748$, $p < 0.001$). The within-region access gradient—the coefficient on *access* in non-Awadh districts—is positive and significant ($\hat{\beta}_{\text{access}} = 0.396$, $p < 0.001$), but the gradient is much flatter in Awadh: the interaction term is -0.352 ($p < 0.001$), so the within-Awadh access gradient is approximately $0.396 - 0.352 = 0.044$, statistically and substantively close to zero.[24]

We interpret this as evidence that the Oudh-Settlement disruption was approximately uniform across the Awadh region: within Awadh, additional travel-cost distance from Lucknow buys very little additional "treatment intensity," because the institutional collapse following 1858 affected the region as a whole. The smooth distance gradient in our headline cross-section is therefore a mix of two components: (i) the Awadh-vs-non-Awadh contrast, which captures the fact that the entire Awadh region was treated, and (ii) a within-non-Awadh access effect, which captures heterogeneous exposure to retributive violence radiating outward from Lucknow. We do not view this decomposition as undercutting our cohort-based interpretation: the Awadh-region effect itself fits naturally with the timeline of post-1858 dispossession and the dying-out of the cohort of household heads that lived through it. But it does mean that researchers wishing to identify a distance gradient *within* the affected region (the within-Awadh placebo design the referee suggested as more powerful) will not find one in our data—not because the channel is absent, but because the treatment was approximately uniform in space within the Awadh boundaries. The within-Awadh placebo design awaits the finer-grained taluqdar-confiscation data being assembled by the second author and is the natural extension of the present analysis.

[23] Lucknow, Fyzabad (Faizabad), Sultanpur, Bahraich, Gonda, Hardoi, Sitapur, Kheri, Partabgarh (Pratapgarh), Rae Bareili, Unao (Unnao), and Bara Banki (Mukherjee, 2002, ch. 1). The same twelve districts appear on the United Provinces acquisition-year list digitised from Reeves (1961, p. 14), providing an independent corroboration. Lakhimpur appears merged with Kheri in our administrative coding. The remaining 42 districts in the northwestern province constitute the non-Awadh comparison group.

[24] A complementary robustness check removes the twelve Awadh-1856 districts entirely from the full sample and re-estimates the headline province-fixed-effects specification. The access-to-Lucknow coefficient remains positive across all decades and significant at the five-percent level or better through the 1900s. Relative to the full-sample headline the 1880s coefficient attenuates by about 26 percent and the 1890s by about 25 percent. The headline finding therefore does not require the Awadh districts to be present in the sample.

6.10 Mortality and Welfare

A natural follow-up question is whether the proximity-to-Lucknow gradient in vaccination uptake translates into a corresponding gradient in smallpox mortality. The colonial vaccination reports record annual smallpox deaths per 1,000 inhabitants alongside vaccination counts, providing a direct (if imperfect) test. Two caveats apply. First, the mortality data come from the vaccination reports rather than from census-based death-deviation reconstructions, and reporting was uneven across provinces and years. Nontrivial fractions of district-decade observations have no mortality figure at all. Second, the 1870s contain no mortality coverage in our digitised sample, and 1880s coverage ($N = 48$) is too sparse to identify the controlled specification. We therefore restrict the mortality regression to the 1890s, 1900s, and 1910s.

Reduced-form result. Table 13 reports the headline yes-FE specification with smallpox deaths per 1,000 on the left-hand side. The 1890s access coefficient is -0.412 ($p < 0.01$): a one-standard-deviation increase in access cost (a worsening of accessibility roughly equivalent to moving from the 25th to the 75th percentile of the 1890s sample) is associated with $0.41 \times 0.27 \approx 0.11$ fewer smallpox deaths per 1,000 inhabitants annually, where 0.27 is the standard deviation of the 1890s mortality outcome. Equivalently, districts closer to Lucknow had *more* smallpox deaths in the 1890s, a direction consistent with the chain implied by our main results: proximity to Lucknow depressed vaccination uptake, and the resulting unvaccinated cohorts faced higher smallpox mortality in childhood. The 1900s and 1910s point estimates are essentially zero and not statistically distinguishable from zero. The 1900s and 1910s mortality outcomes also have heavily right-skewed distributions (sample standard deviations of 1.03 and 5.14, against medians of 0.34 and 0.22 respectively), which limits what can be inferred from those decades.

Translating to absolute welfare via medical parameters. A formal instrumental-variables strategy that uses access to Lucknow as an instrument for vaccination is unattractive in this setting because access plausibly affects smallpox mortality through channels other than vaccination—economic disruption following the 1858 Oudh Settlement, post-1857 displacement, differences in pre-existing variolation prevalence, and other public-health infrastructure—each of which would violate the exclusion restriction. We therefore proceed by triangulation. Three quantities are well-established in the medical and epidemiological literature on smallpox: the case fatality of *variola major* in childhood, approximately 25–30% (Dixon, 1962; Banthia and Dyson, 1999); the lifetime smallpox attack rate in unvaccinated populations in pre-eradication settings, approximately 70–90% (Hopkins and Lythcott, 1983; Fenner et al., 1988); and the efficacy of vaccinia against severe disease, approximately 95% with some waning over five to ten years (Fenner et al., 1988; Hammarlund et al., 2003). Combining these implies that an unvaccinated infant in colonial India faced roughly a 20–25% lifetime probability of dying from smallpox in childhood, of which vaccination would prevent about 95%, or 0.19–0.24 averted child deaths per vaccination.

Applying these parameters to our 1880s vaccination shortfall—the headline gradient implied by Section 6.7 is approximately 14 to 17 fewer infant vaccinations per 1,000 inhabitants annually in high-mistrust districts—yields a predicted mortality elevation of $14 \times 0.19/1000 \approx 0.06$ to $17 \times 0.24/1000 \approx 0.10$ excess smallpox deaths per 1,000 inhabitants annually in the years that follow. We observe an elevation of 0.11 deaths per 1,000 in the 1890s—essentially within the predicted range. The lag is consistent with the underlying epidemiology: smallpox kills predominantly in childhood, so a vaccination shortfall in the 1880s produces an elevated child-mortality cohort over the next five to ten years, peaking in the 1890s data.

We do not interpret this convergence as proof that vaccination is the only mechanism through which 1857 violence affected mortality; the alternative channels listed above could in principle generate part of the 1890s mortality gradient on their own. What the convergence does establish is that the order-of-magnitude welfare cost implied by combining our reduced-form vaccination gradient with independent medical-literature parameters is essentially the same as the mortality gradient we directly observe. That coincidence is hard to engineer through alternative mechanisms unless they happen to scale precisely with vaccine efficacy and case-fatality parameters that are estimated entirely outside our setting. To put the magnitude in perspective: across the 1890s North-India provinces with populations on the order of 50 million, the implied welfare cost in the high-mistrust subset is roughly 4,000 excess smallpox deaths per year attributable to the 1857 mistrust shock—approximately 40,000 over the decade in our sample, concentrated in children under five.

Variance-inflation factors, the bivariate access–vaccination scatter, and the nested-specification progression that shows how the headline coefficient builds with each block of controls are reported in Appendix Section 8.1.

7 Conclusion

This paper examines how historical experiences of state violence shape trust in public health interventions by analyzing the legacy of the Great Revolt of 1857 on smallpox vaccination adoption in colonial India. Using a newly constructed district-level panel dataset spanning 1870–1919, we find that districts with greater geographic and linguistic proximity to Lucknow—the epicenter of both the rebellion and its violent suppression—showed persistently lower infant vaccination rates in subsequent decades. These effects were economically substantial: a one standard deviation improvement in access to Lucknow was associated with approximately 14 to 17 fewer infant vaccinations per 1,000 inhabitants annually during the 1880s—equivalent to a 1.4 to 1.7 standard deviation decrease in vaccination rates—when the cohort that directly experienced the revolt was making vaccination decisions for their children. The temporal pattern of our results, with effects peaking in the 1880s–1890s and becoming weak by the 1900s–1910s as the affected generation

aged, provides evidence that exposure to state violence created lasting barriers to the adoption of welfare-enhancing technologies.

Our findings contribute to several strands of literature. The marginal contribution is that *non-medical* state violence—a political-military episode with no direct medical content—can persistently depress engagement with subsequent and unrelated medical campaigns. Lowes and Montero (2021) and Alsan and Wanamaker (2018) show that medical episodes generate medical mistrust; we extend this by documenting that mistrust transmits across domains, from the suppression of an armed rebellion to the take-up of childhood vaccination decades later. A welfare interpretation follows by triangulation against medical-literature parameters: the case fatality of *variola major* in childhood (25–30%), the lifetime smallpox attack rate in unvaccinated populations (70–90%), and the efficacy of vaccinia (95%) together imply 0.19–0.24 averted child deaths per vaccination, so our 1880s vaccination shortfall predicts a 0.06–0.10 elevation in smallpox deaths per 1,000 inhabitants annually in subsequent years. We directly observe a 0.11 elevation in the 1890s, consistent in both magnitude and timing with the implied chain. While we do not pursue a formal instrumental-variables design (access to Lucknow plausibly affects mortality through channels other than vaccination), the agreement between the two reduced-form gradients and the independently estimated medical parameters is hard to generate through any single alternative mechanism.

Second, our results speak to the broader literature on state capacity and legitimacy in development. The colonial Indian experience reflects a tension in state-building: the same coercive capacity that enables states to establish authority and extract resources can undermine the legitimacy required for effective public goods provision. In districts near Lucknow, the colonial state's technical ability to supply vaccines mattered less than households' willingness to accept them. This finding challenges development approaches that prioritize building state capacity without equal attention to state legitimacy. When states lack legitimacy rooted in trust, even well-intentioned interventions may fail to achieve their objectives.

Third, we contribute to understanding vaccine hesitancy and medical mistrust by identifying colonial state violence as a persistent barrier to adoption. Our findings complement recent work showing how historical medical campaigns and unethical experiments generate lasting mistrust, but we demonstrate that even non-medical state violence can spill over to reduce engagement with health interventions. This broader source of mistrust may help explain persistent gaps in health technology adoption across many post-colonial contexts where historical grievances continue to shape state-society relations.

The policy implications extend beyond historical interest. Contemporary public health campaigns continue to face resistance in communities with histories of state violence, and our results suggest that technical improvements to vaccines or supply chains may have limited impact where the fundamental barrier is mistrust rather than access.

Several limitations of our analysis point to directions for future research. First, while we document a 1890s mortality gradient consistent with the lower-vaccination → higher-mortality chain, the colonial-vaccination-report mortality data are sparse and uneven across decades and provinces, and we therefore avoid stronger quantitative welfare claims; the natural extension is to combine our panel with census-deviation-based mortality reconstructions. Second, our Awadh-vs-NWP decomposition shows that the apparent distance-to-Lucknow gradient is partly an Awadh-vs-non-Awadh contrast: the within-Awadh access gradient is essentially flat, consistent with the Oudh Settlement having affected the entire Awadh region uniformly. The natural follow-up is a within-Awadh design exploiting heterogeneity in the intensity of post-1858 land confiscations across Awadh districts, which the second author is currently assembling. Third, our focus on aggregate district-level patterns necessarily obscures within-district heterogeneity in both exposure to violence and vaccination decisions; household-level analysis could reveal whether certain groups—defined by caste, religion, or economic status—were differentially affected. Fourth, our linguistic-distance results suggest two parallel channels of cultural transmission: a close-kin channel (sensitive to nearby branches of the language tree) that fades with the dying-out of the cohort that lived through 1857, and a broader-cultural channel (sensitive to distance across language families) that does not depend on the survival of any particular birth cohort and may even widen as later events reinforce inter-group cleavages. Disentangling these channels is a natural extension that the present data can only partially support.

Our findings ultimately highlight how the means by which states establish and exercise authority can have profound and lasting consequences for their ability to provide public goods and implement welfare-enhancing policies. The British colonial state's violent suppression of the 1857 revolt, intended to secure imperial control, inadvertently created barriers to its own public health campaigns that persisted for decades. The historical experience also speaks to contemporary state-building efforts, where coercive strategies may establish control in the short term, but the resulting erosion of trust and legitimacy can undermine even the most beneficial government programs. Colonial violence can constrain state capacity long after political order is restored, not by weakening institutions, but by leaving behind a population unwilling to trust them.

References

Acemoglu, Daron, Simon Johnson and James A Robinson (2001), 'The colonial origins of comparative development: An empirical investigation', *American economic review* **91**(5), 1369–1401.

Acemoglu, Daron, Simon Johnson and James A Robinson (2002), 'Reversal of fortune: Geography and institutions in the making of the modern world income distribution', *The Quarterly journal of economics* **117**(4), 1231–1294.

Acemoglu, Daron, Simon Johnson and James A Robinson (2005), 'Institutions as a fundamental cause of long-run growth', *Handbook of economic growth* **1**, 385–472.

Alsan, Marcella and Claudia Goldin (2019), 'Watersheds in child mortality: The role of effective water and sewerage infrastructure, 1880–1920', *Journal of Political Economy* **127**(2), 586–638.

Alsan, Marcella and Marianne Wanamaker (2018), 'Tuskegee and the health of black men', *The quarterly journal of economics* **133**(1), 407–455.

Anderson, James E and Eric Van Wincoop (2004), 'Trade costs', *Journal of Economic literature* **42**(3), 691–751.

Arnold, David et al. (1993), *Colonizing the body: State medicine and epidemic disease in nineteenth-century India*, Univ of California Press.

Asher, Ronald E (2008), 'Language in historical context', *Language in South Asia* pp. 31–48.

Banerjee, Abhijit and Lakshmi Iyer (2005), 'History, institutions, and economic performance: The legacy of colonial land tenure systems in india', *American economic review* **95**(4), 1190–1213.

Banthia, Jayant and Tim Dyson (1999), 'Smallpox in nineteenth-century india', *Population and Development Review* **25**(4), 649–680.

Bass, Frank M (1969), 'A new product growth for model consumer durables', *Management science* **15**(5), 215–227.

Bennett, Michael (2007), 'Passage through india: global vaccination and british india, 1800–05', *Journal of Imperial and Commonwealth History* **35**(2), 201–220.

Besley, Timothy and Torsten Persson (2011), *Pillars of prosperity: The political economics of development clusters*, Princeton University Press.

Bhattacharya, Sanjoy (1998), 'Re-devising jennerian vaccines?: European technologies, indian innovation and the control of smallpox in south asia, 1850-1950', *Social Scientist* pp. 27–66.

Bhattacharya, Sanjoy, Mark Harrison and Michael Worboys (2005), *Fractured States: Smallpox, Public Health and Vaccination Policy in British India 1800-1947*, Vol. 11, Orient Blackswan.

Bisin, Alberto and Thierry Verdier (2001), 'The economics of cultural transmission and the dynamics of preferences', *Journal of Economic theory* **97**(2), 298–319.

Bose, Sugata and Ayesha Jalal (2022), *Modern South Asia: history, culture, political economy*, Taylor & Francis.

Census of Bombay City (1872), Technical report, Bombay.

Chandra, Bipan, Mridula Mukherjee, Aditya Mukherjee, K Panikkar and S Mahajan (1987), *India's struggle for independence 1857-1947*, Viking New Delhi.

Chaudhary, Latika (2009), 'Determinants of primary schooling in british india', *The Journal of Economic History* **69**(1), 269–302.

Chaudhary, Latika and Manuj Garg (2015), 'Does history matter? colonial education investments in i ndia', *The Economic History Review* **68**(3), 937–961.

Cutler, David and Grant Miller (2005), 'The role of public health improvements in health advances: the twentieth-century united states', *Demography* **42**(1), 1–22.

Dalrymple, William (2007), *The Last Mughal*, Vintage.

David, Julian Saul Markham (2001), The Bengal Army and the outbreak of the Indian mutiny, PhD thesis, University of Glasgow.

de la Fuente, José Andrés Alonso (2005), 'Krishnamurti, bhadriraju," the dravidian languages", cambridge, cambridge university press, 2003, xxiv+ 574 págs.', *Revista Española de Lingüística* **35**(2), 639–642.

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

Dell, Melissa and Benjamin A Olken (2020), 'The development effects of the extractive colonial economy: The dutch cultivation system in java', *The Review of Economic Studies* **87**(1), 164–203.

Dixon, CW (1962), 'Vaccination against smallpox', *British Medical Journal* **1**(5287), 1262.

Donaldson, Dave (2018), 'Railroads of the Raj: Estimating the impact of transportation infrastructure', *American Economic Review* **108**(4-5), 899–934.

Dubé, Eve, Maryline Vivion and Noni E MacDonald (2015), 'Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: influence, impact and implications', *Expert review of vaccines* **14**(1), 99–117.

Dupas, Pascaline (2011), 'Health behavior in developing countries', *Annu. Rev. Econ.* **3**(1), 425–449.

Eberhard, David, Gary Simons and Charles Fennig (2025), Ethnologue: Languages of the world, Technical report, SIL International, Dallas, Texas.

Egger, Peter H and Andrea Lassmann (2012), 'The language effect in international trade: A meta-analysis', *Economics Letters* **116**(2), 221–224.

Erikson, Erik H (1968), *Identity youth and crisis*, number 7, WW Norton & company.

Esteban, Joan, Laura Mayoral and Debraj Ray (2012), 'Ethnicity and conflict: An empirical study', *American Economic Review* **102**(4), 1310–1342.

Fearon, James D and David D Laitin (2003), 'Ethnicity, insurgency, and civil war', *American political science review* **97**(1), 75–90.

Fenner, Frank, Donald Ainslie Henderson, Isao Arita, Zdeněk Ježek and Ivan D Ladnyi (1988), *Smallpox and its eradication continued.*

Fenske, James and Namrata Kala (2021), 'Linguistic distance and market integration in india', *The Journal of Economic History* **81**(1), 1–39.

Frawley, William (2003), *International encyclopedia of linguistics: 4-volume set*, Vol. 1, Oxford university press.

Fremont-Barnes, Gregory (2014), *The Indian Mutiny 1857–58*, Vol. 68, Bloomsbury Publishing.

Frey, Uwe and Richard GM Morris (1997), 'Synaptic tagging and long-term potentiation', *Nature* **385**(6616), 533–536.

Galor, Oded and Ömer Özak (2016), 'The agricultural origins of time preference', *American economic review* **106**(10), 3064–3103.

Geddes, Alasdair M (2006), 'The history of smallpox', *Clinics in dermatology* **24**(3), 152–157.

Giuliano, Paola and Antonio Spilimbergo (2025), 'Aggregate shocks and the formation of preferences and beliefs', *Journal of Economic Literature* **63**(2), 542–597.

Glenn, Norval D (1980), 'Values, attitudes, and beliefs', *Constancy and change in human development* **596**, 640.

Gupta, Bishnupriya (2014), 'Discrimination or social networks? industrial investment in colonial india', *The Journal of Economic History* **74**(1), 141–168.

Haak, Wolfgang, Iosif Lazaridis, Nick Patterson, Nadin Rohland, Swapan Mallick, Bastien Llamas, Guido Brandt, Susanne Nordenfelt, Eadaoin Harney, Kristin Stewardson et al. (2015), 'Massive migration from the steppe was a source for indo-european languages in europe', *Nature* **522**(7555), 207–211.

Habib, Irfan et al. (1982), 'An atlas of the mughal empire: Political and economic maps with detailed notes, bibliography, and index', *(No Title)* .

Hammarlund, Erika, Matthew W Lewis, Scott G Hansen, Lisa I Strelow, Jay A Nelson, Gary J Sexton, Jon M Hanifin and Mark K Slifka (2003), 'Duration of antiviral immunity after smallpox vaccination', *Nature medicine* **9**(9), 1131–1137.

Hardiman, David (2006), *Healing bodies, saving souls: medical missions in Asia and Africa*, Vol. 80, Rodopi.

Harrison, Mark (1994), *Public Health in British India: Anglo-Indian Preventive Medicine 1859-1914*, Cambridge University Press.

Hopkins, Donald R and George I Lythcott (1983), *Princes and peasants: smallpox in history*, University of Chicago Press Chicago.

Jackson, Matthew O. (2008), *Social and Economic Networks*, Princeton University Press.

James, S P (1909), *Smallpox and Vaccination in British India*, Thacker, Spink and Co.

Kolsky, Elizabeth (2010), *Colonial Justice in British India*, Vol. 17, Cambridge University Press Cambridge.

Krosnick, Jon A and Duane F Alwin (1989), 'Aging and susceptibility to attitude change.', *Journal of personality and social psychology* **57**(3), 416.

Lahariya, Chandrakant (2014), 'A brief history of vaccines & vaccination in india', *Indian Journal of Medical Research* **139**(4), 491–511.

Larson, Heidi J, Richard M Clarke, Caitlin Jarrett, Elisabeth Eckersberger, Zachary Levine, Will S Schulz and Pauline Paterson (2018), 'Measuring trust in vaccination: A systematic review', *Human vaccines & immunotherapeutics* **14**(7), 1599–1609.

Llewellyn-Jones, Rosie (2014), *Last King in India: Wajid Ali Shah*, Hurst.

Lowes, Sara and Eduardo Montero (2021), 'The legacy of colonial medicine in central africa', *American Economic Review* **111**(4), 1284–1314.

Malleson, George Bruce (1891), *The Indian Mutiny of 1857*, Seeley and Company.

Metcalf, Thomas R (2015), *Aftermath of revolt: India 1857-1970*, Princeton University Press.

Montaut, Annie (2005), 'colonial language classification, post-colonial language movements and the grassroot multilingualism ethos in india', *Living Together Separately. Cultural India in History and Politics* pp. 75–116.

Mukherjee, Rudrangshu (2002), *Awadh in revolt, 1857-1858: a study of popular resistance*, Orient Blackswan.

Naraindas, Harish (2003), 'Preparing for the pox: A theory of smallpox in bengal and britain', *Asian Journal of Social Science* **31**(2), 304–339.

Nunn, Nathan (2008), 'The long-term effects of africa's slave trades', *The Quarterly journal of economics* **123**(1), 139–176.

Nunn, Nathan and Diego Puga (2012), 'Ruggedness: The blessing of bad geography in africa', *Review of Economics and Statistics* **94**(1), 20–36.

Oldenburg, Veena Talwar (2014), 'The making of colonial lucknow, 1856-1877'.

Pandey, Gyanendra et al. (2001), *Remembering partition: Violence, nationalism, and history in India*, Vol. 7, Cambridge University Press Cambridge.

Reed, Megan Nicole (2022), Marriage and Family in India, PhD thesis, University of Pennsylvania.

Reeves, Peter D. (1961), Agrarian legislation and rural society in Uttar Pradesh, India: an historical study with especial reference to the period 1921–1958, PhD thesis, University of Tasmania.

Rogers, Leonard (1945), 'Smallpox and vaccination in british india during the last seventy years', *Proceedings of the Royal Society of Medicine* **38**(3), 135.

Roy, Tirthankar (2020), *The economic history of India, 1857–2010*, Oxford University Press.

Sen, Surendra Nath (1957), 'Eighteen fifty-seven'.

Spolaore, Enrico and Romain Wacziarg (2018), 'Ancestry and development: New evidence', *Journal of Applied Econometrics* **33**(5), 748–762.

Stokes, Eric (1978), 'The peasant and the raj: Studies in agrarian society and peasant rebellion in colonial india'.

Visaria, Leela and Pravin Visaria (1982), *The Cambridge Economic History of India*, Vol. 2, Cambridge University Press, chapter Population (1757–1947), pp. 463–532.

Yang, Anand A (1999), *Bazaar India: Markets, society, and the colonial state in Bihar*, Univ of California Press.

8 Appendix: For Online Publication

8.1 Multicollinearity and Inference Diagnostics

We report three additional diagnostics for the 1880s yes-FE specification: variance-inflation factors (VIFs), the bivariate access–vaccination relationship, and a nested-specification progression.

The variance-inflation factor for access to Lucknow is approximately 9 in the headline specification, just below the conventional concern threshold of 10. The next-largest VIF is on distance to sea (≈ 6); all other regressors have VIFs below 4. Multicollinearity is therefore present but not dominant.

The bivariate (no-controls) relationship between standardized access to Lucknow and the standardized infant vaccination rate in the 1880s sample is shown in Figure 18. The slope is positive and the line traces the same direction as our controlled estimates, but the bivariate slope is markedly smaller than the controlled coefficient.

Table 8 makes this contrast explicit. With no controls, the access coefficient is 0.50 and statistically significant only at the ten-percent level. Adding geographic controls—ruggedness, caloric suitability, and distance to sea and rivers—more than doubles the coefficient to 1.05 and pushes significance to the one-percent level. Subsequent additions (rail access, population, spending, and finally province fixed effects) raise the coefficient further, ultimately to 1.74. The substantive interpretation: geographic and ecological characteristics correlated with access to Lucknow are also negatively correlated with infant vaccination rates—higher caloric suitability and shorter distance to the sea both predict lower infant vaccination, and they tend to co-occur with worse access to Lucknow. Without controls, this offsetting variation masks the access effect; with controls, the access effect emerges clearly. The nested progression confirms that the headline coefficient is identified by within-province, within-geography variation in accessibility rather than by any single confounding axis.

8.2 Additional Tables and Figures

Decade	Birth Cohort Age	Age range midpoint	Conditional Survival (est.)
1870–79	23–47	35	$e^{-(35-20)\lambda_{20}}$
1880–89	33–57	45	$e^{-(45-20)\lambda_{20}}$
1890–99	43–67	55	$e^{-(55-30)\lambda_{30}}$
1900–09	53–77	65	$e^{-(65-30)\lambda_{30}}$
1910–19	63–87	75	$e^{-(75-30)\lambda_{30}}$

Table 1: Conditional Survival Estimates by Decade and Age Cohort used to create Figure 2. From Visaria and Visaria (1982) we know $\lambda_{20} = \frac{1}{38}$ and $\lambda_{30} = \frac{1}{30}$.

Table 2: Descriptive Statistics

Variable	Mean	SD	Min	Max	N
Vaccination Rate Under Age 1	0.02	0.01	0.00	0.17	7622
Ruggedness (50km)	37561.17	63752.76	630.12	494843.44	7746
Caloric Avg (50km)	3232.31	829.88	0.00	4732.34	7746
Min Distance to Sea (m)	506135.70	347331.35	92.92	1259194.97	7746
Min Distance to River (m)	53670.59	51467.02	408.08	253496.63	7746
Total Population	925889.37	648094.64	2853.00	9138818.00	7694
Spending per Capita	0.01	0.01	0.00	0.15	6358
Distance to Rail (1870)	122202.51	128888.87	20.63	650085.13	7746
Distance to Rail (1880)	83098.38	109358.55	20.63	537276.94	7746
Distance to Rail (1890)	37774.42	60891.85	20.63	326556.30	7746
Distance to Rail (1900)	19303.14	28446.67	20.63	173759.48	7746
Distance to Rail (1910)	12832.09	21618.30	20.63	166421.10	7746
Access to Lucknow (1870)	66.09	37.09	0.00	170.04	7746
Access to Lucknow (1880)	57.28	32.62	0.00	145.33	7746
Access to Lucknow (1890)	51.53	29.64	0.00	131.66	7746
Access to Lucknow (1900)	47.52	27.45	0.00	123.05	7746
Access to Lucknow (1910)	44.59	25.86	0.00	117.21	7746
Vaccination Rate	0.04	0.02	0.00	0.39	7694
Distance to Lucknow	748871.71	421194.42	2191.69	2043062.67	7746
Linguistic Distance to Lucknow	0.43	0.29	0.01	1.00	5044
Smallpox Deaths per 1000	1.13	9.18	0.00	349.00	5831

Table 3: Archival Sources for Vaccination Reports

Province	**Years**	**URL**
Assam	1875–1894	https://digital.nls.uk/91524026
Assam	1895–1927	https://digital.nls.uk/91524027
Bengal	1869	https://digital.nls.uk/91533951
Bengal	1869–1873	https://digital.nls.uk/91534624
Bengal	1875–1889	https://digital.nls.uk/91536150
Bengal	1887–1898	https://digital.nls.uk/91537851
Bengal	1896–1909	https://digital.nls.uk/91539671
Bengal	1910–1929	https://digital.nls.uk/91541450
Bombay and Sindh	1870–1873	https://digital.nls.uk/91022597
Bombay and Sindh	1874–1876	https://digital.nls.uk/91022598
Bombay and Sindh	1890–1901	https://digital.nls.uk/91029430
Bombay and Sindh	1902–1912	https://digital.nls.uk/91029431
Bombay and Sindh	1913–1923	https://digital.nls.uk/91029432
Central Provinces and Berar	1868–1885	https://digital.nls.uk/91229505
Central Provinces and Berar	1885–1899	https://digital.nls.uk/91229506
Central Provinces and Berar	1899–1911	https://digital.nls.uk/91229507
Central Provinces and Berar	1911–1920	https://digital.nls.uk/91229508

Province	Years	URL
Eastern Bengal and Assam	1905–1912	https://digital.nls.uk/91528315
Hyderabad Assigned Districts	1872–1889	https://digital.nls.uk/91713999
Hyderabad Assigned Districts	1888–1903	https://digital.nls.uk/91714000
Madras Presidency	1894–1903	https://digital.nls.uk/9174264
Madras Presidency	1903–1920	https://digital.nls.uk/9174265
Mysore	1878–1880	https://digital.nls.uk/9171470
North-West Frontier Province	1904–1920	https://digital.nls.uk/90586831
North-Western Provinces and United Provinces	1866–1877	https://digital.nls.uk/87224066
North-Western Provinces and United Provinces	1895–1901	https://digital.nls.uk/87224068
North-Western Provinces and United Provinces	1899–1922	https://digital.nls.uk/87230620, https://digital.nls.uk/87230621
Oudh	1871	https://digital.nls.uk/90721052
Oudh	1872	https://digital.nls.uk/90721053
Oudh	1873	https://digital.nls.uk/90721054
Punjab	1867–1880	https://digital.nls.uk/85747399
Punjab	1883–1896	https://digital.nls.uk/85747400
Punjab	1896–1918	https://digital.nls.uk/85747401, https://digital.nls.uk/85747402
Punjab	1918–1920	https://digital.nls.uk/85747403

Province	Years	URL
Rajputana	1909	https://digital.nls.uk/90721632
Rajputana	1910	https://digital.nls.uk/90722003
Rajputana	1911	https://digital.nls.uk/90722470
Rajputana	1912	https://digital.nls.uk/90722852
Rajputana	1913	https://digital.nls.uk/90723240
Rajputana	1914	https://digital.nls.uk/90723677
Rajputana	1915	https://digital.nls.uk/90724055
Rajputana	1916	https://digital.nls.uk/90724638
Rajputana	1917	https://digital.nls.uk/90725042
Rajputana	1918	https://digital.nls.uk/90725453
Rajputana	1919	https://digital.nls.uk/90725872

Table 4: Access to Lucknow on Infant Vaccination Rates by Decade (no FE's)

Dependent Variable:	vax_rate_under_1				
	1870s	1880s	1890s	1900s	1910s
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
Constant	0.6186***	0.7606**	0.2341*	0.2613+	0.1394
	(0.1637)	(0.2489)	(0.1085)	(0.1405)	(0.1099)
Access to Lucknow	0.7114*	1.412**	0.6689***	0.2793	0.1925
	(0.3148)	(0.4116)	(0.1829)	(0.2225)	(0.1792)
Distance to railroad	-0.2449**	-0.3751*	-0.1258	0.0770	-0.1388
	(0.0824)	(0.1622)	(0.1194)	(0.1189)	(0.1008)
Ruggedness	0.2880	0.0680	-0.2391***	-0.3376***	-0.2906**
	(0.1772)	(0.1748)	(0.0628)	(0.0665)	(0.0943)
Caloric suitability	-0.1994	-0.8541***	-0.1622	0.1063	0.1067
	(0.1657)	(0.2041)	(0.1207)	(0.1334)	(0.0987)
Distance to sea	0.0988	0.8340*	0.6125***	0.3576*	0.3570***
	(0.2221)	(0.3767)	(0.1326)	(0.1451)	(0.0846)
Distance to river	0.0815	0.2831*	0.3723***	0.1074	0.1189
	(0.1469)	(0.1237)	(0.0680)	(0.0993)	(0.0934)
Total population	-0.4586***	-0.3612**	-0.3957***	-0.1855*	-0.1995*
	(0.1192)	(0.1275)	(0.0810)	(0.0733)	(0.0911)
Spending per capita	0.0630**	-0.0355	0.1248	0.3665**	0.2131
	(0.0198)	(0.0945)	(0.1082)	(0.1166)	(0.1489)
Fit statistics					
Observations	79	78	146	203	206
R^2	0.58037	0.62665	0.56690	0.29004	0.26933

*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, +: 0.1*

Table 5: Access to Lucknow on Infant Vaccination Rates by Decade (yes FE's)

Dependent Variable:	vax_rate_under_1				
	1870s	1880s	1890s	1900s	1910s
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
Access to Lucknow	1.057***	1.743**	0.8148***	0.5870*	0.3674+
	(0.2639)	(0.5336)	(0.1790)	(0.2554)	(0.2081)
Distance to railroad	-0.3689***	-0.4894*	-0.0686	0.0406	-0.1530
	(0.1059)	(0.2100)	(0.1061)	(0.1119)	(0.0934)
Ruggedness	0.1062	0.0094	-0.2362**	-0.3221***	-0.2159*
	(0.1026)	(0.1694)	(0.0879)	(0.0831)	(0.0925)
Caloric suitability	-0.0680	-0.8562***	-0.2276+	-0.1285	-0.0696
	(0.1030)	(0.2057)	(0.1368)	(0.1189)	(0.0758)
Distance to sea	-0.1381	0.8396+	0.8887***	0.9191***	0.6208***
	(0.3475)	(0.4413)	(0.2352)	(0.2036)	(0.1687)
Distance to river	0.1855*	0.2366+	0.2261***	-0.0055	-0.0208
	(0.0793)	(0.1358)	(0.0638)	(0.0899)	(0.0878)
Total population	-0.4442***	-0.3440*	-0.3006***	-0.1476*	-0.1365*
	(0.0935)	(0.1545)	(0.0553)	(0.0687)	(0.0681)
Spending per capita	0.0926+	0.0006	0.0822	0.2575*	0.0923
	(0.0476)	(0.0586)	(0.0989)	(0.1253)	(0.0967)
Fixed-effects					
province_id	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	78	76	144	202	204
R^2	0.64831	0.68708	0.62989	0.43658	0.44297
Within R^2	0.41669	0.51420	0.31452	0.23061	0.14287

*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, +: 0.1*

Table 6: Linguistic Distance from Lucknow on Infant Vaccination Rates by Decade (no FE's)

Dependent Variable:	vax_rate_under_1				
	1870s	1880s	1890s	1900s	1910s
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
Constant	-0.1110	-0.2228	-0.1283	-0.0198	-0.0143
	(0.1445)	(0.1817)	(0.1329)	(0.1431)	(0.1489)
Linguistic Distance from Lucknow	0.1844	0.4619*	0.1722	0.0049	0.0165
	(0.1541)	(0.1852)	(0.1078)	(0.0982)	(0.1123)
Distance to railroad	-0.2459***	-0.1667	-0.1665+	-0.0828	-0.2636*
	(0.0689)	(0.1149)	(0.0993)	(0.1182)	(0.1197)
Ruggedness	-0.3939	-0.0268	-0.1700	-0.2203	-0.3094*
	(0.2568)	(0.1736)	(0.1420)	(0.1446)	(0.1497)
Caloric suitability	-0.1588	-0.2530	-0.0952	0.0346	0.1062
	(0.1713)	(0.2502)	(0.0636)	(0.0831)	(0.0824)
Distance to sea	0.0054	0.5982*	0.4613***	0.2935**	0.2671**
	(0.1786)	(0.2321)	(0.1237)	(0.0914)	(0.0943)
Distance to river	0.0638	0.1613*	0.2027**	0.1386*	0.1705**
	(0.1204)	(0.0800)	(0.0756)	(0.0588)	(0.0517)
Total population	-0.7297***	-0.4391**	-0.3532**	-0.3709**	-0.3985***
	(0.1527)	(0.1298)	(0.1141)	(0.1133)	(0.0982)
Spending per capita	0.2033*	0.0116	0.2391+	0.1049	-0.0080
	(0.0833)	(0.1149)	(0.1268)	(0.1023)	(0.0901)
Fit statistics					
Observations	80	78	136	164	156
R^2	0.46502	0.44198	0.50606	0.39371	0.42522

*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, +: 0.1*

Table 7: Pearson Correlation Matrix (p-values in parentheses)

	Access 1870	Access 1880	Access 1890	Access 1900	Access 1910	Straight Dist	Ling. Dist
Access 1870	1.000 (NA)	0.998 (0.000)	0.995 (0.000)	0.994 (0.000)	0.993 (0.000)	0.977 (0.000)	0.797 (0.000)
Access 1880	0.998 (0.000)	1.000 (NA)	0.999 (0.000)	0.998 (0.000)	0.997 (0.000)	0.970 (0.000)	0.803 (0.000)
Access 1890	0.995 (0.000)	0.999 (0.000)	1.000 (NA)	1.000 (0.000)	0.999 (0.000)	0.966 (0.000)	0.806 (0.000)
Access 1900	0.994 (0.000)	0.998 (0.000)	1.000 (0.000)	1.000 (NA)	1.000 (0.000)	0.968 (0.000)	0.809 (0.000)
Access 1910	0.993 (0.000)	0.997 (0.000)	0.999 (0.000)	1.000 (0.000)	1.000 (NA)	0.971 (0.000)	0.811 (0.000)
Straight Dist	0.977 (0.000)	0.970 (0.000)	0.966 (0.000)	0.968 (0.000)	0.971 (0.000)	1.000 (NA)	0.793 (0.000)
Ling. Dist	0.797 (0.000)	0.803 (0.000)	0.806 (0.000)	0.809 (0.000)	0.811 (0.000)	0.793 (0.000)	1.000 (NA)

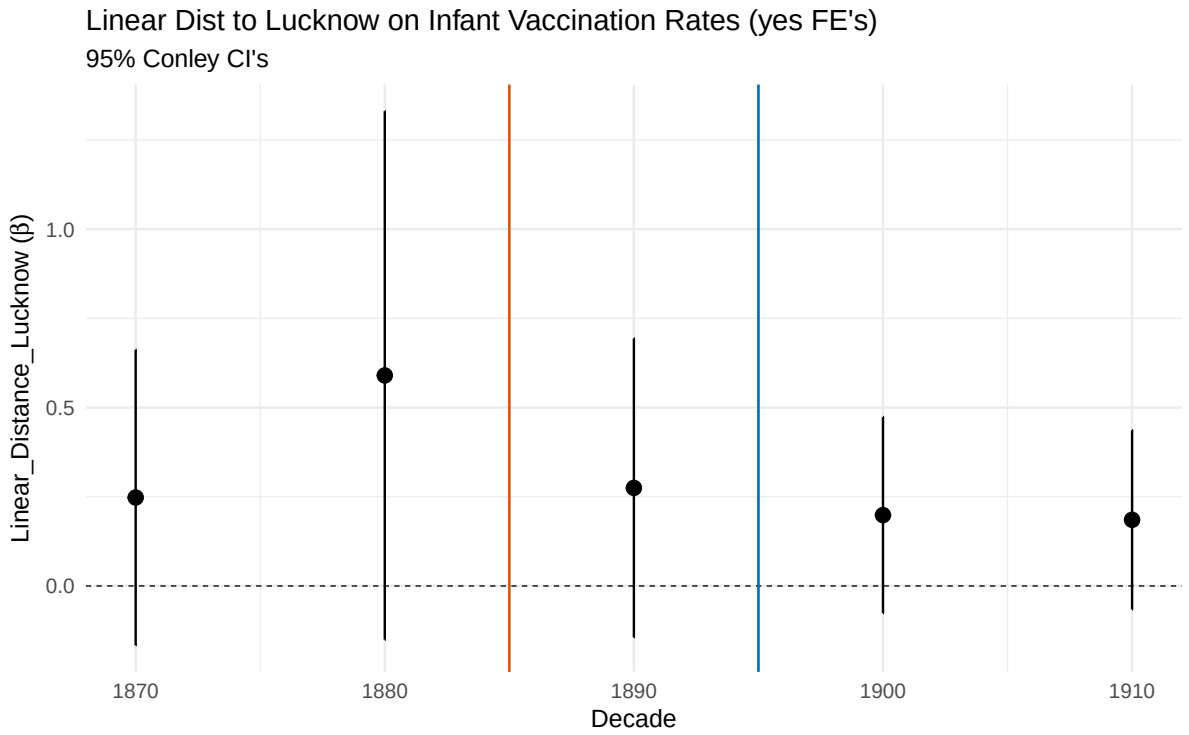


Figure 15: Linear Distance to Lucknow

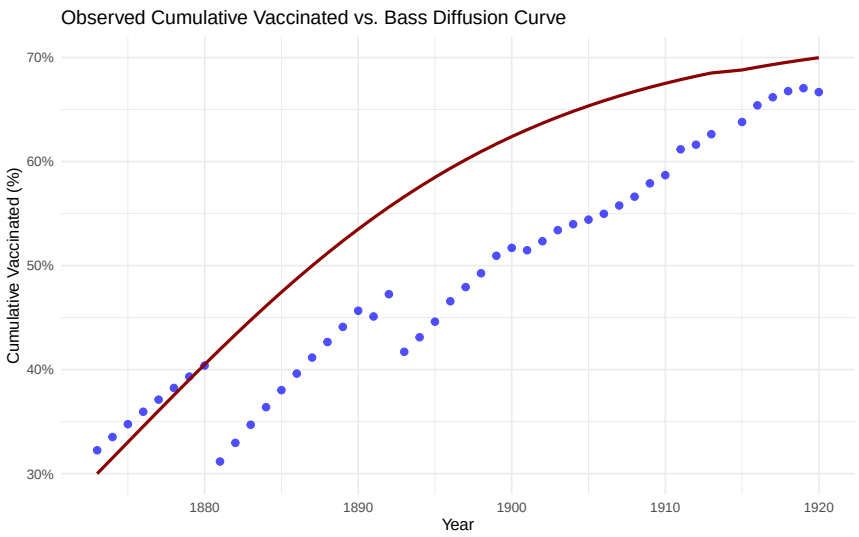


Figure 16: Bass Diffusion Curve Using Bombay Data

Table 8: Nested-Specification Progression for 1880s, with Province Fixed Effects in (5)

	(1) No ctrls	(2) +Geog.	(3) +Infra.	(4) +Pop/Spend	(5) +Prov. FE
Access to Lucknow	0.501^{+}	1.048^{***}	1.121^{***}	1.412^{***}	1.743^{***}
	(0.328)	(0.215)	(0.228)	(0.412)	(0.534)
Geographic ctrls	No	Yes	Yes	Yes	Yes
Infrastructure	No	No	Yes	Yes	Yes
Pop & Spending	No	No	No	Yes	Yes
Province FE	No	No	No	No	Yes
Observations	144	144	144	78	76
R^2	0.091	0.466	0.496	0.627	0.687

Conley (spatial HAC) standard errors in parentheses.

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 9: Horse-Race Specification: Access to Lucknow and Access to Delhi, 1880s (yes FE)

Regressor	Coefficient (SE)
Access to Lucknow (1880)	1.454***
	(0.290)
Access to Delhi (1880)	-1.268***
	(0.343)
Geographic ctrls	Yes
Distance to railroad	Yes
Population & spending	Yes
Province FE	Yes
Observations	76
R^2	0.745

Conley (spatial HAC) standard errors in parentheses.

$^{+}\,p < 0.10$, $^{*}\,p < 0.05$, $^{**}\,p < 0.01$, $^{***}\,p < 0.001$.

Table 10: Linguistic Distance to Lucknow on Infant Vaccination Rates: Sensitivity to Branch-Decay Parameter δ

	$\delta = 0.5$ (Fearon-Laitin)	$\delta = 0.05$ (Fenske-Kala)
Decade	coef (SE)	coef (SE)
1870s	0.184 (0.154)	0.425* (0.209)
1880s	0.462* (0.185)	0.507 (0.344)
1890s	0.172 (0.108)	0.436* (0.216)
1900s	0.005 (0.098)	0.633+ (0.365)
1910s	0.016 (0.112)	0.826** (0.281)

Conley (spatial HAC) standard errors in parentheses.

Specifications without province fixed effects, mirroring the linguistic-distance specification in the main results.

$^{+}\,p < 0.10$, $^{*}\,p < 0.05$, $^{**}\,p < 0.01$, $^{***}\,p < 0.001$.

Table 11: Linguistic Distance and Cohort Survival, Pairwise Placebo Comparisons

	Dacca	Poona	Lahore	Karachi	Bombay	Calcutta
Lucknow LD × survival	0.629**	0.637**	0.639**	0.656*	0.696*	0.373
	(0.225)	(0.208)	(0.204)	(0.263)	(0.309)	(0.408)
Placebo LD × survival	0.204	0.121	-0.756***	-0.126	-0.191	0.284
	(0.303)	(0.160)	(0.197)	(0.140)	(0.225)	(0.410)
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	452	452	452	452	452	452
Districts	120	120	120	120	120	120

The dependent variable is the standardized infant vaccination rate. Each column reports a separate panel regression that includes the interaction of standardized linguistic distance to Lucknow with the predicted survival probability of the cohort aged 10–25 in 1857, and the analogous interaction for one comparison city. Linguistic distances are computed under the $\delta = 0.5$ (Fearon-Laitin) parameterization. Survival probabilities follow the calculation reported in Appendix Table 1. The main effects of linguistic distance are absorbed by district fixed effects. The main effect of survival is absorbed by decade fixed effects. Controls are distance to the rail network, total population, and vaccination spending per capita, all standardized. Standard errors are Conley spatial HAC, reported in parentheses. A positive coefficient on the Lucknow interaction means that, in decades when the 1832–47 cohort is more likely to be alive, districts linguistically closer to Lucknow have lower infant vaccination rates. The row label LD abbreviates linguistic distance. The sample is restricted to districts within 1,000 km of Lucknow.

$^{+}\, p < 0.10$, $^{*}\, p < 0.05$, $^{**}\, p < 0.01$, $^{***}\, p < 0.001$.

Table 12: Interaction of Access to Lucknow with Local Sepoy Activity in 1857

	1870s	1880s	1890s	1900s	1910s
Panel A. Sepoy exposure within 50 km					
Access × Sepoy 50 km	0.310	0.793***	0.017	0.220	0.228
	(0.261)	(0.184)	(0.113)	(0.136)	(0.144)
Sepoy-exposed districts	45	46	81	87	93
Non-exposed districts	34	32	65	116	113
Panel B. Sepoy exposure within 100 km					
Access × Sepoy 100 km	-0.170	1.015*	0.332	0.318	0.095
	(0.657)	(0.390)	(0.227)	(0.271)	(0.254)
Sepoy-exposed districts	59	60	115	129	138
Non-exposed districts	20	18	31	74	68
Geographic controls	Yes	Yes	Yes	Yes	Yes
Distance to railroad	Yes	Yes	Yes	Yes	Yes
Population & spending	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes
Observations (Panel A)	78	76	144	202	204
Observations (Panel B)	78	76	144	202	204

The dependent variable is the standardized infant vaccination rate. Each column reports a separate decade-specific regression with the access to Lucknow main effect, the sepoy exposure main effect, their interaction, and the full set of geographic and decadal controls. Access to Lucknow is standardized within decade. Sepoy exposure equals one if the district lies within the stated radius of at least one documented sepoy mutiny, disarmament, or disbandment event in 1857. Jail-break events are excluded. Event coordinates are drawn from David (2001) and geocoded by hand. A positive interaction coefficient means that the access-to-Lucknow gradient is steeper in sepoy-exposed districts than in non-exposed districts. Standard errors are Conley spatial HAC, reported in parentheses. The sample is restricted to districts within 1,000 km of Lucknow.

$^{+}\,p < 0.10$, $^{*}\,p < 0.05$, $^{**}\,p < 0.01$, $^{***}\,p < 0.001$.

Table 13: Suggestive Mortality Evidence: Smallpox Deaths per 1,000 on Access to Lucknow

	1890s	1900s	1910s
Access to Lucknow	-0.412**	0.008	0.009
	(0.167)	(0.010)	(0.009)
Geographic ctrls	Yes	Yes	Yes
Province FE	Yes	Yes	Yes
Observations	136	165	169

Conley (spatial HAC) standard errors in parentheses.

Outcome is smallpox deaths per 1,000 inhabitants annually, from the vaccination reports. The 1870s and 1880s are omitted: 1870s have no mortality coverage in our digitised sample, and 1880s coverage ($N = 48$) is too sparse to support the controlled specification.

$^{+}\, p < 0.10$, $^{*}\, p < 0.05$, $^{**}\, p < 0.01$, $^{***}\, p < 0.001$.

Table 14: Awadh-vs-NWP Within-Province Falsification: Pooled Decade Spec (Northwestern Province Only)

Regressor	Coefficient (SE)
Access to Lucknow	0.396***
	(0.108)
Awadh dummy	-0.748***
	(0.178)
Access × Awadh	-0.352***
	(0.105)
Geographic ctrls	Yes
Distance to railroad	Yes
Population & spending	Yes
Decade FE	Yes
Awadh districts (unique)	12
Non-Awadh NWP districts (unique)	42
Observations	245

Conley (spatial HAC) standard errors in parentheses. Awadh districts are the 12 historical districts of the Oudh region present in our panel: Lucknow, Fyzabad (Faizabad), Sultanpur, Bahraich, Gonda, Hardoi, Sitapur, Kheri, Partabgarh (Pratapgarh), Rae Bareili, Unao (Unnao), and Bara Banki (Mukherjee, 2002, ch. 1; Lakhimpur appears merged with Kheri in our administrative coding). Sample restricted to the northwestern province (no province FE; decade fixed effects only).

$^{+}\, p < 0.10$, $^{*}\, p < 0.05$, $^{**}\, p < 0.01$, $^{***}\, p < 0.001$.

Table 15: Within-District Effect of Access to Lucknow on Infant Vaccination, by Decade

	1880s	1890s	1900s	1910s
Panel A. Unbalanced panel (all districts observed in each decade)				
Access to Lucknow × decade	0.259***	0.144*	-0.002	-0.100
	(0.068)	(0.073)	(0.086)	(0.094)
Panel B. Balanced panel (districts observed in all five decades)				
Access to Lucknow × decade	0.198***	0.094	0.001	-0.059
	(0.051)	(0.057)	(0.072)	(0.084)
District FE	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations (Panel A)	681			
Observations (Panel B)	310 (62 districts)			

The dependent variable is the standardized infant vaccination rate. Reported coefficients are interactions between standardized cumulative least-cost travel cost to Lucknow and decade indicators. The 1870s is the omitted baseline. Access is standardized across the pooled district-decade panel, so one unit corresponds to one standard deviation of pooled access. District fixed effects absorb all time-invariant district characteristics, including permanent geography, fixed cultural composition, and time-invariant institutional differences. Decade fixed effects absorb shocks common to all districts. Controls are distance to the rail network, total population, and vaccination spending per capita, all standardized. Standard errors clustered by district appear in parentheses. The sample is restricted to districts within 1,000 km of Lucknow.

$^{+}\ p < 0.10$, $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$.

Table 16: Access to Lucknow in Restricted Regional Samples

	1870s	1880s	1890s	1900s	1910s
Full sample	0.635***	1.058**	0.480***	0.420*	0.273+
	(0.158)	(0.324)	(0.105)	(0.183)	(0.155)
Observations	78	76	144	202	204
NWP/Oudh only	0.574***	0.806***	0.739***	0.564***	0.650***
	(0.092)	(0.176)	(0.171)	(0.055)	(0.069)
Observations	47	49	52	49	48
NWP/Oudh + Punjab	0.574***	0.806***	0.750*	0.921**	0.760+
	(0.092)	(0.176)	(0.352)	(0.336)	(0.401)
Observations	47	49	94	91	85
NWP/Oudh + Punjab + Bengal	0.468***	0.961**	0.682*	0.615*	0.478
	(0.118)	(0.292)	(0.276)	(0.277)	(0.343)
Observations	65	70	116	113	102

The dependent variable is the standardized infant vaccination rate. Each row reports the headline province-fixed-effects specification estimated on a progressively restricted sample of provinces. The NWP/Oudh sample contains only one province in our coding, so the specification omits province fixed effects for that row. Punjab contributes no additional usable observations in the 1870s and 1880s under our 1,000 km baseline, so the NWP/Oudh and the NWP/Oudh plus Punjab estimates coincide in those decades. Standard errors are Conley spatial HAC, reported in parentheses. The sample is restricted to districts within 1,000 km of Lucknow.
$^{+}\, p < 0.10$, $^{*}\, p < 0.05$, $^{**}\, p < 0.01$, $^{***}\, p < 0.001$.

Table 17: Instrumenting Least-Cost Access with Straight-Line Distance to Lucknow

	1870s	1880s	1890s	1900s	1910s
OLS access to Lucknow	0.635***	1.058**	0.480***	0.420*	0.273+
	(0.158)	(0.324)	(0.105)	(0.183)	(0.155)
IV access to Lucknow	0.829***	1.567***	0.600***	0.568**	0.372*
	(0.136)	(0.313)	(0.100)	(0.190)	(0.166)
First-stage F-statistic	340.5	281.0	439.6	411.3	483.4
Observations	78	76	144	202	204

The endogenous regressor is standardized cumulative least-cost travel cost to Lucknow. The instrument is standardized straight-line distance to Lucknow. Both specifications include the headline controls and province fixed effects. Standard errors are Conley spatial HAC, reported in parentheses. The instrument isolates the component of least-cost access predicted by raw geographic distance rather than by the placement of railways. The sample is restricted to districts within 1,000 km of Lucknow.
$^{+}\, p < 0.10$, $^{*}\, p < 0.05$, $^{**}\, p < 0.01$, $^{***}\, p < 0.001$.

Table 18: Log Count of Under-One Vaccinations on Access to Lucknow

	1870s	1880s	1890s	1900s	1910s
Access to Lucknow	1.214***	1.320**	0.474***	0.114	0.078
	(0.298)	(0.384)	(0.140)	(0.094)	(0.078)
Log total population control	Yes	Yes	Yes	Yes	Yes
Other geographic controls	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes
Observations	78	76	144	202	204

The dependent variable is the standardized log count of vaccinations among children under one year of age, log(1 + under one vaccinations). The specification replaces the standardized total population control with the standardized log total population control. All other controls match the headline specification. Standard errors are Conley spatial HAC, reported in parentheses. The sample is restricted to districts within 1,000 km of Lucknow.

$^{+}\, p < 0.10$, $^{*}\, p < 0.05$, $^{**}\, p < 0.01$, $^{***}\, p < 0.001$.

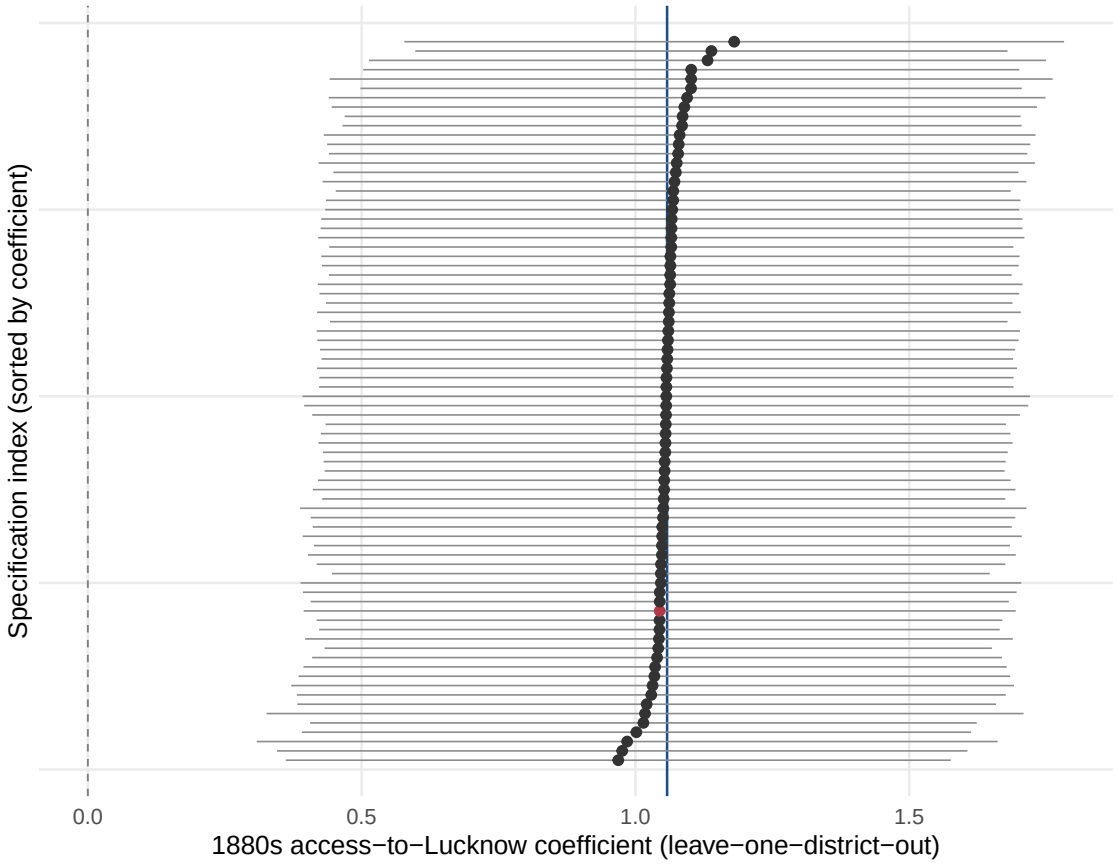


Figure 17: Leave-one-district-out robustness of the 1880s access-to-Lucknow coefficient. Each point is the access coefficient from the headline province-fixed-effects specification after dropping one district from the 1880s estimating sample, with horizontal segments representing 95% Conley confidence intervals. The vertical line marks the full-sample 1880s estimate. The red highlighted point is the estimate obtained after dropping Lucknow district itself. The estimates remain positive across all 78 leave-one-out samples.

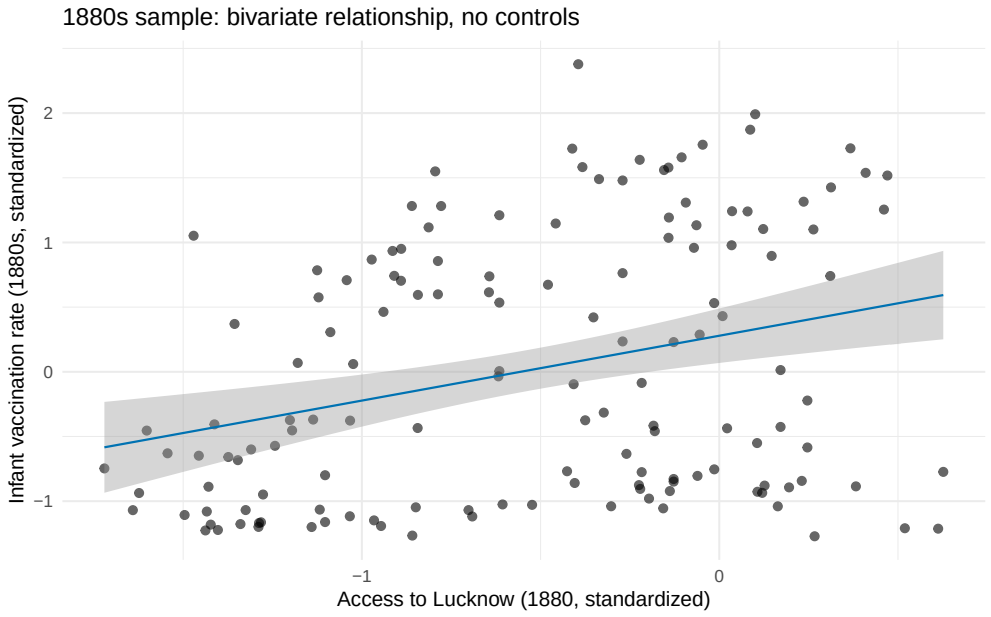


Figure 18: Bivariate relationship between access to Lucknow and infant vaccination rate, 1880s sample, no controls. The slope is positive and statistically significant at the ten-percent level (Conley HAC). Compare with Table 8 for the controlled estimates.

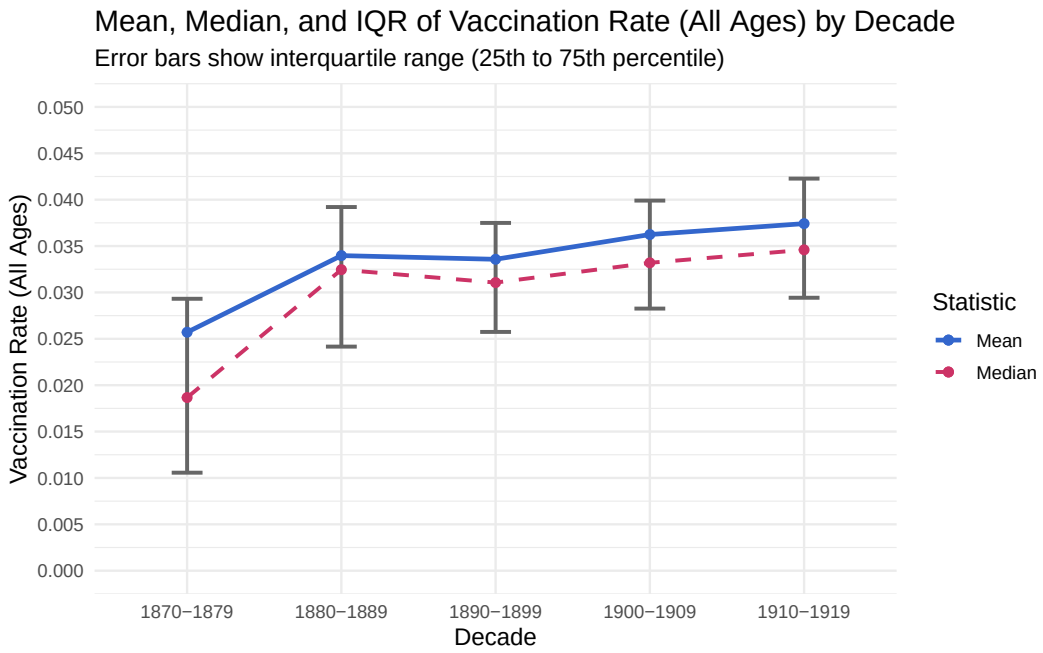


Figure 19: Vax Rate All Ages by Decade

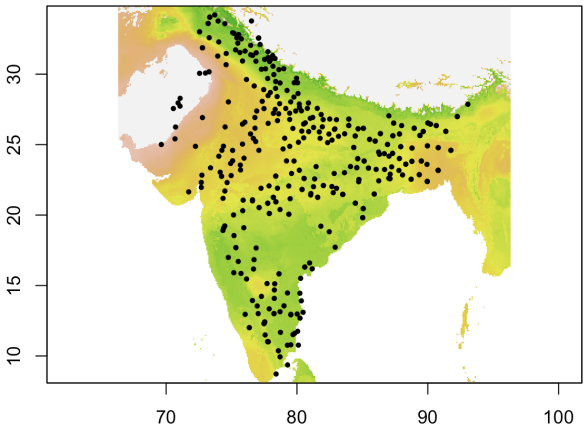


Figure 20: Caloric Suitability (lighter worse)

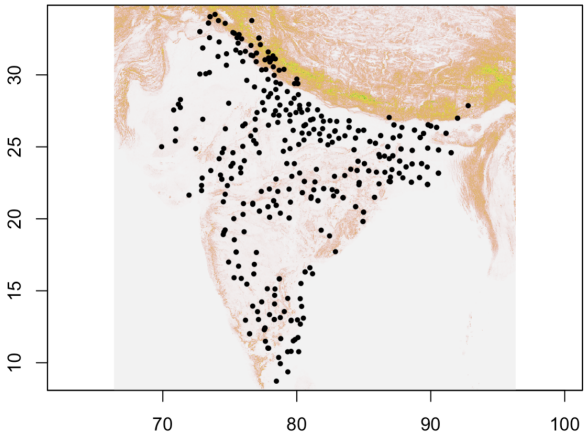


Figure 21: Ruggedness (lighter less rugged)

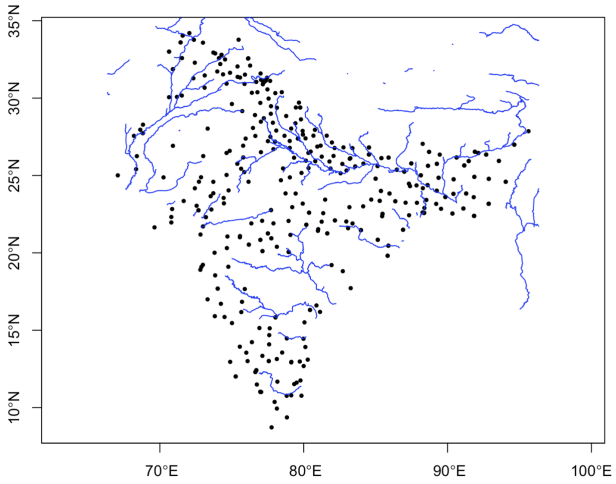
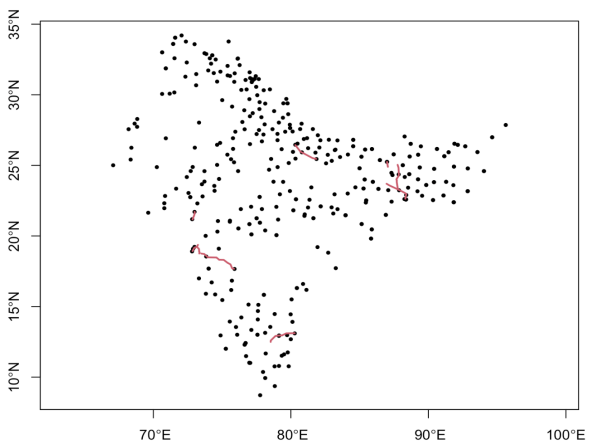
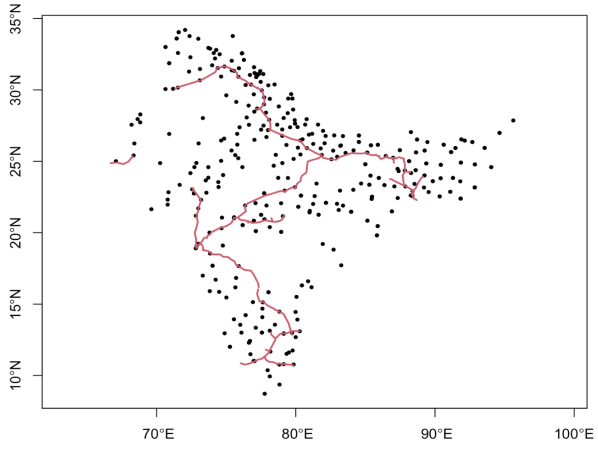


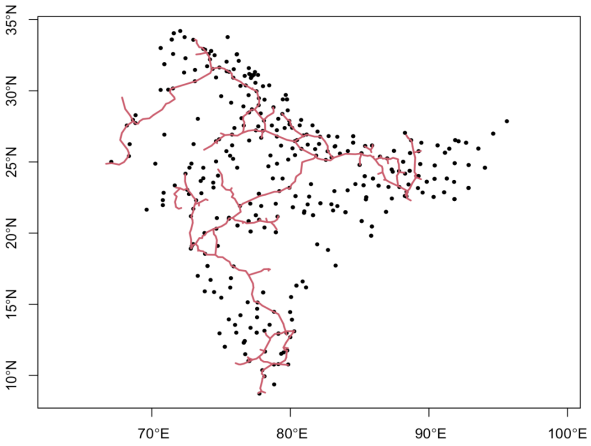
Figure 22: Navigable Rivers



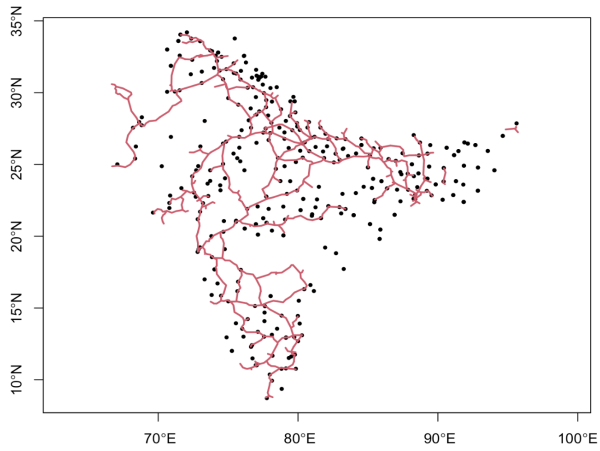
(a) 1860



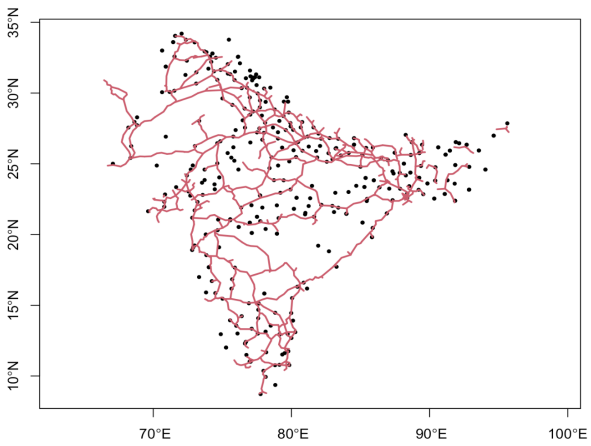
(b) 1870



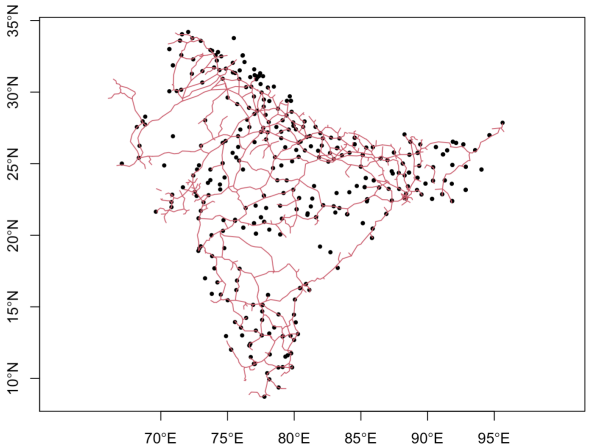
(c) 1880



(d) 1890



(e) 1900



(f) 1910

Figure 23: Growth of the Rail Network

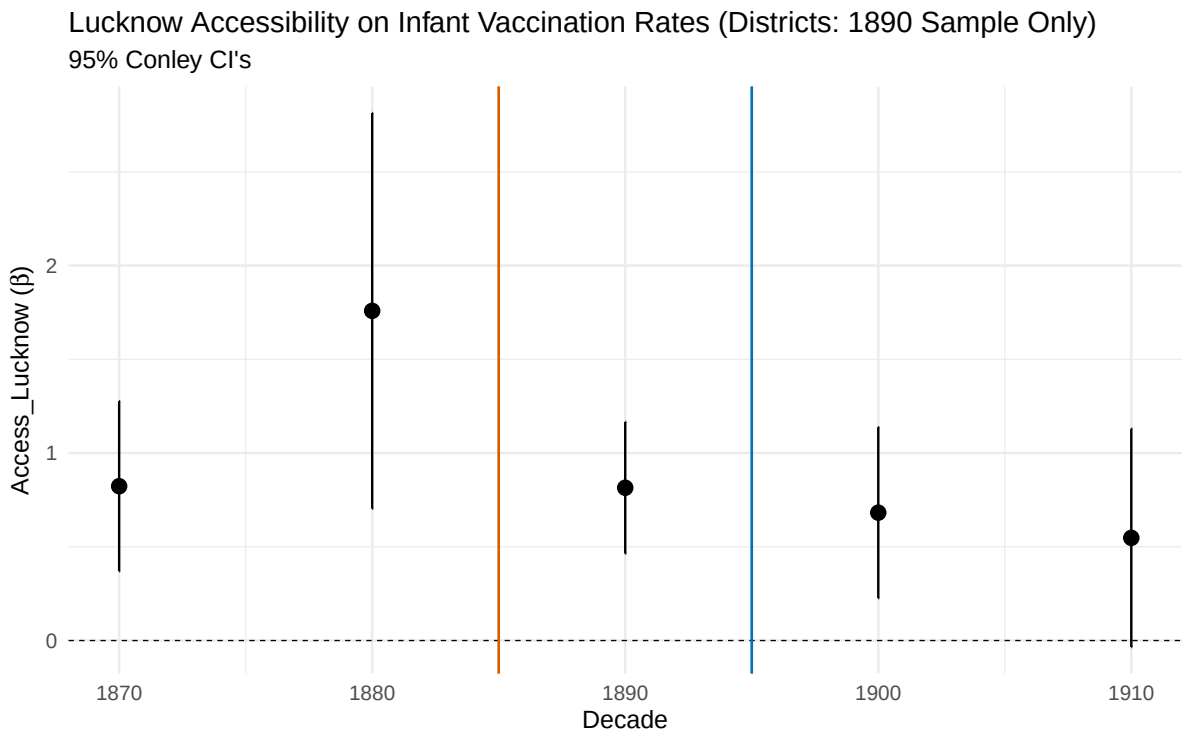


Figure 24: Sample restricted to 1890-99 Districts

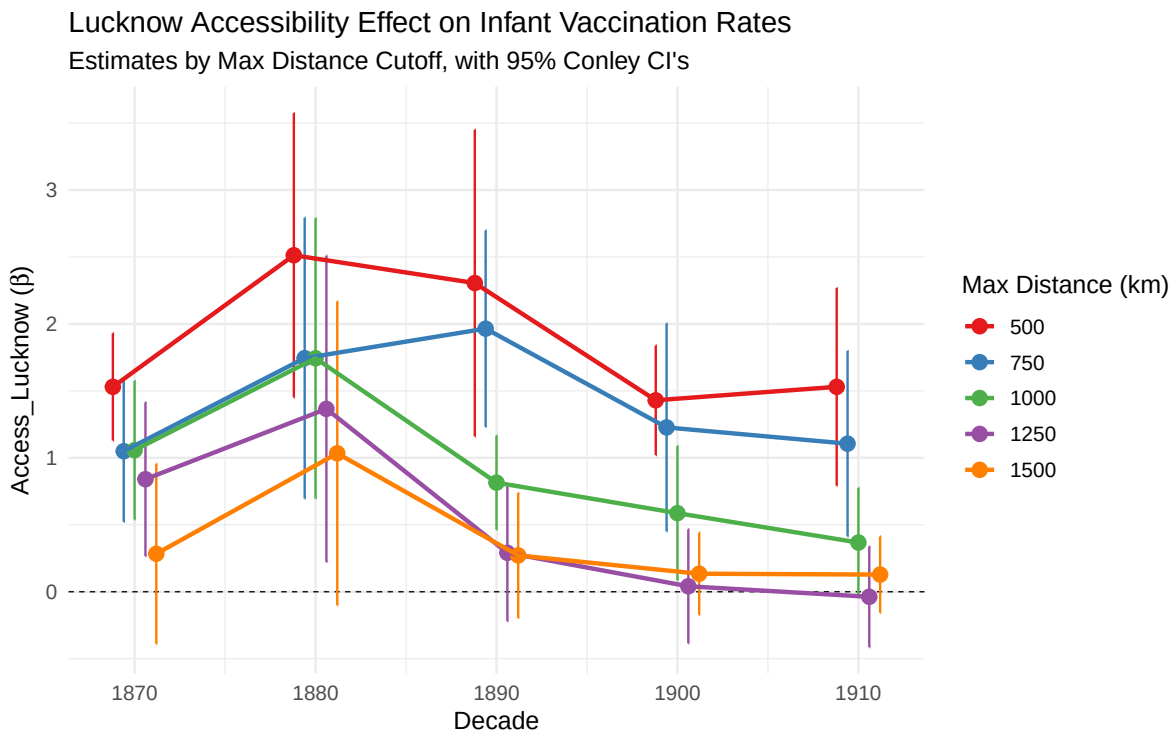


Figure 25: Varying Max Distance to Lucknow

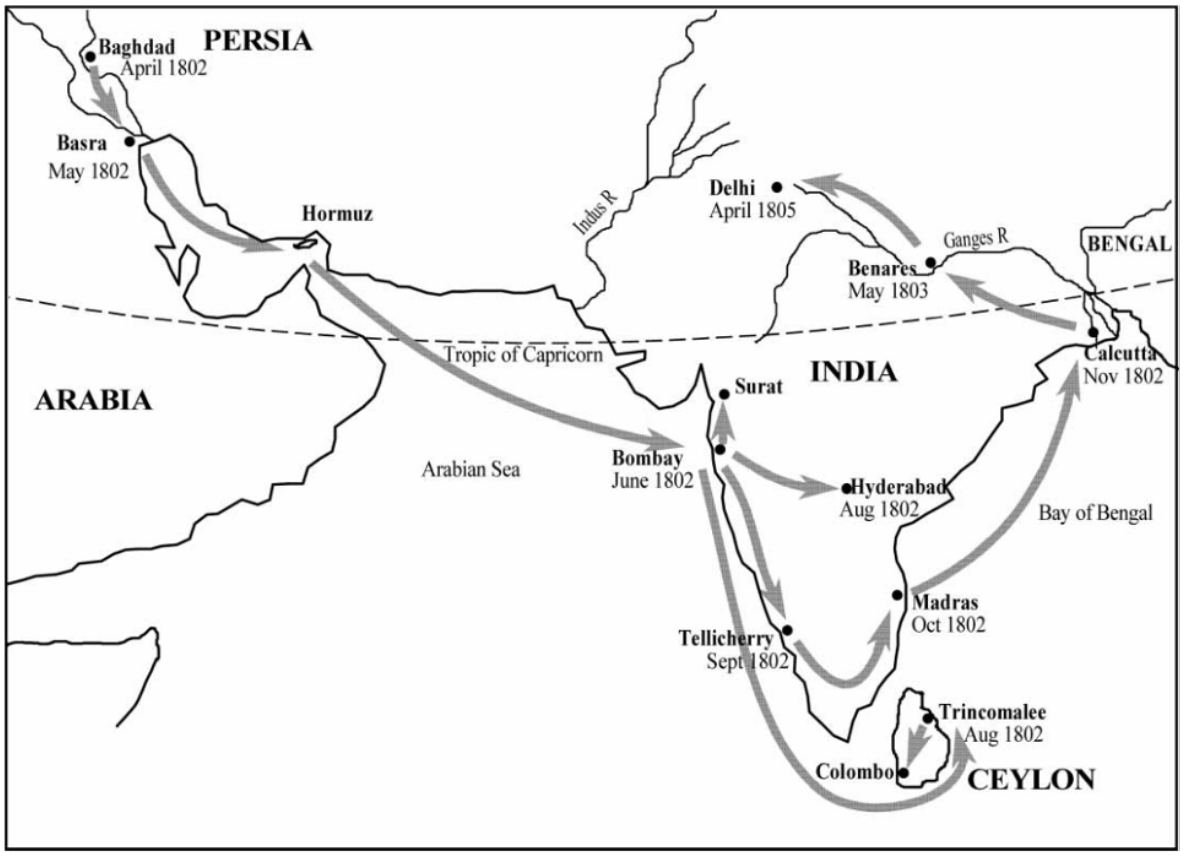
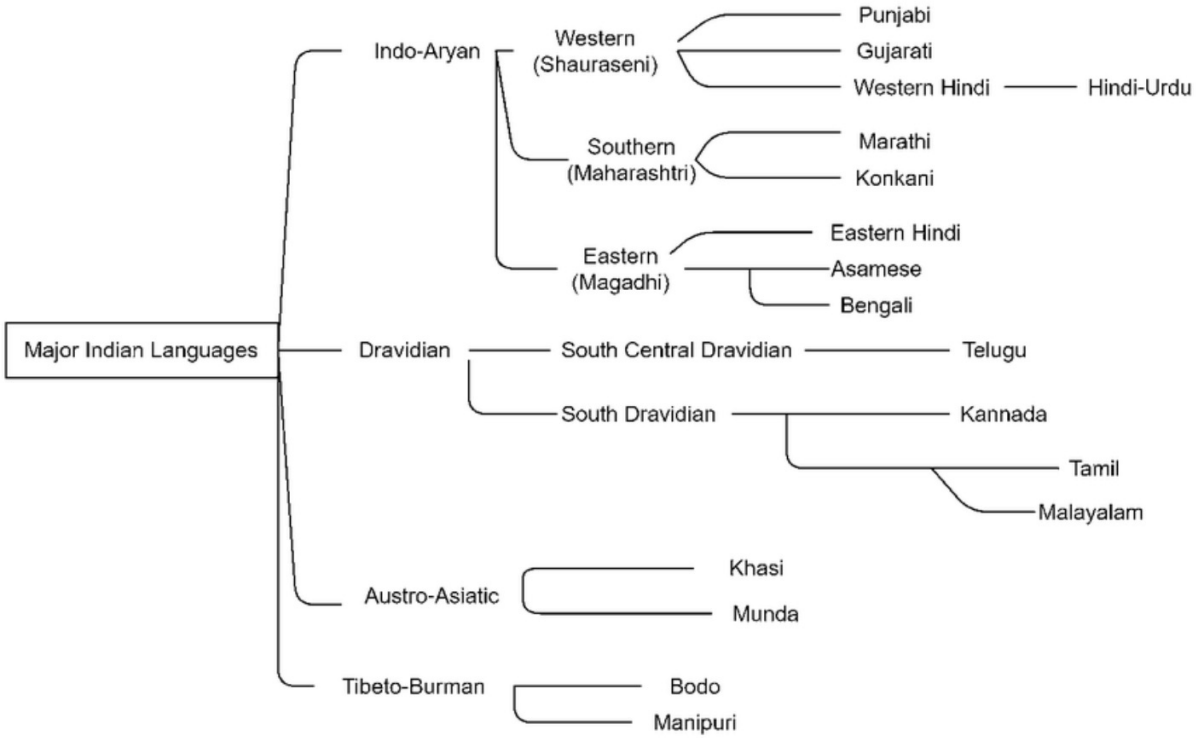
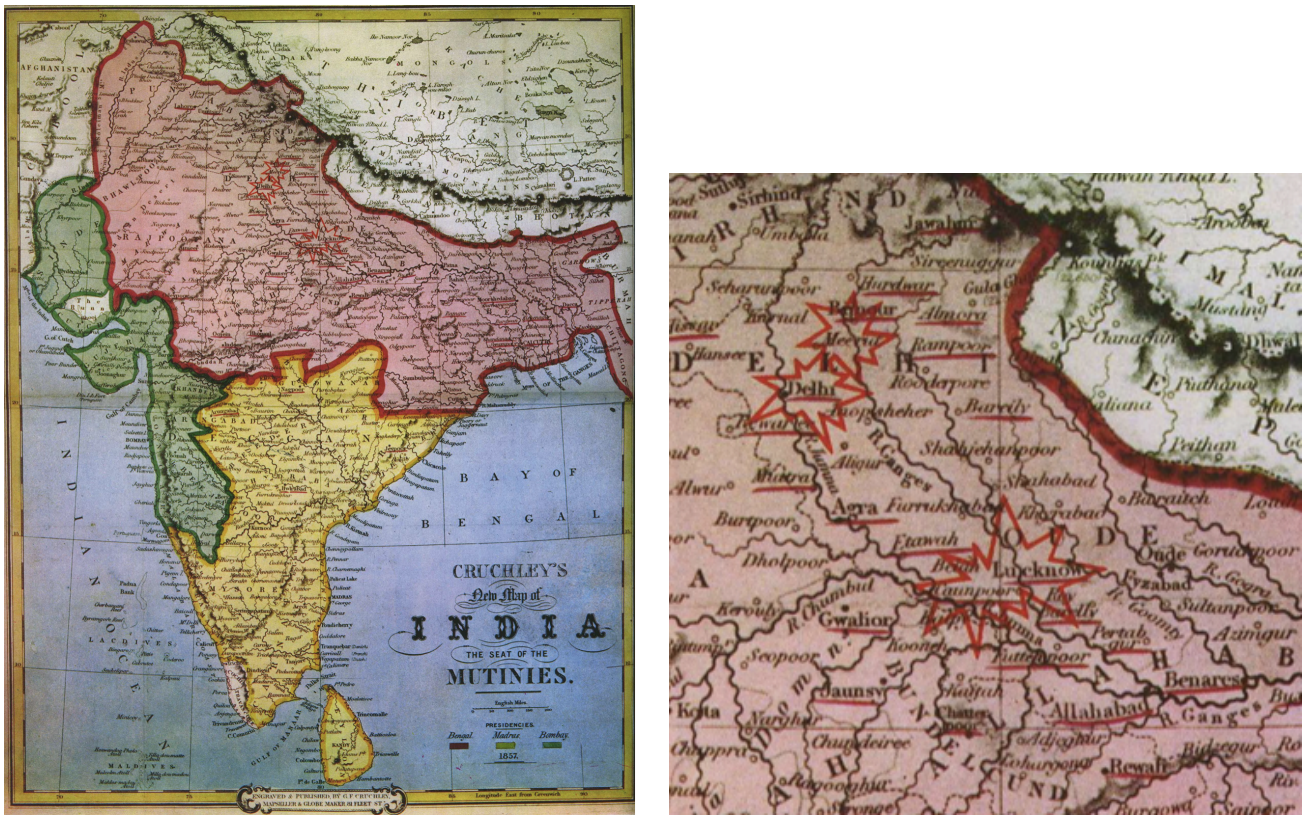


Figure 26: Routes for the transmission of smallpox vaccine lymph from Britain to Persia and India, 1802–1805. Adapted from Bennett (2007).



Tree diagram to illustrate the language closeness of major Indian languages

Figure 27: Indian Language Branches



(a) Full Map (b) Detail

Figure 28: Crutchley 1857 Conflict Map