Plague, Planning, and Persistence: Residential Segregation and Education in India

Khadija Aslam*

Abstract

This paper estimates the causal effect of residential segregation on inter-generational educational mobility among Muslims in India. I exploit the bubonic plague of the late nineteenth and early twentieth centuries as a source of exogenous variation in segregation. Plague-driven mortality, mass evacuation, and the colonial town planning that followed reshaped the spatial organization of cities in Bihar, West Bengal, and Uttar Pradesh in ways that have proved durable: plague-affected districts remain 6 to 11 percentage points more residentially segregated today. Using plague exposure as an instrument for contemporary segregation, I find that a one standard deviation increase in the dissimilarity index reduces years of completed schooling relative to the district mean by 0.48 years and lowers the conditional rank by 5.27 percentile ranks for Muslims with below-median parental education. These effects operate at the intensive and persist across the parental distribution, suggesting that segregation imposes a group-level constraint that household wealth alone cannot offset. Preliminary evidence points to three mechanisms: higher land use inequality in plague-affected districts, systematic under-provision of public goods in Muslim neighborhoods, and a higher opportunity cost of schooling arising from the concentration of household enterprises within these enclaves.

JEL classification: J15, N35, R23, I24, O15

Keywords: Residential segregation, inter-generational mobility, colonial institutions, bubonic plague

*Corresponding author: Khadija Aslam, Department of Economics, Delhi School of Economics, University of Delhi, Delhi 110007, India. E-mail address: khadija@econdse.org.

Disclaimer: Work in progress. Please do not quote.

I am thankful to my supervisor Anirban Kar (Delhi School of Economics) and Suresh Naidu (Columbia University) for all their inputs and constant guidance. I am also thankful to faculty and PhD students at Columbia University for their feedback at the Development Economics Colloquium.

1 Introduction

Muslims are India's largest religious minority, accounting for roughly 14 percent of the population, yet they rank among its most economically disadvantaged groups. Decades of survey evidence document that Muslim households earn less, own less land, and are substantially less educated than the national average, gaps that persist and have, widened over time. A growing literature attributes part of this disadvantage to local factors and area-based determinants. Muslim neighborhoods in Indian cities are systematically deprived of public goods, and residential concentration is associated with markedly lower inter-generational mobility in education. Yet the causal status of residential segregation as a driver of these outcomes remains unresolved. Segregation is endogenously determined by the same economic and political forces that shape educational attainment directly, making it difficult to separate the effect of spatial concentration from the characteristics of the households who sort into segregated areas.

A distinctive feature of Muslim segregation in India is that the enclaves are economically diverse. Because of housing discrimination, communal tensions, and dietary restrictions, even affluent Muslims often find it difficult to move into mixed neighborhoods. Using NFHS data matched to dissimilarity indices from the Sixth Economic Census, I show that wealth dispersion within urban Muslim neighborhoods rises steadily with district-level segregation, whereas in urban non-Muslim neighborhoods it falls. No such pattern appears in rural areas, which suggests that the economic diversity of Muslim enclaves is largely an urban phenomenon. This sets Muslim segregation in India apart from Black residential segregation in the United States, where spatial concentration and economic marginalization tend to reinforce each other (Cutler and Glaeser, 1997; Logan and Parman, 2025). It also raises the central question of this paper. Living near more affluent co-religionists could generate positive spillovers through networks, social capital, and role models, or the deprivation and state neglect that come with ghettoization could deepen the disadvantage of the poorer Muslims who live in these enclaves. Which of these forces dominates is an empirical question, and answering it requires separating the effect of segregation from the characteristics of the households who live in segregated areas.

This paper addresses that identification problem by exploiting the bubonic plague of the late nineteenth and early twentieth centuries as a source of exogenous variation in residential segregation. The epidemic, which arrived in India in 1896 and killed over ten million people, set off a chain of consequences with lasting spatial implications. Plague-driven mortality and mass evacuation disrupted settled patterns of residential cohabitation across cities in Bihar, West Bengal, and Uttar Pradesh. In the aftermath, the British colonial administration established City Improvement Trusts in the major affected cities, undertaking large-scale demolition, slum clearance, and the planned resettlement of displaced populations on the urban periphery. These interventions, framed as a public health response to overcrowding, encoded the social hierarchies of colonial governance into the built environment: lower-caste communities and Muslims were systematically relocated to peripheral neighborhoods, hardening the city's social geography along lines of religion and caste. The lines of segregation institutionalized in this period have proved durable. Plague-affected districts remain more segregated today, more than a century after the epidemic subsided.

The key feature that enables identification is the geographic distribution of the plague itself. The incidence of the plague was determined by the species of rat flea prevalent in a given locality, the climatic conditions at the time of the first outbreak, and the local ecology that governed the rat's ability to thrive and spread. Since none of these factors bear any systematic relationship to the educational attainment, caste composition, or institutional quality of affected districts prior to the epidemic, plague exposure constitutes a plausibly exogenous source of variation in the colonial town planning it subsequently triggered and, through that channel, in the residential segregation that persists today. While planning intervention and its intensity was itself correlated with city size and economic characteristics, using plague exposure as the instrument addresses this concern directly.

The paper makes the following contributions. The first is a new panel dataset of religious dissimilarity indices at the subdivision level for Bihar, West Bengal, and Uttar Pradesh, spanning 1901 to 1961 and constructed from thana- and ward-level population tables in the decennial Census of India. This dataset allows me to trace the evolution of residential sorting from before the plague through the first decade of Indian independence, a period that encompasses the plague itself, the

town planning interventions that followed, communal violence in the 1940s, and the demographic dislocations of partition. The second contribution is to show that residential segregation along religious lines increased more sharply in plague-affected subdivisions than in unaffected ones. Measured as the change in the dissimilarity index among plague-affected subdivisions relative to the change among unaffected subdivisions over the same census years, the gap between 1911 and 1961 reaches 6.6 percentage points in Bihar and 9.7 percentage points in West Bengal. In both states the steepest rise came between 1931 and 1961, when the gap widened by 5.9 percentage points in Bihar and 7.2 percentage points in West Bengal, consistent with segregation consolidating as colonial town planning was absorbed into the built form of cities and as partition-related violence accelerated an already ongoing process of spatial sorting. For Uttar Pradesh, where I have so far curated data for three major plague-affected districts, the dissimilarity index rose by 8.5 percentage points in Kanpur and 8.1 percentage points in Allahabad between 1911 and 1931, while Lucknow saw a decline of 2.3 percentage points. Contemporary data from the Sixth Economic Census confirm that plague-affected districts remain more segregated today, with dissimilarity indices 6 to 11 percentage points higher in urban areas across all three states.

The second contribution is methodological. While the dissimilarity index is the standard measure of residential segregation, it captures only the level of unevenness at a single point in time. It cannot show whether segregation is deepening in the same wards or shifting elsewhere, whether it is driven by Muslims concentrating or non-Muslims leaving, or whether segregated wards form a single bloc or lie scattered across the city. I augment the index with two measures built from the same ward-level census data. The first is a dynamic segregation index, which weights the change in each ward's distance from even representation by its distance at baseline, showing whether segregation is entrenching in the historically segregated wards or whether those wards are dissolving as new enclaves form elsewhere. It also splits the change in the dissimilarity index exactly into a minority and a majority channel, separating minority flight from majority consolidation. Finally, I use a spatial clustering index, to measure whether wards of the same type are grouped together. Applied to Calcutta, Kanpur, and Lucknow, these measures show that the dissimilarity index alone can hide very different processes. The index rose by similar margins in Calcutta and Kanpur, but the

rise was driven mainly by the movement of non-Muslims in Calcutta and almost entirely by the movement of Muslims in Kanpur. In Lucknow the index barely changed, yet this concealed large and offsetting shifts, as Muslims dispersed across the city while non-Muslims consolidated around the former civil lines. Spatial clustering rose in all three cities, including Lucknow, suggesting that the communal geography of a city can harden even when the conventional measure of segregation stays flat.

I then employ an instrumental variable specification to estimate the causal effect of residential segregation on educational outcomes, using district-level plague exposure as the instrument.A one standard deviation increase in the dissimilarity index reduces years of completed schooling relative to the district mean by 0.48 years and lowers the conditional rank by 5.27 percentile ranks for Muslims with below-median parental education, with the conditional rank penalty persisting (a decline of 4.06 percentile ranks) for those with above-median parental backgrounds. These effects operate primarily at the intensive margin: net attendance ratios are largely unaffected, but children in more segregated districts complete fewer years of school and fall further behind the district average. Since better-off Muslim families also tend to remain within these enclaves, the educational penalty extends across the parental distribution, holding back children even from comparatively advantaged households. Non-Muslim children from Other Backward Classes, who face broadly similar socioeconomic conditions as Muslims on observable indicators, face no significant educational penalty from segregation, and OBC children whose parents are in the top half of the education distribution in fact complete about 0.4 more years of schooling in more segregated districts. Residential segregation thus operates as a group-level constraint for Muslims: the enclave imposes costs that cut across class lines, and upward mobility remains bounded by the collective conditions of the neighborhood rather than the resources of any individual family.

Preliminary evidence points to three channels through which segregation operates. First, plague-driven displacement and colonial town planning generated higher land use inequality in affected districts: using thana-level census data on the number of houses and the area of each thana in square miles, preliminary results show that residential density became more unequally distributed

across thanas in plague-affected districts over time. Second, more segregated districts expose Muslim households to the systematic under-provision of public goods, as has been documented in earlier work. A one standard deviation increase in the dissimilarity index reduces the probability of a neighborhood having a primary school by 8.6 percentage points, and the shortfall is sharpest for public facilities, with public primary and secondary schools being significantly less likely to be found in the more segregated districts. Segregation's bite is concentrated in public provision, as private facilities are largely unaffected. Third, the co-location of small-scale enterprises and residential households within Muslim enclaves raises the opportunity cost of schooling: children in these neighborhoods face a standing demand for cheap labor from nearby workshops, which plausibly contributes to dropout rates that are well above the national average.

The paper relates to several strands of the literature. A large body of work on residential segregation in the United States documents significant negative effects of racial concentration on earnings and educational attainment (Cutler and Glaeser 1997; Logan and Parman 2025). The mechanisms and empirical context in India differ in important ways. As documented here, Muslim enclaves in India are economically diverse, which distinguishes segregation observed here from the canonical American pattern in which racial and economic marginalization reinforce one another. The paper also contributes to a literature on the long-run effects of colonial institutions on contemporary outcomes, and specifically to work connecting colonial urban planning to present-day spatial inequality. More directly, it builds on (Asher et al. 2024), who document the correlation between Muslim residential concentration and low inter-generational mobility in education, and on (Asher et al. 2026), where the authors establish the prevalence and scale of segregation in India. Finally, it adds to the literature that uses historical epidemics (Nunn 2010) as instruments for long-run institutional or development outcomes.

The remainder of the paper is organized as follows. Section 2 describes the historical background and the identification strategy. Section 3 presents the empirical framework. Section 4 describes the data. Section 5 presents results on the evolution of segregation from 1901 to 1961, including the dynamic and spatial clustering measures, and the first-stage relationship between plague exposure

and contemporary segregation. Section 6 presents the 2SLS estimates of the effect of segregation on educational outcomes. Section 7 discusses mechanisms, and Section 8 concludes.

2 Historical Background and Identification

2.1 Identification Strategy

In this paper I examine the change in residential segregation over time and estimate its effect on educational outcomes for Muslims in India. Estimating this relationship poses substantial identification challenges, as residential segregation is endogenously determined by household sorting, local economic conditions, political processes, and historical institutions, all of which may directly influence educational attainment. Formally, there are two potential sources of endogeneity that must be addressed: omitted variable bias and endogenous migration. Omitted variable bias may occur if we fail to account for factors such as the political influence that certain groups may exert in specific areas, which affect both segregation and economic outcomes. Endogenous migration, by contrast, occurs when individuals have preferences regarding their neighborhood, and when people with specific characteristics move into or out of segregated areas, potentially biasing the estimated effects. Both sources of endogeneity can lead to selective sorting and confound the causal relationship between segregation and the outcome variable. To overcome these concerns, the paper exploits plausibly exogenous variation in bubonic plague exposure and the consequent town planning undertaken in its aftermath, both of which generated and expanded residential segregation.

Historical accounts document that plague exposure prompted massive deaths and out-migration. The immediate response was evacuating homes and even demolishing houses or clearing slums infected by the plague. In the aftermath of the plague, extensive town planning was initiated by the British administration under the pretext of creating sanitary cities. While the scale of town planning was largely correlated with the size of the city, the outbreak of the plague was determined by the species of the rat and the climate in which the first case of bubonic plague was detected. In

this paper I exploit this variation in exposure to the plague to account for the change in residential segregation.

2.2 Bubonic Plague of the 19th century

The Bubonic Plague ravaged India in the late nineteenth and early twentieth centuries. The first case was registered in 1896 in Mandvi, a locality in Mumbai, and gradually spread to other parts of the country. Figure 1 records the outbreak and severity of the plague across different regions in India. It is believed to have caused the death of over 10 million people and the displacement of hundreds of thousands. High mortality rates were recorded in Bombay, Punjab, United Provinces, and the Bengal Presidency. Infection rates were relatively modest in Mysore, Hyderabad, Central India, Rajputana, and the Central Provinces, except in the city of Nagpur (the capital of the Central Provinces).

This variation is partly explained by geography (Yu and Christakos, 2006). The hills and mountains of Central India were a relatively unfavorable environment for fleas and rats compared to the flat plains of the Bombay Presidency, Punjab, the United Provinces, and the Bengal Presidency. Rivers also delayed its progress further north, and the eastern and southern regions never became strongholds of the disease. In addition, the severity of the plague was highly sensitive to the dominant species of rat and the season of outbreak in a particular region. Xenopsylla cheopis was the dominant rat flea in the plague-infected areas of Bombay. Since X. astia was the dominant rat flea in southern and eastern India, fewer cases were observed in those parts of the country (Sarkar N, 2024).

Bubonic plague in India typically peaked during the post-monsoon or autumn months (August to October) and sometimes in late winter or early spring (January to June), when warm temperatures and high humidity were ideal for X. cheopis to thrive. Although the outbreak was entirely contingent upon the presence of rats, the spread was faster in densely populated areas. Railways had some effect on the spread of the plague, and British India was more affected than princely states due to greater connectivity. Although Kanpur and Bhagalpur, both centers of trade, were affected more

than neighboring regions, no considerable tendency of the plague to move along trade routes was observed. The plague remained prevalent across India for the two next decade and the spread was sporadic, slow, and difficult to predict. In this paper I will be focusing on the states of Bengal, Bihar and Uttar Pradesh. Figure 2 maps plague exposure across the study districts in these three states.

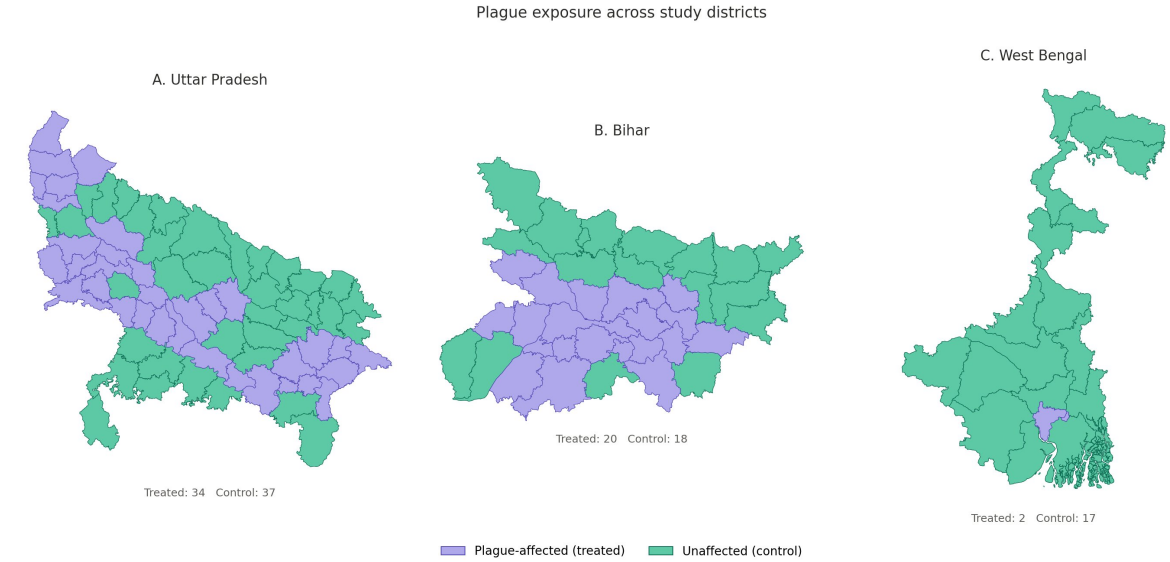
Figure 1: Spread of the Bubonic Plague in India in the Late Nineteenth and Early Twentieth Centuries

Phase	Date / Season	Regions Affected	Key Locations	Severity / Notes
Phase I: Origin & Initial Outbreak	Late 1896	Western India (Port City)	Mumbai (Bombay)	Point of origin in India; severe outbreak linked to port trade
	September 1896	Western & Eastern India	Ahmedabad; Calcutta	Ahmedabad heavily affected; Calcutta minimal impact
	December 1896	Western & Northwestern India	Poona; Karachi	Spread along railways and maritime routes
Phase II: Early Expansion	March 1897	Western India; Konkan; Gujarat; Deccan	Ratnagiri; Kutch; Kathiawar; Surat; Solapur	Rapid inland diffusion
	April 1897	Northern India (Punjab region)	Jalandhar; Hoshiarpur	Severe rural outbreaks
Phase III: Seasonal Resurgence & Southern Spread	April 1898	Eastern India	Calcutta	Plague gained foothold due to seasonality effects
	September 1898	Southern India	Hyderabad; Mysore	Limited spread
Phase IV: Diffusion into Eastern India	March 1900	Eastern India	Bihar; Orissa	Spread present but generally mild
Phase V: Northern Plains Intensification	1901	Northern & Eastern Indo-Gangetic Plains	Punjab; United Provinces; Bihar	Flat plains and seasonal conditions facilitating wider spread and higher mortality
Phase VI: Limited Presence in Madras Presidency	1901	Southern India (East Coast)	Madras Presidency	Sporadic cases and minor outbreaks
Phase VII: Central Indian Penetration	December 1902	Central India	Central Provinces	Moderate spread
Phase VIII: Residual / Localized Persistence	1904	Western India (Gujarat)	Surat region	Localized recurrence

2.2.1 Bengal, Bihar and Uttar Pradesh

Within the Bengal Presidency, the plague followed an uneven geography that maps closely onto the variation in housing structure and rat ecology documented in the 1911 Census Report (O'Malley, 1913). Plague first appeared in Bengal in 1898, with two outbreaks, one in Calcutta and one in Backergunge. By the cold weather of 1900-01 the disease had spread over a larger area, but thereafter its course diverged sharply between Bengal proper and Bihar. In Bihar, the plague established itself with 'wonderful regularity,' recurring each cold season and attaining its greatest virulence in the first three months of the year, before practically disappearing through the hot and rainy seasons. Over the decade 1901-1910, Bihar and Orissa recorded 545,450 deaths from plague, the vast majority concentrated in Bihar, with Orissa and Chota Nagpur remaining almost entirely

Figure 2: Plague Exposure across Study Districts in Uttar Pradesh, Bihar, and West Bengal



Purple denotes plague-affected (treated) districts; green denotes unaffected (control) districts.

immune. The districts of Patna, Saran, and Shahabad suffered most severely, with plague deaths over the decade representing 90, 80, and 80 deaths per mille of the 1901 population respectively.

The marked contrast in severity between Bihar and Bengal, particularly its spread in rural areas, is directly attributable to differences in the rat ecology of the two regions. In Bihar, villages were compact and overcrowded clusters of thick mud-walled houses with narrow passages between walls, fostering ideal conditions for rats to shelter and breed. In Bengal, houses were built on raised ground, each in its own compound, allowing better ventilation that offered little harborage to rats. Rural Bengal consequently suffered little from bubonic plague and its spread was largely confided to Calcutta and the urban towns around Calcutta. The initial spread of the plague within Bihar was also shaped by the railway network. Grain godowns, infested by plague rats, were naturally most common in towns and stations along the railway, facilitating the movement of infected rodents alongside goods. In twelve towns situated on the Bihar railway, there were some 60,000 plague deaths after 1901, representing one-eighth of their combined 1901 population. Although the plague did not establish an annual foothold in Calcutta, it exhibited signs of population pressure and displacement. The number of persons born in Calcutta and resident there fell by nearly 34,000 between 1901 and 1911, as the population spread outward to Howrah and the suburban municipalities. This dispersal

was partly a consequence of the health conditions, poor sanitation and overcrowding of the city itself.

The consequences for urban population were severe and lasting. The continued presence of plague in Bihar was the principal cause of a 2.5 percent decline in the number of persons living in Bihar and Orissa towns between 1901 and 1911. The plague not only caused serious mortality but also affected trade and industries in the affected towns. In Gaya, for instance, 30 percent of the population had abandoned their work and fled their homes at the time the census was taken. The population displacement caused by the plague extended well beyond temporary flight. In Bihar, the balance of migration was heavily negative: emigrants outnumbered immigrants by 36 per mille in 1901 and 49 per mille in 1911. This out-migration was the combined result of plague mortality, famine, and the absence of industrial development in plague affected towns.

When plague broke out, the immediate and nearly universal response across Bihar was the physical evacuation of homes. The disease was understood to thrive in congested dwellings, and populations responded by abandoning their houses and encamping in the open in temporary huts. In villages attacked by epidemic, inhabitants would demolish their houses entirely and relocate to a new site. Even in towns, plague prompted the partial or complete desertion of residential quarters. In Gaya, Jehanabad and Chapra in Saran, a large proportion of the population had fled to villages in the interior at the time of the 1911 census.

The United Provinces recorded 1,315,252 or 2.76 per 1000 of the total deaths from plague over the decade 1901-1910, producing an overall population decline of 1.1 percent in British territory. At the district level, the mortality was catastrophic in places like Ballia, Ghazipur, Azamgarh, Jaunpur, Lucknow, Allahabad, and Unao. Women bore a disproportionate share of mortality on account of being confined to the domestic space which maximised their exposure to the rat flea.

Here as well the immediate population response to plague was the evacuation of infected houses, a phenomenon so widespread and so disruptive that the Census Report treats it as a distinct category of demographic movement ‘temporary plague migration’. People abandoned their homes

and encamped in temporary dwellings or fled to uninfected neighbouring districts, in the periphery of the plague centres. In Mirzapur city, five-sixths of the population had fled before the census was taken. In Mainpuri, Hardoi, Jaunpur, Sultanpur, Bara Banki, and Ballia, too observed lowered immigrant counts and out-migration.

2.3 Town planning following the plague

Following the spread of the plague, the colonial government undertook a series of town planning and improvement reforms to address overcrowding and sanitary deficiencies held responsible for spread of the disease. The institutional vehicle for these reforms was the City Improvement Trust, a model first developed in Bombay and subsequently replicated across British India. In Bombay, The Trust acquired nearly two-ninths of the city's land area by the eve of the First World War and by 1918 had displaced approximately 200,000 residents through its demolition operations (Arnold, 1993). The Calcutta Improvement Trust was eventually established in 1911, with broad powers to drive roads through congested streets, create open spaces, and develop suburban areas for the displaced population. In the United Provinces, under the Town Improvement Act of 1919 trusts were constituted in Lucknow, Kanpur, Allahabad, and other cities, completing a network of improvement institutions spanning the major urban centres of British India (Town Improvement Trusts in India, 1945, p. 285). As a part of the interventions, insanitary dwellings were demolished, sewage systems were rebuilt, streets were widened, and new residential areas were planned on the periphery. Improvement trust interventions frequently demolished more houses than what was rebuilt, and what was framed as a public health intervention carried a spatial logic that was inseparable from the social hierarchies of colonial governance (Singh, 2025). Planned suburbs were created in former plague camps with sewage systems and key civil amenities to create self-sustaining neighborhoods which further reduced the frequency of interactions between the residents of the city and those inhabiting the peripheries (Hibbard, 2024).

Town planning measures were evidently not uniform across cities but largely proportional to the size of the city, and the effect of the plague and whether it had a British administered cantonment area. Town improvement trusts were established and elaborative planning measures were adopted

in big cities like Mumbai, Kolkata, Patna, Lucknow, Allahabad, Kanpur and Bangalore which were densely populated and severely affected by the plague. In a number of other cities affected by the plague modest interventions through slum clearance, drainage improvements, and sanitation checks were administered through the existing municipal bodies.

The main part of the city was occupied by the British, with upper castes inhabiting the immediate neighborhood. The lower caste groups and other marginalized communities were relegated to live in the outer periphery made accessible by cheap conveyance (Meller H, 1979, Ghosh N, 2020). The outcome was a progressive hardening of the city's social geography along racial, class and caste lines. The dual structure of colonial cities, the European civil lines and cantonment on one side, the native town on the other, was reinforced following the plague. In Ahmedabad, housing cooperatives built for mill workers during the same period constructed two separate sets of residential quarters, one for Hindus and one for Dalits and Muslims, thereby encoding caste segregation directly into the planned city (Jamhoor, 2023). Colonial planning and housing schemes formed the basis for postcolonial cities, and were subsequently expanded by upper-caste elite power brokers in municipal government (Jamhoor, 2023). Once the British left India, the enclaves they had occupied were inherited by the Indian affluent, and the lines of segregation, already well-established, became progressively absolute. Although colonial urban planning following the plague was undertaken to ensure sanitation and improved public health, by segregating space for the elite and displacing the poor to the periphery, it reinforced spatial fragmentation and exacerbated social and economic hierarchies. It is no wonder that cities planned by the British improvement trusts in colonial India continue to be highly segregated even today (Meller, 1979; Ghosh, 2020; Kidambi, 2001; Dossal, 1991).

3 Empirical Strategy

In this paper, I employ an instrumental variable (IV) strategy to examine the causal impact of residential segregation on educational outcomes. Before turning to the causal estimates, I first document how residential segregation evolved over the period 1901 to 1961, tracing the trajectory

of spatial sorting within districts from the onset of the plague era through the first decade of independence. This descriptive exercise serves two purposes: it establishes that the post-plague town planning interventions were associated with reorganization of residential space, and it motivates the use of historical plague exposure as a source of identifying variation in contemporary segregation. I employ a two-stage least squares (2SLS) framework, specified as follows:

$$\begin{aligned} Y_{ids} &= \beta_0 + \beta_1 \widehat{DI_{ds}} + \beta_2 X_{ids} + \beta_3 \text{Year}_y + \beta_4 \text{State}_s + \epsilon_i \\ DI_{ds} &= \gamma_0 + \gamma_1 \text{Plague}_{ds} + \upsilon_{ds} \end{aligned} \tag{1}$$

where Y_{ids} is the educational outcome for individual i in district d in state s. DI_{ds} is a measure of the dissimilarity index, capturing the extent of residential segregation in district d. The vector X_{ids} includes individual and household-level characteristics, such as the father's years of education, a household wealth index, number of siblings, age, and sex of the child. Since the analysis pools two rounds of the National Family Health Survey (NFHS), I additionally control for the year of interview using survey-year fixed effects. State fixed effects are included to absorb time-invariant differences across states.

I measure residential segregation with two indices. The dissimilarity index is the share of Muslims (or non-Muslims) who would need to relocate to achieve an even distribution across sub-district units, and ranges from 0 (no segregation) to 1 (complete segregation). The isolation index is the probability that a randomly drawn Muslim shares a neighborhood with another Muslim, and so captures concentration within blocks. I study four educational outcomes. The net attendance ratio is the share of children of official school age attending primary or secondary school, measured separately for the two levels. Years of schooling are the child's completed years of education, expressed relative to the district mean. The conditional rank is a child's percentile rank in the district education distribution, within sector and survey round, computed using survey weights and scaled from 0 to 100, with higher values indicating a higher rank (Asher, Novosad and Rafkin, 2024; Chetty and Hendren, 2018). I stratify the sample by parental education, assigning children to the

bottom half if their parent's years of schooling fall below the district median and to the top half otherwise.

Current residential segregation is instrumented with a dummy variable Plague_{ds}, which captures historical exposure to the bubonic plague in district d in state s. As documented in the historical record, the outbreak and severity of the plague across districts was primarily determined by the dominant species of rat flea present in the local environment, the ecology of the terrain, and the climatic conditions prevailing at the time of the first outbreak. None of these ecological and climatic determinants bear any systematic relationship to the pre-existing human capital endowments, caste compositions, or institutional quality of affected districts, making plague exposure plausibly orthogonal to underlying city characteristics. Plague-induced mortality, displacement, and the town planning interventions that followed generated the patterns of residential segregation that lend credibility to the relevance condition. The exclusion restriction requires that historical plague exposure affects contemporary educational outcomes only through its effect on residential segregation. The plague was an acute demographic shock that spanned over two decades, and while its direct demographic effects had largely dissipated by the middle of the century, the town planning interventions it prompted continue to shape cities even today. Town planning intensity was itself correlated with the economic characteristics of cities; using plague exposure as an instrument, rather than town planning activity directly, addresses this concern. While I acknowledge that the plague may well have altered other city characteristics that independently affect educational outcomes today, the identifying assumption does not require that it left no other trace on these cities. It requires, instead, that any such channel did not differentially affect the two religious communities, and that whatever influence it exerted on the overall level of educational attainment in a city had dissipated by the end of the century. This is not an assumption I can defend unconditionally as yet. The preliminary results reported here rest on a simple IV specification, without district-level controls, and I am working towards a specification that is more robust to any such threats to identification.

4 Data Source

4.1 Segregation Data: 1901-1961

The primary source for the historical segregation measures is the decennial Census of India, spanning the census years 1901, 1911, 1921, 1931, and 1961. I draw on the provincial census tables reporting population by religion, at the thana and ward level (where available) for Bihar, West Bengal, and Uttar Pradesh. The census years 1941 and 1951 are excluded in this draft on account of enumeration issues and the population displacements associated with partition respectively. It must also be noted that the Census does not report religious demography below the sub-district level, at the thana or ward, for any year after 1961, which limits the period over which the dissimilarity index can be constructed.

4.2 Plague Exposure Data

Data on plague exposure are drawn from two primary sources. The first is the decennial Census of India, which in its district-level administrative volumes records in considerable detail the events that transpired in each district over the preceding decade. These accounts document whether a district was affected by the plague, mortality estimates attributable to the epidemic, any other major demographic or public health events of note, and the population's response to the outbreak, providing both a binary indicator of plague exposure and a qualitative record of its severity for each town and, in a number of cases, at the thana level, allowing me to interpret plague intensity for each subdivision. Additional contextual information on the geographic spread of the plague, the dominant species of rat flea across different regions, and the timing of first recorded outbreaks is drawn from the Indian Medical Gazette (1906).

4.3 Town Planning Data

Data on post-plague town planning interventions are drawn from multiple sources. The primary source is Town Improvement Trusts in India (1945), which documents the formation, powers, and activities of improvement trusts across British India, including the Bombay Improvement Trust (1898),

the Calcutta Improvement Trust (1911), and the trusts constituted under the United Provinces Town Improvement Act (1919) for Lucknow, Kanpur, and Allahabad. This source provides information on the timing of trust formation, the scale of slum clearance and land acquisition operations, and the geographic scope of improvement schemes. I supplement this with the works of historians examining town planning in specific cities (Home, 1997; Kulkarni, 1979; Kosambi and Brush, 1988). Additional epidemiological and geographic context on plague-led displacement and planning across British India is drawn from Sarkar (2024). For Bihar, supplementary information on public health measures and sanitary improvements in the post-1931 period is drawn from the District Gazetteers of Bihar and Orissa.

4.4 Segregation Data: 2011

The contemporary measure of residential segregation is constructed from the Sixth Economic Census of India (2013-14). The Economic Census provides establishment-level data geo-referenced to the enumeration block, where each block consists of fewer than 200 establishments. I use the religion of the owner of each establishment to construct town and Tehsil-level measure of segregation.

4.5 Individual and Household Data: National Family Health Survey

The primary data source for the analysis of educational outcomes is the National Family Health Survey (NFHS). I pool the fourth round (NFHS-4, 2015-16) and the fifth round (NFHS-5, 2019-21), which together provide nationally representative data on individual and household characteristics including educational attainment, household wealth, and demographic composition.

5 Impact on Residential Segregation

5.1 Using the Dissimilarity Index

I first document the evolution of residential segregation across districts in Bihar, West Bengal, and Uttar Pradesh from 1901 to 1961. The geographic unit of analysis for the 1901-1961 period is the thana, the jurisdiction of a single police station and the primary sub-district unit used for census

enumeration (Blunt, 1915). Each district is divided into multiple subdivisions, and each subdivision into multiple thanas, with the average thana population close to 40,000 persons. For a couple of cities in West Bengal and Uttar Pradesh, I additionally have ward-level data over time, which allows for a finer-grained picture of within-city residential sorting in those cases. I compute the Dissimilarity Index (DI) following Duncan and Duncan (1955) as defined in equation (2) using thana as the unit of enumeration. I aggregate across all thanas within a sub-division to obtain the DI for each subdivision. Data on plague exposure are once again recorded for each subdivision and, in a number of cases, also at the finer thana level, allowing me to capture both the presence and intensity of the epidemic within each administrative unit. For Bihar and West Bengal, I have so far computed the subdivision-level dissimilarity indices for 1901 (Bihar only), 1911, 1931, and 1961. For Uttar Pradesh, at present I have indices only for the three major plague-affected cities: Lucknow, Kanpur, and Allahabad, spanning 1901 to 1931, and the figures should be interpreted accordingly. While the town planning activities were largely contained in urban areas, the preliminary findings consider the change in segregation across the entire district. This is because I have not yet disaggregated the urban and rural components of each thana or adjusted for changes in district boundaries over time.

$$DI = \frac{1}{2}\sum_{i}^{N}\left|\frac{Muslim_i}{Muslim_{total}} - \frac{Non\text{-}Muslim_i}{Non\text{-}Muslim_{total}}\right| \tag{2}$$

$$II = \sum_{i}\left(\frac{Muslim_i}{Muslim_{total}} \times \frac{Muslim_i}{N_i}\right) \tag{3}$$

where $Muslim_i$ ($Non\text{-}Muslim_i$) (N_i) is the population of Muslims (Non-Muslims) (total population) in block i, and $Muslim_{total}$ ($Non\text{-}Muslim_{total}$) is the population of Muslims (Non-Muslims) in a particular town.

To compare the two groups, I take the change in the average DI among plague-affected subdivisions between two census years and subtract the change among unaffected subdivisions over the same years. For the years 1911 and 1931, for instance, this is $(\overline{DI}^{T}_{1931} - \overline{DI}^{T}_{1911}) - (\overline{DI}^{C}_{1931} - \overline{DI}^{C}_{1911})$, where T denotes plague-affected and C unaffected subdivisions, and I refer to it as the relative

change in DI. A positive value means that segregation rose by more, or fell by less, in plague-affected subdivisions.

In Bihar the average DI among the 23 plague-affected subdivisions was 0.116 in 1901 and 0.132 among the 11 unaffected controls. However, the 1901 census in Bihar is subject to enumeration irregularities since recording the census was a relatively new administrative process at that time. Consequently, the quality of sub-district enumeration is uneven and I therefore treat the 1901 baseline with caution. Between 1901 and 1911, the relative change in DI is -2.8 percentage points, consistent with plague-driven out-migration temporarily disrupting settled residential patterns in both affected and non-affected areas. Between 1911 and 1931, segregation in plague-affected subdivisions rose modestly relative to unaffected areas, with a relative change of 0.68 percentage points. The sharpest divergence occurs between 1931 and 1961, when the DI in plague-affected subdivisions rose by 5.94 percentage points more than in unaffected controls. The increase in segregation among plague-affected subdivisions was more than twice that observed among unaffected controls, suggesting that the gap in residential sorting between the two groups consolidated in the decades following partition. Over the full 1911-1961 period, the relative change in DI amounts to 6.6 percentage points.

In West Bengal, the average DI among the 3 plague-affected subdivisions was 0.229 in 1911 and 0.183 among the 33 unaffected controls. Between 1911 and 1931, segregation in plague-affected subdivisions rose by 2.23 percentage points relative to unaffected areas, before accelerating sharply to 7.20 percentage points between 1931 and 1961. Over the full 1911-1961 period, the DI in plague-affected subdivisions rose by 9.73 percentage points more than in unaffected controls, with the treated group registering an increase in DI approximately twice as large as that of the control group.

The pronounced rise in segregation between 1931 and 1961 in both Bihar and West Bengal warrants further study. The eve of partition was marked by severe communal violence across both provinces. In Bihar, riots erupting between October and November 1946 were triggered by the Great

Calcutta Killings and the Noakhali violence. Hindu mobs targeted Muslim communities across the Bihar province, with casualty estimates ranging from 2,000 to 30,000 and the Muslim League reporting that its members remained afraid to leave their homes in the aftermath (Damodaran, 1990; Ghosh, 2009). In West Bengal, the Noakhali riots of October 1946 and the subsequent partition of Bengal drove millions of Hindu refugees out of East Bengal into West Bengal in a sustained reorganization of residential patterns that continued well into the 1950s (Chatterji, 1995; Chatterji, 2007). These episodes of violence and forced displacement plausibly contributed to the sharp rise in measured segregation by 1961. However, it is important to note that the treated group comprises all plague-affected districts, irrespective of how severely any individual district was struck by partition-related violence. Segregation rose across this group as a whole, including in districts where the primary historical disruption was induced by plague and not partition. Partition, in this reading, accelerated a process of residential sorting that had already been set in motion by the plague and institutionalized in built form by colonial improvement trusts. The consistent positive differential between treated and control districts across both Bihar and West Bengal, and across the full 1911-1961 period, is more plausibly attributed to the structural legacy of plague-era spatial reorganization, which was then expedited by communal violence.

In Uttar Pradesh, I have so far curated data for the three major plague-affected cities, Lucknow, Kanpur, and Allahabad, for the period 1901-1931. Across the three cities, the average DI rose by 8.1 percentage points between 1901 and 1931. The change between 1901-1911 was small, with the DI rising by 0.02 percentage points in Allahabad, 2.04 percentage points in Kanpur, and 1.71 percentage points in Lucknow. The larger increases came after 1911. Between 1911 and 1931 the DI rose by 8.5 percentage points in Kanpur and 8.13 percentage points in Allahabad, consistent with severe plague mortality and substantial improvement trust intervention under the United Provinces Town Improvement Act of 1919. Kanpur thus exhibits the most pronounced trajectory among the three cities, with its DI rising by 10.5 percentage points over the three decades. Lucknow, by contrast, saw its DI fall by 2.3 percentage points between 1911 and 1931. The aftermath of the 1857 uprising had prompted the British to undertake large-scale demolition and spatial reorganisation of Lucknow, driving wide roads through dense Muslim neighbourhoods and relocating populations

along lines of religion and perceived political loyalty, well before the plague-era town planning that is the focus of this paper. This earlier reorganisation likely explains why Lucknow does not exhibit the same increase in segregation observed in other plague-affected cities.

Taken together, segregation rose sharply between 1911 and 1931 in the cities with active plague exposure, by 8.1 percentage points in Allahabad, 8.5 percentage points in Kanpur, 7.9 percentage points in Calcutta, and 2.1 percentage points in Patna, with Lucknow the only exception. The most segregated cities are those that underwent extensive improvement trust interventions, which is consistent with colonial town planning being the primary mechanism.

5.2 Segregation dynamics

The dissimilarity index records the level of residential unevenness at a single point in time, but gives little information on the underlying process. Furthermore, the aggregate measure may conceal which wards are becoming more dissimilar, what the underlying mechanism of this change is, and the geography of separation. On the question of persistence, when aggregate unevenness rises, are the wards that were already segregated at baseline deepening further, or is the locus of unevenness migrating to new areas even as old enclaves dissolve? The DI is identical in both cases, yet the two describe opposite trajectories of communal separation. With regard to mechanism, a ward that moves away from even representation may do so because Muslims are concentrating, because non-Muslims are withdrawing, or through some combination of the two, which the index does not isolate. Finally, the DI treats wards as exchangeable cells and is invariant to any relabeling of which ward sits next to which, so it cannot register whether the wards carrying the unevenness form a consolidated bloc or lie scattered across the urban fabric. I address the first two questions with a dynamic decomposition of the index (Section 5.2.1) and the third with a spatial clustering statistic following Moran's I statistic (Moran 1950; Anselin 1995) (Section 5.2.2).

5.2.1 The dynamic segregation index

Let the signed deviation of ward j at period τ be

$$\psi_j^\tau \;=\; \alpha_j^\tau - \beta_j^\tau \;=\; \frac{a_j^\tau}{A^\tau} - \frac{b_j^\tau}{B^\tau}, \tag{4}$$

where α_j^τ and β_j^τ are the ward's shares of the city's Muslim and non-Muslim populations, respectively. A ward is Muslim-overrepresented when $\psi_j > 0$ and non-Muslim-overrepresented when $\psi_j < 0$; by construction $\sum_j \psi_j^\tau = 0$, and the dissimilarity index is $\mathrm{DI}^\tau = \frac{1}{2}\sum_j |\psi_j^\tau|$. Comparing DI^0 with DI^1 reveals whether unevenness rose or fell, but not whether the rise was concentrated in the historically most segregated wards. To recover persistence, I weight each ward's trajectory by its baseline distance from evenness,

$$D^* \;=\; \sum_j |\psi_j^0|\,\Delta|\psi|_j, \qquad \Delta|\psi|_j = |\psi_j^1| - |\psi_j^0|, \tag{5}$$

so that $D^* > 0$ when the wards already most uneven at baseline are the ones diverging further, segregation entrenching in the same geography, and $D^* < 0$ when those same wards converge on net. Because D^* re-weights the ward-level changes by their baseline unevenness rather than summing them uniformly, as DI does, its sign need not track that of DI: the aggregate index can rise while $D^* < 0$. This would happen when a city's historically segregated wards are dissolving even as new enclaves form elsewhere, the center of unevenness shifting rather than deepening.[1] The normalised counterpart,

$$\widetilde{D}^* \;=\; \frac{\sum_j |\psi_j^0|\,|\psi_j^1|}{\sum_j |\psi_j^0|} \in [0,1], \qquad r_0 = \frac{\sum_j (\psi_j^0)^2}{\sum_j |\psi_j^0|}, \tag{6}$$

is the baseline-magnitude-weighted mean of the period-1 distances from evenness. It exceeds the no-change reference r_0 when the historically segregated geography diverges on net.

Because $\psi_j = \alpha_j - \beta_j$, every movement of the deviation is the exact sum of a minority-share movement and a majority-share movement, which permits a decomposition into a minority and a

[1] D^* uses the change in the magnitude of the deviation, $\Delta|\psi|_j$, rather than the raw signed change $\Delta\psi_j$. The distinction matters only for wards that cross even representation between periods: a ward tipping from $\psi_j^0 = +0.30$ to $\psi_j^1 = -0.40$ has in fact grown more uneven by 0.10, which the magnitude-based weight records correctly while a naive signed product would register strong integration.

majority channel with no residual. Writing $\theta_j = \text{sign}(\psi_j^0)$ for the side of evenness a ward occupies over the transition,[2] the minority and majority channels are

$$M_j = \theta_j \, \Delta\alpha_j, \qquad W_j = -\theta_j \, \Delta\beta_j, \qquad M_j + W_j = \Delta|\psi|_j, \tag{7}$$

and they decompose the aggregate change without residual, $\Delta\text{DI} = \frac{1}{2}\sum_j M_j + \frac{1}{2}\sum_j W_j$. The sign of each channel is read against the ward's baseline type: in a Muslim-overrepresented ward a positive minority channel records deepening Muslim concentration and a positive majority channel records non-Muslim flight, whereas in a non-Muslim-overrepresented ward a positive minority channel records Muslim exit and a positive majority channel records non-Muslim consolidation. Table 1 summarises the city-level dynamics and Table 2 reports the channel split by ward type for each city. I have so far carried out this analysis for three cities, Calcutta, Kanpur, and Lucknow, each of which underwent extensive improvement-trust intervention, over the decades bracketing those interventions (1911-1931), in order to trace how the movements of the two groups diverged in their aftermath.

In Calcutta the ward-level dissimilarity index rose from 0.345 to 0.417 between 1911 and 1931, and the dynamic component confirms that this was entrenchment rather than relocation: $D^* = +0.005$ is positive and the normalised $\widetilde{D}^* = 0.046$ exceeds its no-change reference of 0.039, so the wards already most uneven in 1911 were, on average, the ones diverging further by 1931. The channel decomposition (Table 2, Panel A) locates the source of the rise on the majority side: the non-Muslim channel contributes +0.041 of the +0.073 increase against +0.032 from the Muslim channel, so the deepening of segregation in Calcutta is more of a story of non-Muslim redistribution than of Muslim enclosure. This is visible even within the Muslim-overrepresented wards, where the largest contribution comes from wards such as Calootola, in which the Muslim share of the city barely moved while the non-Muslim share fell sharply from 4.0 to 2.2 per cent: the unevenness deepened because non-Muslims departed, not because Muslims concentrated. Bhawanipur ward

[2]For the few wards that cross even representation between periods, θ_j is the time-weighted side $\int_0^1 \text{sign}(\psi_j(t))\,dt$ along the straight-line path between the two deviations, lying strictly inside $(-1, 1)$; this is the adjustment that keeps the channel sum equal to $\Delta|\psi|_j$ for crossing wards.

illustrates the same logic on the non-Muslim side, its non-Muslim city-share rising from 6.6 to 9.1 per cent as non-Muslims relocated southward.

Kanpur produced an almost identical aggregate movement through the opposite mechanism. The index rose from 0.386 to 0.443, a gain of +0.057, and $D^* = +0.001$ again indicates entrenchment in the baseline-segregated circles. Here, however, the rise is almost entirely a minority-side phenomenon: the Muslim channel accounts for +0.056 of the +0.057 increase while the non-Muslim channel contributes a negligible +0.001 (Table 2, Panel B). The action is concentrated not in the already Muslim-dominated circles, where the Muslim channel is essentially zero, but in the near-even and non-Muslim-overrepresented circles, where Muslim in-concentration of +0.025 and Muslim exit from non-Muslim areas of +0.030 together drive the change. The Sisamau circles are the sharpest case, flipping from slightly majority-leaning to strongly Muslim-concentrated as their Muslim population grew to 2.3 times the count expected from the city-wide rate while the non-Muslim population fell in absolute terms.

Lucknow is the most instructive of the three, because its near-static index conceals the largest underlying movements. The DI fell marginally, from 0.360 to 0.356, and $D^* = -0.001$ is negative, with $\widetilde{D}^* = 0.011$ falling below its reference of 0.012: the historically most uneven circles were, if anything, converging. Yet this near-stasis is not quiet. The minority and majority channels are large and almost exactly offsetting, at −0.034 and +0.030 respectively (Table 2, Panel C), so the aggregate change of −0.004 is the near-cancellation of two opposing redistribution rather than evidence of a settled geography. Muslims spread more evenly across the city as their numbers more than doubled over the decade, from 87,558 to 200,532. Against this dispersal, non-Muslims concentrated in the near-even circles, a majority channel of +0.018 falling almost entirely in that group, drawn disproportionately to Hazratganj and the adjacent areas of the former British civil station and turning what had been mixed circles into coherently non-Muslim-dominated ones. A change in the dissimilarity index of −0.4 per cent thus conceals the formation of a consolidated non-Muslim zone around the old civil lines and the even spread of Muslims across the city.

Taken together, the three cities make the case for the decomposition. An identical aggregate rise reflects non-Muslim redistribution in Calcutta and Muslim redistribution in Kanpur, while a flat index in Lucknow masks two of the largest group movements in the sample.

5.2.2 The spatial clustering index

The dissimilarity index and its dynamic component are both aspatial: each is invariant to any permutation of the wards, so neither can distinguish a city in which the Muslim-overrepresented wards form a single consolidated bloc from one in which the same wards lie scattered across the urban fabric, even at identical values of the index. The distinction matters because a resident of a minority-concentrated ward in a strongly clustered city remains within minority-concentrated space as she moves about her local area, whereas in a city of equal unevenness but interspersed geography the same movement would routinely carry her into majority-dominated parts of the city. Clustering therefore governs the physical exposure that the dissimilarity index leaves unmeasured, that is, whether the communal geography raises barriers to incidental inter-community contact, and with it differential access to schools, public-health facilities, and cross-community networks. To capture it I construct a block clustering index on the same signed deviations.

The historical censuses assign each ward to a larger administrative block: the geographical region in Calcutta, a charge in Kanpur and Lucknow, but do not record ward-to-ward boundaries, so co-membership within a block is the proximity relation the data certify. Setting $w_{jk} = 1/(n_R - 1)$ when wards $j \neq k$ share block R and zero otherwise, the index is

$$I \;=\; \frac{\sum_R L_R}{\sum_j \psi_j^2}, \qquad L_R = \frac{\Psi_R^2 - \sum_{j \in R} \psi_j^2}{n_R - 1}, \qquad \Psi_R = \sum_{j \in R} \psi_j, \tag{8}$$

a Moran-type autocorrelation of the signed deviations. Here $I > 0$ means that Muslim-overrepresented wards are found together in the same blocks, and non-Muslim-overrepresented wards likewise among themselves, more often than a random assignment of deviations to wards would produce. Because the block partition mechanically forces some pairs of wards to share a block and others not to, the reference value under random reassignment is $\mathbb{E}[I] = -1/(n-1)$, slightly negative rather than zero; I assess significance by permutation, reassigning the deviations

to wards $R = 9{,}999$ times while holding the block structure fixed, and locate the observed I in the resulting reference distribution. Table 3 reports the results.

Clustering rose in all three cities, including the one where unevenness did not. In Calcutta the index climbed from $I^0 = 0.282$ to $I^1 = 0.330$, both significant against the -0.034 benchmark, with the non-Muslim-majority North supplying roughly 62 per cent of the clustering in both periods as its constituent wards grew more internally coherent. Kanpur rose from 0.210 to 0.293, the most precisely estimated of the three, driven by the Anwarganj charge, the Muslim occupational quarter that is 79 per cent Muslim, whose share of the clustering grew from 70 to 78 per cent. Lucknow is the decisive case: although its dissimilarity index fell, its clustering index rose from 0.073 to 0.099. This was concentrated in the Hazratganj charge around the former civil station, whose share of the clustering rose from 71 to 74 per cent. In every city, then, a single block dominates the spatial signal and its internal coherence strengthened over the interval. These are the non-Muslim North in Calcutta, the Muslim quarter of Anwarganj in Kanpur, and the colonial civil station of Hazratganj in Lucknow.

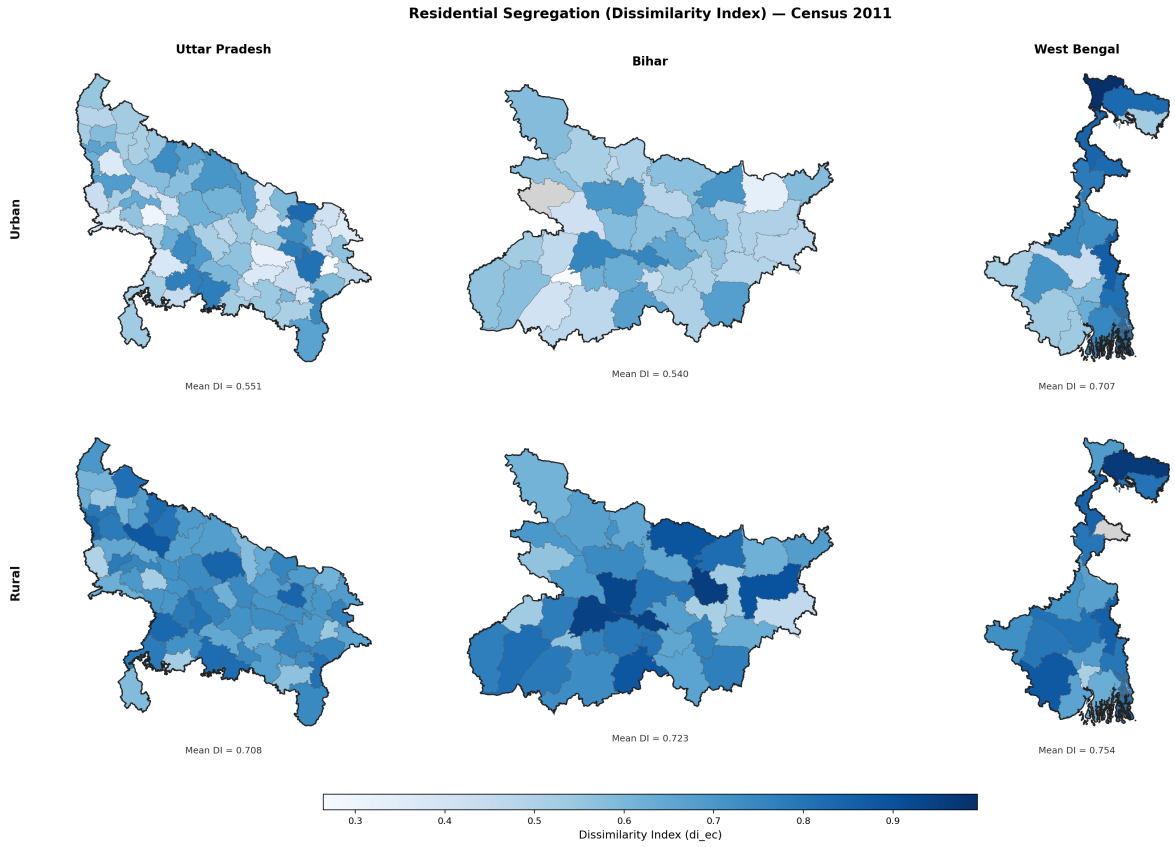
That clustering rose even where the dissimilarity index was flat is the empirically important point for what follows. A geography of high clustering means that residents whose daily movement is local are unlikely to cross communal lines, limiting their exposure to the other community and the scope for incidental inter-community interaction. It can also allow for disparity in access to shared resources, since the same school, clinic, or network that lies within reach of a household in the dominant bloc is effectively inaccessible to one in the opposing bloc, irrespective of the city-wide level of unevenness.

5.3 Impact on Contemporary Segregation

To understand contemporary trends in segregation, I use data from the Sixth Economic Census of India (2013-14). Here, the neighborhood is the enumeration block, each consisting of fewer than 200 establishments, which is a considerably finer spatial unit than the thana. Using this, we can measure segregation at the census town level in urban areas and at the tehsil level in rural areas. Although the economic census records information on commercial establishments, approximately

64% of establishments owned by Muslims in the three states under consideration are residential-based. With this caveat, in 2011 plague-affected districts exhibit higher DI than unaffected districts across all six comparisons. In Bihar, West Bengal, and Uttar Pradesh, segregation along religious lines in urban areas is about 6.4 pp, 10.8 pp, and 7.6 pp higher in plague-affected areas. We observe a similar trend in rural areas, where tehsils comprising plague-affected villages are on average 4 pp more segregated. The consistency of direction across states and sectors supports the hypothesis that plague-induced colonial town planning left a durable imprint on residential sorting. Figure 3 maps the dissimilarity index across districts in the three states, separately for urban and rural areas.

Figure 3: Residential Segregation across Districts in Uttar Pradesh, Bihar, and West Bengal, 2011



The dissimilarity index is computed from the religion of establishment owners, with the enumeration block as the neighborhood unit. The top row shows urban areas and the bottom row rural areas, and the mean DI for each state and sector is reported beneath each map. Districts shaded grey have no data. Data source: Sixth Economic Census (2013-14).

Table 4 presents regression estimates of the association between plague exposure and the 2011 segregation indices, estimated separately for urban areas and rural areas. In urban areas, plague-

affected districts exhibit significantly higher segregation on both measures. The Dissimilarity Index is 7.4 percentage points higher in affected towns, and the Isolation Index is 7.3 percentage points higher. Both point estimates are identified from within-state variation and are consistent in direction and magnitude across the two indices, one capturing distributional unevenness and the other measuring contact probabilities, which corroborates the role of the plague is structurally segregating the urban fabric. In rural areas, the Dissimilarity Index is 4.0 percentage points higher in plague-affected tehsils. The Isolation Index, however, shows no significant difference in rural areas. This divergence indicates that plague increased the spatial unevenness of Muslim settlement across rural blocks without raising within-block concentration, a pattern consistent with the more dispersed settlement structure of rural Bengal and the United Provinces, where Muslim households remain embedded in mixed-religion villages even where aggregate sorting has increased.

6 Education among Muslims in India

Residential segregation along religious and caste lines is a well-documented phenomenon in India. In urban areas, such segregation can have significant implications for economic and social well-being. What distinguishes Muslim segregation in India is the economic diversity of these enclaves. Anthropologists and political scientists have shown that due to factors such as dietary restrictions, communal tensions, and discrimination in housing, even affluent Muslims often find it difficult to move into culturally mixed neighborhoods. Using NFHS data matched to dissimilarity indices from the Sixth Economic Census, the analysis shows that wealth dispersion within Muslim urban neighborhoods rises monotonically as we move from less segregated districts to more segregated districts, confirming the prior that segregated neighborhoods in India exhibit the assortment of different economic classes. This urban asymmetry, however, does not extend to, non-Muslim urban enclaves or rural India, where both Muslim and non-Muslim neighborhoods exhibit non-monotonic, noisy patterns across segregation quartiles. This economic diversity concentrated within Muslim neighborhoods distinguishes Muslim segregation in India from patterns observed elsewhere, such as Black residential segregation in the United States, where spatial concentration and economic marginalization tend to reinforce one another (Cutler and Glaeser, 1997; Logan and Parman, 2025).

This feature motivates the central question of this paper: whether proximity to more affluent co-religionists within these spaces generates positive externalities through improved networks, social capital, role model effects, and shared religious identity, or whether the entrenched deprivation and state neglect associated with ghettoization ultimately exacerbate the marginalization of the poor Muslims who reside within them. Figures A1 and A2 in the Online Appendix illustrate the pattern of economic diversity across segregation quartiles for urban and rural neighborhoods respectively.

Existing research indicates that areas predominantly inhabited by Muslims and Scheduled Castes often suffer from substandard public services and lag substantially in educational attainment (Asher et al. 2024). While poor provision of public amenities is itself detrimental, this paper aims to examine the broader spillover effects of residential segregation on economic and social outcomes. Although prior literature highlights that inter-generational mobility is particularly limited among Muslims, who have in fact fallen behind other historically disadvantaged groups (Asher, Novosad and Rafkin, 2024), there is limited empirical evidence on the net effects of segregation as a structural condition.

6.1 Segregation and Education

Here I examine the impact of segregation on educational outcomes. Using individual-level data from the NFHS on years of schooling and net attendance ratios, I document how educational attainment varies with the share of Muslims in a neighborhood. In urban areas, mean years of education decline monotonically as the Muslim share of a sampling unit increases, and this decline is steeper in more segregated districts. To fix ideas, a Muslim child residing in a neighborhood with a greater share of Muslims completes 2.03 fewer years of education; this figure rises to 2.6 years in a neighborhood belonging to a more segregated district. Comparing two neighborhoods both with a hundred percent Muslim population share, years of completed education are 0.14 years lower in the more segregated district, isolating the independent role of district-level segregation over and above neighborhood composition. An analogous pattern holds for net attendance rates at both the primary and secondary level, with the negative association again more pronounced in highly segregated districts. In the rural sector, years of education and attendance rates similarly

decline with Muslim neighborhood share; however, the gradient is steeper in less segregated districts. Taken together, these patterns suggest that the relationship between neighborhood composition and individual educational attainment is strong and pervasive, and that residential segregation in the urban setting amplifies the magnitude of this association. Figures A3 to A6 in the Online Appendix document these gradients for years of education and net attendance ratios across urban and rural samples, split by above- and below-median district dissimilarity index.

Table 5 presents the causal effect of residential segregation on educational outcomes for Muslims, estimated via 2SLS using district-level plague exposure as an instrument for the dissimilarity index, following the specification in equation (1). The estimates discussed below pertain to urban areas. Results for the rural sector are qualitatively similar and are reported in Tables A1 and A2 in the Online Appendix. The outcome variables are net attendance ratios at the primary and secondary level, years of completed schooling expressed relative to the district mean, and the conditional rank mobility index following Asher, Novosad and Rafkin (2024). The sample is stratified by parental educational background, with children assigned to the bottom half of the parental distribution if their parent's years of schooling fall below the district median, and to the top half otherwise. For Muslims whose parents belong to the bottom half of the education distribution, a one standard deviation increase in the dissimilarity index reduces years of education relative to the district mean by 0.48 years and lowers the conditional rank by 5.27 percentile ranks, significant at the five and one percent levels respectively. Attendance ratios at the primary and secondary level are statistically indistinguishable from zero for this group, thus segregation operates primarily at the intensive rather than the extensive margin of schooling, as children are broadly enrolled but complete fewer years and fall further behind the district average. For Muslims with parents in the top half of the distribution, the effect on years of completed schooling is no longer statistically significant, yet the mobility penalty persists and remains substantial: the conditional rank falls by 4.06 percentile ranks, significant at the one percent level. These estimates indicate that residential segregation imposes a substantial and persistent educational penalty on Muslims across both parental backgrounds, constraining advancement even among children from relatively advantaged households.

Table 6 presents the analogous estimates for non-Muslim OBCs, a group that serves as a natural comparison given their broad socioeconomic proximity to Muslims on a range of welfare indicators. The results stand in sharp contrast to those for Muslims. For OBCs with parents in the bottom half of the distribution, segregation has no statistically significant effect on any of the four outcomes. OBC children with parents in the top half of the distribution face no discernible educational penalty in more segregated districts either. The effect on years of education relative to the district mean is positive and statistically significant, though small in magnitude. Plague-induced segregation is thus associated with inequality that is not only exacerbated along class lines but also across communities, with segregation systematically widening the educational gap between Muslims and other groups. The near-absence of comparable effects for OBC Hindus confirms that the large negative estimates for Muslims are not a generic feature of plague-affected districts but reflect the specific consequences of Muslim residential concentration.

7 Mechanisms

The preliminary results from the causal estimates establish that plague-induced residential segregation imposes a large and persistent educational penalty on Muslims, operating primarily at the intensive margin of schooling. This section sketches three channels through which segregation plausibly translates into reduced educational attainment. The evidence is preliminary and should be interpreted accordingly. The first channel is higher land use inequality in plague-affected districts, on account of plague-driven displacement and colonial town planning. The second is the under-provision of educational and health infrastructure to Muslim households in more segregated districts. The third is a higher opportunity cost of schooling, since small-scale enterprises and residential households are located together within Muslim enclaves, which can draw children into work. I present evidence for the first two channels below, and the third is work in progress.

7.1 Land Inequality

A first channel operates through land inequality. Plague-driven out-migration led to the abandonment of residential quarters, and the improvement trust interventions that followed demolished insanitary dwellings at a scale that frequently surpassed rehabilitation. To assess how this spatial reorganization evolved, I exploit thana-level census data to construct, for each thana, a measure of residential density: the number of houses relative to the area of the thana in square miles. Comparing how this density varies across thanas within a district and how that variation differs between plague-affected and unaffected districts over time reveals the degree to which land access became unequally distributed across urban and semi-rural locations as a consequence of the plague and the town planning that followed. Preliminary results indicate that this measure of land inequality, is higher in plague-affected districts than in unaffected controls, and that the gap widens over time as town planning interventions consolidated a spatial hierarchy in which marginalized groups were concentrated in densely packed peripheral settlements while other areas were opened up for planned, development. Overall in Bihar, the land inequality in 1961 was 0.34 percentage points (or 3.4%) higher in the plague-affected districts relative to their levels in 1901. If we take a look at the major towns where town improvement trust were established or elaborative town planning took place, we observe a steep rise in inequality. In Patna, between 1901 and 1961, the gini coefficient of land inequality rose by 24 percentage points. In Calcutta, the gini coefficient of land inequality rose by 3.1 percentage points between 1901 and 1921 and by a further 4 percentage points between 1921 and 1931. Following partition in 1947 and with over 300000 migrants settling in Calcutta as per the 1961 census, we observe a decline in the gini coefficient. Between 1931 and 1961, the coefficient fell by 9 percentage points.

7.2 Provision of Public Goods

A second channel operates through the under-provision of public services in Muslim neighborhoods. Tables 7 and 8 present 2SLS estimates of the effect of neighbourhood-level segregation and its interaction with Muslim neighbourhood share on access to educational facilities, instrumenting both regressors with historical plague occurrence and its interaction with Muslim neighbourhood

share. The specification is

$$Y_{nds} = \beta_0 + \beta_1 \widehat{DI}_{ds} + \beta_2 \big(\widehat{DI}_{ds} \times MuslimShare_{nds}\big) + \beta_3 MuslimShare_{nds} + \beta_4 \log Population_{nds} + \alpha_s + \varepsilon_{nds} \tag{9}$$

where Y_{nds} is an indicator for the presence of a facility in neighbourhood n of district d in state s, or $\ln(1 + N_{nds})$ where N_{nds} is the number of such facilities. DI_{ds} is the dissimilarity index, $MuslimShare_{nds}$ is the Muslim population share of the neighbourhood, and α_s are state fixed effects. Both DI_{ds} and its interaction with Muslim share are instrumented with $Plague_{ds}$ and $Plague_{ds} \times MuslimShare_{nds}$, and standard errors are clustered at the district level.

Table 7 shows that a one standard deviation increase in the dissimilarity index reduces the probability of a neighbourhood hosting a primary school by 8.6 percentage points and the number of primary schools by 4.7 percent, both significant at the five percent level. Higher Muslim neighbourhood share independently reduces access to education facilities at every level (primary, secondary, and college), indicating that the deficit is concentrated in the most Muslim-dense parts of segregated districts. Table 8 restricts to public facilities and yields sharper results: the direct effect of segregation is negative and significant for public primary and secondary schools, reducing the probability of a public primary school by 4.7 percentage points and a public secondary school by 0.9 percentage points, and the number of public primary and secondary schools by 3.0 and 0.5 percent respectively, confirming that the shortfall is driven by gaps in public provision. These estimates connect directly to the main results in Table 5: the absence of nearby schooling infrastructure raises the effective cost of attendance and reduces years of completed schooling, consistent with the intensive-margin pattern documented there. Table 9 provides an additional check using health facilities. A one standard deviation increase in segregation has no significant effect on the presence of any health facility, but higher Muslim neighbourhood share sharply reduces it, lowering the probability of any health facility by 13.5 percentage points. For public health facilities specifically, the direct effect of segregation is negative and significant, as a one standard deviation increase in segregation reduces the probability of access by 1.4 percentage points and the number of facilities by 1.1 percent (-0.014 and -0.011), corroborating the educational findings. As with education,

segregation's bite is confined to public provision, which is the primary margin along which Muslim enclaves are underserved. Segregated Muslim neighbourhoods are thus under provided by the state, with fewer public schools and public health facilities, consistent with their political exclusion from public goods provision. The analogous estimates for rural neighbourhoods are reported in Tables A3 to A5 in the Online Appendix.

7.3 Opportunity Cost of Schooling

A third channel, which remains work in progress, concerns the opportunity cost of schooling within Muslim enclaves. Muslim dropout rates in India are high: NFHS-5 places the net attendance ratio for Muslims at the secondary level at 62.5 percent, the lowest of all religious groups, and UDISE data record a secondary school dropout rate of 28.3 percent for the 2020-21 cohort. Muslim residential areas are characterized by a concentration of small-scale, household-based enterprises such as tailoring, leather goods, and other craft trades, whose production units are typically located within or immediately adjacent to the home. This spatial co-location of labor demand and supply creates a standing pull on children from poor families, for whom the immediate income from a workshop next door is readily comparable to the expected return from an additional year of school, particularly where returns to formal education are suppressed by apprehension and perceived discrimination in formal labor markets. The mechanism is consistent with the finding that segregation operates at the intensive rather than the extensive margin: children enrol but drop out early.

Preliminary evidence from the Time Use Survey of India is consistent with this account. In more segregated districts, children from poorer households are both more likely to take up paid work and spend more minutes of each day in it. Is this poverty or proximity? The poverty account, in the tradition of Basu and Van (1998), reads the pattern as 'scarcity doing what scarcity does', raising the margin at which a household sends its children to work. How children's time is distributed across activities can somewhat separate the two explanations. Conditional on household income, time spent in domestic and unpaid activity is not affected by segregation, while time in paid work continues to rise with it. Thus scarcity alone cannot explain this asymmetry, since a binding income

constraint should raise every margin of child work rather than the remunerated one selectively. What segregation appears to alter is the cost of finding work. Nonetheless, I cannot entirely rule out the possibility that poor households in more segregated districts might be more severely deprived than those in less segregated districts, thus the finding should be read as descriptive. Formal testing of this channel, including direct evidence on child labour participation and household income across segregation quartiles, is left for subsequent work, and all three mechanisms should be treated as preliminary.

8 Conclusion

This paper examines the long-run consequences of the bubonic plague of the late nineteenth and early twentieth centuries on residential segregation and educational outcomes for Muslims in India. Using variation in plague exposure across districts in Bihar, West Bengal, and Uttar Pradesh, I document that plague-driven mortality, mass evacuation, and the colonial town planning that followed significantly increased religious residential segregation, with dissimilarity indices in plague-affected areas rising by 6.6 percentage points more than in unaffected controls in Bihar and by 9.7 percentage points more in West Bengal between 1911 and 1961, with the steepest rise in both states between 1931 and 1961. This spatial reorganization has proved durable: plague-affected districts remain 6-11 percentage points more segregated in urban areas today. Instrumenting contemporary segregation with historical plague exposure, I find that a one standard deviation increase in the dissimilarity index reduces years of completed schooling relative to the district mean by 0.48 years and lowers the conditional rank by 5.27 percentile ranks for Muslims with below-median parental education, with the conditional rank penalty persisting for those from more advantaged backgrounds. These effects operate at the intensive rather than the extensive margin as children are enrolled but complete fewer years of school and persist across the parental distribution, consistent with the finding that even relatively affluent Muslims are unable to exit segregated neighborhoods. The near-absence of comparable effects for OBC Hindus, a group with similar observable socioeconomic characteristics, confirms that these penalties are specific to Muslim residential concentration rather than a generic feature of plague-affected or low-income districts.

The findings carry implications that extend beyond the historical case studied here. They demonstrate that a short, acute demographic shock an epidemic spanning barely two decades can leave a spatial imprint on cities that persists for over a century, shaping inter-generational inequality long after its direct effects have dissipated. The mechanism runs not through the epidemic itself but through the institutional response it triggered: colonial town planning encoded existing social hierarchies into the built environment, and those hierarchies, were eventually consolidated by communal animosity. The results also contribute to a broader understanding of the costs of residential segregation in developing country cities, where the relationship between spatial concentration and economic outcomes has received far less attention. Preliminary evidence on mechanisms points to three compounding channels, unequal distribution of residential land use, systematic underprovision of public goods in Muslim neighborhoods, and a higher opportunity cost of schooling arising from the co-location of household enterprises and residential areas within enclaves each of which warrants further investigation. More broadly, the paper suggests that the spatial organization of cities, shaped by historical contingency and colonial policy, remains a first-order determinant of who gets ahead and who does not in contemporary India.

References

Anselin, L. (1995). Local indicators of spatial association: LISA. *Geographical Analysis*, 27(2), 93-115.

Arnold, D. (1993). *Colonizing the Body: State Medicine and Epidemic Disease in Nineteenth-Century India*. University of California Press.

Asher, S., Jha, K., Novosad, P., Adukia, A., and Tan, B. (2026). Residential segregation and unequal access to local public services in India: Evidence from 1.5m neighborhoods. NBER Working Paper 34818.

Asher, S., Novosad, P., and Rafkin, C. (2024). Intergenerational mobility in India: New measures and estimates across time and social groups. *American Economic Journal: Applied Economics*, 16(2), 66-98.

Banerjee, A., and Iyer, L. (2005). History, institutions, and economic performance: The legacy of colonial land tenure systems in India. *American Economic Review*, 95(4), 1190-1213.

Blunt, E. A. H. (1915). *The Census of India, 1911: United Provinces of Agra and Oudh*. Government Press.

Chatterji, J. (1995). *Bengal Divided: Hindu Communalism and Partition, 1932-1947*. Cambridge University Press.

Chatterji, J. (2007). *The Spoils of Partition: Bengal and India, 1947-1967*. Cambridge University Press.

Chetty, R., and Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. *Quarterly Journal of Economics*, 133(3), 1107-1162.

Cutler, D., and Glaeser, E. (1997). Are ghettos good or bad? *Quarterly Journal of Economics*, 112(3), 827-872.

Damodaran, V. (1990). *Broken Promises: Popular Protest, Indian Nationalism, and the Congress Party in Bihar, 1935-1946*. Oxford University Press.

Dossal, M. (1991). *Imperial Designs and Indian Realities: The Planning of Bombay City, 1845-1875*. Oxford University Press.

Duncan, O. D., and Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review*, 20(2), 210-217.

Ghosh, N. (2009). Partition and communal violence in Bihar. In S. Bandyopadhyay (Ed.), *Bengal: Rethinking History*. Manohar Publishers.

Ghosh, N. (2020). *Urban Planning and Colonial India: Calcutta and Beyond*. Routledge.

Hibbard, M. (2024). Colonial planning, plague camps, and the making of the urban periphery. *Journal of Urban History*.

Home, R. (1997). *Of Planting and Planning: The Making of British Colonial Cities*. E & FN Spon.

Indian Medical Gazette (1906). Reports on the epidemiology of plague in India.

Iyer, L. (2010). Direct versus indirect colonial rule in India: Long-run consequences. *Review of Economics and Statistics*, 92(4), 693-713.

Jamhoor (2023). Caste, housing, and the colonial city: Evidence from Ahmedabad. *Economic and Political Weekly*.

Kidambi, P. (2001). *The Making of an Indian Metropolis: Colonial Governance and Public Culture in Bombay, 1890-1920*. Ashgate.

Kosambi, M., and Brush, J. E. (1988). Three colonial port cities in India. *Geographical Review*, 78(1), 32-47.

Kulkarni, S. (1979). Town planning and improvement trusts in British India. *Indian Economic and Social History Review*, 16(3), 281-305.

Logan, T., and Parman, J. (2025). Racial residential segregation and economic outcomes in the United States. *Journal of Economic Perspectives*.

Meller, H. (1979). *The Ideal City*. Leicester University Press.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2), 17-23.

Nunn, N. (2010). Religious conversion in colonial Africa. *American Economic Review: Papers and Proceedings*, 100(2), 147-152.

O'Malley, L. S. S. (1913). *Bengal District Gazetteers*. Bengal Secretariat Book Depot.

Sarkar, N. (2024). Rat ecology, climate, and the geography of plague in colonial India. *Economic History Review*.

Singh, P. (2025). Public health and spatial inequality: Colonial improvement trusts and the urban poor. *Journal of Development Economics*.

Town Improvement Trusts in India (1945). Report of the Central Advisory Council on Town Improvement Trusts. Government of India Press.

Yu, H.-L., and Christakos, G. (2006). Spatiotemporal modelling and mapping of the bubonic plague epidemic in India. *International Journal of Health Geographics*, 5(1), 12.

Tables

Table 1: Evenness and Dynamics Summary

City	Interval	DI^0	DI^1	ΔDI	D^*	$\bar{D}^*$ (r_0)
Calcutta	1911-31	0.345	0.417	+0.073	+0.0049	0.046 (0.039)
Kanpur	1921-31	0.386	0.443	+0.057	+0.0008	0.018 (0.017)
Lucknow	1921-31	0.360	0.356	−0.004	−0.0009	0.011 (0.012)

DI^0 and DI^1 are the dissimilarity indices at the start and end of the interval, and ΔDI is their difference. D^* is the dynamic segregation index, which weights the change in each ward's distance from evenness by its baseline distance; $D^* > 0$ indicates that the wards most uneven at baseline diverged further. $\bar{D}^*$ is its normalised counterpart, with the no-change reference r_0 in parentheses. Units are wards in Calcutta and circles in Kanpur and Lucknow. Data source: Author's calculations from ward-level (Calcutta) and circle-level (Kanpur, Lucknow) population tables of the Census of India, 1911, 1921, and 1931.

Table 2: Channel Decomposition by Period-0 Ward Type

Ward type (period 0)	n	Minority channel	Majority channel	Net
Panel A: Calcutta, 1911-1931				
Muslim-overrepresented	10	+0.021	+0.010	+0.030
Near-even	10	−0.011	+0.018	+0.007
Non-Muslim-overrepresented	10	+0.022	+0.013	+0.036
Total	30	+0.032	+0.041	+0.073
Panel B: Kanpur, 1921-1931				
Muslim-overrepresented	25	+0.000	+0.007	+0.007
Near-even	35	+0.025	+0.006	+0.032
Non-Muslim-overrepresented	45	+0.030	−0.012	+0.018
Total	105	+0.056	+0.001	+0.057
Panel C: Lucknow, 1921-1931				
Muslim-overrepresented	36	−0.014	+0.005	−0.009
Near-even	35	−0.004	+0.018	+0.014
Non-Muslim-overrepresented	34	−0.017	+0.008	−0.009
Total	105	−0.034	+0.030	−0.004

Entries are $\frac{1}{2}\sum_{j\in\text{group}} M_j$ and $\frac{1}{2}\sum_{j\in\text{group}} W_j$. Ward types are defined by the period-0 deviation: the bottom tercile of $|\psi_j^0|$ is near-even, the upper two terciles are split by sign. The minority channel is the Muslim-share movement, the majority channel the non-Muslim movement. Data source: Author's calculations from ward-level (Calcutta) and circle-level (Kanpur, Lucknow) population tables of the Census of India, 1911, 1921, and 1931.

Table 3: Clustering Summary

City	n	Blocks	$\mathbb{E}[I]$	I^0	I^1	Dominant block
Calcutta	30	5	−0.034	0.282^{**}	0.330^{***}	North (non-Muslim-majority, 63%)
Kanpur	105	7	−0.010	0.210^{***}	0.293^{***}	Anwarganj (Muslim quarter, 79%)
Lucknow	105	8	−0.010	0.073^{**}	0.099^{**}	Hazratganj (former civil lines, 74%)

Permutation inference, R = 9,999, against the benchmark $\mathbb{E}[I] = -1/(n-1)$. $^{***}p < 0.01$, $^{**}p < 0.05$. Calcutta: $p^0 = 0.011$, $p^1 = 0.005$. Kanpur: $p^0, p^1 < 0.001$. Lucknow: $p^0 = 0.033$, $p^1 = 0.012$. Data source: Author's calculations from ward-level (Calcutta) and circle-level (Kanpur, Lucknow) population tables of the Census of India, 1911, 1921, and 1931.

Table 4: Plague Exposure and Contemporary Residential Segregation (2011)

	Urban		Rural	
	(1) Dissimilarity Index	(2) Isolation Index	(3) Dissimilarity Index	(4) Isolation Index
Plague Exposure	0.0736***	0.0733***	0.0398***	−0.0003
	(0.0149)	(0.0248)	(0.0130)	(0.0016)
Observations	2,177	2,177	1,138	1,136
R^2	0.186	0.133	0.045	0.785
State FE	Yes	Yes	Yes	Yes
Pop. Weights	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes

Robust standard errors in parentheses, clustered at the district level. All regressions include state fixed effects and are weighted by town population (columns 1-2) or tahsil population (columns 3-4). Plague exposure is a binary indicator to account for severe plague-related exposure over time. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Data source: Sixth Economic Census (2013-14) for the segregation indices; Census of India district volumes for plague exposure.

Table 5: Effect of Segregation on Educational Outcomes: Muslims (Urban), 2SLS

	(1) NAR Primary	(2) NAR Secondary	(3) Yrs. Educ. (dist. mean)	(4) Conditional Rank
Panel A: Below-median parental education				
Dissimilarity Index	0.009	0.018	−0.481**	−5.270***
	(0.069)	(0.056)	(0.232)	(1.672)
Observations	3,375	4,775	11,976	11,976
Panel B: Above-median parental education				
Dissimilarity Index	−0.021	0.030	−0.102	−4.064***
	(0.059)	(0.050)	(0.268)	(1.525)
Observations	1,846	2,732	6,771	6,771
State FE	Yes	Yes	Yes	Yes
Survey-Year FE	Yes	Yes	Yes	Yes
Household Controls	Yes	Yes	Yes	Yes

2SLS estimates instrumenting the Dissimilarity Index with district-level plague exposure. Standard errors clustered at the district level in parentheses. Household controls include father's years of education, wealth index, number of siblings, child age, and child sex. Survey-year fixed effects control for differences across NFHS rounds. All specifications include survey weights. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Data source: NFHS-4 (2015-16) and NFHS-5 (2019-21), matched to district dissimilarity indices from the Sixth Economic Census (2013-14).

Table 6: Effect of Segregation on Educational Outcomes: OBC Hindus (Urban), 2SLS

	(1) NAR Primary	(2) NAR Secondary	(3) Yrs. Educ. (dist. mean)	(4) Conditional Rank
Panel A: Below-median parental education				
Dissimilarity Index	−0.012	−0.048	−0.175	−0.913
	(0.031)	(0.036)	(0.170)	(0.904)
Observations	1,980	3,024	7,654	7,654
Panel B: Above-median parental education				
Dissimilarity Index	0.002	−0.018	0.399**	1.382
	(0.026)	(0.028)	(0.187)	(0.929)
Observations	2,033	3,765	8,913	8,913
State FE	Yes	Yes	Yes	Yes
Survey-Year FE	Yes	Yes	Yes	Yes
Household Controls	Yes	Yes	Yes	Yes

2SLS estimates instrumenting the Dissimilarity Index with district-level plague exposure. Standard errors clustered at the district level in parentheses. Household controls include father's years of education, wealth index, number of siblings, child age, and child sex. Survey-year fixed effects control for differences across NFHS rounds. All specifications include survey weights. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Data source: NFHS-4 (2015-16) and NFHS-5 (2019-21), matched to district dissimilarity indices from the Sixth Economic Census (2013-14).

Table 7: Neighborhood-level Education Facilities vs. Segregation and Muslim Share (Urban)

	Primary School		Secondary School		College	
	(1)	(2)	(3)	(4)	(5)	(6)
	Indicator	$\ln(1+N)$	Indicator	$\ln(1+N)$	Indicator	$\ln(1+N)$
Dissimilarity Index	-0.086^{**}	-0.047^{**}	0.003	0.001	-0.003	-0.003
	(0.042)	(0.020)	(0.025)	(0.012)	(0.005)	(0.003)
DI × Muslim Share	0.020	0.011	-0.062	-0.034	-0.026	-0.008
	(0.052)	(0.028)	(0.058)	(0.030)	(0.019)	(0.007)
Muslim Share	-0.117^{***}	-0.062^{***}	-0.107^{***}	-0.060^{***}	-0.039^{***}	-0.019^{***}
	(0.023)	(0.013)	(0.029)	(0.015)	(0.011)	(0.004)
Observations	135,780	135,780	135,780	135,780	135,780	135,780
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Instrument	Plague	Plague	Plague	Plague	Plague	Plague

IV (2SLS) estimates of the effect of neighborhood-level segregation (*Dissimilarity Index*), the interaction of Muslim neighborhood share with segregation (*DI × Muslim Share*), and Muslim share on access to education facilities (all types) in urban neighborhoods. Indicator columns report whether the neighborhood has at least one such facility; $\ln(1+N)$ columns use the log of one plus the number of facilities. Both endogenous regressors are instrumented using historical plague occurrence (*plague*) and its interaction with Muslim neighborhood share (*plague × Muslim share*). All regressions include state fixed effects and control for log neighborhood population (enumeration block). Standard errors in parentheses are clustered at the state-district level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Data source: Sixth Economic Census (2013-14); plague exposure from Census of India district volumes.

Table 8: Neighborhood-level *Public* Education Facilities vs. Segregation and Muslim Share (Urban)

	Primary School		Secondary School		College	
	(1)	(2)	(3)	(4)	(5)	(6)
	Indicator	$\ln(1+N)$	Indicator	$\ln(1+N)$	Indicator	$\ln(1+N)$
Dissimilarity Index	-0.047^{***}	-0.030^{***}	-0.009^{***}	-0.005^{***}	-0.001	-0.002
	(0.015)	(0.009)	(0.003)	(0.002)	(0.004)	(0.002)
DI × Muslim Share	0.030	0.016	0.007	0.004	-0.013	-0.001
	(0.020)	(0.012)	(0.007)	(0.005)	(0.012)	(0.003)
Muslim Share	-0.026^{***}	-0.016^{***}	-0.019^{***}	-0.012^{***}	-0.019^{***}	-0.007^{***}
	(0.010)	(0.006)	(0.003)	(0.002)	(0.007)	(0.001)
Observations	135,780	135,780	135,780	135,780	135,780	135,780
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Instrument	Plague	Plague	Plague	Plague	Plague	Plague

IV (2SLS) estimates of the effect of neighborhood-level segregation (*Dissimilarity Index*), the interaction of Muslim neighborhood share with segregation (*DI × Muslim Share*), and Muslim share on access to *public* education facilities in urban neighborhoods. Indicator columns report whether the neighborhood has at least one such facility; $\ln(1+N)$ columns use the log of one plus the number of facilities. Both endogenous regressors are instrumented using historical plague occurrence (*plague*) and its interaction with Muslim neighborhood share (*plague × Muslim share*). All regressions include state fixed effects and control for log neighborhood population (enumeration block). Standard errors in parentheses are clustered at the state-district level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Data source: Sixth Economic Census (2013-14); plague exposure from Census of India district volumes.

Table 9: Neighborhood-level Health Facilities vs. Segregation and Muslim Share (Urban)

	Overall Health		Public Health	
	(1)	(2)	(3)	(4)
	Indicator	$\ln(1+N)$	Indicator	$\ln(1+N)$
Dissimilarity Index	0.002	-0.017	-0.014^{***}	-0.011^{***}
	(0.018)	(0.022)	(0.005)	(0.004)
DI × Muslim Share	-0.061	-0.078	0.001	0.003
	(0.048)	(0.061)	(0.008)	(0.007)
Muslim Share	-0.135^{***}	-0.198^{***}	-0.021^{***}	-0.018^{***}
	(0.025)	(0.031)	(0.003)	(0.003)
Observations	135,780	135,780	135,780	135,780
State FE	Yes	Yes	Yes	Yes
Instrument	Plague	Plague	Plague	Plague

IV (2SLS) estimates of the effect of neighborhood-level segregation (*Dissimilarity Index*), the interaction of Muslim neighborhood share with segregation (*DI × Muslim Share*), and Muslim share on access to overall and public health facilities in urban neighborhoods. Indicator columns report whether the neighborhood has at least one such facility; $\ln(1+N)$ columns use the log of one plus the number of facilities. Both endogenous regressors are instrumented using historical plague occurrence (*plague*) and its interaction with Muslim neighborhood share (*plague × Muslim share*). All regressions include state fixed effects and control for log neighborhood population (enumeration block). Standard errors in parentheses are clustered at the state-district level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Data source: Sixth Economic Census (2013-14); plague exposure from Census of India district volumes.

Online Appendix

Plague, Planning, and Persistence: Colonial Institutions, Residential Segregation, and Educational Opportunity in India

Not for print publication

Appendix Tables

Table A1: Effect of Segregation on Educational Outcomes: Muslims (Rural), 2SLS

	(1) NAR Primary	(2) NAR Secondary	(3) Yrs. Educ. (dist. mean)	(4) Conditional Rank
Panel A: Below-median parental education				
Dissimilarity Index	0.148*	0.157**	1.488	−5.718**
	(0.086)	(0.080)	(0.763)	(2.859)
Observations	6,114	8,329	17,245	17,245
Panel B: Above-median parental education				
Dissimilarity Index	0.067	0.162*	1.423	−7.130**
	(0.074)	(0.096)	(0.772)	(3.224)
Observations	3,604	5,203	10,749	10,749
State FE	Yes	Yes	Yes	Yes
Survey-Year FE	Yes	Yes	Yes	Yes
Household Controls	Yes	Yes	Yes	Yes

2SLS estimates instrumenting the Dissimilarity Index with district-level plague exposure. Standard errors clustered at the district level in parentheses. Household controls include father's years of education, wealth index, number of siblings, child age, and child sex. Survey-year fixed effects control for differences across NFHS rounds. All specifications include survey weights. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Data source: NFHS-4 (2015-16) and NFHS-5 (2019-21), matched to district dissimilarity indices from the Sixth Economic Census (2013-14).

Table A2: Effect of Segregation on Educational Outcomes: OBC Hindus (Rural), 2SLS

	(1) NAR Primary	(2) NAR Secondary	(3) Yrs. Educ. (dist. mean)	(4) Conditional Rank
Panel A: Below-median parental education				
Dissimilarity Index	0.053*	0.082*	0.836*	−4.188**
	(0.029)	(0.048)	(0.477)	(1.854)
Observations	13,170	18,045	37,759	37,759
Panel B: Above-median parental education				
Dissimilarity Index	0.065*	0.117*	1.289**	−3.232
	(0.036)	(0.062)	(0.600)	(2.202)
Observations	12,590	19,018	40,220	40,220
State FE	Yes	Yes	Yes	Yes
Survey-Year FE	Yes	Yes	Yes	Yes
Household Controls	Yes	Yes	Yes	Yes

2SLS estimates instrumenting the Dissimilarity Index with district-level plague exposure. Standard errors clustered at the district level in parentheses. Household controls include father's years of education, wealth index, number of siblings, child age, and child sex. Survey-year fixed effects control for differences across NFHS rounds. All specifications include survey weights. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Data source: NFHS-4 (2015-16) and NFHS-5 (2019-21), matched to district dissimilarity indices from the Sixth Economic Census (2013-14).

Table A3: Neighborhood-level Education Facilities vs. Segregation and Muslim Share (Rural)

	Primary School		Secondary School		College	
	(1)	(2)	(3)	(4)	(5)	(6)
	Indicator	$\ln(1+N)$	Indicator	$\ln(1+N)$	Indicator	$\ln(1+N)$
Dissimilarity Index	-0.242^{**}	-0.122^{**}	0.017	0.012	0.006	0.004
	(0.116)	(0.061)	(0.033)	(0.021)	(0.007)	(0.004)
DI × Muslim Share	0.458	0.241	−0.170	−0.107	−0.013	−0.010
	(0.442)	(0.241)	(0.165)	(0.104)	(0.021)	(0.014)
Muslim Share	-0.255^{***}	-0.147^{***}	-0.044^{***}	-0.028^{***}	-0.007^{***}	-0.004^{***}
	(0.035)	(0.018)	(0.014)	(0.009)	(0.002)	(0.001)
Observations	471,991	471,991	471,991	471,991	471,991	471,991
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Instrument	Plague	Plague	Plague	Plague	Plague	Plague

IV (2SLS) estimates of the effect of neighborhood-level segregation (*Dissimilarity Index*), the interaction of Muslim neighborhood share with segregation (*DI × Muslim Share*), and Muslim share on access to education facilities (all types) in rural neighborhoods. Indicator columns report whether the neighborhood has at least one such facility; $\ln(1+N)$ columns use the log of one plus the number of facilities. Both endogenous regressors are instrumented using historical plague occurrence (*plague*) and its interaction with Muslim neighborhood share (*plague × Muslim share*). All regressions include state fixed effects and control for log neighborhood population (enumeration block). Standard errors in parentheses are clustered at the state-district level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Data source: Sixth Economic Census (2013-14); plague exposure from Census of India district volumes.

Table A4: Neighborhood-level *Public* Education Facilities vs. Segregation and Muslim Share (Rural)

	Primary School		Secondary School		College	
	(1)	(2)	(3)	(4)	(5)	(6)
	Indicator	$\ln(1+N)$	Indicator	$\ln(1+N)$	Indicator	$\ln(1+N)$
Dissimilarity Index	-0.115^{*}	−0.063	0.000	0.000	0.002	0.001
	(0.068)	(0.040)	(0.016)	(0.011)	(0.003)	(0.002)
DI × Muslim Share	0.155	0.087	−0.065	−0.040	0.009	0.006
	(0.207)	(0.123)	(0.064)	(0.041)	(0.012)	(0.008)
Muslim Share	-0.186^{***}	-0.113^{***}	-0.029^{***}	-0.019^{***}	-0.004^{***}	-0.003^{***}
	(0.017)	(0.010)	(0.006)	(0.004)	(0.001)	(0.001)
Observations	471,991	471,991	471,991	471,991	471,991	471,991
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Instrument	Plague	Plague	Plague	Plague	Plague	Plague

IV (2SLS) estimates of the effect of neighborhood-level segregation (*Dissimilarity Index*), the interaction of Muslim neighborhood share with segregation (*DI × Muslim Share*), and Muslim share on access to *public* education facilities in rural neighborhoods. Indicator columns report whether the neighborhood has at least one such facility; $\ln(1+N)$ columns use the log of one plus the number of facilities. Both endogenous regressors are instrumented using historical plague occurrence (*plague*) and its interaction with Muslim neighborhood share (*plague × Muslim share*). All regressions include state fixed effects and control for log neighborhood population (enumeration block). Standard errors in parentheses are clustered at the state-district level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Data source: Sixth Economic Census (2013-14); plague exposure from Census of India district volumes.

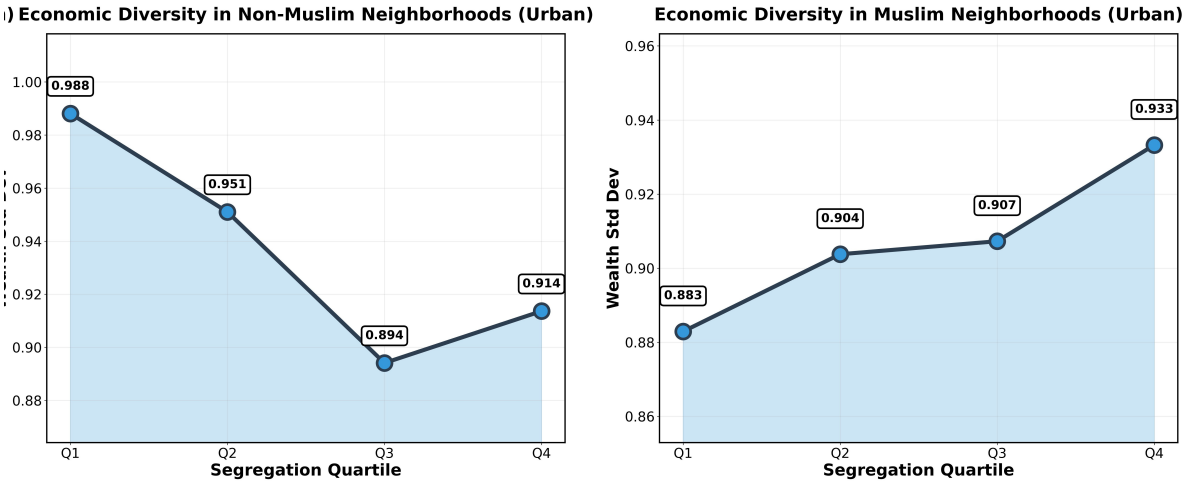
Table A5: Neighborhood-level Health Facilities vs. Segregation and Muslim Share (Rural)

	Overall Health		Public Health	
	(1)	(2)	(3)	(4)
	Indicator	$\ln(1+N)$	Indicator	$\ln(1+N)$
Dissimilarity Index	0.015	0.001	-0.019^{*}	-0.016^{**}
	(0.035)	(0.036)	(0.010)	(0.008)
DI × Muslim Share	0.274	0.313	0.066	0.050
	(0.214)	(0.255)	(0.058)	(0.044)
Muslim Share	−0.001	−0.006	-0.028^{***}	-0.023^{***}
	(0.020)	(0.023)	(0.005)	(0.004)
Observations	471,991	471,991	471,991	471,991
State FE	Yes	Yes	Yes	Yes
Instrument	Plague	Plague	Plague	Plague

IV (2SLS) estimates of the effect of neighborhood-level segregation (*Dissimilarity Index*), the interaction of Muslim neighborhood share with segregation (*DI × Muslim Share*), and Muslim share on access to overall and public health facilities in rural neighborhoods. Indicator columns report whether the neighborhood has at least one such facility; $\ln(1+N)$ columns use the log of one plus the number of facilities. Both endogenous regressors are instrumented using historical plague occurrence (*plague*) and its interaction with Muslim neighborhood share (*plague × Muslim share*). All regressions include state fixed effects and control for log neighborhood population (enumeration block). Standard errors in parentheses are clustered at the state-district level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Data source: Sixth Economic Census (2013-14); plague exposure from Census of India district volumes.

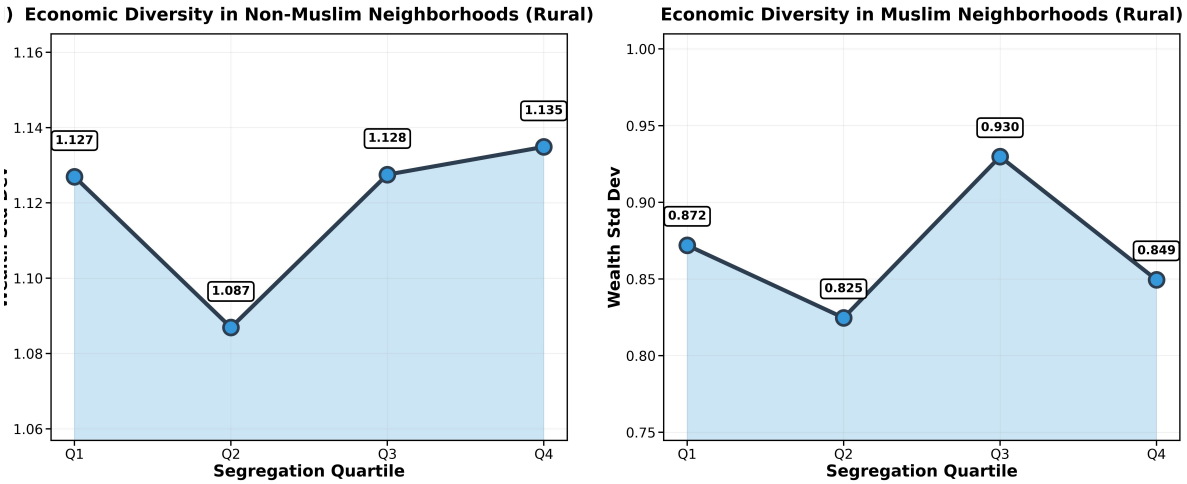
Appendix Figures

Figure A1: Economic Diversity by Segregation Quartile, Urban Neighborhoods



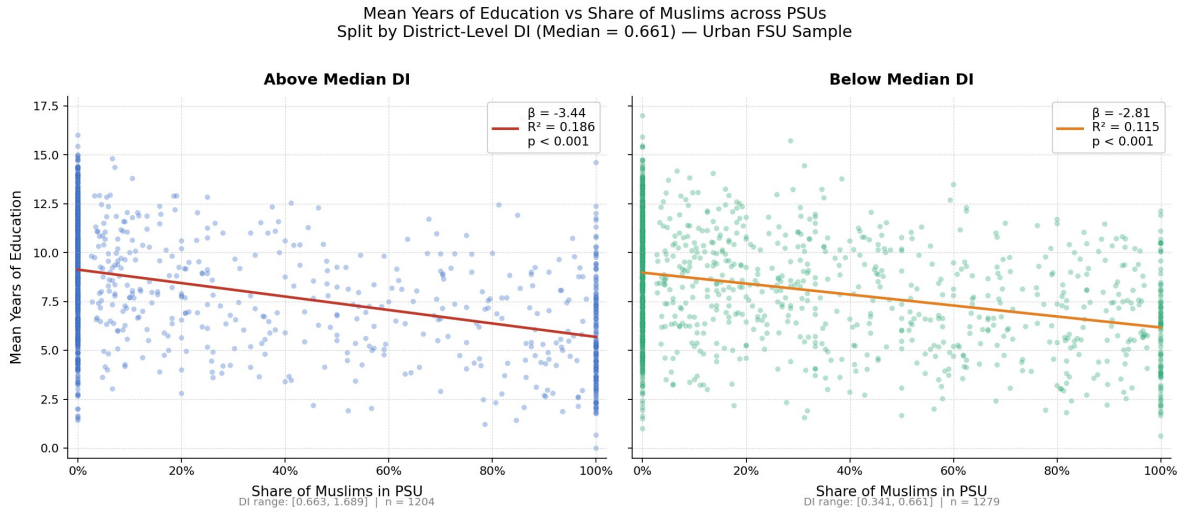
The left panel shows non-Muslim neighborhoods; the right panel shows Muslim neighborhoods. The y-axis reports the mean standard deviation of the household wealth index within a neighborhood, averaged across districts in each segregation quartile. Data source: NFHS-4 (2015-16) and NFHS-5 (2019-21), matched to district dissimilarity indices from the Sixth Economic Census (2013-14).

Figure A2: Economic Diversity by Segregation Quartile, Rural Neighborhoods



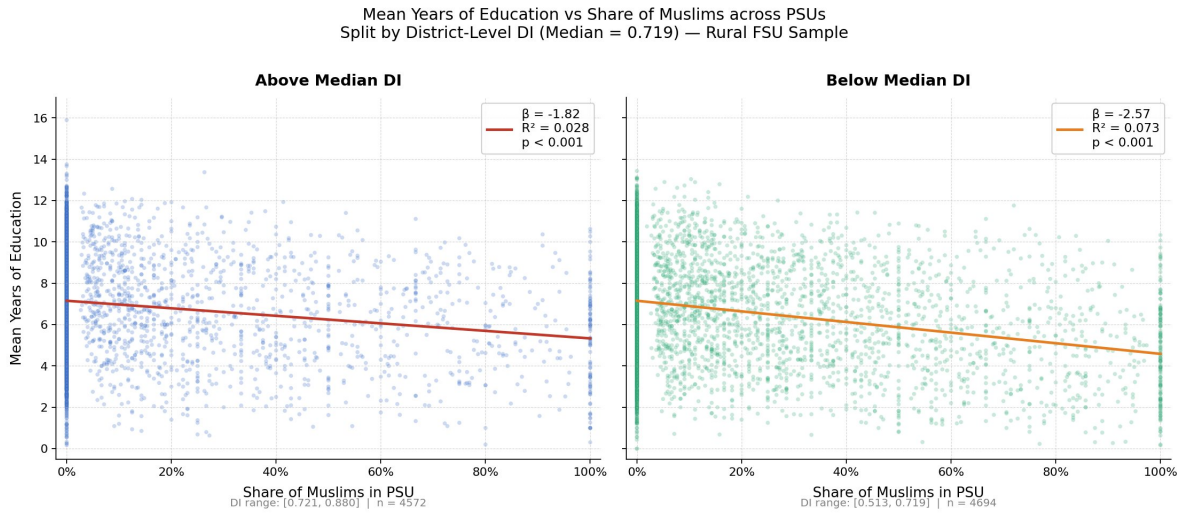
The left panel shows non-Muslim neighborhoods; the right panel shows Muslim neighborhoods. Data source: NFHS-4 (2015-16) and NFHS-5 (2019-21), matched to district dissimilarity indices from the Sixth Economic Census (2013-14).

Figure A3: Years of Education and the Muslim Share of the Primary Sampling Unit, Urban Sample



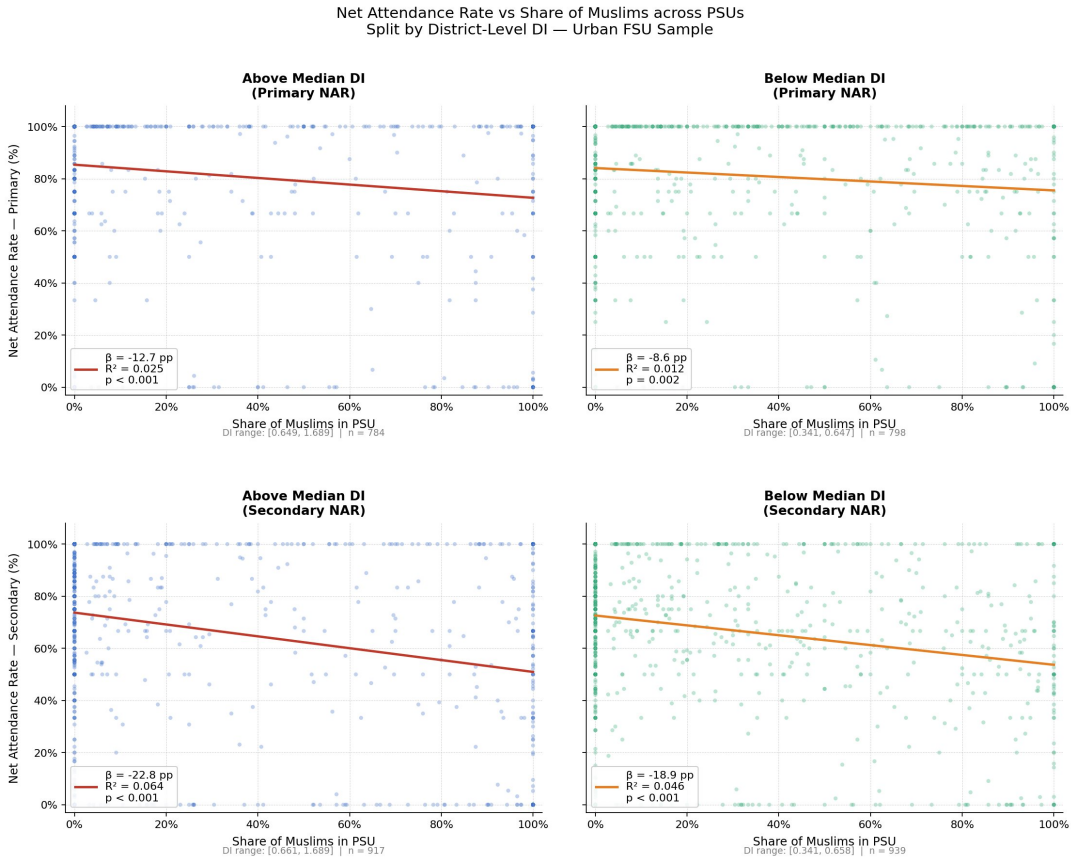
Mean years of education against the share of Muslims in the primary sampling unit (PSU), split by above- and below-median district dissimilarity index. Each point is a PSU; fitted lines from OLS regressions. Data source: NFHS-4 (2015-16) and NFHS-5 (2019-21), matched to district dissimilarity indices from the Sixth Economic Census (2013-14).

Figure A4: Years of Education and the Muslim Share of the Primary Sampling Unit, Rural Sample



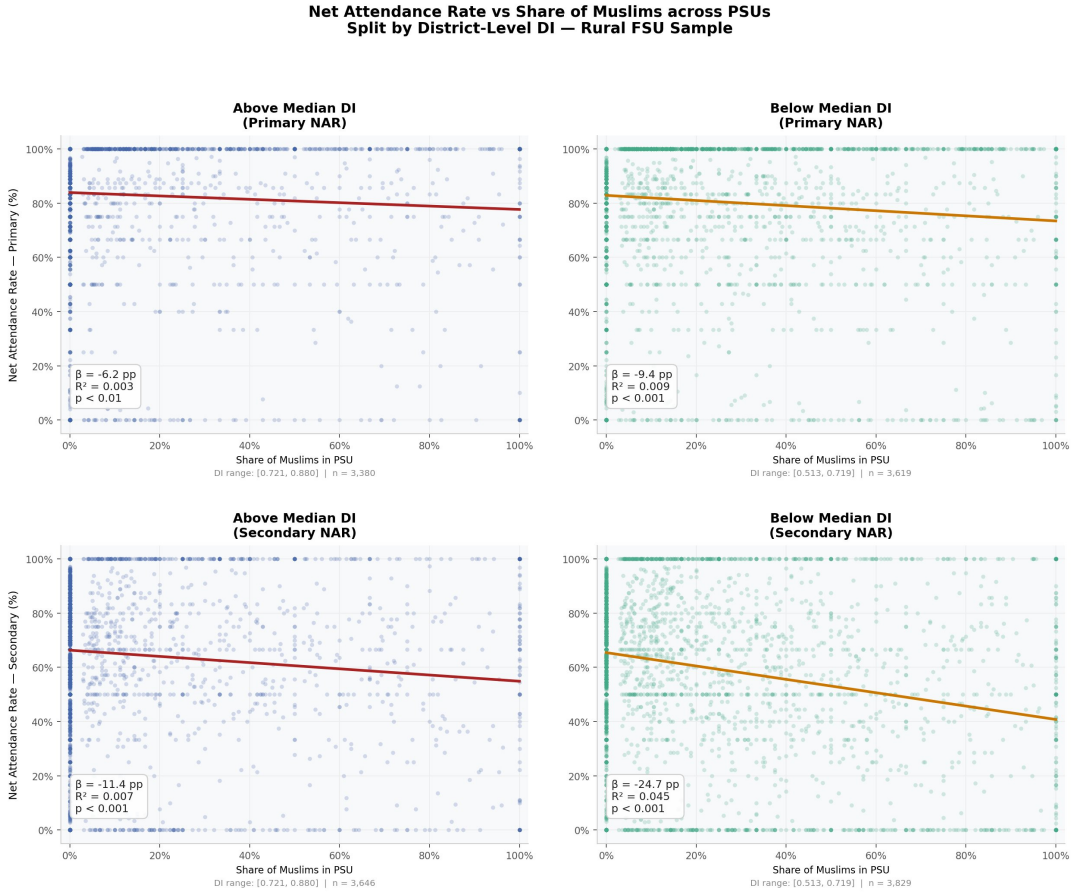
Mean years of education against the share of Muslims in the primary sampling unit (PSU), split by above- and below-median district dissimilarity index. Data source: NFHS-4 (2015-16) and NFHS-5 (2019-21), matched to district dissimilarity indices from the Sixth Economic Census (2013-14).

Figure A5: Net Attendance Ratio and the Muslim Share of the Primary Sampling Unit, Urban Sample



Net attendance ratio (primary and secondary) against the share of Muslims in the primary sampling unit (PSU), split by above- and below-median district dissimilarity index. Data source: NFHS-4 (2015-16) and NFHS-5 (2019-21), matched to district dissimilarity indices from the Sixth Economic Census (2013-14).

Figure A6: Net Attendance Ratio and the Muslim Share of the Primary Sampling Unit, Rural Sample



Net attendance ratio (primary and secondary) against the share of Muslims in the primary sampling unit (PSU), split by above- and below-median district dissimilarity index. Data source: NFHS-4 (2015-16) and NFHS-5 (2019-21), matched to district dissimilarity indices from the Sixth Economic Census (2013-14).