RATIONALITY
& COMPETITION
CRC TRR 190

Private Interest in the Public Good: Cholera and the Origins of Sanitary Reform

Daniel Gallardo-Albarrán (Wageningen University)
Kalle Kappner (HU Berlin)

Discussion Paper No. 573

May 19, 2026

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, Ludwig Maximilian University of Munich, 80539 Munich, Germany
info@rationality-and-competition.de

Private Interest in the Public Good: Cholera and the Origins of Sanitary Reform*

Daniel Gallardo-Albarrán,† Kalle Kappner‡

This version: May 19, 2026

Abstract

European countries paved the way for modern economic growth in the 19th century with large-scale reforms facilitating human capital accumulation. The literature has looked at the role of elites in broad education investments, but less attention has been devoted to reforms promoting workers' health, a key component of human capital. This paper studies the impact of the 1866 cholera outbreak on modern waterworks construction in the German Empire to test the hypothesis whether concerns about workers' health may have compelled elites to invest in health-enhancing public goods. We find that the epidemic raised the annual probability of building waterworks by about 35%. Exogenous variation relying on the cholera-spreading effect of military movements during the Austro-Prussian War in 1866 further underpins this result. Quantitative and qualitative evidence indicates that non-agricultural elites employing more productive capital and better-skilled workers pushed for reform to avoid the costly prospect of future labour shocks. Long-term analyses show sizeable and persisting effects of the epidemic on public health and development outcomes shortly before the First World War.

Keywords: political economy, public goods, sanitation, cholera, elites

JEL Codes: H41, H54, I18, N33, N93.

* For comments on earlier drafts we thank Erik Hornung, Davide Cantoni, Jonathan Chapman, Eric Schneider, Uwe Sunde, Stephan Heblich, Ewout Frankema, Sandra de Pleijt, Isabelle Devos and Esther Gehrke. We also thank seminar participants at Cambridge University, Humboldt University, Mainz University, Valencia University, Wageningen University, Essen University, ifo Institute; and participants at the EHS Conference (Newcastle, 2024), GGDC Conference (Groningen, 2024), GSWG Conference (2025), UEA Conference (Berlin, 2025), EHES Conference (Hohenheim, 2025). Daniel Gallardo-Albarrán acknowledges financial support from the Netherlands Organization for Scientific Research for a project entitled "Global health inequality and the diffusion of sanitation since 1850" (NWO; grant number VI.Veni.201H.048). Kalle Kappner is grateful for financial support from Deutsche Forschungsgemeinschaft (CRC TRR 190, project number 280092119).

† Corresponding author. Wageningen University, daniel.gallardoalbarran@wur.nl

‡ Humboldt-Universität zu Berlin, kallekappner@googlemail.com

1 Introduction

Western countries paved the way for modern economic growth during the nineteenth century with large-scale reforms facilitating human capital accumulation. The literature has investigated the drivers of this process by looking at the role of elites in broad education investments (Sokoloff and Engerman 2000; Acemoglu and Robinson 2000; Galor et al. 2009). However, less attention has been devoted to understanding the provision of infrastructures promoting workers' health, an important component of labour productivity at the time. Frequent sickness diminished work effort and cognitive functioning, and it often led to the premature loss of human capital in the form of experience and skills, particularly in urban areas with higher mortality (Fogel 1994; Costa 2015).[1]

Earlier research argues that private commercial interests often aligned with the promotion of health-related urban amenities that were used in manufacturing processes, such as clean water, due to complementarities between public spending and capital (Aidt et al. 2010). It is, however, unclear whether concerns about workers' health compelled elites to invest in health-enhancing public goods. On the one hand, reducing the incidence of infectious diseases could mitigate financial losses from labour shortages, especially for capital owners employing valuable equipment and more skilled labourers (Krieger 2024; Galor and Moav 2006). On the other hand, the rich may have opposed public health investments when the marginal cost of taxation became too high (Chapman 2024), and they only engaged in redistributive policies when the risk of social unrest became too threatening (Acemoglu and Robinson 2000). Also, even if there was interest in public goods among the wealthy, collective action problems could undermine their decentralized provision if most of their benefits were not captured by a small group facing low coordination costs (Olson 1965).

We test whether workers' health concerns motivated elites in local councils to invest in health infrastructures by leveraging variation from the 1866 cholera epidemic in more than 2,500 cities throughout the German Empire. This epidemic resulted in the highest mortality rate of the modern era in the country (36 deaths per 1,000

[1] For studies highlighting the role of health in economic productivity in more recent periods, see Weil (2007) and Bloom et al. (2022).

inhabitants, see figure 1), causing an exceptional and sudden labour shock, as working-age individuals accounted for about two thirds of cholera deaths. Some areas were heavily affected, with more than ten percent of the population dying, while others were spared. At the same time, elites' physical capital was unaffected, which allows us to explore the consequences of a labour scarcity shock (Franck 2024), while keeping labour demand constant. In addition, the rapid diffusion of this disease coupled with unhygienic standards at the time made the outbreak highly unpredictable and deadly, unlike other potential labour shocks.[2]

We relate the mortality impact of cholera to the most important public health technology before the advent of modern medicine: piped water supplies. Modern waterworks transformed the urban public environment and households, ultimately paving the way for the epidemiological transition (Alsan and Goldin 2019).[3] The decision to build them required a great deal of political consensus and long-term commitment due to their extraordinary costs (Cutler and Miller 2006), particularly in a context in which public ownership was the norm.[4] In addition, piped water supplies were at the forefront of efforts to improve the urban disease environment in the 1870s and 1880s when the effectiveness of healthcare institutions, such as hospitals, was limited in the absence of effective treatments. Elites were aware of some of the health benefits of centralized water supplies through the lobbying efforts of sanitarians and reformers (Mokyr and Stein 1996).

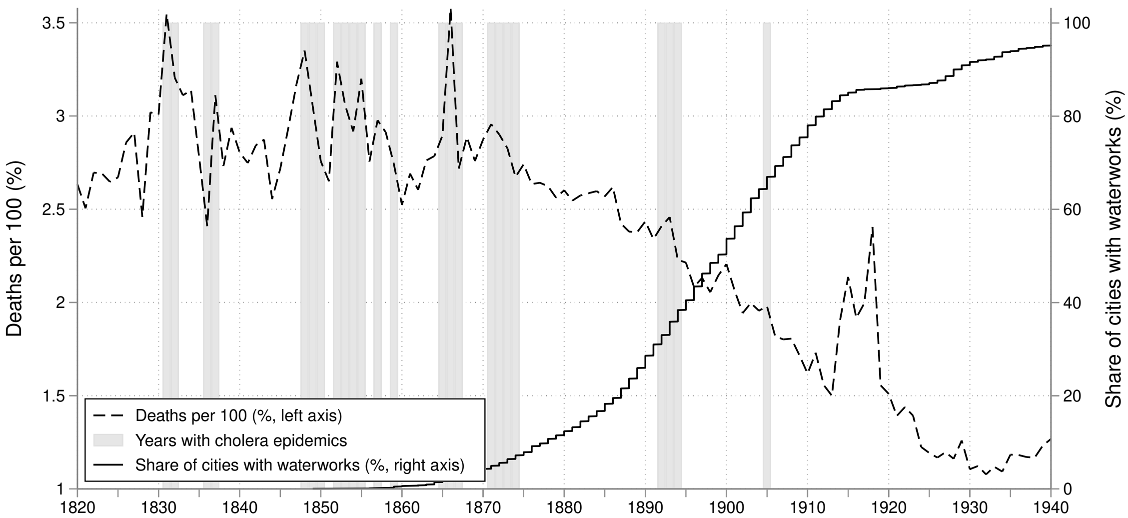
The context of the German Empire during the 19th century makes for an appropriate setting to analyse whether workers' health concerns motivated elites to support public health investments. First, the political system in place until the First World War granted disproportionate decision-making power to the wealthy (Becker and Hornung 2021), whose preferences for public spending following a labour shock can be more readily observed than in countries with a wider franchise. Second, the 1866 cholera epidemic happened at a time when major constraints on building waterworks

[2]Famines had largely disappeared by the time of Germany's high industrialization (Pfister and Fertig 2020), and military conflicts had lost importance after unification in 1871.

[3]In German cities, Gallardo-Albarrán (2020) finds that piped water led to significant declines in mortality, particularly in places that also constructed sewerage systems.

[4]The vast majority of waterworks in Germany were public (Grahn 1898-1902). Sewerage systems were also part of the sanitary reform, though their implementation in the German Empire was delayed by the prohibition to dispose untreated sewage into rivers in 1874 (Hennock 2000, 281).

Figure 1: Mortality and waterworks in Germany, 1820-1940



Note: This figure shows the crude death rate (all-cause deaths per 100) for Prussia (1820-1870) and the German Empire (1871-1940), and the share of cities with waterworks, according to our dataset. Grey background shading marks years with cholera epidemics (Kappner 2025). Crude death rates are computed from Kraus (1980) and Rahlf (2016).

had been lifted, unlike several decades earlier when the availability of the technology and associated engineering knowledge were limited. Third, the vast geographic scope of the German Empire, comprising large parts of Central Europe beyond modern-day Germany, offers substantial spatial city-level variation in exposure to cholera as well as temporal variation in the timing of waterworks construction that we leverage in our analyses (see figures 2 and B.2). And fourth, the unusually rapid transition to modern water supplies in Germany provides an interesting case of how a country can successfully set in motion a major public reform with far-reaching consequences for both the economy and overall welfare. Most medium- and large-size cities were supplied by waterworks by 1900, while just about ten percent had these water systems in the immediate years following the epidemic (Gallardo-Albarrán 2025, 572).

We build several datasets to analyse the impact of the 1866 cholera epidemic on investments decisions at the most suitable level of aggregation: the city. The first traces the spread of the epidemic throughout the German Empire for more than 2,500 cities. This gives us a precise indicator to pinpoint which places were impacted by cholera while others were left untouched. The second dataset tracks the piped water transition for all German cities until 1914, by recording the date at which waterworks started operating. In addition, in our long-run analyses we use a measure of within-city access to the pipe network and various development indicators to assess the indirect impact of cholera–through sanitary reform–on health, education, business and innovation outcomes at the eve of the First World War. Our third dataset introduces a new measure of local decision-making elites' economic background, derived from tax revenues. This allows us to explore how the link between cholera exposure and public health investment differed between commercial-industrial (henceforth: commercial) elites and those deriving their income from traditional agricultural activities.

Our results show that epidemic increased the annual probability of opening waterworks by about 35% relative to non-shocked cities up to 1891. Though this effect persisted until the eve of the First World War (more on this below), our main analyses stop in the early 1890s due to the occurrence of another high-profile cholera outbreak and local tax reforms. Our results are robust to a large number of tests, controlling for multiple relevant confounders and alternative model specifications. We support a causal interpretation of this effect through two complementary approaches exploiting the plausibly exogenous spread of cholera along military routes as the 1866 Austro-Prussian War unfolded on German territory. First, by employing an instrumental variable (IV) approach, we estimate a local causal effect based on cities' direct exposure to troops. Second, we recover a universally representative causal estimate from a calibrated, spatially structured epidemiological model of the 1866 cholera outbreak dynamics and its contagion processes through military movement. From a counterfactual simulation of this model–where the 1866 military campaign and associated contagion are "switched off"–, we isolate the marginal epidemiological effect of exogenous troop movements on all cities, not just those directly visited by soldiers. Our causal estimates are qualitatively similar to the OLS ones, though 25-50% larger.

Why did local elites push for sanitary reform in the aftermath of the cholera crisis? We find evidence that exposure to cholera led to greater public health investments in cities with better-skilled workers and a higher productive capital stock relative to agriculturally used land. This finding is in line with the idea that commercial elites stood to lose more from disruptive labour shocks than agricultural elites, since they employed a workforce equipped with greater human capital and realized important complementarity gains due to their more productive physical capital. Importantly, we rule out that these results are driven by cross-city differences in fiscal capacity, population density, social unrest potential or industrial demand for water. Substantial qualitative evidence underpins our interpretation, indicating that health reformers appealed to economically-minded councils by arguing that healthier workers would increase economic productivity and prosperity.

Our last set of results examines the long-run effects of the epidemic before the onset of the First World War. By 1911, cholera exposure had increased the probability of waterworks adoption by 11 percentage points, but did not significantly impact the within-city expansion of piped water networks. We show that the cholera-induced sanitary reforms had considerable effects on urban development shortly before the First World War. The increased probability of building waterworks triggered by cholera correlates positively with levels of health, human capital in the form of education, economic activity and innovation. This finding suggests that commercial elites pushing for expensive public health reform ultimately obtained high economic returns.

Our paper contributes to various strands of the literature. The first studies the political-economy drivers of broad societal reform in industrializing regimes, such as the extension of the political franchise to voters that demand redistribution (Miller 2008; Aidt et al. 2010; Chapman 2024), threat of social unrest (Acemoglu and Robinson 2000) or the preferences of elites for progressive policies when these align with their own economic interests (Lizzeri and Persico 2004; Galor and Moav 2006; Ashraf et al. 2024). Our findings align with the last of these theories, as health concerns about the workforce prompted elites to invest in waterworks. We also contribute to earlier research on the Prussian context indicating that industrial elites pushed for modernization (Brown 1988; Becker and Hornung 2021; Hollenbach 2021; Krieger 2024), in four novel ways. First, we show that, in additional to slow-moving trends so far studied in the literature, one-off high-profile labour supply shocks had long-term repercussions for the provision of public goods, highlighting fears of labour scarcity as a key mechanism. Second, we offer a new approach to causal identification by leveraging the cholera-spreading effect of military movements. Importantly, our approach allows us to recover a sample-wide causal effect. Third, we move beyond counties, a common administrative unit analysed in the literature (e.g., Becker and Hornung 2021; Krieger 2024), and consider cities by drawing on unexplored sources. This is essential in our context, since cities were responsible for the construction of waterworks and the most relevant political economy dynamics in this respect occurred at the local level. And fourth, we measure the initiation of public health reform over time with a *stock* indicator (opening events) instead of using aggregate expenditure *flows*. This is important to create a cross-city comparable indicator of public health reform at a time when different technologies and building approaches coexisted.

Our article is also related to research looking at the impact of epidemics. This body of work has devoted substantial attention to economic performance and labour market outcomes,[5] and less so to public goods provision. Only a few recent studies have considered the lasting impact of plague epidemics and the 1918 Influenza on overall public goods provision and healthcare (e.g. Dittmar and Meisenzahl 2020; Esteves et al. 2022). Our work adds to these by considering a different public good (water infrastructures), historical context and epidemic disease; and a political setting that was much more responsive to demands from local commercial elites, whose political power greatly weakened after the First World War when secret ballot and universal franchise were introduced. Our study also relates to a historical literature that has long argued that the spread of infectious disease caused the rise of the public health movement during the 19th century,[6] though scholars disagree on whether cholera was a major driving force, or just another disease among the many afflicting the population (Evans 1988; Hamlin 2009). To the best of our knowledge, our study is the first to test this hypothesis quantitatively.

On the methodological level, our simulation-based identification strategy con-

[5]See Voigtländer and Voth (2013) on the Black Death; Barro et al. (2022), Correia et al. (2022) and Ager et al. (2024) on the 1918 Influenza pandemic; Young (2005) on the AIDS epidemic; and Chetty et al. (2023) on COVID-19. Jordà et al. (2022) consider a large number of pandemics since the 14th century.

[6]This argument has been made both in general form by Rosen (1993, 202, 209), and with specific reference to cholera by, among others, Rosenberg (1966, 455), McNeill (1976, 271–273), Hennock (2000, 279), Harper (2021, 434) or Alfani (2022, 32). In a review of the historical literature, Gallardo-Albarrán (2025, 588–589) mentions city-specific examples throughout the world where the cholera narrative plays a prominent role.

tributes to recent debates on causal inference from natural experiments (Borusyak et al. 2025). We estimate a representative causal effect by combining knowledge of initial location-specific exogenous shocks (army exposure) with a calibrated model of their subsequent–but not directly observable–spatio-temporal contagion through the full sample. This allows us to simulate the plausible counterfactual general equilibrium in the absence of the exogenous shock. Structurally similar "formula-based" identification approaches are established in other fields where contagion processes through social and natural pathways are well understood, such as trade (Autor et al. 2013), economic geography (Donaldson and Hornbeck 2016) and environmental economics (Keiser and Shapiro 2019). To our knowledge, this paper is the first to demonstrate that calibrated compartmental epidemiological models can be used to identify the causal effects of a labour supply shock–an approach that can be applied to other settings as well. In this way, we complement recent studies integrating compartmental models into macroeconomic and trade frameworks (e.g., Avery et al. 2020; Antràs et al. 2023).

Moreover, we add to a body of work that has quantified the mortality impact of public health interventions in industrialized countries (e.g. Alsan and Goldin 2019; Anderson et al. 2022; Chapman 2022), including Germany (Gallardo-Albarrán 2020). Our study provides a deeper understanding of how economic interests influenced political dynamics that ultimately determined investments in large-scale infrastructures with broad societal benefits.

Our last contribution refers to research on the conditions leading to the success or failure of policies in current lower-income contexts, such as trust, culture, social conflict or local state capacity (e.g., Nunn and Wantchekon 2011; Acemoglu et al. 2015; Michalopoulos and Papaioannou 2016; Banerjee and Iyer 2005). Our findings from historical Central Europe indicate that political capture by elites in polities without strong central decision makers shapes public goods provision and can create long-lasting inequalities. In particular, the prospects of labour scarcity and their differential impact on elites endowed with varying types of physical capital are essential in creating political momentum for expensive infrastructural reform. Such reforms, as we show, have important long-term developmental consequences that are relevant for cities in contemporary low-income countries.

2 Historical context

2.1 Cholera and labour

Cholera is often regarded as the classic epidemic illness associated to European industrialization (Evans 1988, 125). Overcrowded living conditions and poor access to safe drinking water provided fertile ground for the rapid spread of this disease that thrives in unsanitary, warm and humid environments. Upon infection and initial signs of massive vomiting and diarrhea, as short as twelve hours could pass until the victim's death.

Cholera visited the German Empire on various occasions during the 19th century and the outbreak between 1865 and 1867 was the most deadly (1866 epidemic henceforth). It led to the highest crude death rates in the modern era (35.80 deaths per 1,000), even if we consider the impact of the 1918 Influenza (see figure 1). At the same time, this national average masks substantial variation. For instance, Kappner (2024, 4–5) estimates that mortality rose by 50% in Berlin, when almost 6,000 people lost their lives in 1866, while other cities were left untouched both nearby Berlin (e.g., Wittenberg) and further away in southern Germany (e.g., Munich or Strasbourg).

The 1866 epidemic caused a major labour shortage, while leaving capital untouched. Table 1 illustrates this point with evidence from 14 large cities. In Breslau, where cholera killed three percent of the population in 1866, individuals between 20 and 60 accounted for about 55% of cholera mortality (this figure would be higher if we consider the range 15 to 65, but data limitations make this impossible). In the remaining cities, this figure lies above 40%. Among working-age adults, those without a specific trade, often earning their living in factories and workshops, were particularly at risk. For Königsberg, Schiefferdecker (1868, 91–92) shows that this group represented 43% of all male adult deaths. In Stettin, 62% workers and servants accounted for all deaths (Goeden 1867, 10–12). This is perhaps unsurprising given that cholera was a disease of the proletariat living in cramped and overcrowded buildings. Underpinning the idea that cholera induced a measurable labour shock, we provide suggestive evidence in the appendix (section A.2.1) that wages increased substantially up to four years after local outbreaks in Germany throughout the 19th century.

Table 1: Total and working-age cholera deaths in 14 large cities, 1865-1867

City	All ages	20-60 yrs.	% Working-age
Berlin	5,457 (0.88%)	3,129 (0.88%)	57.34%
Breslau	5,012 (3.06%)	2,792 (2.92%)	55.71%
Königsberg	2,017 (1.99%)	859 (1.45%)	42.59%
Danzig	1,945 (2.18%)	819 (1.58%)	42.08%
Stettin	1,752 (2.48%)	937 (2.35%)	53.48%
Leipzig	1,658 (1.98%)	945 (1.98%)	57.00%
Halle	1,594 (3.47%)	776 (3.30%)	48.70%

City	All ages	20-60 yrs.	% Working-age
Posen	1,398 (2.62%)	694 (2.19%)	49.64%
Barmen	1,204 (2.02%)	576 (1.98%)	47.80%
Erfurt	927 (2.31%)	527 (2.38%)	56.80%
Elberfeld	912 (1.47%)	574 (1.85%)	62.97%
Cologne	851 (0.70%)	423 (0.65%)	49.66%
Essen	791 (2.52%)	328 (1.88%)	41.41%
Magdeburg	626 (0.63%)	323 (0.58%)	51.55%

Note: The table includes the 15 German cities with the largest number of cholera deaths between 1865-1867, except for Hamburg (962 deaths), for which no age-specific information is available. Percentages in parentheses are deaths in relation to the relevant population (total population or population of age 20-60). Age-specific death counts were computed based on Schmieder (1867), Goeden (1867), Delbrück (1867), Liczbińska (2021), Schiefferdecker (1868), Müller (1867), Pfeiffer (1871), Grätzer (1867), Sander (1872), Lent (1868), and KSBiB (1870); age-specific population counts come from KSBiB (1867) and KSSL (1866). Estimates are based on age-bracket-specific excess mortality for Danzig, Elberfeld, Essen and Magdeburg.

Preventing a cholera outbreak was a challenging, nearly impossible, endeavour at the time. There was no systematic national plan against it and cities had a large degree of autonomy over public health matters. Some local governments implemented traditional containment measures upon its arrival, such as home quarantine or the evacuation of buildings for disinfection (Baldwin 1999, 158), while others reacted late or only slightly. In Aachen, for instance, the medical beliefs of the local health official explain varying levels of response over time (Schmitz-Cliever 1952, 264). Ultimately, imperfect knowledge on disease diffusion coupled with idiosyncratic characteristics of local councils and varying degrees of civic compliance rule out the existence of a systematic relationship between public health responses and economic and demographic factors.

2.2 Public health and local politics

Waterworks began spreading at a large scale in the decades after the 1866 cholera epidemic. Initially, a handful of waterworks were built by British experts with the human and physical capital necessary to embark on such challenging engineering projects, such that only about 20 of these were built in large cities before the mid-1860s (Grahn 1904, 310). From the 1870s onwards modern water supplies became increasingly available along with sewerage systems, though the latter followed a delayed trajectory due to the prohibition of the Prussian state to dispose of untreated sewage into rivers in 1874 (Hennock 2000, 281). By 1913, virtually all cities over ten thousand inhabitants were supplied by waterworks, while less than a third of them had piped water systems in the mid-1870s (Gallardo-Albarrán 2025, 572). At a time when treatments for infectious diseases were broadly ineffective, the centralized distribution of piped water to buildings and communities played an important role in protecting the population against water- and foodborne ailments, such as dysentery or typhoid fever (Gallardo-Albarrán 2020).

Public health concerns provided a general motivation for councils to invest in waterworks during the second half of the 19th century, as medical knowledge had improved substantially since the first visitation of cholera (Rosenberg 1966). The connection between this disease and water supplies was acknowledged by 1866,[7] even if epidemiological knowledge was far from perfect and the miasma theory of disease prevailed for a long time. At the same time, policy recommendations based on the avoidance of miasmas often also entailed sanitary improvements (Mokyr and Stein 1996). For instance, a note published in June of 1867 in the British Medical Journal after a cholera conference stated that "[i]t is very probable, that the specific cause of cholera may be contained in the drinking water" (BMJ 1867, 625).[8]

A growing *general* appreciation of the relation between disease and contaminated water, however, was not sufficient to prompt waterworks construction, since the *local* decision to build them was heavily influenced by municipal political economy considerations. This was due to the interplay of three main factors: local responsibility to provide public services, the high costs of waterworks, and the preferences of an

[7] An important breakthrough happened in 1854 when John Snow established that cholera was a waterborne disease by showing that its incidence in London was greater nearby street water pumps fed with contaminated sources. Using a similar method, William Farr traced the origin of cholera in London during the 1866 epidemic to a company that had supplied polluted water illegally (Szreter 1988, 13).

[8] This recognition is particularly remarkable if we consider that this conference was organized by Max von Pettenkofer, a renowned German expert that proposed a miasmatic theory that considered drinking water as secondary in cholera transmission. A few years later, with the impending threat of another cholera epidemic in 1873, Pettenkofer and others present at the Weimar conference created an expert commission that openly advocated the priority, along other measures, of ensuring clean water as fundamental for public hygiene (Cholera-Kommission 1876, 13).

economic elite that held disproportionate political power. We briefly elaborate on these.

German towns took up the responsibility to provide public services at the local level, a development that has been termed "municipal socialism", and expensive waterworks were no exception. Private initiative had a small role in the construction of water infrastructures, so the vast majority of waterworks were public (Grahn 1898-1902). This meant that local councils, and by extension the voters they represented, decided whether and when to embark on building a system of piped water. This decision was not taken lightly, since their associated costs–construction, operation and maintenance–were substantial. Municipal tax-based income was often not enough to fund piped water systems and cities had to rely on long-term municipal bonds. Large cities entered capital markets independently with a solid collateral, while smaller places relied on insurance institutes, the funds of professional associations and mortgage banks (Reulecke 1985, 111). Municipal debt increased to such an extent that sanitary infrastructures accounted for one third of spending on debt retirement between 1869 and 1908 in large Prussian cities (Brown 1988, 307).

Another factor influencing the political economy of public health spending concerns the preferences local elites, which controlled council spending decisions within a political system that granted disproportionate power to the wealthy. Before the introduction of universal and equal suffrage after the First World War, only a small class of citizens that paid enough direct taxes qualified as voters. While there was not a single set of rules defining who could vote across German states (and even within them regulations varied), franchise was limited. In Bielefeld, for instance, less than 20% of citizens could vote in 1911, while just 4% of the population in Hamburg in 1892 had citizenship rights that allowed them to participate in local politics (Krabbe 1989, 64–65). On top of this, the three-class system–in states such as Prussia or Baden–biased decisions even more in favour of the wealthiest voters, by dividing the electorate into three classes.[9] The first class was populated by those that paid the top one third of total direct taxes, the second class with those who paid the second

[9] In some instances, electoral systems changed to keep unwanted groups outside city councils. For example, Leipzig moved from list-based suffrage to a three-class system with the incorporation of working class suburbs in 1894. Other cities (e.g. Cologne, Kiel or Krefeld) raised the tax payment needed to qualify as voter (Reulecke 1985, 133).

third, and so on. Each class had the same number of votes assigned, but the number of voters per class could greatly differ, especially in cases where a small number of individuals paid most taxes. In this way, a small group of extremely rich individuals could easily control a local council. The system was so unequal that as late as 1907 still one or two percent of voters populated the first class, and between 8 and 14 percent the second class (Silbergleit 1908, 184). On top of this, the percentage of voters in the first and second classes declined over time, thus contributing to higher political inequality, due to the relaxation of citizenship payments, inflation and growing nominal wages (Tilly 1995). Adding to these issues of unequal representation, municipal legislation in some states established that property owners should comprise at least half of the municipal council members. In places where this group represented a small share of the population, the regulation ensured that the background of politicians and economic elites was similar.

2.3 The economic calculus of political elites

The political decision-making process leading to municipal reform was intimately linked to the private interests of wealthy individuals. Whether the impact of cholera in 1866 resulted in expensive health-related investments, therefore, depended on the costs and benefits of such investments from their perspective.

Economic elites had financial incentives to promote the construction of waterworks following the labour-shock created by cholera. Workers were an important input in the production process and shielding them from sickness (and early death) presented clear benefits to employers as a sudden decline in labour provoked idle capital and thus lower returns. Such considerations, however, varied among elites of different economic backgrounds. Among the wealthy, industrialists probably had the greatest interest in promoting waterworks. They employed a more productive and human capital intensive workforce that was more expensive to replace in the event of an epidemic (Semrad 2015). Their incentives were even stronger in local economies more reliant on human capital due to greater capital-skill complementarities (Galor and Moav 2006).

While our article puts the focus on the labour shock caused by cholera, elites had other financial motivations to invest in waterworks. Following an epidemic, commercial interests could have taken advantage of the epidemic-induced political momentum to push in favour of public spending that benefitted them. An illustrative example is that of manufacturing production processes that profited from the provision of large amounts of soft and pure piped water (Aidt et al. 2010), particularly in textile and heavy industries (Brown 1988). This complementarity between spending in piped water and private capital created incentives for elites to use public health arguments with the ulterior motive of advancing their interests.

The maintenance of social order may have been another reason the wealthy supported some type of reform. Threat of revolution pushed elites to engaged in redistributive policies (Acemoglu and Robinson 2000), and the social unrest spurred by cholera may have heightened the perceived risks of revolution and its associated costs. This issue may have been particularly salient in urban areas where workers and capital owners lived closer to each other. In our setting, however, this seems less important because, while the earlier cholera epidemics were associated with significant civil protests and discontent, the responses to the epidemic by public authorities became much less strict during the second half of the 19th century (Baldwin 1999). In line with this, Krieger (2024) argues that pressure by workers' movements did not promote policies that prevent disease outbreaks.

In contrast to the arguments above, there are several reasons why elites may have *failed* to invest in public health in the aftermath of cholera. First, a single shock may have been insufficient to trigger permanent change in an epidemiological context characterized by frequent mortality events. Instead, health spending may have increased temporarily on relatively low-cost measures–street cleaning, slum clearing, establishment of health boards–as it had been the case in preceding decades. In contrast to those, modern water supplies were very expensive and the tax requirements to fund them may have been prohibitive even if their benefits were salient (Chapman 2024). Second, even if clean water supplies were considered a worthy use of public funds, municipal spending decisions entailed allocating scarce resources to other areas that elites valued perhaps even more, such as transport infrastructure (Becker and Hornung 2021) or human capital in the form of education (Hollenbach 2021). And third, in a context where the provision of public goods was highly decentralized, elites may have faced high coordination costs (Olson 1965). These could have undermined or substantially delayed waterworks construction, particularly in densely populated regions.

A final consideration concerns the type of cholera exposure that most influenced local policymaking. On the one hand, only high-mortality outbreaks resulted in directly observable costs associated with labour scarcity and broader disruptions in the local economy. On the other hand, cholera was a dreaded disease with a long history of recurrence up to the mid-1860s, so even relatively mild outbreaks–coupled with the fear that future outbreaks may turn out worse–could have generated enough political momentum by making public health concerns salient. Beyond the magnitude of local outbreaks, the experience of nearby places being affected may have independently contributed to public health awareness. In our empirical framework, we take these considerations into account. In sum, expectation about the magnitude and significance of cholera's effect on local public health investments are *a priori* ambiguous. Our empirical framework focusses on the labour shock dimension, but we consider the other pathways just discussed where necessary.

3 Data and descriptive statistics

Our sample is based on a broad notion of Germany's urban landscape in the 19th century. It contains 2,552 municipalities that either acquired a town charter or are listed in the *Deutsches Städtebuch*, a multi-volume historical encyclopedia of German cities and city-like municipalities.[10] Importantly for the representativeness of our analyses, our dataset is not restricted to places with formal privileges, but also captures the experiences of a diverse range of places, including large "industrial villages", which we refer to as cities for simplicity.

[10]The initial target sample comprised 2,685 locations, but lack of information for some of them resulted in a 95% coverage of our target. See figure 2 for the spatial distribution of the city sample.

3.1 Waterworks

We measure cumulative investments in waterworks by gathering information on the date when they started operating. This strategy has several advantages. First, from a historical perspective we capture an event that was clearly meaningful to contemporaries and therefore well recorded. For instance, when Munich finished the main works of its new water supply system in 1883 the *Allgemeine Zeitung* reported on May 2nd that the city will "[...] be in possession of a public facility that many cities, not just in Germany, will rightly envy–a facility that will bring lasting honour to those who conceived it, as well as to those who built it." Second, the construction and initial operation of waterworks were a proud engineering achievement, representing an important milestone in urban development. Consequently, this information is often reported by multiple sources, such as contemporary newspapers, engineering journals or local histories. Third, the dates we put together reflect a comparable phenomenon across cities and time, as they indicate the timing of a major urban transition. Opening dates were also the focus of 19th century engineers and educated individuals interested in tracking the expansion of modern water supplies, as illustrated by the work of Baker (1888) for the United States or Grahn (1883) for Germany. In addition, a similar date approach has been used in recent studies estimating the health impacts from access to piped water (Gallardo-Albarrán 2020; Alsan and Goldin 2019; Beach et al. 2016). This suggests that the events we capture were meaningful at the time both in public discourse and with respect to their effect on citizens' health.

We draw on a large array of sources including contemporary compilations by experts (e.g., Grahn 1898-1902; Salomon 1906-1911), the journal of the German association of gas and water experts (DVGWF, various years), fiscal reports (KPSLiB 1914-1920), and a variety of local histories from lesser known publishers, local authorities, water supply companies or enthusiasts of the topic. Section B.2.1 in the appendix details our sources and procedures and it elaborates on several conceptual aspects of our outcome variable. Figure B.2a shows the diffusion of piped water supplies in Germany by city size (at the time of access). We observe a clear acceleration of waterworks completion in the late 1860s, a faster growth rate from the mid-1880s onward and a peak around 1900. At the eve of the First World War, more than 20 million German citizens were living in communities supplied by waterworks.[11]

3.2 Cholera

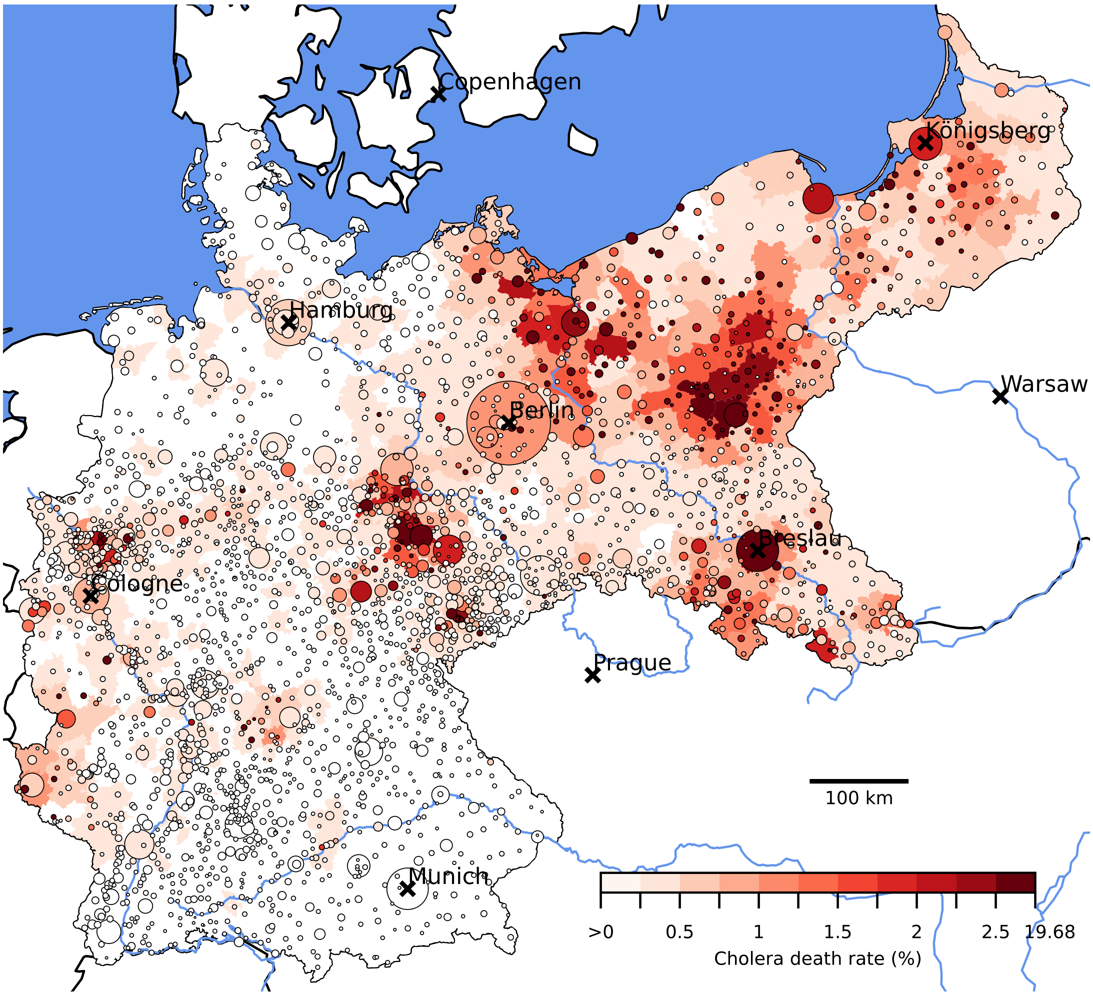
We reconstruct the mortality impact of cholera across all states and territories that formed the German Empire in 1871 by drawing on a variety of sources.[12] Whenever possible, we rely on archival or official published sources. Where such sources are not available, we use medical and public health journals or newspapers. Our procedures and sources are spelled out in detail in Kappner (2025), though we provide additional relevant information in section B.2.2 of the appendix. The clear symptoms of cholera, coupled with the attention public authorities paid to effectively detecting infected individuals (e.g., compulsory notification of relevant symptoms to authorities) give us confidence that our dataset accurately depicts the incidence of cholera throughout the analysed territories. A systematic validation against independently measured excess-mortality data and alternative sources is provided by Kappner (2025).

Figure 2 shows that the 1866 epidemic exhibits substantial spatial variation. About 36% of our sample (921 cities) experienced a cholera epidemic and, among shocked locations, the 25th, 50th, 75th and 90th percentiles of cholera death rates are 0.11, 0.52, 1.66 and 3.43 percent (descriptive statistics are provided in table B.1). Some of the epidemic clusters reflect the different ways in which the disease originally entered Germany. According to Krehnke (1937, 399–407), the initial signs of cholera appeared in 1865 in the Thuringian city Altenburg, when the child of a woman coming from Odessa with diarrhoeal symptoms died in August; by mid-December, when the epidemic was over, 92 people had died in the city. At around the same time, the disease hit the Kingdom of Saxony. The main wave of 1866 was introduced from Luxembourg. Cholera appeared first in Westphalia and the Rhineland and it reached

[11] The First World War ended this trend abruptly and cities struggled to raise the necessary funds in the economic depression of the post-war years. In the mid-1920s construction activity picked up again, now driven by centrally enforced and subsidized programs. Due to the radically changed political environment of the post-war years, we do not consider this period in our analysis.

[12] We focus on mortality rather than morbidity because, while we get a fairly complete picture of the former, we cannot reconstruct non-lethal cases for a fraction of our sample due to data restrictions. Moreover, the focus on mortality reduces measurement error as non-lethal (and particularly mild) cholera cases were often not reported for those who stayed at home (Guttstadt 1906, 274), while reliable records of lethal cholera are the norm due to its clear symptoms.

Figure 2: The spatial spread of the 1866 cholera epidemic



Note: The map shows cities and rural counties in Germany, with colour shading indicating cholera deaths per 100 in 1865-1867. Circles are cities, with circle size proportional to their 1866 population. Cross markers place the cities of Cologne, Hamburg, Munich, Berlin, Breslau, Königsberg, and foreign capitals. Black lines show the German Empire in its 1871 borders and blue lines show major rivers.

some harbour cities in northern Germany, such as Stettin. Then, East Prussia was affected in June as a result of the mobilization of troops and from there cholera reached different territories, such as the Prussian province of Saxony and the Duchy of Brunswick in July, and Oldenburg in August, spreading widely across North and

West Germany. In 1867, a third wave reached, among other places, the Rhineland, Breslau, Mecklenburg-Schwerin and Baden. Notably, while it had been subject to previous cholera outbreaks, South Germany remained largely untouched in 1865-1867, with major exceptions associated with Prussian troops' penetration along the border territories.

3.3 City characteristics

Confounders: Pre-determined socioeconomic and hydrogeological characteristics We collect additional data to measure various pre-determined (i.e., pre-1866) city-level attributes that block confounding pathways in our regressions (a detailed overview is given in appendix B.3). We control for the pre-1866 crude death rate to address that generally less healthy cities might have been more likely to suffer from cholera, while simultaneously more incentivized to invest into waterworks. We include 1866 population levels to take into account that larger cities can exploit economies of scale to build waterworks (Gallardo-Albarrán 2025), while simultaneously exhibiting a higher contagion potential. With an industry-structured-based measure of economic development, we address the concern that cities with greater fiscal capacity were more capable of building waterworks (Brown 1988), while at the same time more likely to mitigate major health shocks through non-observable mechanisms and bureaucratical capability. Market access is also included to address that central, better connected places might have gotten waterworks earlier due to their representative functions, access to ideas and supply networks (Andersson et al. 2023), while at the same time being more exposed to cholera due to trade integration (Zimran 2020). Finally, we compute six time-constant hydrogeological properties (e.g., groundwater salinity, surface water access and precipitation) that simultaneously shaped technical aspects of water supplies and cholera epidemiology.

Main mechanisms: Elites' economic background and workers' human capital The variables we describe in the following only refer to Prussian cities due to data constraints. These localities make up more than half of our sample and cover all of Germany, except the South (see table B.1 for some descriptive statistics). Further information is provided in appendix section B.5.

To test whether the extent of political momentum generated by cholera varies across cities with different economic elites and human capital levels (as discussed in section 2.2), we construct two city-level measures, again based on data measured before or shortly after 1866. The first measures the economic background of local elites based on fiscal information on their key tangible assets: the value of their productively used land and buildings. In particular, we compute the ratio of the value of the urban non-residential building stock to the combined value of this building stock and agriculturally used land held by urban residents. The underlying city-level data was produced in 1861 as a result of a reform of Prussia's land and building tax code (see section B.5 in the appendix for detailed information on the sources).[13] A higher ratio indicates a stronger presence of commercial-industrial elites both in the local economy and the city council. In figure B.6, we show that this measure exhibits a coherent spatial pattern with high values in industrial and commercial regions, such as the Rhine area, Upper Silesia, Hamburg or Berlin. Moreover, in appendix section A.4.1 we validate it for a smaller sample of Prussian cities. We show that our asset-value-based measure correlates positively with the relative importance of capital in individual income taxes, the actual share of commercial-industrial elites in local city councils, and liberal parties' vote shares.

Our second measure proxies workers' human capital through officially measured literacy rates, following Franck (2024) and Ashraf et al. (2024). From Becker and Cinnirella (2020), we obtain the share of literate inhabitants at age 10 or older in all inhabitants in 1871. Literacy rates exhibit substantial spatial variation (see figure B.7) ranging from below 50% to universal literacy. Also, there is a clear East-West gradient that may partly reflect an anti-Polish bias in language proficiency measurement, though the comparatively low literacy rates in German-speaking East Prussia suggest that our variable also reflects genuine differences in literacy. In line with this, our validation exercise in the appendix (section A.4.2) shows that our literacy-based human capital measure correlates strongly with other education indicators, such as the number of teachers in public elementary schools from censuses conducted in 1849,

[13] We do not use elite-composition measures established in earlier studies because those are only available at the county level, such as vote inequality or landownership inequality (Becker and Hornung 2021).

1864 and 1878.

For the empirical regressions, we discretize the two continuous ratios discussed above into dummies. We define cities exhibiting above-average values as *commercial* and *literate*, respectively.

Other mechanisms We measure three more pre-determined city-level properties to capture alternative mechanisms that could explain the link between cholera exposure, local elites and public health spending (see the appendix section B.5 for more details). The first is the ratio of inhabitants to inhabited buildings, based on the 1871 building census. This density measure proxies the direct cholera contagion risk of elites living in more crowded cities, closer to poor inhabitants and unsanitary quarters. We use it to rule out that public health investments simply reflected elites' worries about their *own* health. Second, we use Tilly (1990) to measure whether any popular protest, riots or insurgencies took place in the 50 years before 1866. This measure addresses the idea that local protest potential–possibly amplified during mortality crises–prompted elites to provide public goods. Third, we build an indicator capturing potential lobbying activity from industrialists that desired a steady, clean water supply as direct production input. Following Brown (1988), we focus on dye works, textile finishing and steel and metalworks, whose presence in the early 1860s we measure at the city level.

Long-run developmental outcomes Our analyses of the long-term effects of the 1866 epidemic is based on several city characteristics measured closely before the First World War. The first is within-city piped water coverage, defined as the percentage of properties connected to the waterworks network in 1911. In addition, we measure developmental proxies for health (crude death rates and infant mortality rates), human capital (students in higher education per capita), economic development (stock-market listed firms per capita) and innovation (patents per capita). The sources are detailed in section B.6 of the appendix.

4 Empirical framework

4.1 Baseline model

We estimate the impact of exposure to the 1866 cholera epidemic on the adoption of waterworks using a linear panel regression framework that captures differential opening dynamics over time:

$$\mathrm{W}_{it} = \beta C_i \mathbf{1}_{t>1866} + \mathbf{X}_i' \gamma \mathbf{1}_{t>1866} + \alpha_i + \delta_{t \times r(i)} + \epsilon_{it}, \tag{1}$$

where city i in region $r(i)$ is observed annually from 1848 to 1891. Our baseline results refer to this period because 1849 marks the construction of the first modern large-scale waterworks in Germany (Grahn 1904, 303); 1891 is the year before another high-profile cholera epidemic took place, leading to the takeover of Hamburg by the Prussian government through Robert Koch and thus marking the superiority of the bacteriological approach to public health. Also, major municipal tax reforms took place in the early 1890s that altered local finances in a way that makes those years less comparable with the preceding decades.[14] An additional advantage of our cutoff is that the control variables refer to cities' characteristics around 1866 and thus become less meaningful beyond 1891 when urbanization and industrialization were in full swing. These considerations notwithstanding, our later long-run analyses extend the study period through 1911, ending before the institutional rupture of the First World War.

The dependent variable, W_{it}, is a binary indicator equal to 1 in the year t a city opens a modern waterworks. As this reflects an absorbing state, similar to Cantoni et al. (2024), we drop the city from the sample from $t+1$ on. The regressor of interest is the interaction of a city's exposure to the 1866 cholera epidemic, C_i, with the a post-1866 indicator, $\mathbf{1}_{t>1866}$. We operationalize C_i along two dimensions to measure distinct ways in which cholera could prompt sanitary reform, as discussed in section 2.3. The first is a binary variable for whether a city experienced any

[14] German cities heavily relied on surcharges on state income tax, which provided limited resources, particularly in municipalities with large working classes. In the 1890s, there were efforts to modernize the fiscal system with progressive taxes and clearer tax assessment regulations with the Miquel reform (Reulecke 1985).

cholera deaths between 1865 and 1867. This approach delivers information on how experiencing an outbreak, regardless of its intensity, may have prompted elites to react. Given the periodical reappearance of cholera since the 1830s, even if death rates were low in 1866 in some locations, elites may have feared that future epidemics could turn out more disruptive. The second metric is cholera mortality per 100 inhabitants. With this variable, we test whether the intensity of an epidemic and the actual extent of the associated labour shock influenced sanitary reform decisions. An advantage of the former approach is that we measure a clear event subject to less measurement error and with a clearer interpretation, while the latter provides a greater range of potentially relevant variation to understand spatial and temporal patterns of waterworks diffusion.

We include city fixed effects (α_i) to control for all time-invariant city characteristics, and region-specific arbitrary time trends ($\delta_{t \times r(i)}$) to capture differential regional trajectories in public investment (see appendix section A.3.1 for a full list of the 75 administrative regions we define). The interaction $C_i \mathbf{1}_{t>1866}$ is therefore identified solely from within-region-year variation across cities, conditional on fixed city-level factors. We cluster standard errors at the level of these fixed effects, and also account for their potential spatial correlation (see appendix section A.3.5).

The timing of sanitary investments may be correlated with city characteristics that simultaneously affected exposure to the 1866 epidemic. We address such spurious correlation by controlling for the set of potentially confounding, pre-determined (pre-1866) city characteristics discussed in section 3.3: pre-1866 crude death rates, population in 1866, economic development in the early 1860s, market access in 1866, and six (hydro-)geological properties (see also appendix B for details, sources and descriptive statistics). Like the the cholera shock, we interact the vector of time-invariant confounders, $\mathbf{X}_i$, with the post-1866 dummy. We do not attempt to control for *post*-1866 changes in these city characteristics, since they may themselves be affected by waterworks adoption, making them "bad controls" (Angrist and Pischke 2009).

Our framework is structured like a difference-in-differences regression, but differs from standard treatment-effect estimation in two important ways. First, by dropping cities from the sample after they open waterworks, we deliberately unbalance the panel to estimate the annual hazard of *transition* into the absorbing state of having waterworks. Second, as pre-1866 waterworks opening activity was negligible (see figure 1), differential pre-trends are not a first-order concern. Thus, β quantifies the average difference in the probability of opening waterworks in a given post-1866 year between cholera-affected and unaffected (or less affected) cities, conditional on city fixed effects, regional trends, and baseline characteristics. In section 5, we also estimate a dynamic model with time-varying opening probabilities and show robustness to alternative estimation approaches. We also show robustness to using a traditional balanced-panel model, and differential pre-trends.

4.2 IV and simulation-based approaches

Despite extensive controls for observable city characteristics, our baseline regressions could suffer from endogeneity due to non-classical measurement error and unobserved confounders. First, under-reporting of cholera outbreaks in 1866 may reflect unobserved attributes such as institutional capacity, bureaucratic salience, or political centrality–factors likely correlated with public health investment. In such cases, our OLS estimates would be attenuated. Second, these same latent attributes–as well as unobserved qualitative differences in medical knowledge, health-conducive behaviour or scientific engagement–may simultaneously reduce cholera incidence and increase the likelihood of waterworks adoption, leading to further downward bias. Third, local investment decisions may have been made well before the actual opening of waterworks. If those decisions reflect long established modernization trajectories that also affected the 1866 cholera outbreak, subtle forms of reverse causality could arise. We overcome these potential sources of endogeneity with two strategies relying on exogenous cholera variation from a military conflict that took place in the same year.

Exogenous variation from the 1866 Austro-Prussian War We exploit plausibly exogenous variation in cholera outbreaks driven by troop movements during the Austro-Prussian War of summer 1866. This short but intense conflict involved Austria, Prussia, and their small-state German allies, prompting military mobilization throughout the German lands. The northern and central smaller states sided with Prussia, while the larger southern ones, such as Bavaria and Württemberg, sided with

Austria. Constituting Europe's last "cabinet war" (Salewski 2003, 57), battles took place in the countryside and along strategic lines, minimizing any direct impact on the civil population, city structures, and physical capital. This distinguishes our setting from other recent work that focusses on historical conflicts in which the civilian population and physical capital were deliberately targeted, including the American Civil War (Feigenbaum et al. 2022), the Thirty Years' War (Bosshart and Weigand 2025), and the Spanish Civil War (Tur-Prats and Caicedo 2025). A scenario closer to ours is the 18th-century Great Northern War, where plague-spreading armies caused labour scarcity along the Baltic Sea (Marczinek 2025).

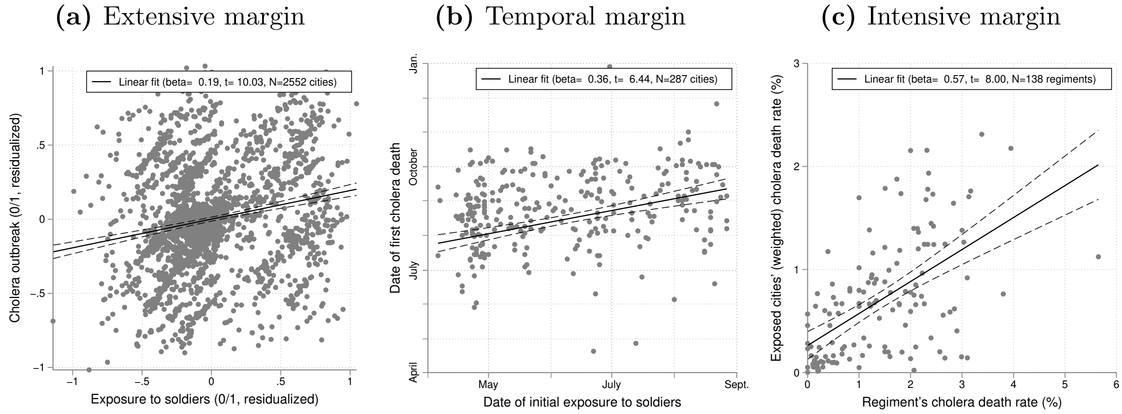
Historians and contemporary observers consistently point to the 1866 war's role in spreading cholera. In 1867, international medical experts gathering in Weimar concluded "beyond doubt" that soldiers were the source of many local outbreaks (BMJ 1867, 625). August Hirsch (1881, 295), co-founder of the German Empire's cholera commission (established in 1873), later described how the disease spread along military routes from western to eastern Prussia and then to southern Germany, an observation later echoed by cholera historian Walter Krehnke (1937, 399–400). Further evidence for the connection between the war and cholera comes from a major field hospital in Bohemia–the central battleground–where cholera accounted for 15, 87 and 89% of soldiers' wartime deaths in June, July and August, respectively (Loeffler 1869, appendix II). City-level and regional sources also link soldier presence to local outbreaks. One example is South German Wertheim, which experienced an outbreak shortly after receiving a garrison from Hamburg where soldiers had shown cholera symptoms (Prinzing 1916, 187).[15]

We use regimental histories (*Regimentsgeschichten*), published since the late 19th century as retrospective tributes to Germany's military, to reconstruct the marching trajectory of each Prussian or allied battalion involved in the 1866 war. The resulting daily location data allows us to systematically document the epidemiological link between troops exposure and cholera outbreaks. 27% of the cities in our sample were visited by soldiers in 1866 (see figure B.5). Figure 3, panel a, shows that these cities

[15] While not all military contact triggered epidemics, the risk was heightened: cholera could persist in barracks and hospitals for weeks before reaching the civil population. For example, Leipzig had repeated contact with soldiers since late June but did not experience an outbreak until mid-August; by November, nearly 2,000 people had died (Schmieder 1867, 1).

were significantly more likely to experience cholera outbreaks. This relationship holds conditional on the full set of fixed effects, socio-economic and geological confounders, and three additional war-related dummy controls that block direct links between military exposure and infrastructure investments: garrison status, capital city status, and local battle activity. Moreover, in the subset of 287 non-garrison cities with both soldier exposure and cholera, the timing of exposure significantly predicts the start of local epidemics (panel b). Finally, regiments with more soldiers dying from cholera visited cities with greater average cholera death rates (panel c).

Figure 3: The propagation of cholera through the 1866 German War



Note: This figure illustrates the relationship between soldier presence and local cholera outbreaks along three dimensions. Panel a correlates 1866 cholera outbreak dummies with soldier presence dummies. Both measures are residualized on the full set of controls. Panel b links initial soldier exposure dates to first cholera deaths in the subset of non-garrison cities. Panel c plots regiment-level cholera mortality against average death rates in visited cities (including cities without cholera outbreaks). Black lines show linear fits with 95% confidence intervals.

We believe that these correlations reflect the genuine epidemiological link between the 1866 military campaign and local cholera outbreaks, rather than any remaining unobserved city characteristics that would also increase outbreak probabilities in the decades before or after. To further support this claim, we confirm that soldiers presence in 1866 does not predict cholera outbreaks in any other epidemic between 1831 and 1895 (figure A.3, panels a to d). Nor does it correlate with above-average outbreak intensities during the 1870–1871 smallpox epidemic among the 106 Prussian cities that housed French prisoners of war in 1870 (panel e), or any pre-1866 historical epidemics, such as plague in the 14th century (panel f).

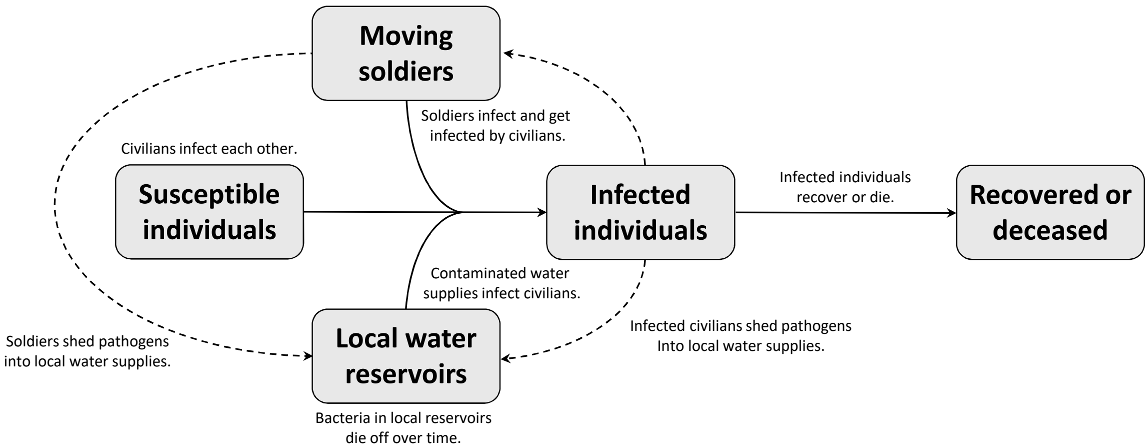
Two-stage-least-squares and control-function estimates We employ two complementary approaches that exploit the epidemiological relevance and (conditional) exogeneity of military exposure to estimate the causal effect of cholera on waterworks adoption. First, we perform a two-stage-least-squares (2SLS) regression by instrumenting observed cholera outbreaks with their *predicted* probabilities, based on city-level exposure to soldiers. In particular, we regress observed cholera deaths on a soldiers-presence dummy in a cross-sectional negative binomial model, controlling for the population at risk as an exposure term. We then use the fitted outbreak probability as excluded instrument in a regular 2SLS regression. This approach, originally recommended by Adams et al. (2009), increases first-stage power by reducing finite-sample bias in binary-on-binary regressions (Wooldridge 2010, 623; Xu 2021). We do, however, obtain similar results with a slightly less powerful first stage when directly using the soldier exposure dummy as excluded instrument (as illustrated in figure 3, panel a). The 2SLS approach delivers a local average treatment effect (LATE) estimate, representative of the "complier cities" whose cholera outbreaks were induced by visiting troops. To estimate a more representative population-wide effect, we additionally run a control-function (CF) regression with a non-linear Probit first stage.[16] The CF estimate is non-local, but biased towards zero if the first-stage link is incorrect and the error terms in the first and second stages are not jointly normal (Basu et al. 2018). We therefore treat it as suggestive evidence rather than a precise estimate.

Simulation-based estimate Our second approach to estimating the impact of the 1866 epidemic on subsequent waterworks investments rests on a counterfactual simulation that isolates exogenous, soldiers-driven variation in local cholera outbreaks across *all* cities, not just those directly visited. We begin by setting up a spatially structured compartmental model tailored to cholera epidemiology (see e.g.,

[16]We follow Basu et al. (2018) by first obtaining the generalised residual from a cross-sectional first-stage Probit regression of the cholera outbreak dummy on the soldiers-predicted outbreak probability. Then, we include this residual in the linear second stage alongside the potentially endogenous observed cholera dummy. In appendix figure A.4 (panels a and b), we show that the CF approach captures more first-stage variation than a linear regression.

Anteneh et al. 2024; Wang 2002). It features realistic temporal outbreak dynamics *within* cities–via person-to-person contact and bacterial contamination of local water reservoirs–and explicit transmission channels *across* cities via trade and migration linkages. Crucially, we go beyond standard modelling practice by adding a pathogen transmission channel via spatially moving battalions–each with their own cholera dynamics–that affect nearby populations and water reservoirs. Figure 4 presents a simplified illustration of the core mechanisms of the epidemiological model. A less stylized version, highlighting additional mechanisms in the model, is available in appendix figure C.1.

Figure 4: Simplified epidemiological model flow chart



Note: This flow chart presents a simplified overview of the compartmental epidemiological model we use in our simulation of the 1866 cholera epidemic. Each box represents a city-specific compartment, except for the "moving soldiers", who form their own entity with compartments. Arrows illustrate the mechanisms governing interaction between the different compartments. See section C for full technical model details and a more detailed flow chart.

Having set up the model, we estimate its 16 endogenous parameters by minimizing the difference between a set of moments in the model-predicted distribution of cholera deaths across cities and their actual observed distribution. Next, we re-simulate the now calibrated model, but switch off the battalions-based transmission channel. This yields a plausible estimate of how each city's epidemic death count would have evolved in the absence of the military campaign. By subtracting these counterfactual estimates from the full-simulation distribution, we isolate the marginal epidemiological contributions of troop movements on each city's outbreak. Finally, we estimate the effect of these counterfactual exogenous shocks on waterworks adoption within our linear regression framework. As our simulation-based approach provides exogenous outbreak variation for *every* city in our sample, we interpret the resulting estimate as a non-local average treatment effect. More details on this simulation approach are discussed in appendix section C.

5 Empirical results

5.1 The impact of cholera on public health reform

Table 2 presents the results of estimating equation 1 using two cholera shock measures: the presence of the disease (panel a) and the intensity of the local epidemic (panel b). Columns 1 to 6 add different control variables individually, with column 7 including all of them. Beginning with panel a, we observe little change in our coefficient of interest when controlling for (hydro-)geological factors, pre-epidemic mortality and market access (columns 2, 4 and 6, respectively). The coefficients become smaller but remain statistically significant when adding population and economic development (columns 3 and 5). We also find that economies of scale and local economic development are important determinants of public health infrastructures (see full regression output in table A.1). Our preferred specification in column 7 implies that the annual probability of opening waterworks after 1866 was about 0.4 percentage points higher in cities stricken by the cholera epidemic, relative to cities unaffected by this disease. This effect is sizeable given that the probability that waterworks opened in non-cholera cities after 1866 was 1.14%, implying a relative increase of 35%.

Panel b repeats the analyses employing cholera death rates. We obtain positive and statistically significant coefficients, except when we control for population size and economic development (columns 3 and 5, respectively). Consequently, the full specification in column 7 yields a positive coefficient, though only significant at the 26% level. Compared with the results in panel a, this suggests that the *presence* of cholera and the fear it may have caused of future shocks was more strongly associated with health reform than its epidemic *intensity*. We take the difference in the statistical

Table 2: The impact of cholera on the opening of waterworks, 1848-1891

	Effect on annual opening probability (percentage points increase)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel a: $C_i \equiv$ **Presence of cholera (0/1)**						
$C_i \mathbf{1}_{t>1866}$	1.087***	1.096***	0.523***	1.087***	0.723***	1.044***	0.393***
	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)	(0.006)
	Panel b: $C_i \equiv$ **Cholera deaths per 100**						
$C_i \mathbf{1}_{t>1866}$	0.0999**	0.105**	0.0609	0.0996**	0.0729	0.0929**	0.0499
	(0.042)	(0.037)	(0.175)	(0.042)	(0.113)	(0.047)	(0.265)
City and region-year FEs	✓	✓	✓	✓	✓	✓	✓
(Hydro-)geology		✓					✓
Population			✓				✓
Pre-epidemic mortality				✓			✓
Economic development					✓		✓
Market access						✓	✓
Control mean (post-1866)	1.143	1.141	1.141	1.141	1.141	1.141	1.141
Cities	2552	2552	2552	2552	2552	2552	2552
R-squared in panel a	0.091	0.090	0.094	0.090	0.091	0.090	0.095
R-squared in panel b	0.090	0.089	0.094	0.089	0.091	0.090	0.095

Note: This table shows the results of estimating equation 1 when incrementally adding control variables. Panel a measures cholera exposure with a binary (dummy) variable, while panel b uses cholera deaths per 100 inhabitants. The dependent variable indicates whether a city opened waterworks in a given year. After opening waterworks, cities exit the sample in the following year. Coefficients have been multiplied by 100, thus measuring percentage points. Numbers in parentheses are p-values based on standard errors clustered at the city and year levels (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$). R-squared reports explained variation after absorbing the fixed effects. See appendix table A.1 for the control variables' coefficients.

significance of the cholera coefficients in table 2 as reason to focus on the dummy measure in the remainder of our analyses. This choice is also supported by the much smaller magnitude of the marginal effects in panel b, if we take the coefficients at face value.[17]

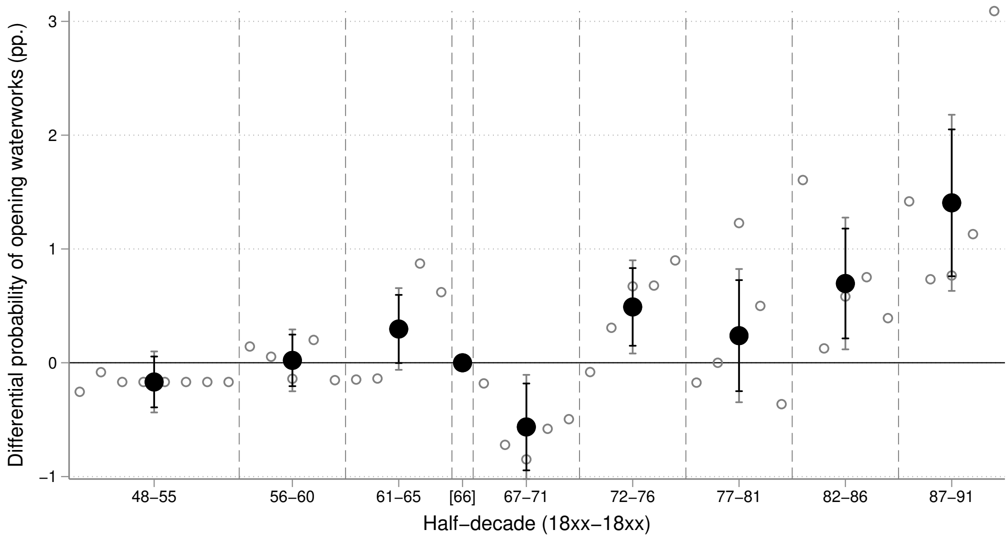
Next, we examine the temporal unfolding of the epidemic shock's effect, given that our estimates in table 2 pool the average annual opening probabilities over the period 1866-1891. We estimate a dynamic version of equation 1 by substituting $\beta C_i \mathbf{1}_{t>1866}$ in equation 1 with a set bin-wise dynamic coefficients, i.e. $(\sum_{\tau \in T} \beta_\tau C_i \mathbf{1}_{t\in\tau})$, where T refers to a set of non-overlapping periods (τ). We let the effect vary by symmetrical 5-year-bins before and after the cholera shock to efficiently trade-off statistical power and temporal resolution, with the reference period limited to the year 1866.[18] The inclusion of pre-1866 bins also allows us to test the assumption that cholera-exposed cities had no differential opening probabilities before 1866. Three mechanisms may undermine parallel trends. First, cities that opened *effective* waterworks before 1866 may have been less vulnerable to cholera in 1866, thus ending up in the control group (the unaffected cities). Second, cities that anticipated vulnerability to a future cholera shock may have proactively opened waterworks before 1866. If, at the same time, those early waterworks shielded only imperfectly against cholera, this would lead to positive selection into the treatment. For instance, Hamburg built waterworks early but had no filtration system, which was key in protecting against cholera (Knutsson 2025, 3). And third, earlier cholera shocks may show some correlation with the 1866 epidemic, while at the same time already having motivated waterworks adoption before 1866, implying positive selection. However, we expect these three selection mechanisms to be weak because our specification controls for many pathways through which selection could operate and, perhaps most importantly, few cities opened waterworks before 1866. Also, these pre-epidemic systems likely still had low within-city coverage rates that provided limited protection by 1866.

Figure 5 presents two interesting patterns on the dynamic effect of cholera. Pre-

[17] More specifically, the extensive margin effect in table 2 is similar to a 6.4 standard deviations increase in the unconditional intensive margin ($\frac{\beta_{\text{extenisve}}}{\beta_{intensive}} \times \text{sd(Cholera death rate)}^{-1} = \frac{0.393}{0.050} \times 1.212^{-1} \approx 6.4$). In the appendix (section A.2.2) we show that the reaction to the cholera shock was similar among spared and mildly-hit cities, suggesting that cholera intensity needed to cross a certain threshold to matter substantially. The regression-coefficient maximizing contrast between mild-and-non-affected and severely-affected cities occurs at a threshold of 0.066 deaths per 100. 21% of the cholera-shocked cities fall below this threshold.

[18] Note that our cholera shock variable is not staggered across time. Thus, estimation problems from heterogenous treatment effects identified by the recent event study literature are not a concern (e.g., Borusyak et al. 2024).

Figure 5: The dynamic impact of cholera on the opening of waterworks, 1848-1891



Note: This graph plots the dynamic effect of cholera on waterworks based on five-year bins around 1866, the reference period. Black (gray) lines show 90% (95%) confidence intervals based on standard errors clustered at the city and year levels. Gray circles in the background show estimates from a regression with annual interactions instead of 5-year bins.

shock coefficients are close to (and statistically equal to) zero, suggesting little net selection into the cholera shock.[19] Also, political momentum took some time to build up after the epidemic, as cholera-shocked cities were *less* likely to open waterworks in the immediate post-shock period (1867-1871). Three factors may account for this: i) short-run economic disruptions caused by the shock (Gallardo-Albarrán and Kappner 2025); ii) opening probabilities between 1867 and 1871 may reflect decisions taken before the shock because it took years to settle the details of the projects (Gallardo-Albarrán 2020); and iii) municipalities may have anticipated the arrival of the disease when it spread out of Bengal in 1863 and gradually approached Western countries. Regardless, from the early 1870s on a clear pattern emerges indicating that cholera-shocked cities were more likely to open waterworks.

[19] The null of joint insignificance for the three pre-shock coefficients is not rejected ($p = 0.182$). However, we can not rule out that both negative and positive selection are significant and cancel each other out.

5.1.1 Robustness tests

The results in table 2 are robust to controlling for temporal trends in several alternative ways, such as moving from 75 districts to 646 counties in the year-space fixed effects structure, or imposing flexibly shaped year-specific latitude and longitude splines to capture spatially clustered trends that transgress administrative boundaries (see table A.11 in section A.3.1). Also, we show in figure A.2 that leaving out individual regions does not systematically influence our results. Finally, our main coefficient remains highly significant when accounting for spatially auto-correlated unobservables following Conley (1999), with cutoffs ranging from 25 to 200 kilometres (see appendix table A.14 in section A.3.5).

Our results are similar when re-estimating equation 1 while keeping cities in the sample after they open waterworks (see table A.12 in section A.2.4). Rather than estimating differential opening hazards over time, this alternative specification estimates differential waterworks *operation* probabilities, i.e. smooth aggregates of the hazards our main approach focuses on. Our results are also robust to employing alternative models that capture distinct dimensions of the post-cholera waterworks adoption process, among them a semi-parametric Cox proportional hazard model for the instantaneous opening risk of waterworks, similar to Alsan and Goldin (2019); a Tobit model to estimate the number of years between 1866 and the opening of waterworks, treating cities that opened after 1891 as right-censored; and a regression of the share of years between 1867 and 1891 in which a city operated waterworks as a continuous outcome, capturing cumulative duration (see section A.3.2, table A.11). Additionally, section A.3.6 in the appendix shows that our results are not driven by earlier cholera epidemics since the 1830s. Finally, we show that our coefficient of interest remains similar in size and statistically significant when taking into account the effect of cholera shocks surrounding a city with spatial lag specifications and when controlling for cross-city imitation processes in sanitary investments (see section A.3.5).

5.1.2 IV and simulation results

We present additional estimates in table 3, based on the causal identification approaches discussed in section 4.2. We limit the sample for the following regressions to cities in regions with variation in cholera outbreaks as the fixed effects structure in our framework absorbs all variation in regions without cholera deaths, leaving no variation for the soldiers-based instrument to explain. We also add three controls that ensure the conditional exogeneity of the instrument, as discussed before.

The re-estimated OLS coefficient (based on the smaller sample and the additional controls, column 1) is in line with our baseline results in table 2. The 2SLS approach also yields a positive and statistically significant coefficient (column 2), although much greater in magnitude than that in column 1. As explained in section 4.2, we interpret this estimate as a local average treatment effect (LATE) for instrument-complying cities–those that experienced cholera outbreaks because of soldiers–which is not representative of the average city in our sample. Moving beyond the LATE interpretation to sample-wide average treatment effects (though with the limitations discussed in section 4.2), column 3 shows that the control-function estimate is closer to the OLS coefficient, but still 50% larger. This possibly reflects downward bias implied by measurement error, as we discussed earlier. We note that our instrument is strong by conventional criteria (see F and χ^2 statistics in columns 2 and 3, and first-stage regression results in appendix table A.2). Finally, the simulation-based reduced-form regression presented in column 4 yields a statistically significant and positive coefficient that is, again, similar in magnitude to that in column 1. In sum, these estimates further support the idea that councils, dominated by economic elites, resented the impact of cholera and demanded the establishment of waterworks.

5.2 Elites' interest in post-epidemic reform

This section explores the mechanisms through which cholera created political momentum to invest in public health infrastructures, presenting quantitative and qualitative evidence on the importance of the local decision-making elites' economic interests. As discussed in section 2.2, elites controlled local councils within a political system that effectively turned taxes into political power. We hypothesized that commercial

Table 3: The impact of cholera on the opening of waterworks, 1848-1891, causal estimates

	(1) OLS	(2) 2SLS	(3) CF	(4) Simulation
$C_i\mathbf{1}_{t>1866}$	0.3273**	4.5780***	0.4803**	0.3722***
	(0.013)	(0.000)	(0.017)	(0.007)
Instrument		$\Pr(C_i\|S_i)$	$\Pr(C_i\|S_i)$	(simulated)
Baseline controls and FEs	✓	✓	✓	✓
IV controls	✓	✓	✓	✓
First stage relevance		98.1 $[F]$	89.7 $[\chi^2]$	
Mean control (post-1866)	0.981	0.981	0.981	0.981
Cities	2125	2125	2124	2125
Interpretation	ATE	LATE	ATE	ATE

Note: This table shows the results of estimating the effect of the binary cholera shock on waterworks in different regressions that use exogenous variation from soldiers exposure. The dependent variable indicates whether a city operated waterworks in a given year. The specification in column 1 is akin to column 7 in table 2, with additional control variables: garrison status, capital city status, and local battle activity. Column 2 employs a two-stage-least squares (2SLS) framework with a linear first-stage; column 3 uses a control-function (CF) regression with a non-linear (Probit) first stage; and column 4 uses the implied marginal epidemiological effect of troop movements on cholera from a compartmental spatial model. See section 4.2 for further details on these three approaches. Numbers in parentheses are p-values based on standard errors clustered at the city and year levels (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. See appendix table A.2 for the control variables' coefficients, first-stage regressions, and additional regressions that use the soldier-presence dummy (S_i) rather than the outbreak probability predicted from it ($\Pr(C_i|S_i)$) as instruments.

elites incurred greater economic losses from labour scarcity than agricultural ones, since they owned more productive capital and employed a human-capital intensive workforce that was more expensive to replace. To test whether this translated into stronger willingness to pay for public health investments, we mean-split our city sample, as discussed in section 3.3, into commercial and non-commercial cities as well as literate and non-literate cities.

Table 4, provides evidence supporting the idea that commercial elites were more likely to push for public health investments. We begin with panel a, contrasting sanitary investment dynamics between commercial and non-commercial cities in the aftermath of the epidemic shock. Columns 1 and 2 show that our baseline results from table 2 (panel a, column 7) remain when we restrict our sample to Prussian cities and control for commercial cities' differential adoption dynamics.[20] In columns 3 to

[20]We control in these regressions for region-specific time trends of commercial cities due to their

7 we allow the impact of cholera to vary between commercial cities and those with predominantly agricultural elites. Column 3 shows that the effect for commercial and cholera-afflicted cities is sizeable, statistically significant and greater than that referring to non-commercial shocked locations (as formally tested through the one-sided t-test presented at the bottom of the table). Given that we control for broad socioeconomic characteristics, the differential effect of the epidemic is not driven by fiscal capacity, size or other factors specific to commercial cities other than their decision-making elites' composition. We provide further evidence to this point by controlling for various alternative channels through which the presence of commercial elites and the construction of public health infrastructures could be linked: i) protest potential (column 4), where elites exposed to greater pressure from the population may have been more likely to invest in public health (Krieger 2024); ii) industry demand (column 5), where cities with a greater presence of heavy and textile industry may have presented more demand for waterworks; iii) direct exposure to epidemics (column 6), where wealthy individuals living in more densely populated cities were more threatened by epidemics and may have demanded more public health infrastructures based on their own feeling of danger. The coefficient contrast between commercial and non-commercial cities remains when adding these confounding mechanisms individually, or simultaneously (column 7). These results also remain when we further interact the three alternative mechanisms' dummies with cholera exposure (see appendix, section A.2.3); and when we consider a balanced panel with more variation by not dropping cities from the sample when waterworks are built (see table A.9).

Table 4, panel b shows evidence consistent with the idea that elites in economies more reliant on skilled workers stood to suffer more from the cholera shock, possibly due to capital-skill complementarities in the production process (Galor and Moav 2006). Columns 2 and 3 repeat the exercise in panel a, but distinguish literate cities from those with a below-mean human capital level. Similar to our results in panel a, the point estimate of the interaction between cholera and the literate-city dummy is statistically significant and greater in magnitude than that for cholera-shocked

heterogeneous nature across Prussia, as these locations include economies reliant on heavy industry (e.g., steelworks), on less-capital intensive manufacturing (e.g., textiles or food processing) and on service-sector activities (e.g., trade or administration). See the discussion in appendix, section A.2.3.

less-literate cities. This result holds when controlling for other relevant mechanisms leading to waterworks construction as well as our baseline controls, including local economic development (columns 4 to 7); as for panel a, the results in panel b are robust to more demanding interaction effects (section A.2.3) and a balanced-panel setup (see table A.10).

The quantitative results presented so far lend broad support our earlier hypotheses: The cholera shock was more likely to trigger public health infrastructure investment in cities with a productive commercial-industrial capital stock and a more educated workforce, reflecting the economic interests of decision-making elites in local councils. Qualitative evidence also supports these ideas. In the aftermath of the epidemic, Hennock (2000, 280) describes how various associations were created with a specific focus on improving health matters, such as the Hygiene Section of the annual Congress of German Scientists and Physicians (1867); the Lower Rhenish Association for Public Health (1869); or the German Association for Public Health (1873). The ideas embodied in these associations reached economic and political circles with decision-making power, since their members often included city officials and council representatives (Lenger 2012, 154). Periodicals also contributed to the diffusion of knowledge on the health conditions of cities and the measures undertaken for their promotion. For instance, the early 1870s mark the beginning of systematic inquiries on the water supplies of cities in the journal of the German association of gas and water experts (DVGW 1870, 49–51).

Health-minded experts and activists often made economic arguments to convince local councils of the benefits of health, and waterworks in particular. To be sure, they also brought forward moral and compassion-based standpoints, but they were aware that moving to a political equilibrium with substantive public-health spending required involving local elites. For instance, Carl Heinrich Reclam, professor for hygiene in Leipzig, stated in the first page of the first volume of the German Public Health Quarterly (*Deutsche Vierteljahrsschrift für öffentliche Gesundheitspflege*) that: "The task of public health care is not to ensure the longevity or well-being of individual people, but rather to secure and increase the productivity of the entire population." He was not the only one who made this connection. James Hobrecht (1869, 70), a reputed Prussian urban planner at the time, stated that "nothing is as

Table 4: The impact of cholera on the opening of waterworks by elite background and human capital level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel a: Effect by local elites' economic background						
$C_i \mathbf{1}_{t>1866}$	0.472***	0.401**					
	(0.008)	(0.029)					
... × Commercial			0.992**	0.999**	0.963**	0.962**	0.950**
			(0.015)	(0.014)	(0.016)	(0.019)	(0.019)
... × Non-commercial			0.195	0.192	0.180	0.168	0.160
			(0.312)	(0.320)	(0.348)	(0.376)	(0.397)
	Panel b: Effect by workers' human capital level						
$C_i \mathbf{1}_{t>1866}$	0.472***	0.476**					
	(0.008)	(0.011)					
... × Literate			0.635***	0.634***	0.618***	0.594**	0.584**
			(0.009)	(0.009)	(0.010)	(0.013)	(0.014)
... × Non-literate			0.0677	0.0664	0.0712	0.00349	0.0137
			(0.709)	(0.718)	(0.694)	(0.984)	(0.939)
All main controls and FEs	✓	✓	✓	✓	✓	✓	✓
City type-specific trends		✓	✓	✓	✓	✓	✓
Protest potential				✓			✓
Industry demand					✓		✓
Direct exposure						✓	✓
p: Commercial > Non-commercial			0.033	0.031	0.035	0.034	0.033
p: Literate > Non-literate			0.022	0.022	0.024	0.019	0.021
Control mean (post-1866)	0.723	0.717	0.717	0.717	0.717	0.717	0.717
Cities	1426	1419	1419	1419	1419	1419	1419
R-squared in panel a	0.099	0.155	0.156	0.156	0.156	0.156	0.156
R-squared in panel b	0.099	0.110	0.110	0.110	0.110	0.110	0.110

Note: This table shows the main estimates from regressions that explore the mechanisms through which cholera exposure and sanitary infrastructure are linked. The dependent variable indicates whether a city opened waterworks in a given year. After opening waterworks, cities exit the sample in the following year. *Commercial*, *Literate*, and their counterparts are dummies that split the sample according to the presence of commercial elites and human capital in each city. See text for details on these definitions. Coefficients have been multiplied by 100, thus measuring percentage points. Regressions include all controls and fixed effects. Numbers in parentheses are p-values based on robust standard errors clustered at the city and year levels. The two additional p-values reported at the bottom of the table refer to one-sided coefficient difference tests. Significance stars follow the conventional coding (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$). See appendix tables A.3 and A.4 for the control variables' coefficients.

costly as illness" when discussing the decline in mortality brought about by sanitary infrastructures. Some authors referred to "sound economics" (*richtige Oekonomie*) when discussing the benefits of reducing preventable illnesses by reducing the loss of workers (Latham 1869, 171), or the amount of sick days (Pettenkofer 1873).

These ideas were not confined to specialized outlets, since already in 1869 various leading experts worked on a petition to the parliament of the North German Confederation (Göttisheim et al. 1869, 639–640). The aim was to reform the administrative organization of public health, including the creation of health committees in urban communities and rural districts applying the latest scientific knowledge. As one core motivation for this initiative, the petition mentions the cholera epidemics which have "harmed family happiness and national prosperity" (Göttisheim et al. 1870, 133). The report also highlights, on the same page, waterworks and sewerage systems in a number of cities as examples of how to improve the state of public health. It even warned: "The northern German states are still as unprepared for the possible outbreak of a new cholera epidemic as they were in 1866" (Göttisheim et al. 1870, 134).

This petition was signed by approximately 3,700 individuals by February 18 of 1870. This included individuals with different formal backgrounds, such as physicians, mayors of both large cities (e.g., Breslau, Frankfurt or Potsdam) and smaller municipalities (e.g., Koblenz , Stralsund, Köslin), members of municipal administrations and representatives. Also, commercial economic elites joined the petition either individually (e.g. merchants, factory owners, managers) or through their trade associations (e.g. in Bochum and Dresden). Even the influential "Lower Rhine Association for Public Health", based in Düsseldorf, which comprised about 1,500 members from 40 cities in the industrialized Lower Rhine region, joined the petition.

Some local public health histories highlight how cholera influenced debates around piped-water infrastructures. Consider the city of Halle, where discussions to improve the water supplies go as far back as the 1850s. This issue became particularly salient in 1866 when cholera killed more than three percent of the population in the city (see table 1). Besides public health considerations, a report by a commission to establish the need of waterworks claimed that this infrastructure would increase the opportunities for industrial enterprises to develop in the city (Fuchs 1910). In 1868, piped water became available. Similarly, the decision to build waterworks in Düsseldorf was made in 1866, coinciding with a decree from the the provincial administrative authority (*Königliche Regierung*) stating that this matter was urgent given that various municipal wells had to be closed due to the threat of cholera in 1866 (Most 1909). The system was finished in 1870. Other cities took certainly longer to improve water supplies due to, among others, financial constraints. However, cholera did bring to the fore the costs that deficiencies of pre-modern water supplies imposed on the local economy.

5.3 The long-run impact of cholera on public health and the economy

We conclude our analysis by exploring the effects of cholera and the sanitary investments it caused on various long-run outcomes around 1911, just before the large disruptions brought about by the First World War. This benchmark is analytically convenient because our sources allow us to measure the extent of within-city water provision and various economic and demographic dimensions at the city-level around this date.

We quantify the contribution of the 1866 cholera epidemic to the *cumulative* sanitary investments realized by 1911, both in terms of waterworks opening probabilities (the extensive margin) and the share of the supplied population (the intensive margin). Rather than simply regressing those 1911 outcomes on the cholera shock, we implement a multi-step estimation procedure consistent with the hazard framework of our baseline regression (equation 1). First, using the differential fitted annual opening hazards for cholera-shocked and spared cities between 1866 and 1911, we estimate the effect of the epidemic on the probability of a waterworks opening until 1911. Next, we estimate the *conditional* intensive margin of water supplies by regressing within-city coverage rates in supplied cities on the cholera shock, controlling for selection into the adoption sample (Heckman 1979). We then combine both margins to recover the effect of cholera on the *unconditional* intensive margin of water supplies. In the process, we quantify uncertainty through cluster-bootstrapped standard errors. Details on the approach are presented in appendix section A.2.5.

The results in table 5, panel a, show that by 1911, the epidemic had increased the probability of adoption by about 11 percentage points (first row), equivalent to roughly one-sixth of the mean adoption rate. At the same time, the within-city coverage rate among cities with waterworks (second row) was only 0.8 percentage points higher in cholera-affected cities–a small, statistically insignificant effect relative to the mean coverage share. If we consider all cities in our sample and combine the two margins (third row), we find that the 1866 epidemic shock raised within-city coverage rates by about 10 percentage points, about 17% of the 1911 mean coverage share. This gain was almost entirely due to the positive effect of cholera on opening probabilities, rather than the deepening of supply coverage within cities.

Next, we examine the second-order effects of the epidemic by linking development outcomes to the adoption of waterworks predicted by the cholera shock. In panel b, each outcome is regressed on the model-predicted probability that a city had waterworks in 1911 under cholera exposure, holding constant the same baseline controls and regional fixed effects as in our hazard model (see section A.2.5 for further details). This approach ensures that we capture the consequences of cholera-induced sanitary investments, rather than spurious long-run correlations driven by unobservable city characteristics. We find that the epidemic-driven acceleration of waterworks adoption was associated with reduced mortality, although the effect on infant mortality is not statistically significant.

We also find positive and statistically significant effects on human capital formation, economic activity and innovation, as measured by students in higher education, listed firms and patents filled, all in per-capita terms. We take these results as evidence that the 1866 cholera epidemic not only accelerated the diffusion of water-supply technology but *through these investments* had lasting effects on important development outcomes by the eve of the First World War. It seems that the elite's calculus was correct in the long run: expensive public health investments ended up paying off in the form of lower worker mortality, higher human capital investments, dynamic economies and innovation activity.

6 Conclusion

This paper explores the political-economy origins of health-enhancing public goods that, alongside education, energy and transport, transformed Western economies in

Table 5: The long-run development effects of the 1866 cholera epidemic

	Change in pp. (p-value)	Mean in 1911 (standard dev.)	N	R^2
Panel a: Cholera's effect on 1911 waterworks adoption				
Waterworks operation probability *(among all cities)*	11.440** (0.019)	69.984% (0.4584)	2,552	
Within-city coverage rate *(among cities with waterworks)*	0.812 (0.652)	80.753% (0.8319)	827	
Within-city coverage rate *(among all cities)*	9.806** (0.018)	56.514% (0.4341)	2,552	
Panel b: Cholera's second-order effect on ca. 1911 development indicators via waterworks adoption				
Crude death rate	-0.1926*** (0.000)	1.6891% (0.5337)	481	0.42
Infant mortality rate	-0.1351 (0.498)	17.0982% (4.9648)	481	0.60
Students in higher education	0.4168*** (0.000)	1.8900% (2.9179)	2,552	0.24
Listed firms	0.0010** (0.019)	0.0139% (0.0228)	2,552	0.12
Patents filed	0.0010*** (0.007)	0.0075% (0.0237)	2,552	0.09

Note: This table summarizes the estimated long-run effects of the 1866 cholera epidemic on sanitary outcomes in 1911 (panel a) and development indicators (panel b). See appendix section A.2.5 for details on the estimation. We observe the 1911 within-city water supply coverage share for 827 adopting cities, representing 46% of all cities with waterworks by then. In panel b, we report the percentage-point change in five outcomes associated with the average increase in the probability of having waterworks by 1911 that our model attributes to the 1866 cholera shock (about 11 percentage points, see first row of panel a). For crude death rates and infant mortality rates, our data is limited to 481 cities. P-values in parentheses are based on a cluster bootstrap with 500 repetitions. All regressions include the same baseline controls and fixed effects as our main regression (equation 1).

the 19th century. Our focus is the 1866 cholera epidemic in Germany, a demographic shock that caused the highest mortality rate in the modern history of the country. We test the hypothesis that concerns about workers' health–and by extension, returns to capital–compelled local decision-making elites to invest into a new and expensive health-enhancing infrastructure: modern waterworks. We argue that understanding elites' economic incentives is essential since, before the First World War, the German political system granted disproportionate political power to wealthy tax-paying individuals, whose interests effectively shaped public spending decisions.

Our first set of results establishes that the cholera epidemic altered the political equilibrium in Germany in favour of more health-related spending: the annual probability of opening waterworks up to 1891 was about 35% higher in cholera-affected cities relative to those spared. This finding is robust to, among many confounder controls and robustness tests, a causal identification approach relying on the cholera-spreading effect of military movements during the Austro-Prussian War in 1866. Exploiting this natural experiment, our identification strategy goes beyond standard 2SLS approaches by employing a calibrated compartmental model that isolates the exogenous contribution of soldiers to local cholera outbreaks.

Turning to the mechanisms linking cholera and public health investments, we provide quantitative evidence that cities dominated by commercial-industrial elites, rather than those with agricultural land owners, pushed for greater public health investments in the aftermath of cholera. In the wake of the epidemic, cities with elites owning relatively more productive capital were more likely to open waterworks than their more agricultural counterparts, and the same applies to locations employing workers with greater levels of human capital. These results align with the idea that industrial and commercial elites stood to loose more from the prospects of a labour shock than agricultural elites and thus pushed for reform. Qualitative evidence from health reformers' efforts underpins this interpretation of our results, as they appealed to economically-minded councils by arguing that healthier workers would increase economic productivity and prosperity.

We also examine the long-run impact of the epidemic and find that it contributed substantially to the improvement of sanitary conditions around 1911, just before the First World War. Our analysis suggests that, by then, cholera-shocked cities were 11 percentage points more likely to have adopted waterworks, but did not achieve more extensive within-city coverage of water supplies. In addition, we uncover considerable second-order effects, as the increased probability of building waterworks triggered by cholera was positively associated with better health, human capital in the form of education, economic performance and innovation. Elites ultimately reaped healthy economic returns from an expensive public health reform, largely funded with public resources. Waterworks played an indispensable role in the transformation of cities from high-mortality demographic sinks to places supporting ever-decreasing mortality.

References

Acemoglu, Daron, Camilo García-Jimeno, and James A. Robinson. 2015. "State Capacity and Economic Development: A Network Approach." *American Economic Review* 105 (8): 2364–2409.

Acemoglu, Daron, and James Robinson. 2000. "Why Did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective." *The Quarterly Journal of Economics* 115 (14): 1167–1199.

Adams, Renée, Heitor Almeida, and Daniel Ferreira. 2009. "Understanding the relationship between founder–CEOs and firm performance." *Journal of Empirical Finance* 16 (1): 136–150.

Ager, Philipp, Katherine Eriksson, Ezra Karger, Peter Nencka, and Melissa A. Thomasson. 2024. "School Closures during the 1918 Flu Pandemic." *The Review of Economics and Statistics* 106 (1): 266–276.

Aidt, Toke S., Martin Daunton, and Jayasri Dutta. 2010. "The Retrenchment Hypothesis and the Extension of the Franchise in England and Wales." *The Economic Journal* 120 (547): 990–1020.

Alfani, Guido. 2022. "Epidemics, Inequality, and Poverty in Preindustrial and Early Industrial Times." *Journal of Economic Literature* 60 (1): 3–40.

Alsan, Marcella, and Claudia Goldin. 2019. "Watersheds in Child Mortality: The Role of Effective Water and Sewerage Infrastructure, 1880-1920." *Journal of Political Economy* 127 (2): 586–638.

Anderson, D. Mark, Kerwin Kofi Charles, and Daniel I. Rees. 2022. "Re-Examining the Contribution of Public Health Efforts to the Decline in Urban Mortality." *American Economic Journal: Applied Economics* 14 (2): 126–157.

Andersson, David, Thor Berger, and Erik Prawitz. 2023. "Making a Market: Infrastructure, Integration, and the Rise of Innovation." *The Review of Economics and Statistics* 105 (2): 258–274.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton: Princeton University Press.

Anteneh, Leul Mekonnen, Bruno Enagnon Lokonon, and Romain Glèlè Kakaï. 2024. "Modelling techniques in cholera epidemiology: A systematic and critical review." *Mathematical Biosciences* 373:109210.

Antràs, Pol, Stephen J. Redding, and Esteban Rossi-Hansberg. 2023. "Globalization and Pandemics." *American Economic Review* 113 (4): 939–981.

Ashraf, Quamrul H., Francesco Cinnirella, Oded Galor, Boris Gershman, and Erik Hornung. 2024. "Structural Change, Elite Capitalism, and the Emergence of Labour Emancipation." *Review of Economic Studies,* 1–29.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103 (6): 2121–2168.

Avery, Christopher, William Bossert, Adam Clark, Glenn Ellison, and Sara Fisher Ellison. 2020. "An Economist's Guide to Epidemiology Models of Infectious Disease." *Journal of Economic Perspectives* 34 (4): 79–104.

Baker, Moses N. 1888. *The Manual of American Water-Works.* New York: Engineering News Press.

Baldwin, Peter. 1999. *Contagion and the State in Europe 1830–1930.* Cambridge: Cambridge University Press.

Banerjee, Abhijit, and Lakshmi Iyer. 2005. "History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India." *The American Economic Review* 95 (4): 1190–1213.

Barro, Robert J., José F. Ursúa, and Joanna Weng. 2022. "Macroeconomics of the Great Influenza Pandemic, 1918-1920." *Research in Economics* 76 (1): 21–29.

Basu, Anirban, Norma B. Coe, and Cole G. Chapman. 2018. "2SLS versus 2SRI: Appropriate methods for rare outcomes and/or rare exposures." *Health Economics* 27 (6): 937–955.

Beach, Brian, Joseph P. Ferrie, Martin Saavedra, and Werner Troesken. 2016. "Typhoid Fever, Water Quality, and Human Capital Formation." *The Journal of Economic History* 76 (1): 41–75.

Becker, Sascha O., and Francesco Cinnirella. 2020. "Prussia disaggregated: the demography of its universe of localities in 1871." *Journal of Demographic Economics* 86 (3): 259–290.

Becker, Sascha O., and Erik Hornung. 2021. "The Political Economy of the Prussian Three-class Franchise." *The Journal of Economic History* 80 (4): 1143–1188.

Bloom, David E., Michael Kuhn, and Klaus Prettner. 2022. "Modern Infectious Diseases: Macroeconomic Impacts and Policy Responses." *Journal of Economic Literature* 60 (1): 85–131.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2025. "Design-based Identification with Formula Instruments: a Review." *The Econometrics Journal* 28 (1): 83–108.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2024. "Revisiting Event-Study Designs: Robust and Efficient Estimation." *The Review of Economic Studies* 91 (6): 3253–3285.

Bosshart, Luis, and Matthias Weigand. 2025. "Crisis, State Capacity, and the Rise of Autocracy. A Design-based Analysis of the Thirty Years' War." Mimeo.

British Medical Journal. 1867. "Notes of the Cholera Conference at Weimar, on April 28th and 29th, 1867." *The British Medical Journal* 1 (335): 625–626.

Brown, John C. 1988. "Coping with Crisis? The Diffusion of Waterworks in Late Nineteenth-Century German Towns." *The Journal of Economic History* 48 (2): 307–318.

Cantoni, Davide, Cathrin Mohr, and Matthias Weigand. 2024. "The Rise of Fiscal Capacity: Administration and State Consolidation in the Holy Roman Empire." *Econometrica* 92 (5): 1439–1472.

Chapman, Jonathan. 2022. "Interest Rates, Sanitation Infrastructure, and Mortality Decline in Nineteenth-Century England and Wales." *The Journal of Economic History* 82 (1): 175–210.

———. 2024. "Gradual Franchise Extensions and Government Spending in Nineteenth-Century England." *Journal of Politics* 86 (1): 369–374.

Chetty, Raj, John N. Friedman, Michael Stepner, and Opportunity Insights Team. 2023. "The Economic Impacts of Covid-19: Evidence from a New Public Database Built Using Private Sector Data." *The Quarterly Journal of Economics* 139 (2): 829–889.

Cholera-Kommission für das Deutsche Reich. 1876. *Untersuchungsplan der Erforschung der Ursachen der Cholera und deren Verhütung.* Berlin: Carl Heymann.

Conley, Timothy G. 1999. "GMM estimation with cross sectional dependence." *Journal of Econometrics* 92 (1): 1–45.

Correia, Sergio, Stephan Luck, and Emil Verner. 2022. "Pandemics Depress the Economy, Public Health Interventions Do Not: Evidence from the 1918 Flu." *The Journal of Economic History* 82 (4): 917–957.

Costa, Dora L. 2015. "Health and the Economy in the United States from 1750 to the Present." *Journal of Economic Literature* 53 (3): 503–570.

Cutler, David, and Grant Miller. 2006. "Water, Water Everywhere: Municipal Finance and Water Supply in American Cities." In *Corruption and Reform: Lessons from America's Economic History,* edited by Edward L. Glaeser and Claudia Goldin, 153–183. Chicago: University of Chicago Press.

Delbrück, Ernst Friedrich Alexander. 1867. *Bericht über die Cholera-Epidemie des Jahres 1866 in Halle, in der Strafanstalt zu Halle und im Saalkreise.* Halle: C. E. M. Pfeffer.

Deutscher Verein von Gas- und Wasserfachmännern. various years. *Journal für Gasbeleuchtung und verwandte Beleuchtungsarten, sowie für Wasserversorgung.* München: R. Oldenbourg.

Dittmar, Jeremiah E., and Ralf R. Meisenzahl. 2020. "Public Goods Institutions, Human Capital, and Growth: Evidence from German History." *The Review of Economic Studies* 87 (2): 959–996.

Donaldson, Dave, and Richard Hornbeck. 2016. "Railroads and American Economic Growth: A "Market Access" Approach." *The Quarterly Journal of Economics* 131 (2): 799–858.

Esteves, Rui, Kris James Mitchener, Peter Nencka, and Melissa A. Thomasson. 2022. *Do Pandemics Change Healthcare? Evidence from the Great Influenza.* NBER Working Paper No. 30643.

Evans, Richard J. 1988. "Epidemics and Revolutions: Cholera in Nineteenth-Century Europe." *Past & Present* 120:123–146.

Feigenbaum, James, James Lee, and Filippo Mezzanotti. 2022. "Capital Destruction and Economic Growth: The Effects of Sherman's March, 1850–1920." *American Economic Journal: Applied Economics* 14 (4): 301–342.

Fogel, Robert W. 1994. "Economic Growth, Population Theory and Physiology: The Bearing of Long-Term Process on the Making of Economic Policy." *American Economic Review* 84 (3): 369–395.

Franck, Raphaël. 2024. "Labor scarcity, technology adoption and innovation: evidence from the cholera pandemics in 19th century France." *Journal of Economic Growth* 29:543–583.

Fuchs, Carl Johannes. 1910. *Die Gemeindebetriebe der Stadt Halle a.S.* Gemeindebetriebe – Neuere Versuche und Erfahrungen über die Ausdehnung der kommunalen Tätigkeit in Deutschland und im Ausland. Zweiter Band, achter Teil, Schriften des Vereins für Socialpolitik, No. 129/VIII. Berlin: Duncker & Humblot.

Gallardo-Albarrán, Daniel. 2020. "Sanitary Infrastructures and the Decline of Mortality in Germany, 1877-1913." *The Economic History Review* 73 (3): 730–757.

———. 2025. "The Global Sanitary Revolution in Historical Perspective." *The Journal of Economic Surveys* 39 (2): 567–598.

Gallardo-Albarrán, Daniel, and Kalle Kappner. 2025. "Epidemics and urban growth: the impact of cholera in Nineteenth-century Prussia." *The History of the Family* 30 (2): 468–499.

Galor, Oded, and Omer Moav. 2006. "Das Human-Kapital: A Theory of the Demise of the Class Structure." *The Review of Economic Studies* 73 (1): 85–117.

Galor, Oded, Omer Moav, and Dietrich Vollrath. 2009. "Inequality in Landownership, the Emergence of Human-Capital Promoting Institutions, and the Great Divergence." *Review of Economic Studies* 76 (1): 143–179.

Goeden, Adolph Ludwig Friedrich. 1867. *Über die Cholera-Epidemie in Stettin im Jahre 1866.* Stettin: Otto Spaethen.

Göttisheim, Hobrecht, Reclaim, Varrentrapp, and Wasserfurh. 1869. *Deutsche Vierteljahrsschrift für öffentliche Gesundheitspflege.* Braunschweig: Friedrich Vieweg / Son. Edited Book.

———. 1870. *Deutsche Vierteljahrsschrift für öffentliche Gesundheitspflege.* Vol. 2. Braunschweig: Friedrich Vieweg / Son.

Grahn, Ernst. 1883. *Die Art der Wasserversorgung der Städte des Deutschen Reiches mit mehr als 5000 Einwohnern.* Munich and Leipzig: Druck und Verlag von R. Oldenbourg.

———. 1904. "Die städtischen Wasserwerke." In *Die deutschen Städte. Geschildert nach dem Ergebnissen der ersten deutschen Städteausstellung zu Dresden 1903,* edited by R. Wuttke, 301–344. Leipzig: Friedrich Brandstetter.

———. 1898-1902. *Die städtische Wasserversorgung im Deutschen Reiche sowie in einigen Nachbarländern.* Munich and Leipzig: R. Oldenbourg.

Grätzer, Jonas. 1867. *Ueber die öffentliche Armen-Krankenpflege Breslau's [1866 and 1867].* Breslau: Max und Komp.

Guttstadt, Albert. 1906. "Die Choleraepidemien in früherer Zeit." *Hygienische Rundschau* 16 (5): 265–280.

Hamlin, Christopher. 2009. "'Cholera forcing'. The Myth of the Good Epidemic and the Coming of Good Water." *American Journal of Public Health* 99 (11): 1946–1954.

Harper, Kyle. 2021. *Plagues Upon the Earth. Disease and the Course of Human History.* Princeton: Princeton University Press.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1): 153–161.

Hennock, Ernest Peter. 2000. "The Urban Sanitary Movement in England and Germany, 1838-1914: a Comparison." *Continuity and Change* 15 (2): 269–296.

Hirsch, August. 1881. "Die allgemeinen acuten Infectionskrankheiten vom historisch-geographischen Standpunkte und mit besonderer Berücksichtigung der Aetiologie." In *Handbuch der historisch-geographischen Pathologie,* vol. 1. Stuttgart: Ferdinand Enke.

Hobrecht, James. 1869. "Ueber Reinigung und landwirthschaftliche Nutzbarmachung des Kanalwassers." In *Deutsche Vierteljahrsschrift für öffentliche Gesundheitspflege,* edited by Göttisheim, Hobrecht, Reclaim, Varrentrapp, and Wasserfurh, 1:65–84. Braunschweig: Friedrich Vieweg / Son.

Hollenbach, Florian M. 2021. "Elite Interests and Public Spending: Evidence from Prussian Cities." *The Review of International Organizations* 16 (1): 189–211.

Jordà, Òscar, Sanjay R. Singh, and Alan M. Taylor. 2022. "Longer-Run Economic Consequences of Pandemics." *The Review of Economics and Statistics* 104 (1).

Kappner, Kalle. 2024. "Did Cholera 'Force' the Reform of Urban Water Infrastructure? Lessons from Nineteenth-Century Berlin." *Journal of Urban History* 51 (3): 545–566.

———. 2025. "The Mortality Impact of Cholera in Germany." *Jahrbuch für Wirtschaftsgeschichte [Economic History Yearbook]* 66 (2): 365–403.

Keiser, David A., and Joseph S. Shapiro. 2019. "US Water Pollution Regulation over the Past Half Century: Burning Waters to Crystal Springs?" *Journal of Economic Perspectives* 33 (4): 51–75.

Knutsson, Daniel. 2025. "The effect of water filtration on cholera mortality." *Journal of Urban Economics* 147:103752.

Königlich Preussisches Statistisches Landesamt in Berlin. 1914-1920. *Preussische Statistik.* Vol. 243: Finanzstatistik der preusischen Städte und Landgemeinden für das Rechnungsjahr 1911. Berlin: Verlag des Königlichen Statistischen Landesamts.

Königlich Sächsisches Statistisches Landesamt. 1866. *Zeitschrift des Statistischen Bueareaus des Königl. Sächs. Ministeriums des Innern.* Vol. 12. Dresden: B. G. Teubner.

Königliches Statistisches Bureau in Berlin. 1867. *Die Ergebnisse der Volkszählung und Volksbeschreibung, der Gebäude- und Viehzählung, nach den Aufnahmen vom 3. December 1864, resp. Anfang 1865.* Vol. X. Preussische Statistik. Amtliches Quellenwerk. Berlin: Ernst Kuehn.

———. 1870. *Die Bewegung der Bevölkerung des Preussischen Staats in den Jahren 1865, 1866 und 1867.* Vol. XVII. Preussische Statistik. Berlin: Verlag des Königlichen Statistischen Bureaus (Dr. Engel).

Krabbe, Wolfgang R. 1989. *Die deutsche Stadt im 19. und 20. Jahrhundert: Eine Einführung.* Göttingen: Vandenhoeck & Ruprecht.

Kraus, Antje. 1980. *Quellen zur Bevölkerungsstatistik Deutschlands, 1815-1875.* Edited by Wolfgang Köllmann. Quellen zur Bevölkerungs-, Sozial- und Wirtschaftsstatistik Deutschlands 1815-1875. Boppard am Rhein: Harald Boldt.

Krehnke, Walter. 1937. "Der Gang der Cholera in Deutschland seit ihrem ersten Auftreten bis heute." *Veröffentlichungen aus dem Gebiete des Volksgesundheitsdienstes* 49 (2): 329–446.

Krieger, Tommy. 2024. "Elites and Health Infrastructure Improvements in Industrializing Regimes." *Journal of Economic Growth* 29:433–468.

Latham, Baldwin. 1869. "Die Kanalisation Danzigs. 1. Gutachten über die Kanalisation von Danzig und das Wiebe'sche Kanalisirungsproject." In *Deutsche Vierteljahrsschrift für öffentliche Gesundheitspflege,* edited by Göttisheim, Hobrecht, Reclaim, Varrentrapp, and Wasserfuhr, 1:168–181. Braunschweig: Friedrich Vieweg / Son.

Lenger, Friedrich. 2012. *European Cities in the Modern Era, 1850-1914.* Leiden and Boston: Brill.

Lent, Eduard. 1868. *Bericht über die zweite Cholera-Epidemie des Jahres 1867 in Köln.* Köln: DuMont-Schauberg.

Liczbińska, Grażyna. 2021. "Spatial and Social Inequalities in the Face of Death: Pilot Research on Cholera Epidemics in Poznań of the Second Half of the 19th Century." *Historical Social Research* Suppl. 33:37–53.

Lizzeri, Alessandro, and Nicola Persico. 2004. "Why did the Elites Extend the Suffrage? Democracy and the Scope of Government, with an Application to Britain's 'Age of Reform'." *The Quarterly Journal of Economics* 119 (2): 707–765.

Loeffler, Gottfried Friedrich Franz. 1869. *Das Preussische Militär-Sanitätswesen und seine Reform nach der Kriegserfahrung von 1866.* Vol. 2. Berlin: August Hirschwald.

Marczinek, Max. 2025. "Labour Scarcity and Productivity: Insights from the Last Nordic Plague." Mimeo.

McNeill, William H. 1976. *Plagues and People.* New York: Anchor Press.

Michalopoulos, Stelios, and Elias Papaioannou. 2016. "The Long-Run Effects of the Scramble for Africa." *American Economic Review* 106 (7): 1802–1848.

Miller, Grant. 2008. "Women's Suffrage, Political Responsiveness, and Child Survival in American History." *The Quarterly Journal of Economics* 123 (3): 1287–1327.

Mokyr, Joel, and Rebecca Stein. 1996. "Science, Health, and Household Technology: The Effect of the Pasteur Revolution on Consumer Demand." In *The Economics of New Goods,* edited by Timothy F. Bresnahan and Robert J. Gordon, 143–206. Chicago: University of Chicago Press.

Most, Otto. 1909. *Die Gemeindebetriebe der Stadt Düsseldorf.* Gemeindebetriebe – Neuere Versuche und Erfahrungen über die Ausdehnung der kommunalen Tätigkeit in Deutschland und im Ausland. II. Band, 2. Teil. Berlin: Duncker & Humblot.

Müller, Eduard H. 1867. *Die Cholera-Epidemie zu Berlin im Jahre 1866.* Berlin: Th. Chr. Fr. Enslin.

Nunn, Nathan, and Leonard Wantchekon. 2011. "The Slave Trade and the Origins of Mistrust in Africa." *American Economic Review* 101 (7): 3221–3252.

Olson, Mancur. 1965. *The Logic of Collective Action.* Cambridge, MA: Harvard University Press.

Pettenkofer, Max von. 1873. *Über den Werth der Gesundheit für eine Stadt.* Braunschweig: Vieweg.

Pfeiffer, Ludwig. 1871. *Die Cholera in Thüringen und Sachsen während der dritten Cholerainvasion 1865-1867.* Jena: Friedrich Mauke.

Pfister, Ulrich, and Georg Fertig. 2020. "From Malthusian Disequilibrium to the Post-Malthusian Era: The Evolution of the Preventive and Positive Checks in Germany, 1730-1870." *Demography* 57 (3): 1145–1170.

Prinzing, Friedrich. 1916. *Epidemics Resulting from Wars.* Oxford: Oxford University Press.

Rahlf, Thomas. 2016. "The German Time Series Dataset, 1834–2012." *Journal of Economics and Statistics* 236 (1): 129–143.

Reulecke, Jürgen. 1985. *Geschichte der Urbanisierung in Deutschland.* Frankfurt am Main: Suhrkamp.

Rosen, George. 1993. *The History of Public Health.* Expanded edition. Baltimore and London: The Johns Hopkins University Press (Original work published in 1958).

Rosenberg, Charles. 1966. "Cholera in Nineteenth-Century Europe: A Tool for Social and Economic Analysis." *Comparative Studies in Society and History* 8 (4): 452–463.

Salewski, Michael. 2003. "Vom Kabinettskrieg zum totalen Krieg: Der Gestaltwandel des Krieges im 19. und 20. Jahrhundert." In *Masse und Macht im 19. und 20. Jahrhundert,* edited by Ulrich Lappenküper, Joachim Scholtyseck, and Christoph Studt, 51–66. München: Oldenbourg.

Salomon, Hermann. 1906-1911. *Die städtische Abwässerbeseitigung in Deutschland. Wörterbuchartig angeordnete Nachrichten und Beschreibungen städtischer Kanalisations- und Kläranlagen in deutschen Wohnplätzen (Abwässer-Lexikon).* Jena: Gustav Fischer.

Sander, Friedrich. 1872. *Untersuchungen über die Cholera in ihren Beziehungen zu Boden und Grundwasser, zu socialen und Bevölkerungs-Verhältnissen, so wie zu den Aufgaben der öffentlichen Gesundheitspflege.* Köln: DuMont-Schauberg.

Schiefferdecker, Wilhelm. 1868. *Die Choleraepidemien in Königsberg.* Königsberg: Pfitzer & Heilmann.

Schmieder, Cölestin. 1867. *Die ärztlich constatirten Cholera-Todesfälle der Stadt Leipzig im Jahr 1866.* Leipzig: J. C. Hinrich.

Schmitz-Cliever, Egon. 1952. "Die Anschauungen vom Wesen der Cholera bei den Aachener Epidemien 1832-1866." *Sudhoffs Archiv für Geschichte der Medizin und der Naturwissenschaften* 36 (4): 261–276.

Semrad, Alexandra. 2015. "Modern secondary education and economic performance: the introduction of the Gewerbeschule and Realschule in nineteenth-century Bavaria." *The Economic History Review* 68 (4): 1306–1338.

Silbergleit, Heinrich. 1908. *Preussens Städte. Denkschrift zum 100 jährigen Jubiläum der Städteordnung von 19. November 1808.* Berlin: Carl Heymanns Verlag.

Sokoloff, Kenneth L., and Stanley L. Engerman. 2000. "Institutions, Factor Endowments, and Paths of Development in the New World." *Journal of Economic Perspectives* 14 (3): 217–232.

Szreter, Simon. 1988. "The Importance of Social Intervention in Britain's Mortality Decline c. 1850-1914: a Re-Interpretration of the Role of Public Health." *Social History of Medicine* 1 (1): 1–38.

Tilly, Richard H. 1990. *Sozialer Protest in Deutschland im 19. Jahrhundert.* GESIS Datenarchiv, Köln. ZA8040 Data file version 1.0.0. https://doi.org/10.4232/1.8040.

———. 1995. "Städtewachstum, Kommunalfinanzen und Munizipalsozialismus in der deutschen Industrialisierung: eine vergleichende Perspektive 1870-1913." In *Die Stadt als Dienstleistungszentrum. Beiträge zur Geschichte der 'Sozialstadt' in Deutschland im 19. und frühen 20. Jahrhundert,* edited by Jürgen Reulecke, 125–152. St. Katharinen: Scripta Mercaturae Verlag.

Tur-Prats, Ana, and Felipe Valencia Caicedo. 2025. "The Long Shadow of the Spanish Civil War." Mimeo.

Verein für Gas- und Wasserfachmänner Deutschlands. 1870. *Journal für Gasbeleuchtung und verwandte Beleuchtungsarten sowie für Wasserversorgung.* Vol. 13. Munich: R. Oldenbourg.

Voigtländer, Nico, and Hans-Joachim Voth. 2013. "The Three Horsemen of Riches: Plague, War, and Urbanization in Early Modern Europe." *The Review of Economic Studies* 80 (2): 774–811.

Wang, Jin. 2002. "Mathematical Models for Cholera Dynamics—A Review." *Microorganisms* 10 (12): 2358.

Weil, David N. 2007. "Accounting for the Effect of Health on Economic Growth." *The Quarterly Journal of Economics* 122 (3): 1265–1306.

Wooldridge, Jeffrey Marc. 2010. *Econometric analysis of cross section and panel data.* Cambridge, MA: MIT press.

Xu, Ruonan. 2021. "On the instrument functional form with a binary endogenous explanatory variable." *Economics Letters* 206:109993.

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

Zimran, Ariell. 2020. "Transportation and Health in the Antebellum United States, 1820-1847." *The Journal of Economic History* 80 (3): 670–709.

A Complementary material

A.1 Full regression output tables

Table A.1: The impact of cholera on the opening of waterworks, 1849-1891, full regression output

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel a: $C_i \equiv$ **Presence of cholera (0/1)**							
$C_i \mathbf{1}_{t>1866}$	0.0109***	0.0110***	0.0052***	0.0109***	0.0072***	0.0104***	0.0039***
	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)	(0.006)
(Hydro-)geology controls: $\mathbf{1}_{t>1866} \times$ ...							
... Elevation (km)		-0.0140					-0.0075
		(0.235)					(0.509)
... Slope (ratio)		-0.0126					0.0083
		(0.553)					(0.674)
... Groundwater productivity (share)		0.0019					0.0004
		(0.455)					(0.862)
... Groundwater salt intrusion (share)		0.0066					0.0100**
		(0.134)					(0.035)
... Distance to fresh water (10 kms)		0.0001					0.0003
		(0.878)					(0.422)
... Monthly precipitation (cm)		0.0035***					0.0033***
		(0.003)					(0.005)
Socio-economic controls: $\mathbf{1}_{t>1866} \times$ ...							
... Population (10^5)			0.0015***				0.0013***
			(0.000)				(0.000)
... All-cause death rate (%)				0.0003			0.0007*
				(0.507)			(0.095)
... Economic development (0-1 index)					0.0391***		0.0226***
					(0.000)		(0.000)
... Market access (10^6)						0.0005***	0.0003***
						(0.001)	(0.010)
Constant	0.0045***	-0.0057*	0.0025***	0.0041***	0.0006	0.0041***	-0.0115***
	(0.000)	(0.088)	(0.001)	(0.000)	(0.468)	(0.000)	(0.004)
Panel b: $C_i \equiv$ **Cholera deaths per 100**							
$C_i \mathbf{1}_{t>1866}$	0.0010**	0.0010**	0.0006	0.0010**	0.0007	0.0009**	0.0005
	(0.042)	(0.037)	(0.175)	(0.042)	(0.113)	(0.047)	(0.265)
(Hydro-)geology controls: $\mathbf{1}_{t>1866} \times$ ...							
... Elevation (km)		-0.0178					-0.0086
		(0.145)					(0.462)
... Slope (ratio)		-0.0045					0.0112
		(0.838)					(0.583)
... Groundwater productivity (share)		0.0019					0.0004
		(0.456)					(0.874)
... Groundwater salt intrusion (share)		0.0067					0.0101**
		(0.118)					(0.032)
... Distance to fresh water (10 kms)		-0.0001					0.0002
		(0.719)					(0.539)
... Monthly precipitation (cm)		0.0035***					0.0032***
		(0.003)					(0.005)
Socio-economic controls: $\mathbf{1}_{t>1866} \times$ ...							
... Population (10^5)			0.0016***				0.0013***
			(0.000)				(0.000)
... All-cause death rate (%)				0.0002			0.0007
				(0.574)			(0.102)
... Economic development (0-1 index)					0.0419***		0.0237***
					(0.000)		(0.000)
... Market access (10^6)						0.0005***	0.0003***
						(0.000)	(0.010)
Constant	0.0063***	-0.0031	0.0033***	0.0060***	0.0014**	0.0059***	-0.0109***
	(0.000)	(0.320)	(0.000)	(0.000)	(0.025)	(0.000)	(0.006)
City and region-year FEs	✓	✓	✓	✓	✓	✓	✓
Control mean (post-1866)	0.011	0.011	0.011	0.011	0.011	0.011	0.011
Cities	2552	2552	2552	2552	2552	2552	2552
R-squared in panel a	0.091	0.090	0.094	0.090	0.091	0.090	0.095
R-squared in panel b	0.090	0.089	0.094	0.089	0.091	0.090	0.095

Note: This table presents the full regression output for table 2 in the main text. See that table's note for further details.

Table A.2: The impact of cholera on the opening of waterworks, 1849-1891, causal estimates, full regression output

	Reduced-form		Instrument: $\Pr(C_i\|S_i)$				Instrument: S_i			
	(1) OLS	(2) Simulation	(3) 2SLS	(4) FS	(5) CF	(6) FS (Probit)	(7) 2SLS	(8) FS	(9) CF	(10) FS (Probit)
Main effects: $\mathbf{1}_{t>1866}\times$ ...										
C_i	0.0033**	0.0037***	0.0458***		0.0048**		0.0569***		0.0043**	
	(0.013)	(0.007)	(0.000)		(0.017)		(0.000)		(0.012)	
Instrument				2.3094***		9.6165***		0.1680***		0.6428***
				(0.000)		(0.000)		(0.000)		(0.000)
CF residual					-0.0002				-0.0002*	
					(0.153)				(0.058)	
(Hydro-)geology controls: $\mathbf{1}_{t>1866}\times$ ...										
... Elevation (km)	-0.0253*	-0.0270**	-0.0045	-0.4174***	-0.0241*	-1.8573***	0.0009	-0.4117***	-0.0252*	-1.8948***
	(0.055)	(0.044)	(0.742)	(0.006)	(0.067)	(0.003)	(0.954)	(0.008)	(0.056)	(0.002)
... Slope (ratio)	0.0334	0.0419	-0.0341	1.4226***	0.0302	7.0010***	-0.0517	1.4402***	0.0330	6.9725***
	(0.174)	(0.101)	(0.332)	(0.001)	(0.216)	(0.000)	(0.228)	(0.001)	(0.181)	(0.000)
... Groundwater productivity (share)	-0.0009	-0.0006	0.0000	-0.0073	-0.0008	-0.0315	0.0003	-0.0226	-0.0006	-0.0817
	(0.680)	(0.784)	(0.995)	(0.828)	(0.719)	(0.814)	(0.926)	(0.507)	(0.779)	(0.539)
... Groundwater salt intrusion (share)	0.0098*	0.0107**	0.0075	0.0940	0.0096*	0.2997	0.0069	0.0626	0.0100**	0.2278
	(0.050)	(0.034)	(0.270)	(0.383)	(0.053)	(0.578)	(0.362)	(0.546)	(0.049)	(0.669)
... Distance to fresh water (10 kms)	0.0004	0.0005	0.0007*	-0.0066	0.0005	-0.0421*	0.0008*	-0.0048	0.0005	-0.0324
	(0.142)	(0.102)	(0.077)	(0.248)	(0.131)	(0.066)	(0.080)	(0.404)	(0.122)	(0.153)
... Monthly precipitation (cm)	0.0042***	0.0042***	0.0058***	-0.0368***	0.0043***	-0.2240***	0.0062***	-0.0401***	0.0044***	-0.2188***
	(0.001)	(0.001)	(0.001)	(0.008)	(0.001)	(0.001)	(0.000)	(0.004)	(0.000)	(0.002)
Socio-economic controls: $\mathbf{1}_{t>1866}\times$ ...										
... Population (10^5)	0.0016***	0.0016***	0.0012***	0.0008	0.0016***	0.0077	0.0011***	0.0075***	0.0020***	0.0594***
	(0.000)	(0.000)	(0.000)	(0.518)	(0.000)	(0.327)	(0.000)	(0.000)	(0.000)	(0.000)
... All-cause death rate (%)	0.0007	0.0007	0.0006	0.0033	0.0007	0.0145	0.0006	0.0023	0.0007	0.0081
	(0.131)	(0.135)	(0.360)	(0.765)	(0.132)	(0.701)	(0.445)	(0.841)	(0.118)	(0.826)
... Economic development (0-1 index)	0.0171***	0.0183***	0.0043	0.0693	0.0168***	0.1024	0.0009	0.2644***	0.0134***	0.7435***
	(0.000)	(0.000)	(0.328)	(0.221)	(0.000)	(0.648)	(0.885)	(0.000)	(0.001)	(0.000)
... Market access (10^6)	0.0004***	0.0004***	0.0004**	0.0003	0.0004***	-0.0013	0.0004**	0.0004	0.0004**	-0.0011
	(0.004)	(0.004)	(0.016)	(0.807)	(0.004)	(0.740)	(0.020)	(0.761)	(0.014)	(0.782)
Instrumental variable controls: $\mathbf{1}_{t>1866}\times$ ...										
... Garrison city (dummy)	0.0013	0.0015	-0.0011	-0.0344	0.0012	-0.1933	-0.0018	-0.0116	-0.0011	-0.1513
	(0.552)	(0.488)	(0.622)	(0.289)	(0.583)	(0.123)	(0.486)	(0.724)	(0.549)	(0.233)
... Capital city (dummy)	0.0153	0.0155	0.0100	0.1151	0.0149	0.3143	0.0086	0.0825	0.0125	0.0603
	(0.339)	(0.331)	(0.530)	(0.309)	(0.350)	(0.370)	(0.603)	(0.460)	(0.442)	(0.874)
... Battle in city (dummy)	0.0257	0.0260	0.0280	-0.1461	0.0256	-0.6870	0.0286	-0.1453	0.0257	-0.7623
	(0.397)	(0.388)	(0.392)	(0.308)	(0.399)	(0.297)	(0.391)	(0.286)	(0.396)	(0.280)
Constant						-2.8106***				0.7750
						(0.000)				(0.182)
Baseline controls and FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
IV controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
First stage relevance				98.1 [F]		89.7 [χ^2]		44.1 [F]		59.1 [χ^2]
Mean control (post-1866)	0.010	0.010	0.010	0.326	0.010	0.304	0.010	0.326	0.010	0.304
Cities	2125	2125	2125	2125	2124	2238	2125	2125	2115	2238

Note: This table presents the full regression output for table 3 in the main text. See that table's note for further details. Columns 1, 2, 3 and 5 reproduce the results shown in table 3. Columns 4 and 6 show the respective first stages, regressing C_i on the soldiers-exposure predicted cholera outbreak probability ($\Pr(C_i|S_i)$). Note that column 6 is a Probit (non-linear) first stage; thus, the regression coefficients are not marginal effects and we present a χ^2 value instead of the conventional first-stage F statistic for the excluded instrument. In column 5, we also show the (insignificant) effect of the first-stage residual. Columns 7-10 refer to the same models as columns 3-6, but we use a dummy for soldiers exposure (S_i) as the excluded instrument.

Table A.3: The impact of cholera on the opening of waterworks by elite background, full regression output

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$C_i \mathbf{1}_{t>1866}$	0.0047***	0.0040**					
	(0.008)	(0.029)					
... × Commercial			0.0099**	0.0100**	0.0096**	0.0096**	0.0095**
			(0.015)	(0.014)	(0.016)	(0.019)	(0.019)
... × Non-commercial			0.0019	0.0019	0.0018	0.0017	0.0016
			(0.312)	(0.320)	(0.348)	(0.376)	(0.397)
(Hydro-)geology controls: $\mathbf{1}_{t>1866}$ × ...							
... Elevation (km)	-0.0414***	-0.0394**	-0.0398**	-0.0406**	-0.0392**	-0.0393**	-0.0396**
	(0.007)	(0.015)	(0.014)	(0.012)	(0.016)	(0.015)	(0.014)
... Slope (ratio)	0.0496	0.0448	0.0448	0.0457	0.0442	0.0443	0.0448
	(0.182)	(0.245)	(0.244)	(0.239)	(0.255)	(0.250)	(0.253)
... Groundwater productivity (share)	-0.0010	0.0000	0.0001	-0.0001	0.0002	0.0001	0.0000
	(0.611)	(0.991)	(0.967)	(0.968)	(0.931)	(0.975)	(0.998)
... Groundwater salt intrusion (share)	0.0101*	0.0122**	0.0123**	0.0122**	0.0116**	0.0128**	0.0119**
	(0.084)	(0.037)	(0.041)	(0.039)	(0.049)	(0.035)	(0.041)
... Distance to fresh water (10 kms)	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005
	(0.149)	(0.165)	(0.175)	(0.164)	(0.136)	(0.190)	(0.139)
... Monthly precipitation (cm)	0.0039**	0.0031*	0.0032*	0.0032*	0.0032*	0.0031*	0.0032*
	(0.023)	(0.078)	(0.072)	(0.065)	(0.071)	(0.079)	(0.069)
Socio-economic controls: $\mathbf{1}_{t>1866}$ × ...							
... Population (10^5)	0.0017***	0.0017***	0.0017***	0.0016***	0.0017***	0.0017***	0.0016***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
... All-cause death rate (%)	0.0009*	0.0007	0.0008	0.0008	0.0007	0.0007	0.0007
	(0.063)	(0.191)	(0.164)	(0.160)	(0.175)	(0.173)	(0.176)
... Economic development (0-1 index)	0.0088*	0.0068	0.0067	0.0059	0.0066	0.0059	0.0053
	(0.096)	(0.208)	(0.213)	(0.263)	(0.220)	(0.283)	(0.325)
... Market access (10^6)	0.0005***	0.0004**	0.0004**	0.0004**	0.0004**	0.0004**	0.0004**
	(0.008)	(0.025)	(0.030)	(0.027)	(0.031)	(0.036)	(0.032)
Mechanism controls: $\mathbf{1}_{t>1866}$ × ...							
... Protest (dummy)				0.0051**			0.0046*
				(0.025)			(0.053)
... Industry demand (dummy)					-0.0028**		-0.0025*
					(0.039)		(0.073)
... High density (dummy)						0.0021	0.0015
						(0.167)	(0.379)
Constant	-0.0140**	-0.0108*	-0.0111*	-0.0112*	-0.0106*	-0.0110*	-0.0107*
	(0.025)	(0.072)	(0.066)	(0.062)	(0.077)	(0.068)	(0.073)
City and region-year FEs	✓	✓	✓	✓	✓	✓	✓
City type-specific trends		✓	✓	✓	✓	✓	✓
p: Commercial > Non-commercial			0.033	0.031	0.035	0.034	0.033
Control mean (post-1866)	0.007	0.007	0.007	0.007	0.007	0.007	0.007
Cities	1426	1419	1419	1419	1419	1419	1419
R-squared	0.050	0.084	0.084	0.085	0.085	0.084	0.085

Note: This table presents the full regression output for table 4, panel a, in the main text. See that table's note for further details.

Table A.4: The impact of cholera on the opening of waterworks by human capital level, full regression output

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$C_i \mathbf{1}_{t>1866}$	0.0047***	0.0048**					
	(0.008)	(0.011)					
... × Literate			0.0064***	0.0063***	0.0062***	0.0059**	0.0058**
			(0.009)	(0.009)	(0.010)	(0.013)	(0.014)
... × Non-literate			0.0007	0.0007	0.0007	0.0000	0.0001
			(0.709)	(0.718)	(0.694)	(0.984)	(0.939)
(Hydro-)geology controls: $\mathbf{1}_{t>1866}$ × ...							
... Elevation (km)	-0.0414***	-0.0409***	-0.0422***	-0.0427***	-0.0418***	-0.0409***	-0.0412***
	(0.007)	(0.009)	(0.008)	(0.007)	(0.008)	(0.010)	(0.010)
... Slope (ratio)	0.0496	0.0520	0.0521	0.0530	0.0516	0.0500	0.0507
	(0.182)	(0.172)	(0.173)	(0.169)	(0.179)	(0.192)	(0.193)
... Groundwater productivity (share)	-0.0010	-0.0006	-0.0007	-0.0009	-0.0008	-0.0009	-0.0010
	(0.611)	(0.757)	(0.720)	(0.673)	(0.716)	(0.680)	(0.644)
... Groundwater salt intrusion (share)	0.0101*	0.0095*	0.0094*	0.0094*	0.0088	0.0103*	0.0097*
	(0.084)	(0.092)	(0.097)	(0.095)	(0.118)	(0.073)	(0.085)
... Distance to fresh water (10 kms)	0.0005	0.0004	0.0005	0.0005	0.0005	0.0005	0.0005
	(0.149)	(0.189)	(0.155)	(0.143)	(0.130)	(0.183)	(0.146)
... Monthly precipitation (cm)	0.0039**	0.0036**	0.0037**	0.0038**	0.0038**	0.0036*	0.0037**
	(0.023)	(0.049)	(0.042)	(0.038)	(0.041)	(0.051)	(0.045)
Socio-economic controls: $\mathbf{1}_{t>1866}$ × ...							
... Population (10^5)	0.0017***	0.0018***	0.0018***	0.0017***	0.0017***	0.0017***	0.0017***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
... All-cause death rate (%)	0.0009*	0.0009*	0.0008	0.0008	0.0008	0.0008	0.0008
	(0.063)	(0.086)	(0.111)	(0.109)	(0.111)	(0.126)	(0.122)
... Economic development (0-1 index)	0.0088*	0.0087	0.0086	0.0076	0.0085	0.0069	0.0061
	(0.096)	(0.100)	(0.100)	(0.139)	(0.103)	(0.189)	(0.233)
... Market access (10^6)	0.0005***	0.0005**	0.0004**	0.0005**	0.0004**	0.0004**	0.0004**
	(0.008)	(0.010)	(0.013)	(0.013)	(0.013)	(0.020)	(0.019)
Mechanism controls: $\mathbf{1}_{t>1866}$ × ...							
... Protest (dummy)				0.0052**			0.0044
				(0.038)			(0.102)
... Industry demand (dummy)					-0.0019		-0.0016
					(0.130)		(0.215)
... High density (dummy)						0.0039**	0.0034*
						(0.013)	(0.055)
Constant	-0.0140**	-0.0130**	-0.0130**	-0.0132**	-0.0127**	-0.0128**	-0.0127**
	(0.025)	(0.044)	(0.043)	(0.041)	(0.048)	(0.048)	(0.048)
City and region-year FEs	✓	✓	✓	✓	✓	✓	✓
City type-specific trends		✓	✓	✓	✓	✓	✓
p: Literate > Non-literate			0.022	0.022	0.024	0.019	0.021
Control mean (post-1866)	0.007	0.007	0.007	0.007	0.007	0.007	0.007
Cities	1426	1419	1419	1419	1419	1419	1419
R-squared	0.099	0.110	0.110	0.110	0.110	0.110	0.110

Note: This table presents the full regression output for table 4, panel b, in the main text. See that table's note for further details.

A.2 Additional results

A.2.1 The wage response to cholera

We examine the dynamic effect of cholera mortality on wages in 19th Germany with distributed lag models that relate log nominal wages to contemporaneous and lagged cholera death rates. The wage data come from Pfister (2018) and it refers to the period 1815-1889, which we use in its raw (nominal) form.[21] The dataset includes 42 distinct wage series for 30 urban labour markets across Germany, covering various occupations and industries. We measure local cholera exposure using both cholera deaths in the city itself and in the surrounding county, reflecting the broader spatial scale of labour markets beyond narrow urban centres alone.[22] While the whole wage panel covers the years between 1815 and 1889, most series set in later and some have gaps.

Given the sparse nature of the wage data, we estimate both standard distributed and Almon lag models. The latter reduce imprecision from multicollinearity across lags but impose a parametric structure on the distributed lag coefficients (Almon 1965). In particular, we estimate:

$$\log(w_{it}) = \sum_{j=0}^{L} \beta_j C_{i,t-j} + \alpha_i + \lambda_t + \varepsilon_{it}, \qquad \text{(Distributed model)}$$

$$\log(w_{it}) = \gamma_0 Z_{0,i,t} + \gamma_1 Z_{1,i,t} + \gamma_2 Z_{2c,i,t} + \alpha_i + \lambda_t + \varepsilon_{it}, \qquad \text{(Almon lag model)}$$

where w_{it} is the nominal wage for series i in year t; α_i and λ_t are series and year fixed effects; and $C_{i,t-j}$ is the cholera death rate lagged j years. In the Almon models, $Z_{p,i,t} = \sum_{j=0}^{L} j^p C_{i,t-j}$ are constructed variables representing the polynomial-weighted sum of the lags. We use a second-degree polynomial to recover the underlying lag

[21] Although the nominal currency and time units (e.g. Mark per day) vary across series, they are constant within. By including series fixed effects, we absorb all time-invariant differences in wage level and measurement across locations and occupations, allowing us to pool data across heterogeneous wage types without explicit conversion. Similarly, nominal wages need not be deflated, as national price trends are captured by year fixed effects; our focus is on local deviations from the national trend, not the evolution of real purchasing power.

[22] 7 of the 42 series refer to a broader region such as the Ruhr area as a whole. In those cases, we use the cholera mortality rate of the whole region.

coefficients as $\beta_j = \gamma_0 + \gamma_1 j + \gamma_2 j^2$. We estimate models for up to five lags as additional lags beyond this are consistently small and statistically insignificant.

The results in table A.5 show that the wage effects are broadly consistent between unrestricted and Almon models: a one-percentage-point cholera death rate was associated with a 0.5 percent to almost 4 percent increase in nominal wages. Given the sparse underlying data, we cannot interpret these estimates as precise effects, though the coefficients for the third and fourth lags are highly statistically significant. Taken together, we interpret this as evidence for a substantial positive and somewhat delayed wage response to cholera epidemics, likely due to labour scarcity in line with Gallardo-Albarrán and Kappner (2025). In further specifications, we were unable to detect any statistically significant effects particular to the 1866 epidemic.

Table A.5: The wage response to cholera in 42 German cities, 1818-1889

	$L = 1$		$L = 2$		$L = 3$		$L = 4$		$L = 5$	
	DL	Almon	DL	Almon	DL	Almon	DL	Almon	DL	Almon
β_0	1.788	1.788	1.715	1.715	1.503	1.558	2.308	1.768	2.399*	1.618
	(0.136)	(0.128)	(0.195)	(0.188)	(0.237)	(0.200)	(0.102)	(0.186)	(0.091)	(0.248)
β_1	1.361	1.361	0.860	0.860	0.713	0.572	0.388	1.195	1.486	2.134*
	(0.243)	(0.236)	(0.422)	(0.417)	(0.524)	(0.629)	(0.718)	(0.259)	(0.206)	(0.072)
β_2			1.867	1.867	1.213	1.361	1.183	1.399	1.085	2.491**
			(0.247)	(0.240)	(0.375)	(0.234)	(0.430)	(0.205)	(0.511)	(0.047)
β_3					3.977**	3.926**	3.480**	2.381	3.963**	2.687*
					(0.035)	(0.033)	(0.043)	(0.069)	(0.033)	(0.051)
β_4							3.612**	4.140**	3.453**	2.724*
							(0.035)	(0.021)	(0.038)	(0.072)
β_5									1.874	2.601
									(0.312)	(0.138)
Cities	42	42	42	42	42	42	42	42	42	42
Series and year FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Within-R2	0.001	0.002	0.002	.002	0.006	0.006	0.010	0.009	0.012	0.010

Note: This table shows the coefficients of distributed lag models and Almon lag models for the effect of cholera shocks on nominal wages in 42 German cities between 1818 and 1889 (see text for further details). Coefficients are multiplied by 100, thus giving the percent increase of the nominal wage associated with a one-percentage-point increase in the cholera death rate (deaths per 100 people) in years t to $t-5$. Nominal wages are obtained from Pfister (2018) and cholera data comes from Kappner (2025). Numbers in parentheses are p-values based on robust standard errors, clustered at the city level. Significance stars follow the conventional coding (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$).

A.2.2 Margins decomposition

In this section we decompose the effect of cholera exposure on the waterworks opening probability into its extensive- and intensive-margin components. Table A.6, columns 1 and 2, replicate our preferred specification with full controls (compare table 2, column 7). In column 3, we include both variables capturing both margins simultaneously to separate the intensive margin effect conditional on exposure to a cholera outbreak. The coefficient on the intensive margin decreases by 90% and becomes statistically insignificant, while the extensive margin remains robust. This indicates that the positive gradient observed in column 2 is largely driven by between-group differences (i.e., exposed vs unexposed cities) and not by variation in cholera severity within the set of exposed cities. This supports the idea of a threshold response mechanism, where the presence of cholera–not its precise severity–triggers political momentum.

Table A.6: The impact of cholera on the opening of waterworks, margins decomposition

	(1)	(2)	(3)	(4)	(5)
Extensive	0.393***		0.389**	0.430***	
	(0.006)		(0.019)	(0.007)	
Intensive		0.0499	0.00443		
		(0.265)	(0.931)		
Log Intensive				0.0290	
				(0.715)	
Threshold extensive					0.486***
					(0.001)
All controls and FEs	✓	✓	✓	✓	✓
Control mean (post-1866)	1.141	1.141	1.141	1.141	1.141
Cities	2552	2552	2552	2552	2552
R-squared	0.095	0.095	0.095	0.095	0.095

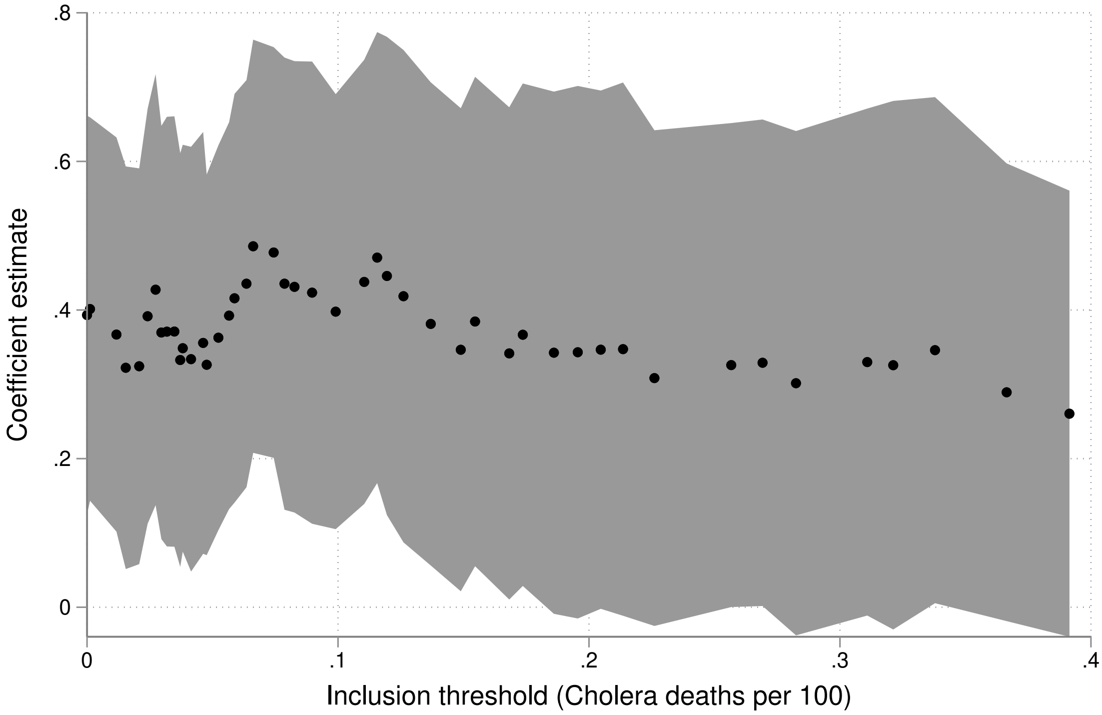
Note: This table shows the results of estimating equation 1 with alternative shock variable definitions. In columns 1, 3 and 4, we measure *extensive-margin* cholera exposure via a binary indicator. In column 5, we redefine the binary extensive-margin dummy using a coefficient-size maximizing threshold (see text for details). In columns 2 and 3 we measure *intensive-margin* cholera exposure as deaths per 100 inhabitants. In column 4, we apply a logarithmic transformation to that measure. Note that columns 1 and 2 replicate our results in table 2, column 7. See that table's notes for further details.

Column 4 tests whether the pattern in column 3 could be due to diminishing sensitivity to cholera intensity, by replacing the death rate with its natural logarithm.

However, this specification also yields a statistically insignificant conditional slope. While we cannot rule out nonlinear responses entirely, the failure of both linear and log-linear specifications to uncover a significant gradient reinforces the idea that the presence, not the intensity of cholera mattered.

Abstracting for a moment from the statistical significance of the intensive margin coefficients, their positive sign in columns 3 and 4 suggest that mildly affected cities may have behaved more like unexposed cities. This raises the possibility that the binary "any exposure" definition underestimates the true discontinuity. We formalize this idea in column 5 using a threshold search framework similar to the method by Hansen (2000). We iterate over candidate cutoffs in the cholera death rate and identify the threshold that maximizes the difference in post-1866 opening probabilities across the split (see figure A.1). This split lies at approximately 0.066%, and the resulting estimate implies a 0.486 percentage point effect for cities above this level (see table A.6, column 5). 21% of cholera-exposed cities fall below this threshold, suggesting that their low exposure was similarly unlikely to trigger political momentum as for entirely unexposed cities.

Figure A.1: Estimated effects around varying cholera death rate cut-off values



Note: The graph plots the estimated β coefficient from equation 1 for the iterative estimation runs of a Hansen (2000) threshold search procedure. In each run, as one moves along the horizontal axis, we include more and more cholera-exposed cities in the unexposed (control) group. For example, at an inclusion threshold of 0.3, cities with a death rate of up to 0.3% are treated as unexposed while cities with a higher death rate form the exposed (treatment) group. Gray areas mark the 95% confidence interval. Coefficients are estimated at each percentile of the death rates distribution. We do not show coefficients estimated above a threshold of 0.4%, as these are always insignificant.

A.2.3 Mechanism regressions with type-specific time trend and triple interactions

In this section, we present additional evidence underpinning our main findings on the mechanisms (section 5.2). To briefly recap, the results in table 4 were obtained estimating the following equation:

$$
\begin{aligned}
\mathrm{W}_{it} &= \beta C_i \mathbf{1}_{t>1866} + \mathbf{X}_i' \gamma \mathbf{1}_{t>1866} + \alpha_i + \epsilon_{it} \\
&+ \beta_2 C_i \mathrm{Type}_i \mathbf{1}_{1>1866} + \delta_{t\times r(i)\times \mathrm{Type}_i} \\
&+ \theta_1 \mathrm{Protest}_i \mathbf{1}_{t>1866} + \theta_2 \mathrm{Industry}_i \mathbf{1}_{t>1866} + \theta_3 \mathrm{Density}_i \mathbf{1}_{t>1866},
\end{aligned}
\tag{2}
$$

where Type_i is a dummy variable for commercial cities in table 4, panel a; and a dummy variable for more literate cities in table 4, panel b. In the second row, we include type-by-region-specific arbitrary time trend $\delta_{t\times r(i)\times \mathrm{Type}_i} = \sum_{\tau=t_0}^{T} \sum_{\rho=r}^{R} (\delta_{\tau\rho} \mathrm{Type}_i \mathbf{1}_{t=\tau,\rho=r})$.

Although we discuss in the main text that we prefer the previous model to account for the large differences in the economic structure of cities (and their evolution) across Germany during the period of high industrialization, we can impose a less parsimonious structure by not letting the effect of city type vary across regions and time, as follows:

$$
\begin{aligned}
\mathrm{W}_{it} &= \beta C_i \mathbf{1}_{t>1866} + \mathbf{X}_i' \gamma \mathbf{1}_{t>1866} + \alpha_i + \epsilon_{it} \\
&+ \beta_2 \mathrm{Type}_i \mathbf{1}_{t>1866} + \beta_3 C_i \mathrm{Type}_i \mathbf{1}_{t>1866} + \delta_{t\times r(i)} \\
&+ \theta_1 \mathrm{Protest}_i \mathbf{1}_{t>1866} + \theta_2 \mathrm{Industry}_i \mathbf{1}_{t>1866} + \theta_3 \mathrm{Density}_i \mathbf{1}_{t>1866}.
\end{aligned}
\tag{3}
$$

Columns 1 to 5 in tables A.7 and A.8 present the outcome of estimating equation 3. In line with our main results, the effect for commercial (literate) cities shocked by cholera is greater and more statistically significant than that referring to their less commercial (literate) counterparts. The coefficient differences are always statistically significant at conventional levels for regressions with the *literate* dummy, though the same does not apply to the regressions with the *commercial* dummy. A simple *commercial* $\times$ $\mathbf{1}_{t>1866}$ interaction does not seem to capture the full extent of non-agricultural city dynamics during the 1870s and 1880s when industrialization was in full swing. We examine this in greater detail in columns 6 to 9 of tables A.7 and A.8, by estimating:

$$
\begin{aligned}
\mathrm{W}_{it} = {} & \beta C_i \mathbf{1}_{t>1866} + \mathbf{X}_i' \gamma \mathbf{1}_{t>1866} + \alpha_i + \epsilon_{it} \\
& + \beta_2 \mathrm{Type}_i \mathbf{1}_{t>1866} + \beta_3 C_i \mathrm{Type}_i \mathbf{1}_{t>1866} + \delta_{t \times r(i)} \\
& + \theta_1 \mathrm{Protest}_i \mathbf{1}_{t>1866} + \theta_2 \mathrm{Industry}_i \mathbf{1}_{t>1866} + \theta_3 \mathrm{Density}_i \mathbf{1}_{t>1866} \\
& + \theta_4 \mathrm{Protest}_i C_i \mathbf{1}_{t>1866} + \theta_5 \mathrm{Industry}_i C_i \mathbf{1}_{t>1866} + \theta_6 \mathrm{Density}_i C_i \mathbf{1}_{t>1866}.
\end{aligned}
\tag{4}
$$

Importantly, besides re-adding the city type-specific trends of our preferred specification, we now include triple-interactions that *also* allow the post-1866 effect of protest potential, industry demand and high density to vary between cholera-exposed and spared cities. Four patterns in columns 6 to 9 provide support to our empirical setup and main results. First, the coefficients for cholera-shocked commercial (literate) cities remain statistically significant and sizeable throughout. Second, the coefficients only vary significantly for the commercial group (compare columns 5 and 9), indicating substantial heterogeneities within this set of cities that should be accounted for. Third, the two patterns discussed above remain even when controlling for the interaction of cholera with the alternative mechanisms–protest potential, industry demand and high density. And fourth, such a large statistically significant difference in coefficients across cholera-interacted mechanism variables is only visible for the commercial and literate dummies.

Table A.7: The impact of cholera on the opening of waterworks by elite background, robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$C_i \mathbf{1}_{t>1866}$									
... × Commercial	0.0062**	0.0063**	0.0060**	0.0058**	0.0057*	0.0096***	0.0113***	0.0110**	0.0124***
	(0.033)	(0.030)	(0.040)	(0.048)	(0.051)	(0.009)	(0.009)	(0.011)	(0.004)
... × Non-commercial	0.0040*	0.0040*	0.0039*	0.0037*	0.0036	0.0018	0.0037	0.0025	0.0042
	(0.071)	(0.075)	(0.077)	(0.096)	(0.100)	(0.373)	(0.148)	(0.247)	(0.129)
Commercial$_i$ × $\mathbf{1}_{t>1866}$	0.0039	0.0037	0.0040	0.0035	0.0034				
	(0.156)	(0.187)	(0.151)	(0.201)	(0.219)				
(Hydro-)geology controls: $\mathbf{1}_{t>1866}$ × ...									
... Elevation (km)	-0.0401***	-0.0406***	-0.0396***	-0.0392**	-0.0393**	-0.0403**	-0.0401**	-0.0395**	-0.0404**
	(0.009)	(0.008)	(0.009)	(0.011)	(0.010)	(0.014)	(0.014)	(0.015)	(0.014)
... Slope (ratio)	0.0451	0.0460	0.0449	0.0439	0.0447	0.0448	0.0453	0.0443	0.0449
	(0.226)	(0.222)	(0.230)	(0.241)	(0.240)	(0.264)	(0.243)	(0.251)	(0.268)
... Groundwater productivity (share)	-0.0012	-0.0013	-0.0012	-0.0012	-0.0013	-0.0001	-0.0000	-0.0000	-0.0003
	(0.540)	(0.513)	(0.530)	(0.535)	(0.506)	(0.978)	(0.985)	(0.989)	(0.887)
... Groundwater salt intrusion (share)	0.0098	0.0098	0.0090	0.0104*	0.0097	0.0120**	0.0116**	0.0128**	0.0119**
	(0.110)	(0.106)	(0.136)	(0.093)	(0.110)	(0.042)	(0.048)	(0.031)	(0.039)
... Distance to fresh water (10 kms)	0.0005	0.0006	0.0006*	0.0005	0.0006*	0.0005	0.0006	0.0005	0.0006
	(0.108)	(0.101)	(0.085)	(0.127)	(0.097)	(0.170)	(0.114)	(0.184)	(0.114)
... Monthly precipitation (cm)	0.0036**	0.0037**	0.0036**	0.0035**	0.0036**	0.0032*	0.0032*	0.0031*	0.0032*
	(0.030)	(0.027)	(0.031)	(0.034)	(0.031)	(0.068)	(0.069)	(0.078)	(0.069)
Socio-economic controls: $\mathbf{1}_{t>1866}$ × ...									
... Population (10^5)	0.0017***	0.0016***	0.0017***	0.0016***	0.0016***	0.0016***	0.0017***	0.0017***	0.0016***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
... All-cause death rate (%)	0.0009*	0.0009*	0.0009*	0.0009*	0.0009*	0.0008	0.0008	0.0007	0.0007
	(0.069)	(0.069)	(0.071)	(0.077)	(0.078)	(0.162)	(0.169)	(0.194)	(0.194)
... Economic development (0-1 index)	0.0066	0.0058	0.0065	0.0055	0.0049	0.0058	0.0064	0.0060	0.0052
	(0.192)	(0.239)	(0.196)	(0.284)	(0.326)	(0.263)	(0.235)	(0.272)	(0.331)
... Market access (10^6)	0.0005***	0.0005***	0.0005***	0.0005**	0.0005**	0.0004**	0.0004**	0.0004**	0.0004**
	(0.009)	(0.010)	(0.009)	(0.013)	(0.013)	(0.030)	(0.031)	(0.039)	(0.037)
Mechanism controls: $\mathbf{1}_{t>1866}$ × ...									
... Protest (dummy)		0.0043*			0.0038	0.0028			0.0024
		(0.071)			(0.140)	(0.637)			(0.680)
... Industry demand (dummy)			-0.0022*		-0.0019		-0.0003		-0.0001
			(0.083)		(0.135)		(0.828)		(0.974)
... High density (dummy)				0.0028*	0.0023			0.0041*	0.0038
				(0.066)	(0.166)			(0.079)	(0.109)
... Protest (dummy) $\times C_i$						0.0036			0.0033
						(0.685)			(0.710)
... Industry demand (dummy) $\times C_i$							-0.0049*		-0.0049*
							(0.066)		(0.067)
... High density (dummy) $\times C_i$								-0.0030	-0.0035
								(0.251)	(0.223)
Constant	-0.0131**	-0.0132**	-0.0127**	-0.0130**	-0.0127**	-0.0110*	-0.0111*	-0.0112*	-0.0113*
	(0.032)	(0.031)	(0.037)	(0.034)	(0.036)	(0.069)	(0.066)	(0.065)	(0.065)
City and region-year FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
City type-specific trends						✓	✓	✓	✓
Triple interactions						✓	✓	✓	✓
p: Commercial > Non-commercial	0.276	0.266	0.290	0.281	0.284	0.027	0.038	0.025	0.023
Mean outcome (post-1866)	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007
Cities	1426	1426	1426	1426	1426	1419	1419	1419	1419
R-squared	0.099	0.099	0.099	0.099	0.100	0.156	0.156	0.156	0.156

Note: This table presents robustness checks to the results in table 4, panel a. Columns 1 to 5 include region-year fixed effects rather than region-year-type effects (where *type* distinguishes commercial and non-commercial cities). In the specifications in columns 6 to 9, we re-include the triple-interaction fixed effects of our main specification, and additionally allow for an interaction effect of the three alternative mechanisms and the cholera shock. See notes of table 4 for further details.

Table A.8: The impact of cholera on the opening of waterworks by human capital level, robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$C_i \mathbf{1}_{t>1866}$									
... $\times$ Literate	0.0059**	0.0059**	0.0057**	0.0055**	0.0053**	0.0063***	0.0084***	0.0059**	0.0081**
	(0.011)	(0.012)	(0.013)	(0.018)	(0.020)	(0.008)	(0.008)	(0.022)	(0.016)
... $\times$ Non-literate	0.0020	0.0020	0.0021	0.0014	0.0015	0.0006	0.0027	0.0000	0.0023
	(0.273)	(0.278)	(0.260)	(0.446)	(0.405)	(0.803)	(0.211)	(0.989)	(0.422)
Literate$_i$ $\times$ $\mathbf{1}_{t>1866}$	-0.0006	-0.0006	-0.0004	-0.0007	-0.0005				
	(0.776)	(0.769)	(0.835)	(0.758)	(0.805)				
(Hydro-)geology controls: $\mathbf{1}_{t>1866}$ $\times$ ...									
... Elevation (km)	-0.0422***	-0.0427***	-0.0417***	-0.0409***	-0.0410***	-0.0426***	-0.0427***	-0.0409***	-0.0420***
	(0.007)	(0.007)	(0.008)	(0.009)	(0.009)	(0.008)	(0.007)	(0.010)	(0.009)
... Slope (ratio)	0.0497	0.0505	0.0495	0.0476	0.0484	0.0526	0.0526	0.0500	0.0514
	(0.181)	(0.178)	(0.185)	(0.203)	(0.201)	(0.194)	(0.172)	(0.193)	(0.208)
... Groundwater productivity (share)	-0.0011	-0.0012	-0.0011	-0.0012	-0.0013	-0.0009	-0.0009	-0.0009	-0.0012
	(0.579)	(0.548)	(0.569)	(0.562)	(0.533)	(0.673)	(0.650)	(0.679)	(0.578)
... Groundwater salt intrusion (share)	0.0097*	0.0098*	0.0090	0.0106*	0.0099*	0.0094	0.0090	0.0103*	0.0099*
	(0.094)	(0.091)	(0.116)	(0.072)	(0.087)	(0.101)	(0.121)	(0.068)	(0.082)
... Distance to fresh water (10 kms)	0.0005	0.0005	0.0006	0.0005	0.0005	0.0005	0.0006	0.0005	0.0006
	(0.133)	(0.122)	(0.106)	(0.157)	(0.121)	(0.143)	(0.103)	(0.184)	(0.114)
... Monthly precipitation (cm)	0.0040**	0.0041**	0.0040**	0.0039**	0.0039**	0.0038**	0.0038**	0.0036*	0.0037**
	(0.023)	(0.021)	(0.023)	(0.028)	(0.025)	(0.039)	(0.040)	(0.051)	(0.046)
Socio-economic controls: $\mathbf{1}_{t>1866}$ $\times$ ...									
... Population (10^5)	0.0017***	0.0017***	0.0017***	0.0017***	0.0016***	0.0017***	0.0017***	0.0017***	0.0016***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
... All-cause death rate (%)	0.0009*	0.0009*	0.0009*	0.0009*	0.0009*	0.0008	0.0008	0.0008	0.0008
	(0.068)	(0.070)	(0.070)	(0.081)	(0.082)	(0.110)	(0.110)	(0.131)	(0.129)
... Economic development (0-1 index)	0.0084	0.0075	0.0084	0.0067	0.0061	0.0076	0.0083	0.0069	0.0059
	(0.108)	(0.144)	(0.110)	(0.203)	(0.240)	(0.141)	(0.115)	(0.189)	(0.255)
... Market access (10^6)	0.0005***	0.0005***	0.0005***	0.0004**	0.0005**	0.0005**	0.0004**	0.0004**	0.0004**
	(0.010)	(0.010)	(0.010)	(0.015)	(0.015)	(0.014)	(0.013)	(0.021)	(0.022)
Mechanism controls: $\mathbf{1}_{t>1866}$ $\times$ ...									
... Protest (dummy)		0.0049**			0.0041	0.0044			0.0040
		(0.039)			(0.102)	(0.475)			(0.510)
... Industry demand (dummy)			-0.0021*		-0.0019		0.0009		0.0013
			(0.078)		(0.141)		(0.546)		(0.449)
... High density (dummy)				0.0037**	0.0032*			0.0039*	0.0037
				(0.012)	(0.051)			(0.091)	(0.100)
... Protest (dummy) $\times C_i$						0.0012			0.0004
						(0.893)			(0.968)
... Industry demand (dummy) $\times C_i$							-0.0059**		-0.0059**
							(0.029)		(0.032)
... High density (dummy) $\times C_i$								0.0000	-0.0004
								(0.997)	(0.872)
Constant	-0.0137**	-0.0138**	-0.0134**	-0.0134**	-0.0132**	-0.0131**	-0.0133**	-0.0128**	-0.0133**
	(0.025)	(0.024)	(0.029)	(0.027)	(0.029)	(0.046)	(0.041)	(0.048)	(0.046)
City and region-year FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
City type-specific trends						✓	✓	✓	✓
Triple interactions						✓	✓	✓	✓
p: Literate > Non-literate	0.082	0.084	0.092	0.073	0.084	0.027	0.022	0.018	0.023
Mean outcome (post-1866)	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007
Cities	1426	1426	1426	1426	1426	1419	1419	1419	1419
R-squared	0.099	0.099	0.099	0.099	0.099	0.110	0.110	0.110	0.110

Note: This table presents robustness checks to the results in table 4, panel b. Columns 1 to 5 include region-year fixed effects rather than region-year-type effects (where *type* distinguishes literate and non-literate cities). In the specifications in columns 6 to 9, we re-include the triple-interaction fixed effects of our main specification, and additionally allow for an interaction effect of the three alternative mechanisms and the cholera shock. See notes of table 4 for further details.

A.2.4 Mechanism regressions estimating operation probabilities in a balanced panel

Table A.9: The impact of cholera on the operation probability of waterworks by elite background

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$C_i \mathbf{1}_{t>1866}$	0.0468***	0.0363***					
	(0.000)	(0.001)					
... × Commercial			0.0980***	0.0934***	0.0950***	0.0886***	0.0848***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
... × Non-commercial			0.0124	0.0107	0.0110	0.0060	0.0052
			(0.204)	(0.263)	(0.255)	(0.522)	(0.572)
(Hydro-)geology controls: $\mathbf{1}_{t>1866}$ × ...							
... Elevation (km)	-0.2393**	-0.2399**	-0.2441**	-0.2608**	-0.2376**	-0.2333**	-0.2463**
	(0.019)	(0.026)	(0.023)	(0.016)	(0.026)	(0.030)	(0.023)
... Slope (ratio)	0.1019	0.1404	0.1579	0.1798	0.1525	0.1591	0.1743
	(0.695)	(0.600)	(0.552)	(0.505)	(0.566)	(0.551)	(0.520)
... Groundwater productivity (share)	0.0003	-0.0015	-0.0011	-0.0056	-0.0002	-0.0018	-0.0050
	(0.985)	(0.923)	(0.944)	(0.707)	(0.991)	(0.902)	(0.734)
... Groundwater salt intrusion (share)	0.0254	0.0230	0.0237	0.0265	0.0180	0.0364	0.0308
	(0.481)	(0.545)	(0.548)	(0.476)	(0.641)	(0.350)	(0.397)
... Distance to fresh water (10 kms)	0.0018	0.0021	0.0020	0.0025	0.0025	0.0017	0.0026
	(0.403)	(0.332)	(0.355)	(0.253)	(0.249)	(0.415)	(0.222)
... Monthly precipitation (cm)	0.0246**	0.0177*	0.0181*	0.0191*	0.0183*	0.0160	0.0177*
	(0.021)	(0.094)	(0.086)	(0.066)	(0.083)	(0.124)	(0.087)
Socio-economic controls: $\mathbf{1}_{t>1866}$ × ...							
... Population (10^5)	0.0010	0.0009	0.0008	0.0006	0.0008	0.0008	0.0006
	(0.392)	(0.389)	(0.398)	(0.507)	(0.407)	(0.401)	(0.511)
... All-cause death rate (%)	0.0016	-0.0002	0.0004	0.0007	0.0002	-0.0002	0.0002
	(0.594)	(0.942)	(0.905)	(0.810)	(0.941)	(0.949)	(0.949)
... Economic development (0-1 index)	0.1914***	0.1537***	0.1480***	0.1193***	0.1460***	0.1257***	0.1047***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
... Market access (10^6)	0.0040***	0.0033***	0.0031***	0.0032***	0.0031***	0.0029***	0.0030***
	(0.000)	(0.001)	(0.002)	(0.001)	(0.002)	(0.003)	(0.002)
Mechanism controls: $\mathbf{1}_{t>1866}$ × ...							
... Protest (dummy)				0.0968***			0.0869***
				(0.000)			(0.001)
... Industry demand (dummy)					-0.0254***		-0.0197**
					(0.003)		(0.017)
... High density (dummy)						0.0459***	0.0330***
						(0.001)	(0.008)
Constant	-0.8538***	-0.8108***	-0.8033***	-0.7911***	-0.8006***	-0.7947***	-0.7841***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
City and region-year FEs	✓	✓	✓	✓	✓	✓	✓
City type-specific trends		✓	✓	✓	✓	✓	✓
p: Commercial > Non-commercial			0.000	0.000	0.000	0.000	0.000
Control mean (post-1866)	0.052	0.052	0.052	0.052	0.052	0.052	0.052
Cities	1426	1419	1419	1419	1419	1419	1419
R-squared	0.535	0.562	0.563	0.567	0.564	0.565	0.568

Note: This table re-estimates the models in table 4, panel a, without dropping cities from the sample after they open waterworks. This balances the panel and the coefficient on $C_i \mathbf{1}_{t>1866}$ now estimates the differential annual operation probability. See notes of table 4 for further details.

Table A.10: The impact of cholera on the operation probability of waterworks by human capital level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$C_i\mathbf{1}_{t>1866}$	0.0468***	0.0460***					
	(0.000)	(0.000)					
... × Literate			0.0576***	0.0532***	0.0552***	0.0478***	0.0445***
			(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
... × Non-literate			0.0145	0.0138	0.0149	0.0034	0.0055
			(0.226)	(0.270)	(0.213)	(0.782)	(0.665)
(Hydro-)geology controls: $\mathbf{1}_{t>1866}$ × ...							
... Elevation (km)	-0.2393**	-0.2207**	-0.2301**	-0.2443**	-0.2251**	-0.2095**	-0.2225**
	(0.019)	(0.030)	(0.025)	(0.020)	(0.029)	(0.041)	(0.033)
... Slope (ratio)	0.1019	0.1232	0.1285	0.1587	0.1238	0.1179	0.1427
	(0.695)	(0.641)	(0.628)	(0.554)	(0.641)	(0.657)	(0.596)
... Groundwater productivity (share)	0.0003	0.0001	-0.0007	-0.0051	-0.0009	-0.0035	-0.0068
	(0.985)	(0.994)	(0.962)	(0.737)	(0.954)	(0.818)	(0.653)
... Groundwater salt intrusion (share)	0.0254	0.0145	0.0142	0.0195	0.0067	0.0322	0.0274
	(0.481)	(0.662)	(0.674)	(0.536)	(0.841)	(0.321)	(0.378)
... Distance to fresh water (10 kms)	0.0018	0.0014	0.0017	0.0024	0.0022	0.0014	0.0023
	(0.403)	(0.519)	(0.427)	(0.278)	(0.312)	(0.522)	(0.276)
... Monthly precipitation (cm)	0.0246**	0.0179*	0.0190*	0.0201*	0.0193*	0.0157	0.0176*
	(0.021)	(0.085)	(0.070)	(0.053)	(0.067)	(0.127)	(0.087)
Socio-economic controls: $\mathbf{1}_{t>1866}$ × ...							
... Population (10^5)	0.0010	0.0010	0.0009	0.0007	0.0009	0.0009	0.0006
	(0.392)	(0.405)	(0.409)	(0.512)	(0.418)	(0.417)	(0.517)
... All-cause death rate (%)	0.0016	0.0008	0.0003	0.0007	0.0003	-0.0003	0.0002
	(0.594)	(0.796)	(0.934)	(0.821)	(0.925)	(0.923)	(0.941)
... Economic development (0-1 index)	0.1914***	0.1921***	0.1908***	0.1538***	0.1881***	0.1536***	0.1280***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
... Market access (10^6)	0.0040***	0.0039***	0.0038***	0.0038***	0.0038***	0.0035***	0.0036***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mechanism controls: $\mathbf{1}_{t>1866}$ × ...							
... Protest (dummy)				0.1037***			0.0889***
				(0.000)			(0.001)
... Industry demand (dummy)					-0.0247***		-0.0178**
					(0.005)		(0.032)
... High density (dummy)						0.0629***	0.0491***
						(0.000)	(0.001)
Constant	-0.8538***	-0.8324***	-0.8342***	-0.8207***	-0.8312***	-0.8167***	-0.8067***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
City and region-year FEs	✓	✓	✓	✓	✓	✓	✓
City type-specific trends		✓	✓	✓	✓	✓	✓
p: Literate > Non-literate			0.007	0.011	0.009	0.006	0.013
Control mean (post-1866)	0.052	0.052	0.052	0.052	0.052	0.052	0.052
Cities	1426	1419	1419	1419	1419	1419	1419
R-squared	0.558	0.568	0.568	0.572	0.569	0.571	0.575

Note: Note: This table re-estimates the models in table 4, panel b, without dropping cities from the sample after they open waterworks. This balances the panel and the coefficient on $C_i\mathbf{1}_{t>1866}$ now estimates the differential annual operation probability. See notes of table 4 for further details.

A.2.5 The long-run impact of cholera: Estimation framework details

In section 5.3, we discuss how exposure to the 1866 cholera shock shaped long-run differences in cities' sanitary investments, and how this, in turn, affected mortality and economic activity just before the First World War. This section details the underlying econometric framework.

Extensive margin: Waterworks adoption status by 1911 We start by re-estimating our linear opening hazards model (equation 1) with annual interaction dummies for the cholera shock. From the fitted model, we obtain for each city–year $\{i, t\}$ two predicted hazards: $\widehat{h}_{1,it}$, the predicted hazard if the city was exposed to the cholera shock, and $\widehat{h}_{0,it}$, the predicted hazard if it was spared. Following Singer and Willett (1993), we map these hazards into discrete-time survival paths

$$\widehat{S}_{d,i,t} = \prod_{\tau \leq t} \left(1 - \widehat{h}_{d,i\tau}\right), \quad d \in \{0, 1\}, \tag{5}$$

so that the probability of having opened waterworks by 1911 is $\widehat{P}_{d,i,1911} = 1 - \widehat{S}_{d,i,1911}$. Averaging over cities, we obtain the mean adoption probabilities $\bar{P}_0$ and $\bar{P}_1$. We define the long-run extensive-margin effect of the 1866 cholera epidemic as

$$\Delta\bar{P} \equiv \bar{P}_1 - \bar{P}_0. \tag{6}$$

Intuitively, $\Delta\bar{P}$ is the average increase–induced by the 1866 shock–in the probability that a city had waterworks by 1911, holding observed covariates and estimated fixed effects constant.

Intensive margin among adopters: Within-city coverage share in 1911 We then estimate the intensive-margin effect conditional on waterworks adoption, using observed 1911 within-city connection rates (the share of connected buildings), $w_i \in [0, 1]$. Because w_i is observed only for cities that had opened waterworks by 1911, we use a fractional logit model (Papke and Wooldridge 1996; 2008) that includes the inverse Mills ratio (IMR_i) from a first-stage selection equation (Heckman 1979).

Defining a dummy for observed waterworks opening by 1911, D_i, we estimate:

$$\text{Selection stage:} \quad \Pr(D_i = 1|\mathbf{X}_i) = \Phi\big(\alpha_s + \beta_s C_i + \mathbf{X}_i'\delta_s + \gamma_{s,r(i)}\big), \quad \text{IMR}_i = \frac{\phi(\hat{z}_i)}{\Phi(\hat{z}_i)}, \tag{7}$$

$$\text{Outcome stage:} \quad \mathbb{E}[w_i|\mathbf{X}_i, D_i = 1] = \text{logit}^{-1}\left(\alpha + \beta C_i + \mathbf{X}_i'\delta + \gamma_{r(i)} + \gamma\,\text{IMR}_i\right), \tag{8}$$

where $\hat{z}_i$ is the first-stage linear index and $\phi(\cdot)$, $\Phi(\cdot)$ are the standard normal probability density and cumulative distribution functions, respectively.

From the outcome stage we obtain fitted connection shares $\hat{\mu}_i$. Averaging over adopters, we compute the average intensive-margin effect of cholera among adopters:

$$\overline{\text{AME}} \;=\; \frac{1}{N_{D_i=1}} \sum_{i:D_i=1} \beta\,\hat{\mu}_i(1-\hat{\mu}_i), \tag{9}$$

which is the sample average of the logit derivative evaluated at $\hat{\mu}_i$. Intuitively, among cities that successfully opened waterworks by 1911, cholera exposure is associated with a $\overline{\text{AME}}$ percentage-point higher connection share at observed covariates and fixed effects.

Combined margins: Estimating the total coverage-share increase We combine the extensive- and conditional intensive-margin effects into an *unconditional* effect on the 1911 coverage share induced by the 1866 cholera shock. Let $\bar{w}$ be the mean fitted connection share among adopters in 1911, and $\bar{P}_1$ the mean adoption probability under cholera exposure. The total effect is

$$\underbrace{\Delta\bar{P}}_{\text{more adopters}} \times \underbrace{\bar{w}}_{\text{typical depth among adopters}} + \underbrace{\bar{P}_1}_{\text{share that adopted}} \times \underbrace{\overline{\text{AME}}}_{\text{deeper coverage among adopters}}. \tag{10}$$

The first term scales the extensive-margin adoption increment by an estimate of the average intensive-margin depth (the within-city coverage share). The second term scales the within-adopter coverage deepening by the share of cities that adopted. Intuitively, the overall effect accounts for (1) how many more cities open waterworks and (2) how much deeper coverage becomes where adoption occurs. Since we do not observe connection shares for the full city sample, we compare the combined margins effect to an estimated average coverage rate based on the fitted connection shares from our intensive-margin model (see above). Formally,

$$\text{Average sample-wide coverage rate} \equiv \frac{1}{N}\sum_{i=1}^{N}\hat{\mu}_i,$$

where $\hat{\mu}_i$ is the fitted connection share from the intensive-margin model and N is the number of cities.

Long-run outcomes associated with the extensive margin Our final step is to estimate the association of cholera-induced sanitary investment with long-run economic and health outcomes. For this, we use the estimated opening probability $\widehat{P}_{1,i,1911}$ as a generated regressor in cross-sectional regressions for outcomes observed around 1911. Specifically, we focus on five measures reflecting demographic and economic performance, all measured at the city level: the crude death rate, the infant mortality rate, students in higher education per capita, stock market–listed firms per capita, and patents filed per capita. For each outcome Y_i, we estimate

$$Y_i = \alpha + \theta_E\,\widehat{P}_{1,i,1911} + \mathbf{X}_i'\delta + \gamma_{r(i)} + u_i, \tag{11}$$

where $\widehat{P}_{1,i,1911}$ is the predicted probability of having adopted waterworks by 1911 under cholera exposure, $\mathbf{X}_i$ are the same baseline controls used in the hazard model, and $\gamma_{r(i)}$ are region fixed effects. We focus exclusively on the extensive margin because, as shown in section 5.3, the 1866 cholera shock had only a negligible influence on the intensive margin (the connection share among adopters) by 1911.

θ_E is the average change in Y_i associated with moving from a zero to full opening probability attributable to 1866 cholera exposure, holding covariates fixed. While this is not a strictly causal effect–unobserved city characteristics could still influence both waterworks adoption and later outcomes–our approach accounts for many confounding pathways through the included covariates and fixed effects. Throughout our analyses, all reported p-values come from a cluster bootstrap at the city level that resamples the data and re-runs the entire estimation (hazard estimation, survivor paths, selection correction, fractional logit, aggregation, and estimation of long-run outcomes) 500 times. This respects the generated-regressor nature of $\widehat{h}_{d,it}$, $\hat{\mu}_i$, and IMR_i terms and the serial correlation within cities.

A.3 Robustness checks

A.3.1 Temporal trends

Table A.11 shows that our coefficient of interest remains statistically significant and relatively similar in size when we change the spatial resolution of the time trends ($\delta_{t\times r(i)}$) in equation 1. Column 1 includes a sample-wide arbitrary time trend via a year fixed effect without capturing regional heterogeneities. Columns 2, 3, and 4 impose state- (25), district- (75),[23] and county- (639) by-year fixed effects, thus absorbing arbitrary time trends common to all cities within these spatial units. The model in column 3 is our baseline specification, identical to column 7 of table 2, relying on a higher number of observations than in columns 2 and 4 because there is a single city in some states and counties.

The last two models are designed to capture spatially clustered trends that transgress administrative boundaries. Column 5, akin to column 1, employs a sample-wide time trend but additionally controls for year specific effects of the planar coordinates and their interaction. In column 6, we replace discrete region-year dummies with a year-specific two-dimensional thin-plate regression spline in the (x, y) spatial coordinates, fitted simultaneously with all other regressors through the Generalized Additive Model framework (Conley and Kelly 2025).

[23]We define 75 regions that reflect regional administrative heterogeneity within the federally structured German Empire, while avoiding stark size differences between regions (in terms of area and cities). While we treat the medium-sized German states as individual regions, we disaggregate Prussia, Bavaria, Saxony, Alsace-Lorraine, Baden, Wurttemberg and Hesse into multiple regions due to their size. We merge the city of Berlin with the surrounding Brandenburg region. We treat the three Hanseatic cities (Hamburg, Bremen, Lübeck) as a single region. We join the duchy of Lauenburg (containing three cities) to the Holstein district, and we add the small Prussian Sigmaringen enclave to the Wurttemberg Schwarzwaldkreis. Finally, we treat the eight Thuringian states as one region.

Table A.11: Alternative regional variation absorption approaches

	(1) Year	(2) Year × state	(3) Year × district	(4) Year × county	(5) Year × (x, y, xy)	(6) Year × $s(x, y)$
$C_i \mathbf{1}_{t>1866}$	0.286**	0.286**	0.393***	0.334**	0.459**	0.405**
	(0.024)	(0.014)	(0.006)	(0.046)	(0.031)	(0.020)
All controls and unit FEs	✓	✓	✓	✓	✓	✓
Spatial granularity of year FEs	1	25	75	646	(flexible)	(flexible)
Control mean (post-1866)	1.143	1.144	1.141	1.024	1.143	1.143
Cities	2552	2551	2552	2363	2552	2552
R-squared	0.049	0.064	0.095	0.362	0.054	0.057

Note: This table shows the results of estimating equation 1 with alternative regional time trend ($\delta_{t \times r(i)}$) specifications. See text and table notes of table 2 for further details.

A.3.2 Alternative models

We test the robustness of our baseline specification with additional models that capture different dimensions of the post-cholera waterworks diffusion. We begin with the closest to our baseline specification by re-estimating equation 1 while keeping cities in the sample after they open waterworks, instead of dropping them. This approach estimates annual *operation* probabilities, rather than opening hazards. It has the advantage of keeping the sample balanced, increasing the variation useable for estimation. Table A.12 shows that cities shocked by cholera had a greater operation probability after 1866 relative to those unaffected by cholera. To make the coefficients comparable with those of our baseline approach (table 2), we compute the implied annual hazard,

$$h^{\text{Operation}} = \beta^{\text{Operation}} \times \frac{T\lambda^2}{1 - (T+1)q^T + Tq^{T+1}},$$

where λ is the average annual hazard in the post-1866 control group, $T = 25$ is the number of post-shock years, and $q = (1 - \lambda)$. We find that these implied hazards are similar to the directly estimated hazard from our baseline model (see also table A.13, column 1).

Next, we employ 5 additional models that depart in different ways from our baseline model:

Table A.12: Alternative model specification: Operation status as outcome

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$C_i \mathbf{1}_{t>1866}$	6.926***	6.873***	5.787***	6.926***	4.217***	6.567***	3.609***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)
City and region-year FEs	✓	✓	✓	✓	✓	✓	✓
(Hydro-)geology		✓					✓
Population			✓				✓
Pre-epidemic mortality				✓			✓
Economic development					✓		✓
Market access						✓	✓
Control mean (post-1866)	10.887	10.887	10.887	10.887	10.887	10.887	10.887
Cities	2552	2552	2552	2552	2552	2552	2552
R-squared	0.545	0.546	0.549	0.545	0.554	0.548	0.559
Implied annual hazard	0.64	0.63	0.53	0.64	0.39	0.61	0.33

Note: This table re-estimates the models in table 2, panel a, without dropping cities from the sample after they open waterworks. This balances the panel and the coefficient on $C_i \mathbf{1}_{t>1866}$ now estimates the differential annual operation probability. See notes of table 2 for further details.

$$
\begin{aligned}
\mathrm{W}_{it} &= \beta^{\text{First-Diff}} \widehat{C_i \mathbf{1}_{t>1866}} + \widehat{\mathbf{X}_i' \mathbf{1}_{t>1866}} \gamma + \epsilon_{it}, && \text{(First-difference linear hazard)} \\
h_i(t) &= h_{0,r(i)}(t) \cdot \exp(\beta^{\text{Cox}} C_i + \mathbf{X}_i' \gamma), && \text{(Cox proportional hazard)} \\
\mathrm{W}_{it} &= \beta^{\text{Pooled}} C_i + \mathbf{X}_i' \gamma + \delta_{t \times r(i)} + \epsilon_{it}, && \text{(Pooled LPM)} \\
T_i &= \beta^{\text{Time}} C_i + \mathbf{X}_i' \gamma + \delta_{r(i)} + \epsilon_i, && \text{(Time-to-opening Tobit)} \\
\mathrm{I}_i &= \beta^{\text{Duration}} C_i + \mathbf{X}_i' \gamma + \delta_{r(i)} + \epsilon_i. && \text{(Duration OLS)}
\end{aligned}
$$

Here, τ_i is the year in which city i opens waterworks (i.e., in which W_{it} turns from 0 to 1). $h_i(t)$ is the opening hazard with region-specific baselines $h_{0,r(i)}(t)$ in a Cox proportional hazards model. $T_i = \min(\tau_i - 1866, \text{Censored})$ is time-to-adoption measured in years since 1866, and $\mathrm{I}_i \in [0, 1]$ is the share of years between 1867-1891 in which city i operated waterworks. The hat-transformed regressors in the first model indicate their instrumenting by their first differences (Farbmacher and Tauchmann 2023). Depending on the cross-sectional or panel nature of the regression, we either include region ($\delta_{r(i)}$) or year-region ($\delta_{t \times r(i)}$) fixed effects.

We now briefly explain the purpose of each model. *First-Diff* re-estimates the linear discrete-time hazard model while instrumenting the interacted cholera shock and control variables with their first differences as suggested by Farbmacher and Tauchmann (2023). *Cox* is a semi-parametric Cox proportional hazards model for the instantaneous opening risk. In that model, we allow for region-specific baseline hazards ($h_{0,r(i)}(t)$) to approximate the regional time trends in our other panel specifications. In the *Pooled* model, we drop city fixed effects and instead estimate the time-constant effects of cholera and all controls. We still include arbitrary regional time trends in this model. *Time* is a cross-sectional Tobit model, estimating the time-to-opening in years since 1866, and treating cities that opened after 1891 as right-censored. Finally, *Duration* estimates the share of years between 1867 and 1891 in which a city operated waterworks.

The results from these models in table A.13 show that cholera exposure is associated with a statistically significant expansion of waterworks.[24] The only exception is *First-Diff* (column 2) for which we obtain a positive coefficient that is similar in size with that from our baseline specification (column 1), but statistically insignificant.[25] Turning to column 3, we find a $e^{0.319} - 1 = 38\%$ higher annual instantaneous adoption hazard among shocked cities, implying a similar annual hazard $b^{\text{Cox}} = 0.39$ as the baseline model. Pooling yields a smaller but directionally consistent coefficient (column 4). The time-to-opening specification (col. 5) indicates that waterworks were established on average 3.1 years earlier in shocked cities, which translates into an implied annual hazard of about 1.13 percentage points. Finally, the duration model (col. 6) shows that shocked cities operated waterworks for a 2.25 percentage-point greater share of years between 1867 and 1891, corresponding to an implied hazard of 0.21 percentage points. Taken together, these estimates confirm that exposure to cholera accelerated and intensified the diffusion of waterworks.

[24] We also compute the implied additive change in the annual opening probability among cholera-shocked cities, directly comparable to the baseline coefficient β^{Baseline} (column 1). Denote this effect by h^{Model}. For the *First-Diff* and *Pooled* models the coefficients are already on this scale, i.e. $h^{\text{First-Diff}} = \beta^{\text{First-Diff}}$ and $h^{\text{Pooled}} = \beta^{\text{Pooled}}$. The other models require transformations. Let λ, T and q be defined as above. Then: $h^{\text{Cox}} \approx \lambda\big(e^{\beta^{\text{Cox}}} - 1\big)$ (col. 3), $h^{\text{Time}} \approx -\beta^{\text{Time}} \cdot \frac{\lambda^2}{(1-q^T)-\lambda T q^{T-1}}$ (col. 5) and $h^{\text{Duration}} \approx \beta^{\text{Duration}} \cdot \frac{T\lambda^2}{1-(T+1)q^T+Tq^{T+1}}$ (col. 6).

[25] This could be due to the nature of the first-differencing linear hazard estimator that is designed for time-variant regressors, while our regressors (including our cholera shock variable) only vary over time through their interaction with the post-1866-dummy.

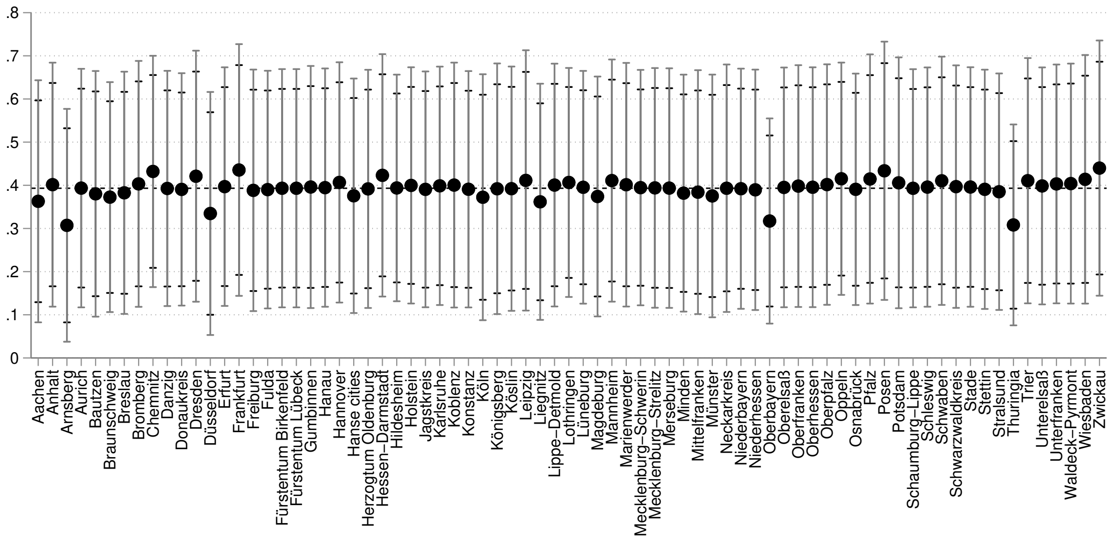
Table A.13: Alternative model specifications

	(1) Baseline	(2) First-Diff	(3) Cox	(4) Pooled	(5) Time	(6) Duration
Coefficient	0.00393***	0.00325	0.319***	0.00197***	-3.070***	0.0225**
	(0.008)	(0.295)	(0.000)	(0.000)	(0.001)	(0.021)
Cities	2552	2552	2498	2552	2510	2498
Implied annual hazard	0.39	0.32	0.39	0.20	1.13	0.21

Note: This table shows the results of estimating equations A.3.2. The control group's average annual post-1866 opening probability is estimated for the sample included in each regression and serves as a baseline for the reported implied annual hazards. See text for calculation of the implied annual hazards. Numbers in parentheses are p-values based on robust standard errors, clustered at the city level for models 1–4 (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$). Note that we do not cluster on years in these regression as several of the reported estimators do not support multi-way clustering.

A.3.3 Sample checks

Figure A.2: Leave-one-region-out coefficient plot

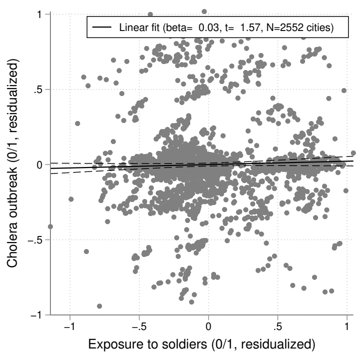


Note: Each coefficient is computed by estimating our preferred specification in column 7 of table 2, panel a, leaving out the region indicated below each coefficient. Black (gray) bars show the 90 (95)% confidence interval.

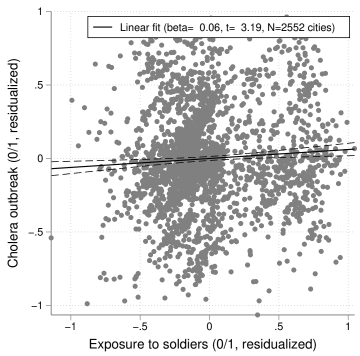
A.3.4 IV diagnostics and IV results

Figure A.3: 1866 troops exposure does not predict other epidemic events

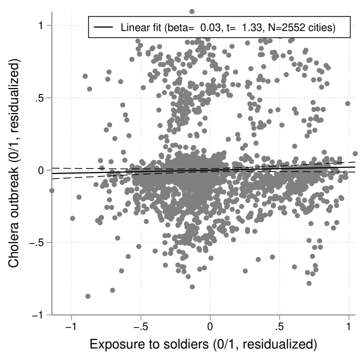
(a) 2nd cholera pandemic (1831-1837)



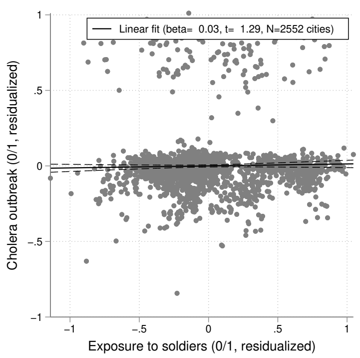
(b) 3rd cholera pandemic (1848-1860)



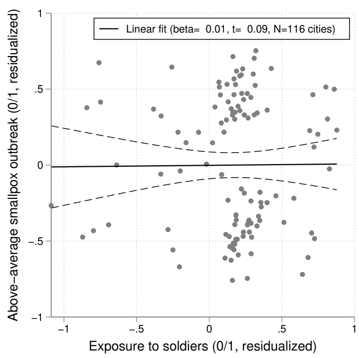
(c) 1871-1874 cholera outbreak



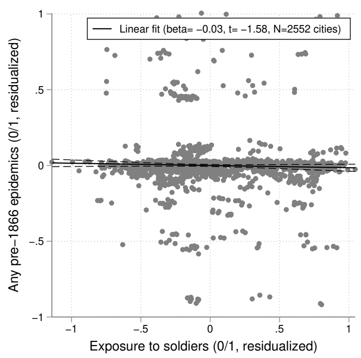
(d) 5th cholera pandemic (1892-1895)



(e) 1870-1871 smallpox outbreak

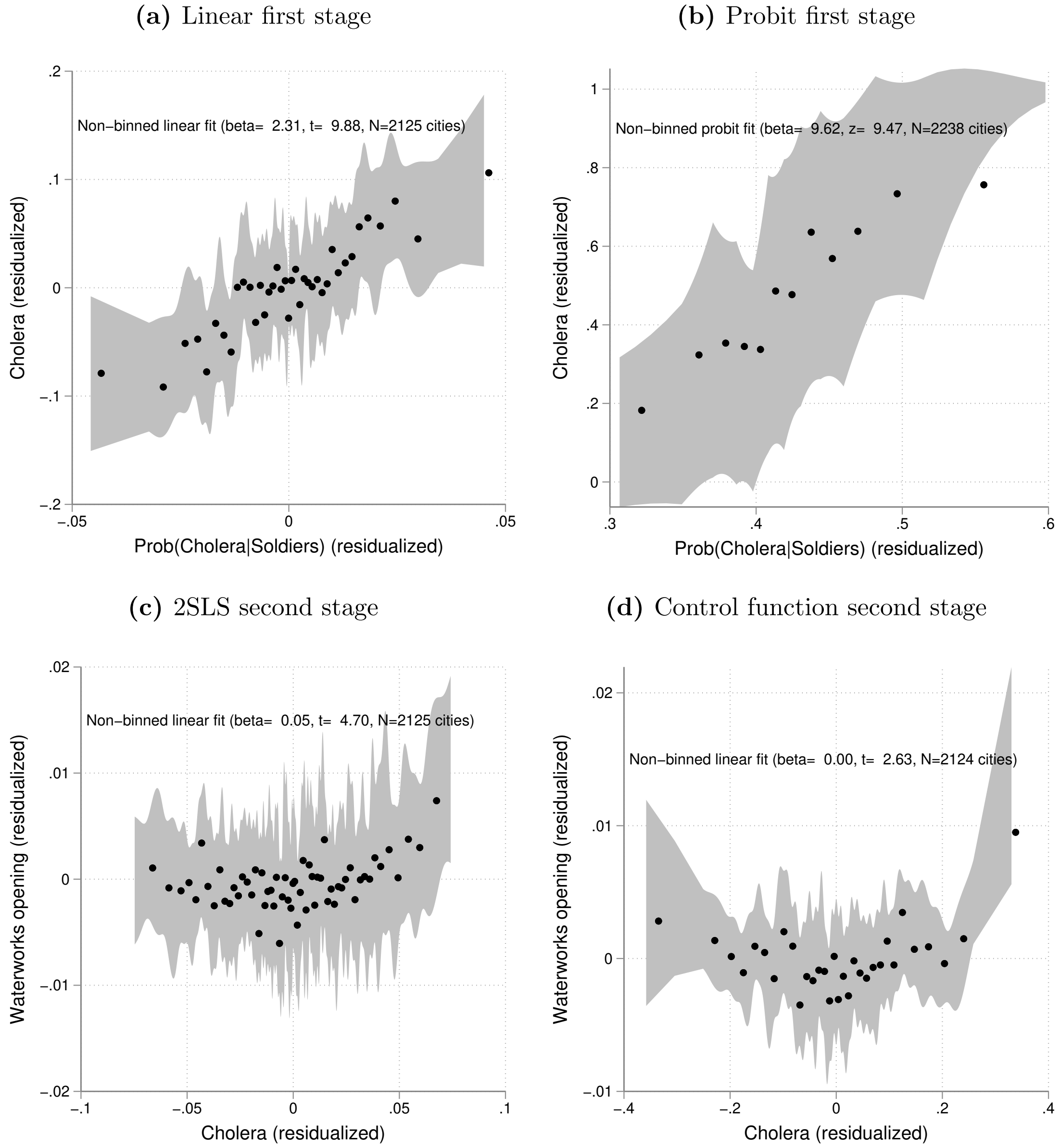


(f) Significant pre-1866 epidemics



Note: These plots show that 1866 troops exposure does not correlate with pre- and post-1866 cholera epidemics (panels a-d), nor does it correlate with above-average smallpox death rates during the 1870-1871 outbreak (panel e) or general significant pre-1866 epidemics (e.g. plague in the 14th century). Panels a-d include all cities in our sample (i.e. the universe of urban Germany). Panel e includes only Prussian cities that housed French prisoners of war during the 1870 Franco-Prussian war; hence, in this panel, we compare cities with above-average to cities with below-average smallpox death rates. Due to the small number of observations, we do not condition on regional fixed effects in this panel. Panel f includes most of our sample, except for cities in Alsace-Lorraine and territories ceded to Poland after the First World War. Data for panels e and f comes from Guttstadt (1873) and Cantoni et al. (2020). See section 4.2 for more details.

Figure A.4: Partial correlation plots for the first and second stages



Note: The binned scatter plot in the upper panel illustrate our two alternative first stages for the 2SLS and control-function (CF) approaches, respectively–linear on the left, Probit on the right. The lower panel shows the resulting second stages–2SLS on the left, and CF on the right. Bins are quantile-spaced and residualized on all controls. Grey-shaded areas signify confidence intervals at the 95 percent level. See section 4.2 for more details.

A.3.5 Spatial inference, the effect of regional cholera shocks and imitation

In this section we perform additional analyses considering the spatial dimension of our data.

Conley standard errors We begin by accounting for spatial autocorrelation in unobservables. In our preferred specification, we absorb common shocks to public–health investment at the region–year level through fixed effects and cluster our standard errors to allow for arbitrary serial correlation within cities and years. However, unobserved spatial correlation could transgress administrative borders so that our baseline specification's standard errors could overestimate the statistical significance of the cholera shock. To address this, we compute Conley (1999) standard errors, allowing the error terms of city-pairs to correlate up to pre-defined cut-off values (25-200 kilometers, in 25km steps) with the correlation decaying linearly in distance ("Bartlett kernel"). We also adjust these standard errors for serial correlation within cities over time (Colella et al. 2023).

Table A.14 reports p-values and 90/95% confidence intervals for our main coefficient of interest (column 7 in table 2, panel a) under different spatial cut-offs. Relative to our baseline clustered errors (column 1), the Conley-based corrections are more conservative, as expected. Crucially, the coefficient of interest remains statistically significant slightly below the 99% level across all distance cut-offs.

Table A.14: Inference under spatially auto-correlated (Conley) standard errors

	Clustering	25km	50km	75km	100km	125km	150km	175km	200km
p-value	0.0063***	0.0093***	0.0103**	0.0109**	0.0122**	0.0133**	0.0141**	0.0147**	0.0147**
[90% CI]	[0.16,0.62]	[0.14,0.64]	[0.14,0.65]	[0.14,0.65]	[0.14,0.65]	[0.13,0.65]	[0.13,0.66]	[0.13,0.66]	[0.13,0.66]
[95% CI]	[0.12,0.67]	[0.10,0.69]	[0.09,0.69]	[0.09,0.70]	[0.09,0.70]	[0.08,0.70]	[0.08,0.71]	[0.08,0.71]	[0.08,0.71]

Note: This table shows p-values (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$) and 90/95% confidence intervals for our main estimate (column 7, panel a in table 2) under different assumptions for spatial correlation in the standard errors. The first column ("Clustering") reproduces the city-and-year clustered standard errors we use in our preferred specification. The next eight columns present Conley (1999) standard errors for different cut-off radii, based on a Bartlett kernel. Standard errors additionally correct for autocorrelation within panel units and over time.

Spatial lag specifications Next, we turn to spatial dependence in the effect of the cholera shock across cities. We estimate a spatial lag model that, in addition to the direct effect of each city's own cholera epidemic, controls for the effect of surrounding cities' (and regions') shocks:

$$\mathrm{W}_{it} = \beta C_i \mathbf{1}_{t>1866} + \rho \left(\sum_{j \neq i} w_{ij} C_j \mathbf{1}_{t>1866} \right) + \mathbf{X}_i' \gamma \mathbf{1}_{t>1866} + \alpha_i + \delta_{t \times r(i)} + \epsilon_{it} \ , \quad (12)$$

where ρ is the marginal effect of the spatial lag of city i's cholera epidemic over all other locations j. The spatial weight matrix elements w_{ij} are computed using inverse-distance weighting or the average across all locations within a 25 kilometres radius. In both cases, we also weight by locations j's population. To make the magnitudes of the indirect (ρ) and direct (β) effects comparable, we report the *average* indirect effect, defined as its marginal effect multiplied by the average sum of weights: $\rho \times \sum_i \sum_{j \neq i} w_{ij} \times N^{-1}$.

The results in table A.15 show that the direct effect only slightly decreases in size and significance once we simultaneously control for the indirect effect, i.e. nearby cholera shocks. This holds true regardless of whether we focus on the rural, urban, or all neighbours. It also does not depend on the specification of the spatial weight matrix. While the indirect effect itself is small, it is also statistically significant. Summing up the direct and indirect effects across specifications roughly accounts for the baseline effect without spatial lags (see column 1). This suggests that the direct effect we focus on also captures the effect of spatially auto-correlated shocks in neighbouring cities and regions, though their contributions are small enough to ignore this dependence in most of our analyses.

Auto-regressive spatial lag models Our final robustness check in this section concerns the spatial diffusion of sanitary investment decisions via city-to-city imitation processes. If early waterworks opening decisions by bigger cities in cholera-struck regions triggered subsequent waterworks investments in nearby smaller cities, a spurious correlation between regionally clustered cholera outbreaks and sanitary investments could arise, confounding the causal effect we aim to measure. Our baseline

Table A.15: The indirect effect of nearby cholera shocks

		Inverse-distance			Neighbors within 25 km		
	(1) Direct impact only	(2) Overall lag	(3) Urban lag	(4) Rural lag	(5) Overall lag	(6) Urban lag	(7) Rural lag
Direct effect	0.393***	0.290**	0.331**	0.321**	0.339***	0.350**	0.352***
	(0.006)	(0.027)	(0.011)	(0.016)	(0.008)	(0.016)	(0.006)
Indirect effect		0.0908***	0.0601**	0.0661***	0.0289*	0.0179	0.0269*
		(0.001)	(0.047)	(0.003)	(0.074)	(0.218)	(0.069)
Cities	2552	2552	2552	2552	2549	2539	2530
R-squared	0.09	0.09	0.09	0.09	0.09	0.10	0.09

Note: This table presents regression results from specifications that include various spatial lags of the cholera shock. The coefficients are estimated as shown in equation 1. However, rather than presenting the raw lag coefficients (ρ), we show their *average* impact, $\rho \times \sum_i \sum_{j \neq i} w_{ij} \times N^{-1}$. The first column is identical to our baseline result in table 2, column 7, panel a. Numbers in parentheses are p-values based on standard errors clustered at the city and year levels (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$).

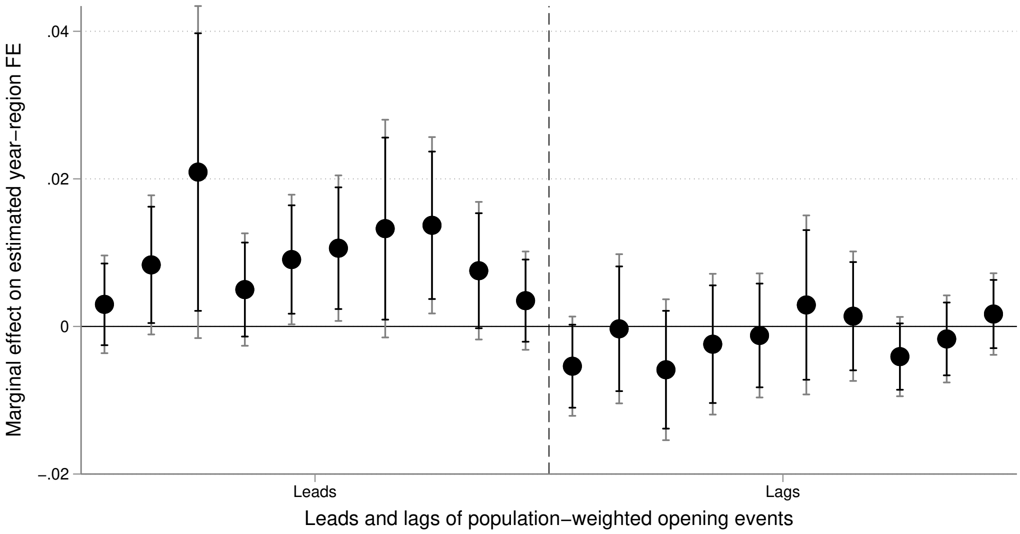
regression approach addresses this issue through year-by-region fixed effects that act as flexible regional baseline hazards. If, for instance, a “role model” city opens waterworks and subsequently gets imitated by close-by cities, the respective year-by-region fixed effect should increase. As our main effects of interest are conditional on these fixed effects, they are not confounded by imitation processes. In figure A.5, we show that the estimated fixed effects indeed perform this role. Waterworks opening events up to six years earlier significantly increase the regional year-fixed effects (i.e. the baseline opening hazard), while their temporal lags are all statistically insignificant and close to 0.

We can model regional imitation processes more directly with a regression of the form:

$$\mathrm{W}_{it} = \beta C_i \mathbf{1}_{t>1866} + \rho \mathrm{V}^s_{i,t-\tau} + \mathbf{X}_i' \gamma \mathbf{1}_{t>1866} + \alpha_i + \delta_t + \epsilon_{it} \ , \tag{13}$$

where $\mathrm{V}^s_{i,t-\tau}$ is the population-weighted share of cities that newly open waterworks in $t-\tau$ among all cities that are still at risk of opening, within a radius of s kilometres around city i (excluding i itself). Therefore, V captures imitation processes operating at various spatial and temporal lags. Note that we relax our fixed effects structure by dropping the region-specific time trends, as these absorb imitation effects, as we

Figure A.5: The effect of waterworks openings on regional baseline hazards



Note: This graph shows the estimated coefficients of 10 annual leads (left side of the dashed line) and 10 annual lags (right side) of a regional waterworks opening indicator on the year-region fixed effects estimated in our preferred model (equation 1, column 7, panel a in table 2). The regional waterworks opening indicator is defined as the population-weighted mean of the number of opening events within a region over the number of cities that have not yet opened waterworks by that year. Black (gray) lines show 90% (95%) confidence intervals based on standard errors clustered at the region level.

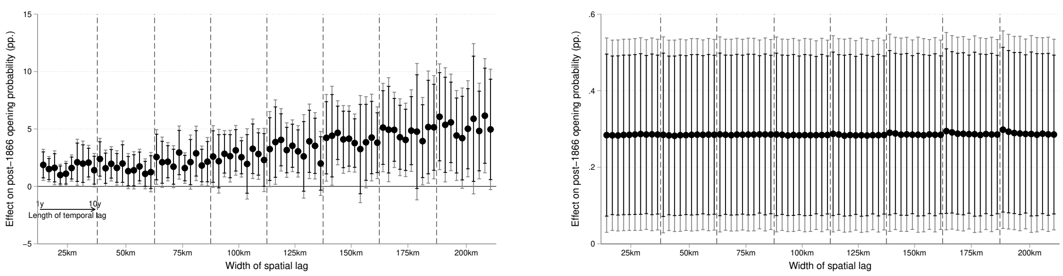
argue above. Since we do not have prior reasons to believe that imitation processes operate at specific spatial and temporal distances, we run this specification for s from 25 to 200 kilometres (in 25km-steps) and for temporal lags τ from 1 to 10 years. In figure A.6, we plot the 80 resulting imitation coefficients (ρ, panel a), and the jointly estimated effects of the cholera shock, which are now conditional on explicitly modelled imitation (β, panel b). We find evidence for imitation processes at various spatio-temporal lags (panel a), and show that explicitly controlling for these effects does not change or main effect estimate (panel b).

A.3.6 Confounding effects of earlier cholera outbreaks

Our baseline result–cities hit by the 1866 cholera epidemic exhibit higher post-1866 hazards of opening waterworks–could in principle be confounded if those cities had

Figure A.6: Imitation effects due to regional waterworks openings

(a) Imitation effect (ρ) **(b)** Cholera's effect (β)



Note: this figure plots the estimated coefficients from regression equation 13, using spatial lag radii from 25 to 200 kilometres, and temporal lag lengths of 1 to 10 years. Coefficients are multiplied by 100, thus denoting percentage point changes. Black (gray) lines show 90% (95%) confidence intervals based on standard errors clustered at the region and year levels. See text for further details.

already been exposed to cholera in earlier pandemics and thus were on a higher adoption trajectory anyway. In this section, we show that explicitly controlling for earlier cholera exposure between the 1830s and 1850s does not affect our estimate for the 1866 outbreak. Moreover, we find that these earlier outbreaks did not systematically increase affected cities' probability of opening waterworks.

We start by estimating the isolated effect of earlier outbreaks on the post-1866 hazard by substituting the 1866 cholera exposure indicator (C_i) for an indicator that measures cholera exposure in 1831-1832, or 1848-1850. Thus, we focus on the most deadly consecutive waves in the 2nd and 3rd pandemics, comparable to our 1866 shock measure (which includes the minor outbreaks in 1865 and 1867).[26] The results in table A.16 show that, in isolation, exposure to the 2nd pandemic had no statistically significant effect on post-1866 waterworks adoption (column 1), while the 3rd and 4th pandemics each did, with comparable magnitudes (columns 2 and 3, respectively).

However, cholera exposure during the third and fourth pandemics are highly correlated ($\rho = 0.45$), such that the respective coefficients partly capture each other's effect. In column 4, we estimate a "horse race" model that includes all three indicators

[26] In our urban sample, 28,479 people died of cholera in 1831-1832 (0.37 per 100 inhabitants) and 68,566 died in 1848-1850 (0.68 per 100). These deaths account for 72.96 and 52.79% of all urban cholera deaths during the 2nd and 3rd pandemic, respectively. For comparison, 75,046 (0.56 per 100) cholera deaths were registered in 1865-1867, accounting for 78.78% of the 4th pandemic.

simultaneously, and find that only the 1866 epidemic has a significant effect, with a point effect only slightly smaller than our main estimate (column 3). In addition, since the coefficients from such a "horse race" are likely partly collinear because exposure across pandemics was not mutually exclusive and many cities were hit repeatedly, we estimate a separate effect for each of eight possible, and mutually exclusive, exposure configurations ($C^{2\text{nd}}, C^{3\text{rd}}, C^{4\text{th}} \in \{0,1\}^3$). We present the results of this exercise in column 5 by reporting three linear combinations of the underlying coefficients, instead of a separate coefficient for each of the eight categories aforementioned. These synthetic coefficients approximate how much flipping a city from unexposed to exposed in a given pandemic changes the post-1866 waterworks adoption hazard, *ceteris paribus*.[27] We find that only the coefficient referring to the 1866 shock is statistically significant. We conclude that correlated exposure to earlier cholera shocks does not drive our main results.

[27]Each linear combination is a weighted average of the four differences between two coefficients that vary exposure to a particular pandemic while holding exposure to the other pandemics constant. For instance, the linear combination for the 2nd pandemic is $w_A(b_{100} - b_{000}) + w_B(b_{110} - b_{010}) + w_C(b_{101} - b_{001}) + w_D(b_{111} - b_{011})$, where b_{100} is the category effect for cities exposed only during the 2nd pandemic (and all other b's defined analogously, with $b_{000} = 0$ being the excluded reference category), and the w's are sample shares for each of the four backgrounds, summing to one. Statistical significance is estimated using the delta method.

Table A.16: The impact of the 1865-67 vs. the 1831-32 and 1848-50 outbreaks

	2nd only	3rd only	4th only	Horse race	Decomposition
$C_i^{2\text{nd}}\mathbf{1}_{t>1866}$	0.3589			0.2292	0.3559
	(0.129)			(0.314)	(0.244)
$C_i^{3\text{rd}}\mathbf{1}_{t>1866}$		0.3295**		0.2269	0.1188
		(0.047)		(0.146)	(0.434)
$C_i^{4\text{th}}\mathbf{1}_{t>1866}$			0.3912***	0.3343**	0.3128**
			(0.008)	(0.017)	(0.018)
Baseline controls and FEs	✓	✓	✓	✓	✓
Mean control (post-1866)	1.141	1.141	1.141	1.141	1.141
Cities	2552	2552	2552	2552	2552
R-squared	0.095	0.095	0.095	0.095	0.095

Note: This table shows the results of estimating various regression models that account for the effects of cholera outbreaks before 1866. The models in columns 1-3 apply equation 1, using cholera exposure dummies (C_i) based on the 1831-32 (col. 1), 1848-50 (col. 2) and 1865-67 (col. 3) outbreaks. The estimate in column 3 is thus identical to our main result (see table 2). In column 4, we include all three indicators simultaneously. The estimates presented in column 5 are linear combinations of eight regression coefficients obtained from a mutually-exclusive category model (see text for details). Coefficients have been multiplied by 100, thus measuring percentage points. Numbers in parentheses are p-values based on standard errors clustered at the city and year levels (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$), and computed via the delta method in column 5. R-squared reports explained variation after absorbing the fixed effects.

A.4 Validation exercises

A.4.1 Building-stock-to-agricultural-land-values as a measure of elite background

In this section, we validate our city-level measure for the relative influence of urban commercial-industrial interests over agricultural ones. As discussed in section 3.3, our proxy for the presence of commercial elites is the ratio of the value of the urban non-residential building stock to the combined value of these assets and agriculturally used land, both measured in the early 1860s following Prussia's comprehensive building and land tax reform (see section B.5 for a list of sources and further details). To validate this measure, we construct a sample with detailed data around the year 1907 for 110 Prussian cities from Silbergleit (1908) for which we observe various alternative indicators indicative of commercial elites. The idea is to test our measure's temporal stability and its correlation with alternative direct and indirect measures of the relative influence of commercial-industrial interests in municipal decision-making. While our validation sample consists of Prussia's largest cities, it contains a mix of places dominated by commercial-industrial elites, and by agricultural elites.[28] Table A.17 presents the correlation of our measure with six variables, always conditional on region fixed effects. All variables are log-transformed so that the coefficients are comparable elasticities.

Table A.17: Elites' economic background measure: Validation regressions

	Outcome: Share of non-residential building stock in total asset value, 1860s		
Correlate:	Same measure, 1905	Mun. tax income p.c., 1905	Eligible voters share, 1907
Elasticity	3.601***	0.330*	0.173
	(0.000)	(0.052)	(0.773)
Constant	-0.789***	-0.0952	-0.964
	(0.000)	(0.880)	(0.426)
R-squared	0.56	0.36	0.32
Cities	106	107	101
Correlate:	Capital income share, 1907	Council's non-rentier share, 1907	Liberal parties' vote share, 1908
Elasticity	1.293**	2.695**	0.189*
	(0.044)	(0.037)	(0.093)
Constant	-0.607*	-0.782***	-1.165***
	(0.096)	(0.001)	(0.000)
R-squared	0.38	0.41	0.45
Cities	109	83	96

Note: This table presents correlations between our measure for the presence of commercial elites in cities–the share of the non-residential building stock in the total value of these assets and agriculturally used land–and six correlates that help to validate our proxy. We base inference on heteroscedasticity-robust standard errors and include region fixed effects in the regressions. While the total city sample for the validation exercise is 110, the number of observations is smaller in each regression due to missing data. See text for more details.

The first regression shows that our measure, based on asset valuations in the 1860s, is highly correlated with the same metric using 1905 valuations, despite an overall increase in value of the non-residential building stock relative to agricultural

[28]The most commercial city in this dataset is Altona, a trading port city neighbouring Hamburg. The most agricultural city is Allenstein in East Prussia, a small agglomeration surrounded by Junker's estates.

land between the 1860s and 1905.[29] We conclude that our measure captures a stable relationship that is not contingent on the specific economic environment of the 1860s. Next, we find a statistically significant correlation (at the 10% level) with municipal fiscal capacity, proxied by tax income per capita. This suggests that the economic background of local elites also affected municipal budgets, underlining the importance of controlling for fiscal capacity (as we do in our main regressions). The third regressions indicates that our measure does not simply reflect the over-representation of economic elites in municipal politics. As discussed in section 2.2, municipal politics were shaped by the three-class voting system favouring highly taxed individuals. While our measure captures the economic background of these decision-makers, it does not correlate with the general bias in representation, as proxied by the share of eligible voters in the total city population.[30]

Our next three validation exercises use alternative proxies for the influence of commercial-industrial interest groups relative to agricultural ones at the municipal level. We use official state tax returns listing the source of taxable income for individuals with an annual income above 3000 marks, who are expected to overlap with the wealthiest municipal voters due to similar income-based voting rights.[31] The tax returns distinguish income from capital assets, land and buildings, business assets, and labour. We find that the share of income from capital and business assets is reasonably correlated with our target measure.[32] We also examine the composition of elected local councils using individual city directories from 1907 or the closest available year. These directories provide the occupations of council members, from which we compute the share engaged in commercial-industrial activities.[33] This share is

[29]Following the 1861 reform, building and land values were reassessed roughly every 15 years (in 1878, 1893, and 1908). Comprehensive city-level data was published only for the original assessments from the 1860s.

[30]Across the 101 cities for which we can compute this measure, the share of eligible voters in the total population in 1907 ranged from 5 to 20%.

[31]The share of state income tax-paying individuals with incomes above 3000 marks varied from 0.08 to 1.7%.

[32]While the fiscal category of income from land and buildings includes earnings from non-agricultural assets, all income from agricultural assets falls under this category.

[33]We match occupations to the official Prussian occupational nomenclature from 1907, classifying those in sectors B (Industry) and C (Trade and Commerce) as commercial-industrial council members, while we classify those in sectors A and F as these who derive wealth from agricultural and residential real assets. In our measure, we ignore council members that fall outside these category,

highly correlated with our measure, suggesting that council members indeed represented the economic interests of their voters. Finally, we calculate the vote share of liberal (free-trading) parties relative to conservative (protectionist) parties in the Prussian state elections of 1908. Based on the cleavage between protectionist agricultural interest groups and more free-trading commercial-industrial interests in 19th century Prussia (Lehmann 2010), the positive correlation of our measure supports its validity as a proxy for the economic background of city elites.[34]

A.4.2 Literacy rates as a measure of human capital

Our mechanism regressions in section 5.2 use literacy rates, measured for Prussian cities in 1871, to proxy the human capital stock at the time of the 1866 cholera epidemic. In this section, we address the concern that Prussian authorities systematically discounted–or outright ignored–literacy in minority languages other than German by, for instance, deeming communities around Posznan, Southern East Prussia, and parts of Upper Silesia and West Prussia as "illiterate" despite their literacy in Polish. A variable with this bias would partly proxy for distinct ethnic, linguistic or regional characteristics that could correlate with sanitary development for reasons other than human capital development. While we directly control for many potential confounders, including market integration and economic development, others remain unobserved, such as credit constraints particular to ethnic minority communities (Suesse and Wolf 2020). Regional fixed effects, included in all our regressions, only partly address this potential bias because cities *within* regions varied widely in their ethnic composition.

We gauge whether low official literacy rates in Polish-speaking territories[35] reflect

i.e. civil servants and professionals. See Albers and Kappner (2023) for more information on these sources. Across cities, the average share of commercial-industrial council members is 84%, ranging from 62 to 100%.

[34] Similar to municipal elections, Prussian state elections were also influenced by a three-class voting system, leading to a similar bias towards wealthy citizens' interests. We do not consider votes for parties that did not align clearly with the protectionist/free-trade cleavage, such as the the Social Democrats and the Catholic Center party, both of which received substantial vote shares across our sample cities.

[35] Figure B.7 shows low officially-measured literacy in the Polish-speaking Eastern territories, the Lithuanian-speaking North-East, and in French-speaking border regions in the West. The other two linguistic minority regions in Prussia, Danish-speaking Northern Schleswig, and Sorbian-speaking

genuine lower human capital levels rather than measurement bias by correlating literacy rates with primary schooling intensity. More specifically, we use the number of public elementary school teachers per eligible child in the three closest comprehensive education censuses published by Prussian authorities in 1849, 1864 and 1878, with different levels of spatial resolution. Primary schooling intensity is less likely to exert a regional-linguistic bias since our sources (KSB 1851, 410-531, for 1849, KPMdUA 1867 for 1864, Petersilie 1882 for 1878) count teachers regardless of their teaching language.

We constructed the relevant variables as follows. For 1849 and 1878 we count all teachers at grade schools (*Elementar- oder Volksschule*), "intermediate" schools for boys (*Mittelschulen für Söhne*) and girls (*Höhere Töchterschulen*). These school forms, with students in the age bracket up to 14 years, focused on basic education, including literacy. For 1864, the source only refers to "elementary school teachers" in total. The eligible populations in 1849 and 1878, children between the age of 6 and 14, was calculated using age-bracket census population counts for 1849 (KSB 1851a, 2–253) and 1875 (Königliches Statistisches Bureau in Berlin 1877). For 1864, the eligible population is reported in the source. Prussian authorities began systematically suppressing Polish-language education facilities after 1871 as part of the *Kulturkampf* (Kersting et al. 2020). This could have resulted in biased statistics after 1871, but is unlikely to be reflected in 1849 statistics.

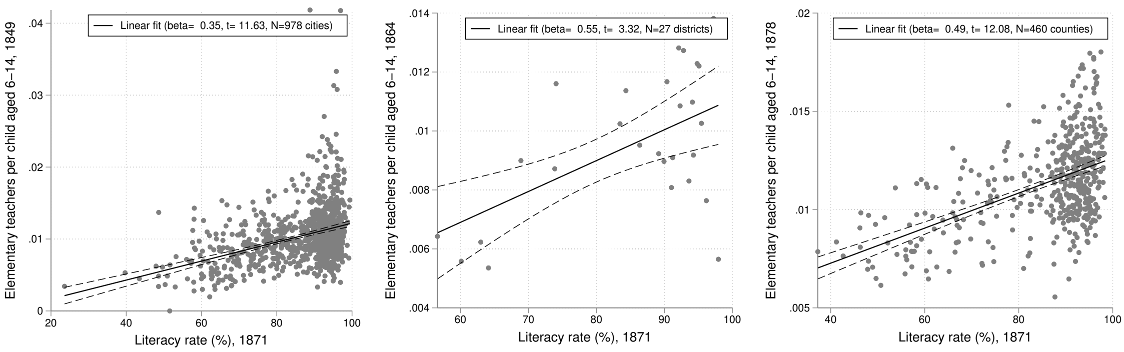
Figure A.7, panel a, demonstrates that literacy rates in 1871 correlate well with elementary schooling intensity 1849 in the ca. 980 cities that belonged to Prussia at that time. Panels b and c present similar correlations at the district-level for 1864, and at the county-level for the whole of post-1866 Prussian territory for 1878. In sum, this evidence indicates that human capital *production* strongly correlates with measured levels in 1871 as indirect evidence that variation in literacy rates primarily reflects variation in human capital.

We provide further evidence for the robustness of our literacy-based human capital measure in table A.18. We repeat the regression approach outlined in section 5.2, while restricting the city sample and adding different controls designed to block confounders that could bias literacy measures. In column 1 we reproduce the main

Lusatia, do not exhibit particularly low literacy rates.

Figure A.7: Validation of literacy rates as a measure of human capital

(a) Schooling intensity, 1849 **(b)** Schooling intensity, 1864 **(c)** Schooling intensity, 1878



Note: The three panels show unconditional correlations between 1871 literacy rates and primary schooling intensity, the number of teachers per eligible child (aged 6-14) at public elementary schools. Panel a plots 978 cities located within the pre-1866 Prussian borders. Panels b and c plot 27 districts an 460 counties, respectively, as city-level information is not available for these years. See text for sources and further details.

result of our main analysis (see table 4, panel B, column 7). In columns 2 and 3, we exclude cities located in counties with an above-10-percent non-German or Polish-speaking population, respectively, using data from the 1867 Prussian census reported in Belzyt (1998). Reassuringly, coefficients barely change when excluding cities in regions where literacy rates could be artificially low due to mismeasurement, in line with our findings above. In columns 4 and 5, we additionally control for the city-level share of Protestants. In the absence of city-level language or ethnicity statistics, religious composition serves as a proxy for the German-speaking population (Kersting et al. 2020). Because Prussia held large Catholic German-speaking territories (such as the Rhineland), we allow the estimate to vary by region in column 5. If literacy rates would primarily proxy for the local prevalence of German language, we would expect their coefficients to change once another, arguably stronger, language proxy is added. This is not the case, however. Finally, in column 6, we use an entirely different measure for the human capital stock: per-capita students in higher schools in 1866.[36] This measure focusses on upper-end human capital production, perhaps

[36]We count students in vocational and university-preparatory high schools (*Realschulen*, *Gewerbeschulen* and *Gymnasien*), and teachers' colleges (*Schullehrer-Seminarien*) listed in

explaining the much larger effects we find relative to our baseline in column 1.

Table A.18: Human capital: Validation regressions

	(1) Baseline	(2) W/o minority counties	(3) W/o Polish counties	(4) Protestantism	(5) Protestantism × region	(6) Higher schooling
$C_i \mathbf{1}_{t>1866} \times$ Literate	0.584**	0.599**	0.579**	0.591**	0.545**	1.704***
	(0.014)	(0.015)	(0.015)	(0.012)	(0.042)	(0.007)
$C_i \mathbf{1}_{t>1866} \times$ Non-literate	0.0137	-0.158	-0.0231	0.0244	0.0184	0.281**
	(0.939)	(0.744)	(0.962)	(0.892)	(0.938)	(0.043)
All main controls and FEs	✓	✓	✓	✓	✓	✓
City type-specific trends	✓	✓	✓	✓	✓	✓
p: Literate > Non-literate	0.021	0.079	0.131	0.022	0.084	0.005
Control mean (post-1866)	0.717	0.831	0.792	0.717	0.717	0.712
Cities	1419	1107	1154	1419	1419	1424
R-squared	0.110	0.115	0.113	0.110	0.111	0.162

Note: The table presents the estimated coefficients from various models testing the robustness of our finding that cities with a larger human capital stock had higher waterworks opening probabilities after the 1866 cholera shock (see table 4, panel B, column 7; here, reproduced in column 1). See text for details on each column's model.

A.4.3 Local economic development indices

We validate our measure of local economic development (see section B.3) by computing its correlation with three other measures capturing important dimensions of development around the same time. These are: (1) Assessed land values plus commercial-industrial building stock values per capita in 1864 (for 1473 Prussian cities), (2) direct state taxes p.c. in 1876 (for 161 Prussian cities, from Herrfurth (1878)), (3) direct state taxes p.c. in 1883 (for 1175 Prussian cities, from Herrfurth and von Tzschoppe (1884)). In table A.19, we document strong correlations with direct state taxes p.c. (columns 2 and 3), explaining up to 25% of the variation in our sample. The correlation with total asset values p.c (column 1) is weaker, though still highly significant.

Mushacke (1866). We do not include students at schools for the deaf and blind (*Taubstummen- und Blindenanstalten*), military schools of various sorts, polytechnical schools, and secondary modern schools (*Hauptschulen*, *Bürgerschulen* and *Töchterschulen*). Due to lack of data, we cannot compute the number of eligible children for these schools around 1866; thus, instead, we compute the ratio in per-capita terms.

Table A.19: Local economic devlopment index: Validation regressions

	Asset values p.c., 1864	Direct state taxes p.c., 1876	Direct state taxes p.c., 1883
Economic development index	0.0846***	0.264***	0.198***
	(0.000)	(0.000)	(0.000)
Intercept	0.0804***	0.122***	0.120***
	(0.000)	(0.000)	(0.000)
adj. R2	0.06	0.21	0.24
Cities	1473	161	1175

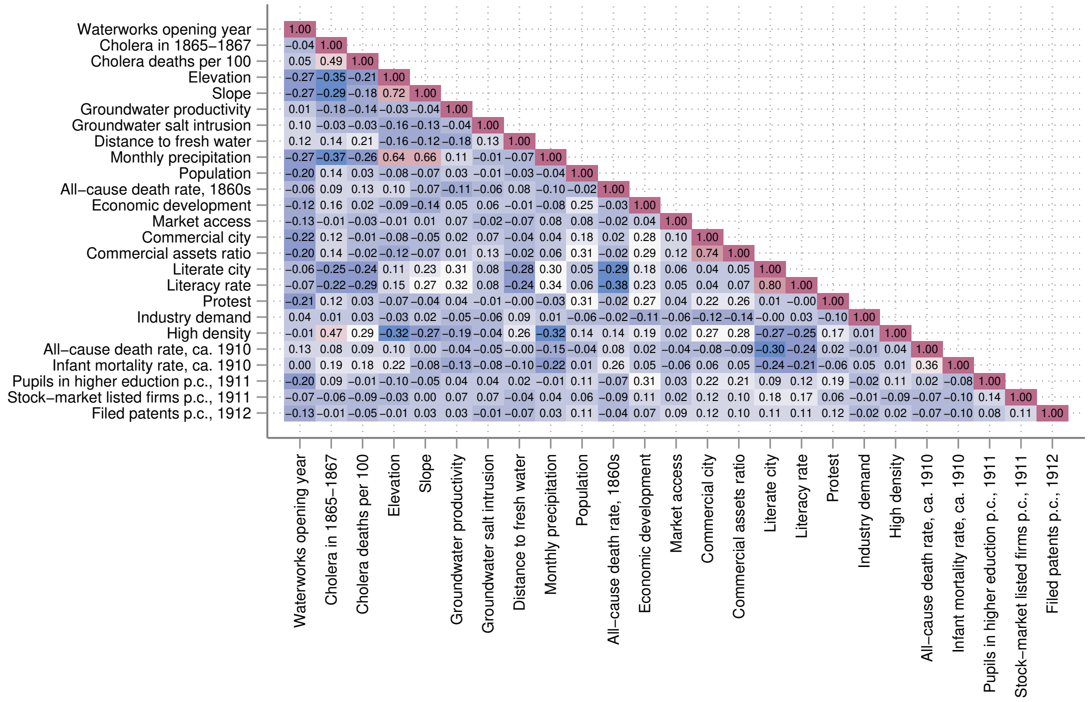
Note: This table shows correlation coefficients for regressions of three alternative measures of local economic development on our preferred measure (as discussed in section B.3). Each of the three measures is normalized between 0 and 1 to allow for a comparison of coefficient magnitudes. See text for further details and sources.

B Data appendix

B.1 Descriptive statistics

We collected data for all 2,685 cities within the German Empire that either held a town charter or are mentioned in the historical encyclopedia of German cities (*Deutsches Städtebuch*, see section 3.1). For about two thirds of these cities–mainly those located within Prussia (the largest German state)–we collect additional data. This section discusses the sources in detail and presents additional summary statistics. Table B.1 show summary statistics for the universe of urban Germany ("All"), our main regression sample ("Main"), and the Prussian cities constituting our mechanism regression sample ("Mechanism"). Figure B.1 visualizes the bivariate correlations between all these variables (for the largest sample available for each variable).

Figure B.1: Correlation matrix



Note: This matrix shows the pairwise linear correlation between variables, based on all available observations. Colour shading follow the strength of correlation with strongly positive (negative) correlations in red (blue).

B.2 Sources

B.2.1 Waterworks

Some conceptual considerations Measuring the transition from traditional sources of water supply to modern waterworks is challenging. One approach considers investment flows measured in monetary terms. This method has the advantage of providing an indicator (e.g. cost of building waterworks) that is directly comparable with local fiscal capacity, which in turn is an important constraint of publicly-funded piped water provision. However, gathering this information systematically for waterworks that differed significantly in their location and timing presents several non-trivial challenges. First, monetary investment flows need to be deflated by appropriate price deflators that might not be available at the local or regional level. Second, the construction costs of waterworks were influenced by technical engineering aspects related to geography (soil ruggedness or elevation), technology (water filtration, pipe material) or capacity (system pressure, water supplied per person). Third, information on costs incurred by cities is not available in a consistent manner for each phase of these projects, such as expenditures on project design, initial exploratory perforations, main works (water storage and distribution), unexpected setbacks, etcetera. As a result of these difficulties, similar projects might exhibit substantially different monetary values, even if they represented the same local achievement: the start of a transition to piped water.[37]

Our *date* approach is intended to measure a comparable phenomenon across cities and time. Consider the waterworks in Hamburg and Berlin as an example. Though both were designed by British engineers and were built around the same time (in 1849 and 1856, respectively), they did differ in some respects: the former did not supply filtered water until 1893 (Grahn 1904, 329) while the latter did so since the beginning (Kappner 2024); Berlin relied on private initiative, while the city of Hamburg owned the system; and city coverage was more comprehensive in Hamburg than in Berlin. By using information about the timing of their opening, we put the focus on the great achievement that modern water infrastructures meant for contemporaries and the dynamics they set in motion that would ultimate result in widespread access to piped water. Indeed, most citizens in Berlin and Hamburg were supplied by their respective water systems a few decades after the their construction (Kappner, 2024, 13; Grahn, 1898-1902, vol. 2, 780). A different issue applies to cities with relatively advanced supply systems relying on gravity and wooden pipes to distribute water. Whenever possible, we obtained information about these systems and assessed whether later ones represented major improvements in terms of water pressure and delivering capacity, as a result of, for instance, the installation of steam-driven pumping stations and iron pipes, or the construction of large reservoirs. Although the issues raised here are not

[37]This is not to say that some of these considerations are unimportant. For instance, taking into account when citizens have access to filtered water is relevant for studying the drivers of mortality (Cutler and Miller 2005; Anderson et al. 2022). However, our goal is to understand how political dynamics influence the establishment of large-scale modern water infrastructures, regardless of their more specific technical characteristics.

unique to our approach, it is important to keep them in mind when interpreting our results.

Sources Starting with big- and medium-sized municipalities we use historical waterworks compilations, such as Grahn (1878), Grahn (1883), Grahn (1898-1902), Grahn (1904) and Salomon (1906-1911). Municipal sanitary descriptions often mention when the construction of waterworks started and when they began supplying water to citizens; we recorded the latter. These encompassing publications cover most big- and medium-sized cities of the German Empire. However, information for some places, especially smaller ones, can be incomplete or non-existent; and towns building waterworks after ca. 1911 are missed altogether. We fill these gaps with additional sources, such as Brix et al. (1934), Guttstadt (1900) and the journal published by the German association of gas and water experts (DVGWF, various years). For the Rhein province and the industrial area of Upper Silesia, we use Selter (1911) and Geisenheimer (1913). To assess the quality and consistency of our dataset, we cross-check dates from these sources with the aforementioned works by Grahn, from which a significant number of dates were retrieved (see appendix B.2.1). We find that they largely overlap. In some instances, we use information on when loans were taken up from the data put together, mostly, by Oskar Tetzlaff in various publications (KPSLiB 1914-1920). We proceed carefully with this source, since it only refers to publicly financed waterworks. Although the vast majority of these infrastructures were public by the turn of the 20th century, some municipalities obtained water from private companies operating in some industrial areas. Finally, we fill remaining gaps with a variety of local histories containing information on sanitary infrastructures, which in some cases have been published by lesser known publishers, local authorities, water supply companies or enthusiasts of the topic; in some cases, the information was retrieved from their websites.

For a reduced number of small cities, these journal publications only mention when construction works were in process or about to begin. Due to lack of better data, we record these. We do not think this introduces a significant measurement bias in our data set, since the experience of towns of similar size suggests that operation mostly started about one to three years after construction start. Also, these dates largely refer to the period beyond 1891, outside of our main analyses. For instance, this assumption has limitations when applied to cities that reported plans to build waterworks shortly before the First World War because, in some case cases, they experienced considerable delays due to the war and post-war turbulence and economic hardships. For these cases, we carefully check alternative sources to arrive at realistic opening dates.

From our target sample (2,685 cities) we could determine an exact opening date for 2,235 locations. For 317 additional cities, we know that their waterworks opened only after 1914. The resulting sample includes 2,552 cities–95% of the target sample. A detailed breakdown of city-specific sources is available from the authors upon request.

As figure B.3 shows, our observations are scattered throughout the German Empire and provide a representative coverage of the experience at the time across a territory that includes, apart from Germany, parts of modern-day France (Alsace-Lorraine), Denmark (Southern Schleswig), Western Poland and the Russian Kaliningrad area (former East Prussia). There are highly urbanized and industrialized clusters (e.g. located in Western Germany, and in particular the Prussian provinces of Rhineland and Westphalia, Berlin, Upper Silesia and, to a lower extent, in the Kingdom of Saxony), as well as rural areas.

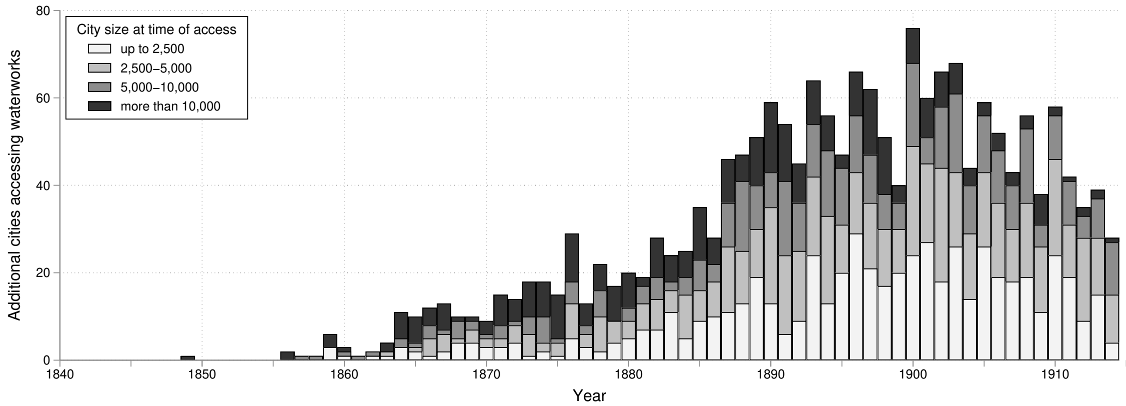
B.2.2 Cholera

We employ a broad collection of historical sources to reconstruct the spatial distribution of deaths from cholera at the municipality level, covering the years 1865 to 1867, and the preceding and subsequent years (from 1831 to 1895). These are discussed in detail in Kappner (2025), although in this appendix we want to highlight several aspects of our sources and procedures.

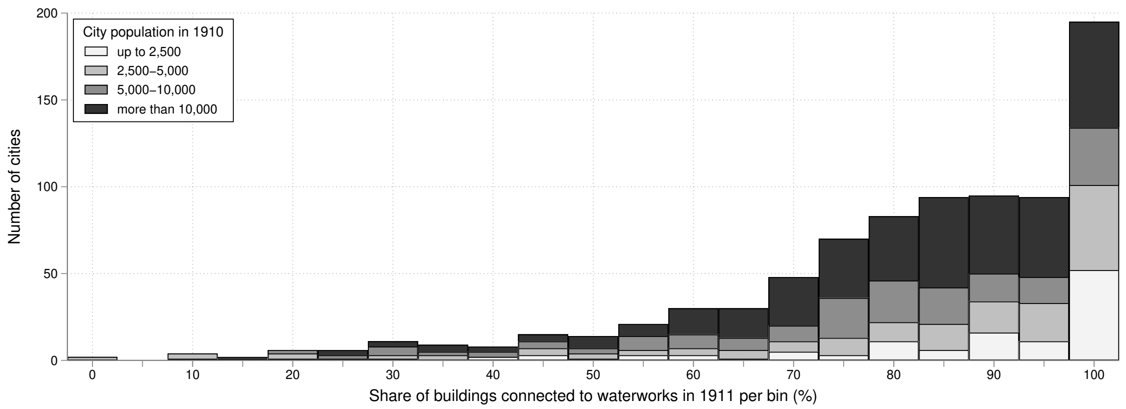
There are no comprehensive city-level epidemic reports or cause-of-death statistics covering all of the German Empire until the end of the 19th century. The official Prussian report on the 1866 cholera epidemic, Engel (1869), only contains county-level data and ignores the the so-called "'new provinces", i.e. the four German states that were annexed by Prussia after the 1866 war (the Kingdom of Hanover, the Duchy of Nassau, the Electorate of Hesse and the Free City of Frankfurt). To recover municipality-level data for Prussia, we identified and transcribed the original archival

Figure B.2: Timing and within-city diffusion of piped water provision by city size, 1840-1914

(a) Timing of access to waterworks by city size, 1840-1914

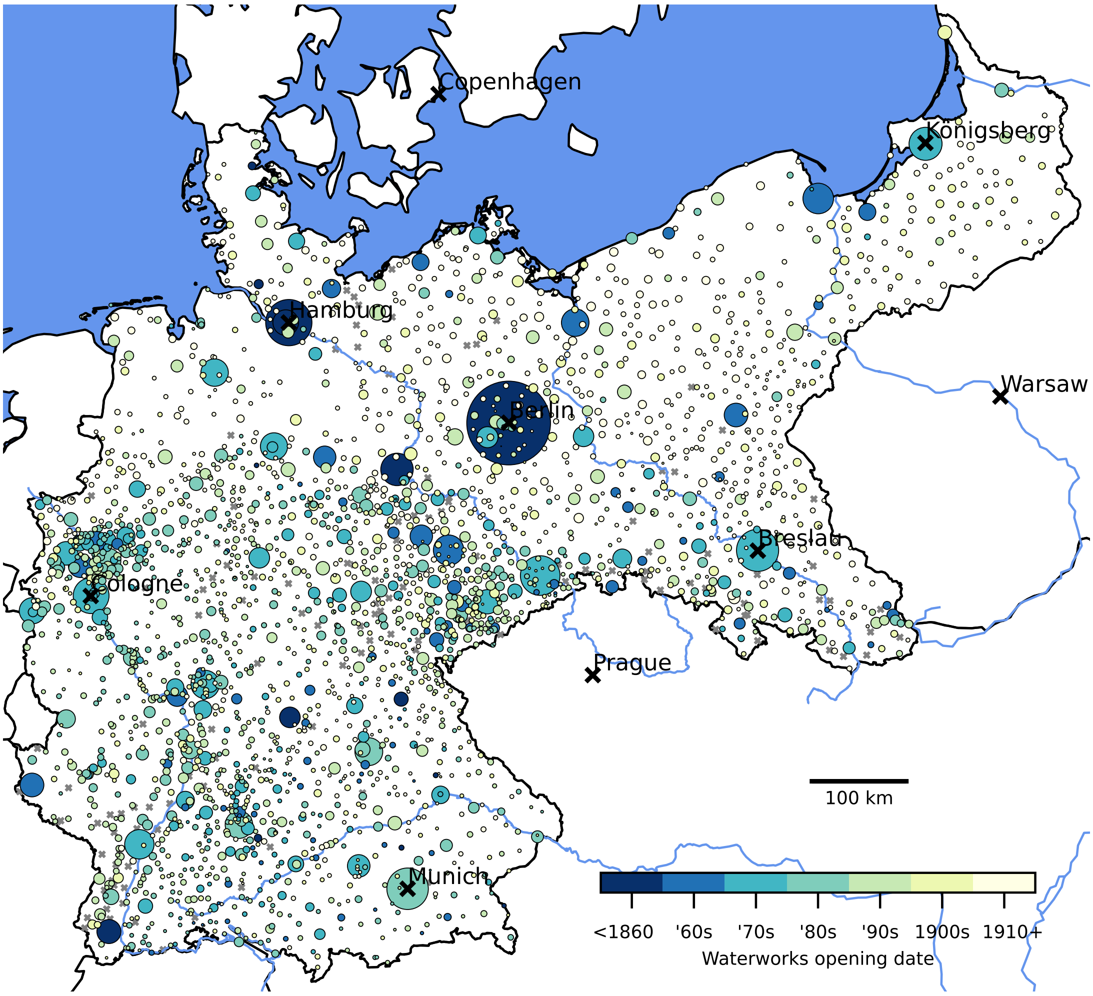


(b) Within-city diffusion, 1911



Note: Panel a shows the number of municipalities that gained access to waterworks in a given year. Cities are classified by their population size at time of access. 1,888 cities acquired access until 1914. 664 cities acquired access after 1914, or at an unknown date. A similar graph for the US context is produced in Beach (2022). Panel b shows the distribution of municipalities that acquired a connection by 1911 with respect to the share of buildings that are connected in that year (in bins of 5 percentage points). Cities are classified by their population size in 1910.

Figure B.3: Cities by waterworks opening date



Note: This map shows the opening timing of waterworks by decadal bins. Circles are cities, with circle size proportional to their 1866 population. Black cross markers place the cities of Cologne, Hamburg, Munich, Berlin, Breslau, Königsberg, and foreign capitals. Gray cross markers show the location of cities where data is missing. Black lines show the German Empire in its 1871 borders and blue lines show major rivers.

records, preserved in the Prussian State Archive (*Geheimes Staatsarchiv Preußischer Kulturbesitz*). Fortunately, these records also cover the "new provinces", effectively yielding municipality-level estimates for about 75% of our target sample.

While the Prussian archival cholera records are uniformly structured and generally well-preserved, three are some limitations in the data. First, in some instances data was missing so we had to infer municipality-level cholera deaths employing county-level information. Second, for some administrative districts (e.g., Bromberg, Gumbinnen or Potsdam) the archival records for the 1866 outbreak report data separately for cities with a legal town charter, but aggregate information for all remaining non-chartered municipalities to the county level. While this reduces somewhat the spatial precision with which we can locate cholera deaths in rural settings, we are still able to measure very precisely the outbreak for almost all cities in the target sample. And third, the archival records are not always consistent with the work by Engel (1869). While summation errors in Engel (1869) account for some discrepancies between his county-level totals and those implied by our sources, we still find minor inconsistencies for the administrative districts of Frankfurt, Koblenz, Marienwerder and Potsdam. In these cases, we deem our archival information of higher quality.

To estimate the extent of the 1866 cholera outbreak in the remaining German states, and the other cholera outbreaks before and after 1866, we rely on a variety of sources, including official reports and pieces in medical journals containing complete municipality-level information. A complete source list can be found in Kappner (2025).

B.3 Controls

Population counts We get population counts for our sample cities at three-to-five year intervals between 1831 and 1895 from Kappner (2025). We add to these population counts for the census years 1900, 1905, and 1910, drawing on various census summary publications published by the German states and the Imperial statistical authorities. In a few cases, we interpolate population figures for some years as official census publications did not always include population figures for non-chartered, small cities. Censuses were typically conducted at the beginning of December. All population counts include military personal, but exclude inhabitants that were not present at census enumeration day. To arrive at a realistic population number in 1866, we take the average of the figures reported for 1864 and 1867, and then add cholera deaths in 1865-1867.

Pre-epidemic crude death rates (CDR) We measure city-level mortality rates in the early 1860s using a wide range of sources. Whenever possible, we trace the total number of deaths in the three years 1862-1864 and divide it by the 1863 population level (which we interpolate from the closest census-measured population levels, usually in 1861 and 1864). In some instances, especially for small or non-chartered cities, we have to approximate the pre-epidemic CDR by using the average CDR of a larger encompassing administrative unit, usually a county. In other cases, we have to chose a different set of reference years because 1862-1864 data is not available. Our CDR estimates for cities located in the old Prussian provinces (about half of the sample) rely on city-level death counts in 1849, which we scale such that each county's CDR in 1849 matches its CDR in 1862-1864. Table B.2 provides an overview.

Economic development We proxy each city's economic development at the onset of the 1866 cholera epidemic with a Germany-wide comparable index derived from reported local industry structure and tax business taxation statistics. Our procedure builds on a two-stage modelling approach. First, we train a predictive model on county-level data from Prussia in 1882–83, mapping sectoral firm composition into assessed per-capita direct state taxes—covering land, buildings, income, and business profits.[38] Second, we apply these estimated elasticities to the industrial structure of cities observed the early 1860s, as recorded in commercial business directories. This allows us to extrapolate a consistent estimate of relative fiscal capacity to cities for which no direct tax records survive.

Our main source are the 18 volumes of J. C. Leuchs' business directory (*Adreßbuch aller Länder der Erde der Kaufleute, Fabrikanten, Gewerbtreibenden, Gutsbesitzer*), published between 1860 and 1867.[39] While business directories are never perfectly complete, their coverage was commercially motivated and revenue-driven, making systematic regional under-reporting unlikely. An additional advantage is their uniform coverage across all German states, helping to overcome the statistical fragmentation

[38] We focus on state taxes rather than municipal taxes because these were assessed in a homogenous way throughout Prussia and thus abstract from local differences in tax systems.

[39] Due to missing sources, we use slightly later volumes for Mecklenburg (1871), Schleswig-Holstein (1869), and Alsace-Lorraine (1871). For a small number of cities in Alsace-Lorraine and Schleswig-Holstein we need to use 1875 and 1877 volumes as those cities were not formally part of Germany in 1869/1871.

of the period.[40]

The first part of our approach consists of training our model on 465 Prussian counties, using per-capita firm counts in 134 standardized industry categories drawn from the 1882 industrial census (KSB 1886). We then match business directory data from the 1860s to the same industry taxonomy and compute per-capita firm densities for each city. Applying the 1882–83 elasticities yields predicted per-capita direct taxes in the early 1860s.

We use a gradient boosting regression model to estimate the nonlinear and potentially interactive relationship between sectoral firm densities and per-capita direct taxes, the latter being reported in Herrfurth and von Tzschoppe (1884). Unlike industry shares, per-capita firm counts capture formal economic density while including the size of the non-reported sectors (agriculture and government) as a residual category. Through gradient boosting, we allow for complex interactions between sectors–for example, complementarities between finance and commerce or diminishing fiscal returns to clustering in low-margin activities. The dependent variable, the log of total assessed direct taxes per capita, encompasses the five core direct taxes levied in Prussia in 1883: *Grundsteuer* (land tax), *Gebäudesteuer* (building tax), *Klassensteuer* (class tax), *Klassifizierte Einkommensteuer* (classified income tax), and *Gewerbesteuer* (trade tax). These taxes are proportional to the value of land and property, personal income, and business profits, therefore closely proxying local economic development.

Our approach rests on several assumptions: (1) that the structural relationship between firm composition and fiscal capacity remained stable between the 1860s and early 1880s; (2) that the model generalizes beyond Prussia (which covers about two thirds of Germany) to other German states; (3) that county-level patterns are informative for city-level variation; and (4) that business directories offer reasonably complete and non-selective coverage of formal economic activity. These assumptions are plausible in the German context. Direct taxes in the 1860s–across most states–were similarly structured around land, property, income, and business activity. Moreover, industry categories, tax logic, and sectoral structure remained broadly stable between

[40] To handle the few cities missing from the commercial directories altogether, we conservatively assign the minimum predicted fiscal capacity observed among reported cities in the same region.

the early industrial era and the census benchmark used for calibration.

Market access We follow Donaldson and Hornbeck (2016) in defining each city i's market access in 1866 as

$$\text{MA}_i = \sum_{j \neq i} P_j t_{ij}^{-\theta},$$

i.e. the sum over all other locations j's populations P_j, each weighted by the inverse of pairwise travel costs t_{ij}^{θ}, where $\theta = 5$ is a trade elasticity parameter and t_{ij} is the travel time between i and j along the fastest route through the transport network.[41] The set of locations with population and travel time estimates includes the 2,685 German cities in our target sample, 851 German counties that aggregate the remaining rural population, and 940 sub-national regions in neighbouring countries.[42]

Population counts for German cities and counties are from Kappner (2025); those for non-German sub-national regions stem from a wide range of contemporary census reports published by European governments.[43] To estimate least-cost paths between locations, we rely on a detailed representation of the German transport network, including all railway lines, train stations, navigable rivers and ports that existed in 1866. Additionally, we approximate pre-industrial road-based travel through Özak (2018)'s *Human Mobility Index* cost layer. Details on sources and assumptions, and validation exercises, are provided in Gallardo-Albarrán and Kappner (2025).[44] We assume a speed of 52.2km/h on rail and 15km/h on inland waterways. The first value is interpolated from 6 decennial estimates of "typical passenger train speeds" for Germany between 1840 and 1890 (Freiherr von Schweiger-Lerchenfeld 1894, 17). The second value is calculated from reported speeds on the Rhine and represent an average of slow upstream and fast downstream speeds (Nasse 1905, 142ff.). Figure B.4, panel

[41] In their foundational approach, Donaldson and Hornbeck (2016, 832) assume an elasticity of $\theta = 8.22$, noting that the median estimate across a broad range of studies is 5.

[42] We calculate travel times from and to the non-city locations using the centroid coordinates of the respective areal units. In doing so, we set the minimum travel time to 15 minutes, thus avoiding unrealistically low distances for cases where some cities are located very close to the centroid of a surrounding county.

[43] Population counts for the exact year 1866 are seldom available. We use counts for the two closest census years and log-linearly interpolate population figures from them.

[44] We extend this transport network by also including all European railways lines outside Germany, based on a GIS dataset provided by Martí-Henneberg (2023).

d, shows the spatial distribution of this measure.

(Hydro-)geological city properties We rely on various sources to capture local (hydro-)geological properties. In particular, we employ the Earth Resources Observation and Science (EROS) Center's Shuttle Radar Topography Mission (SRTM) dataset to calculate the average elevation and slope within a 10km radius around each city.[45] Furthermore, we use the International Hydrogeological Map of Europe (Duscher et al. 2015) to calculate the share of land featuring highly productive aquifers, and the share of land with aquifers subject to seawater intrusion within a 10km radius.[46] We use the *WISE (Water Information System for Europe) Large rivers and large lakes* suite published by the European Environment Agency to calculate the distance to the nearest fresh waterw body.[47] Finally, we use WorldClim's historical climate database to calculate the average precipitation between 1970 and 2000 within a 10km radius around each city (Fick and Hijmans 2017).[48]

B.4 Regimental histories for the instrument

We trace regimental histories (*Regimentsgeschichten*) of almost all Prussian regiments that participated in the 1866 Austro-Prussian war. These regimental histories include information on the initial size and the daily whereabouts of each battalion during the 1866 campaign, either in tabulated form, through a map, or through a narrative report.[49] We use this to calculate how many unique soldiers passed through each city within a 1km radius between June and December of 1866, i.e. during the war. In the

[45] See https://doi.org/10.5066/F7F76B1X for the data source.

[46] This dataset is accessible at the European Commission's website. We code both highly productive aquifers and locally aquiferous porous or fissured rocks as productive (types Ia, IIa and IIIa).

[47] In particular, we draw 100 random points within a 10km radius around each city and calculate the average distance to the closest river or lake. The underlying dataset can be accessed at the EEA's website.

[48] The dataset can be accessed at WorldClim's website.

[49] We locate regimental histories with data on the trajectory for 102 out of the 129 active regiments. For the remaining 27 regiments, we approximate their trajectory using the observed movements of other regiments that served in the same division, adjusting for their known home garrisons. Whenever a regimental history does not clearly state the regiment's initial size, we use the standard magnitude of the respective regiment class, e.g. 1000 soldiers per infantry battalion.

process, we adjust the regiment size according to the number of deaths in each battle reported in Engel (1867). In our IV regressions, we additional control for each city's garrison status, which we transcribe from Niederstetter (1867).

B.5 Mechanism

Land and building values To approximate the relative influence of agricultural and commercial-industrial interest groups at the municipal level (for Prussia), we rely on official estimates of the value of agriculturally used land and the urban non-residential building stock (including built-on land). These estimates were produced as a result of a reform of Prussia's land and building tax code in 1861. In particular, we use the 1885 volume of Prussia's municipal directories (*Gemeindeverzeichnisse*), which reports the cultivated area and the average income less all costs of farming for farmland (*Ackerland*), meadows (*Wiesen*) and forests (*Holzungen*). We use this information to compute the total value of land used for these three purposes.[50]

To measure the value of the urban non-residential building stock, we start with data tabulated in KSBiB (1871). This source reports the value of this asset class for all chartered cities and a small subset of non-chartered cities, covering about 85% of our sample of Prussian cities. For the remaining non-chartered cities, we rely on approximations: For non-chartered cities in the old provinces (about 12% of the sample), we get the tax revenue from the total building stock from a series of official publications that were issued after the 1861 tax reform (e.g. KF 1870 for the district of Aachen). This *total* building stock includes both non-residential and residential buildings. Thus, we back out the approximate value of the purely non-residential building stock by multiplying the total building stock value with the share of non-residentially used over all buildings (as reported in the same publication) and dividing by 0.02, the tax rate for non-residential buildings. Finally, for non-chartered cities in the new provinces (about 3% of the sample), we use the aggregate value of the non-residential building stock for the rural hinterland of the respective county

[50]Unfortunately, the sources do not report the cultivated area and income from four additional classes of agriculturally used land defined by the Prussian tax law: gardens (*Gärten*), pastures (*Weiden*), water (*Wasserstücke*) and wasteland (*Ödland*). To the extent that these sources of income mattered, our measure of agricultural land value is thus incomplete. See Kopsidis and Wolf (2012, 640–645) for a discussion of this source.

(as reported in KSBiB (1871)) and multiply it by the respective city's share in the county's building stock (as reported in the municipal directories for the year 1871). We validate our asset-value-based measure in the appendix, section A.4.1.

Literacy rates We obtain the share of literate inhabitants at age 10 or older in all inhabitants from the 1871 Prussian census (KSB 1873; Becker and Cinnirella 2020). When computing this ratio, we exclude those with unknown literacy status (a rare case). We validate our literacy measure in the appendix, section A.4.2.

Other mechanisms We compute population density as inhabitants per inhabited building, collecting the latter from (KSB 1873). Using 1871 instead of 1866 building counts means that we slightly underestimate density. We measure whether a city experienced a protest, riot or insurgency between 1816 and 1866 using Tilly (1990), who provides a comprehensive city-level dataset based on national newspapers. Finally, we use the same data as for our economic development measure to identify cities that had steel and metal refining industries, dyeworks, textile finishing or paper industries in the early 1860s, following (Brown 1988).

B.6 Long-run outcomes

Within-city piped water coverage We draw on various sources to calculate the share of buildings with access to running water in 1911. The first is the work by Oskar Tetzlaff (KPSLiB 1914-1920) that contains information for Prussian municipalities on the number of connected lots to the water supply network, the length of pipes laid out and the amount of water supplied. For large- and medium-sized non-Prussian cities throughout the German Empire we use Neefe (1920) and DVGWF (1913). The detailed nature of these data does not allow us to cover as many cities and decades as we did with initial operation dates. However, the resulting dataset still contains a sizeable number of municipalities. Of all waterworks built by 1911, we found information for 957 of them. Among cities that had acquired a connection by 1911, on average 81% of the buildings were connected (see figure B.2b, panel b).

Mortality rates We compute average crude death rates and infant mortality rates between 1905-1910 using information from the annual publications of the Veröffentlichungen Des Kaiserlichen Gesundheitsamtes (KG 1907-1911), and specific sources for Baden (SLGB 1907-1913), Hesse (GHZL 1911) and Saxony (KSSL 1912).

Students in higher education The city-specific number of school children enrolled in higher and middle schools in 1911 is reported in (N.N. 1911).

Patents Registered patents per capita for each city in 1912 are computed using a yet unpublished dataset provided by Jochen Streb and Niclas Griesshaber.

Listed firms We compute the number of stock market-listed firms p.c. in 1911 using the 1911 volume of the annual handbook of German stock exchange-listed firms (N.N. 1912).

B.7 Maps showing the spatial distribution of selected variables

Figure B.4: Controls

(a) Population, 1866

Figure has been removed to conform to file size restrictions.

(b) All-cause mortality, early 1860s

Figure has been removed to conform to file size restrictions.

(c) Economic development index, early 1860s

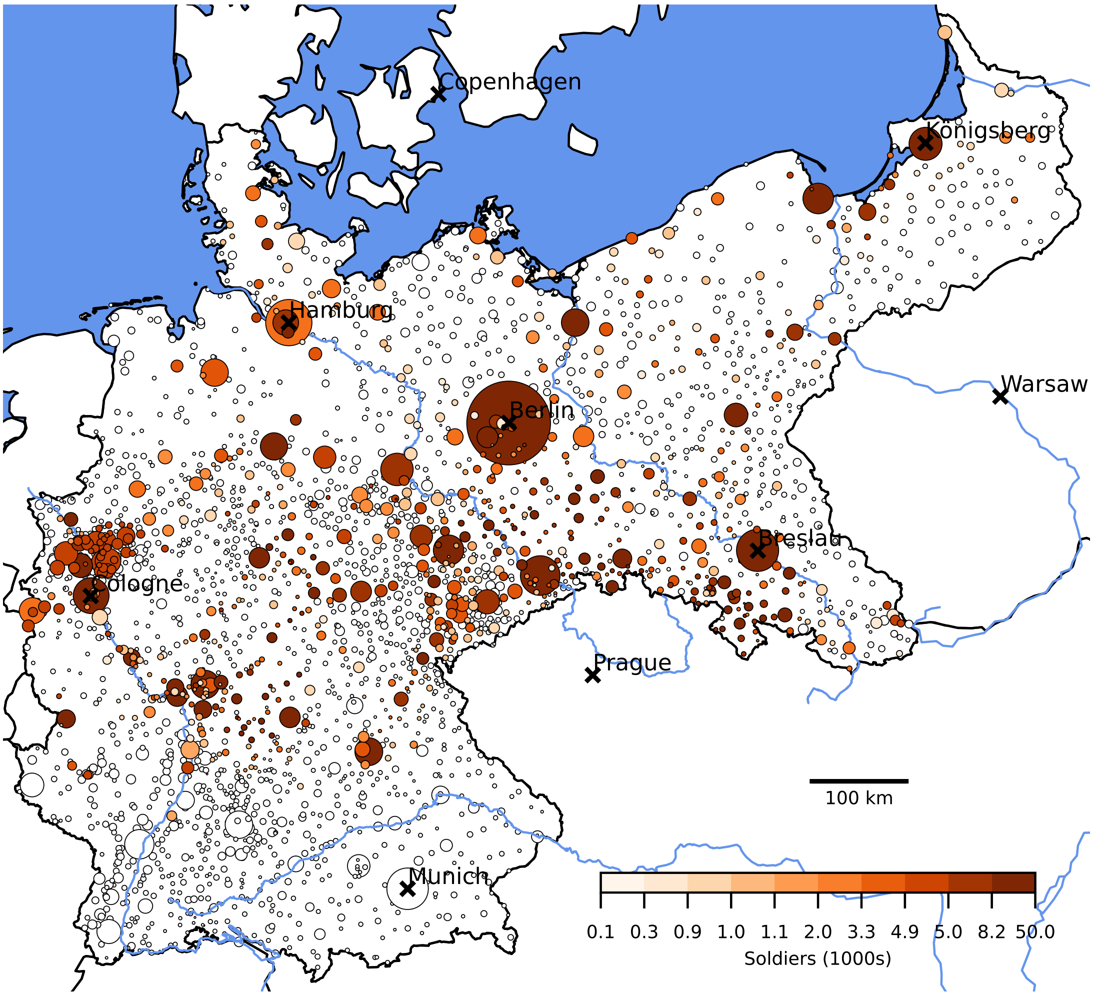
Figure has been removed to conform to file size restrictions.

(d) Market access, 1866

Figure has been removed to conform to file size restrictions.

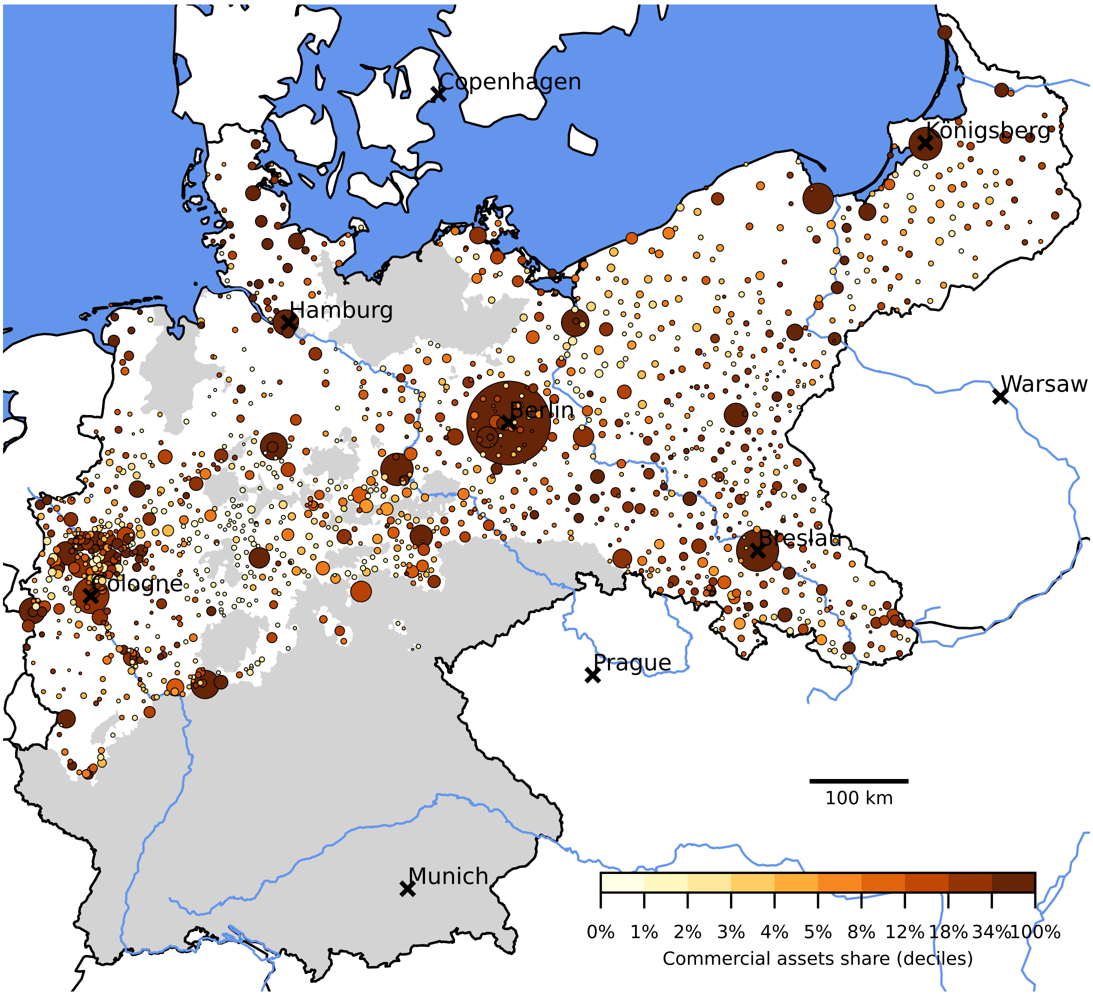
Note: These four maps visualize the distribution of the four socio-economic control variables. Darker colors indicate higher values. Circle sizes are proportional to population. See appendix B for the sources and procedures used to create the variables.

Figure B.5: Number of unique Prussian soldiers crossing each city in 1866



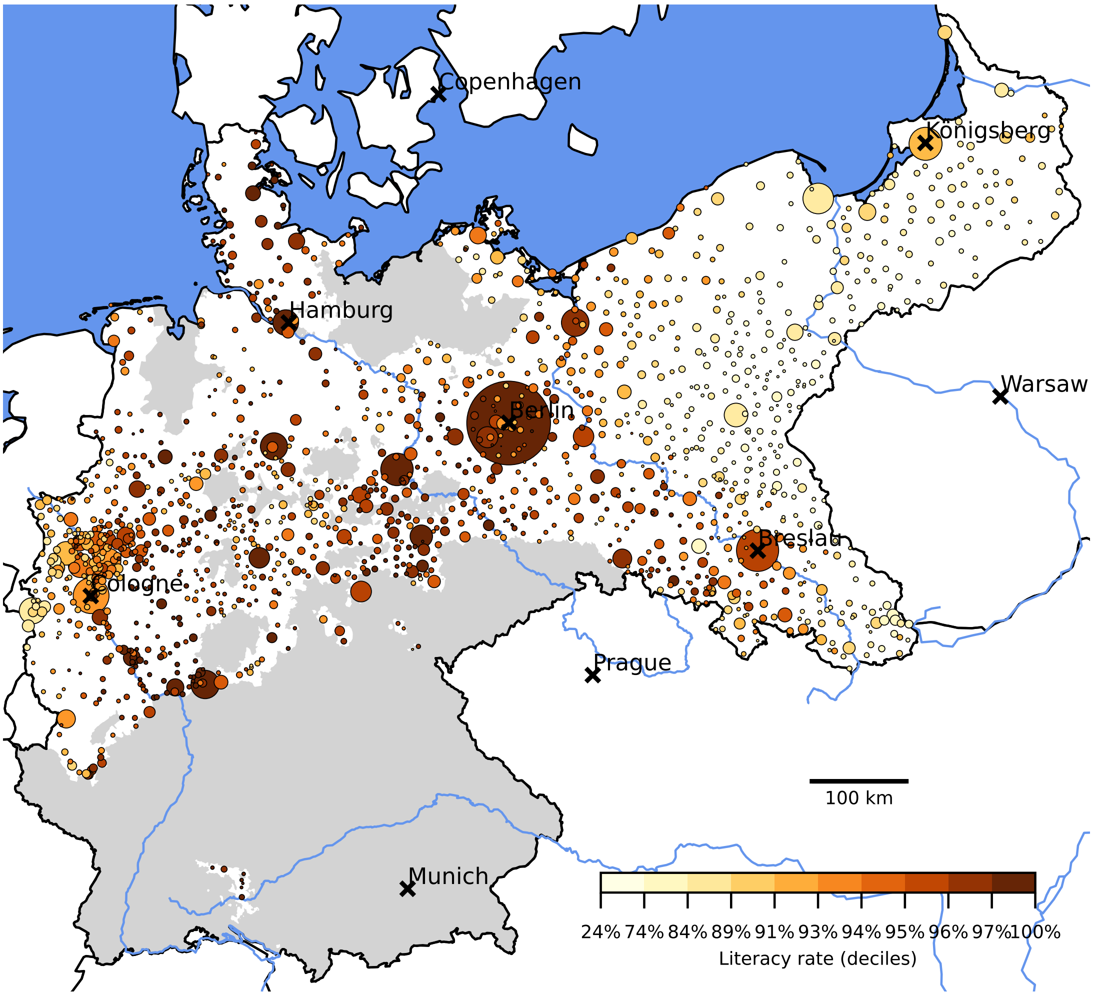
Note: The map shows the number of unique Prussian soldiers that crossed each city during the 1866 Austro-Prussian war. Black lines show Germany in its 1871 borders. See appendix B for the sources and procedures used to create the variables.

Figure B.6: Elites' economic background, early 1860s



Note: The map shows the distribution of our measure for each city's elite background composition: the share of the value of the non-residential building stock in the total value of this asset class plus the value of agricultural land. Grayed-out areas mark areas that are in our main sample, but for which this measure cannot be computed. Black lines show Germany in its 1871 borders. See appendix B for the sources and procedures used to create the variables.

Figure B.7: Human capital levels, 1871



Note: The map shows literacy rates in Prussian cities in the year 1871. Literacy rates are computed excluding children below the age of 10. Grayed-out areas mark areas that are in our main sample, but for which this measure cannot be computed. Black lines show Germany in its 1871 borders. See appendix B for the sources and procedures used to create the variables.

Table B.1: Summary table of variables in data set

		All $N = 2685$		Main $N = 2510$		Mechanism $N = 1410$	
Variable	Range	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.
Waterworks							
Opening year	[1849, 1997]	1901.186	20.505	1901.938	19.955	1905.095	21.101
Cholera shock							
Cholera in 1865–1867	$\{0, 1\}$	0.353	0.478	0.358	0.479	0.521	0.500
Cholera deaths per 100	[0, 19.682]	0.430	1.196	0.441	1.216	0.705	1.501
(Hydro-)geology							
Elevation (km above sea level)	[-.002, 1.325]	0.249	0.198	0.246	0.198	0.145	0.125
Slope (rise-run ratio)	[0.002, 0.442]	0.066	0.050	0.065	0.050	0.049	0.042
Groundwater productivity (share)	[0, 1]	0.511	0.339	0.509	0.338	0.464	0.334
Groundwater salt intrusion (share)	[0, 1]	0.010	0.082	0.010	0.082	0.014	0.091
Distance to fresh water (10 kms)	[0.171, 15.008]	2.457	2.314	2.489	2.346	2.982	2.606
Monthly precipitation (cm)	[4.071, 14.665]	6.209	1.505	6.214	1.516	5.727	1.330
Socio-economic characteristics							
Population (10^3)	[0, 681.910]	4.879	16.735	4.409	9.168	4.715	9.186
All-cause death rate (%)	[0.551, 11.471]	2.842	0.887	2.848	0.902	2.845	1.112
Economic development (index)	[0, 1]	0.221	0.213	0.226	0.213	0.237	0.196
Market access (10^6 population)	[0, 158.778]	1.843	8.204	1.927	8.465	1.460	7.574
Mechanism							
Commercial city	$\{0, 1\}$					0.286	0.452
Commercial assets ratio	(0, 1)					0.123	0.173
Literate city	$\{0, 1\}$					0.713	0.452
Literacy rate	(0.240, 1)					0.890	0.101
Protest	$\{0, 1\}$					0.093	0.291
Industry demand	$\{0, 1\}$					0.379	0.485
High density	$\{0, 1\}$					0.388	0.488
Long-run outcomes							
All-cause death rate, ca. 1910 (%)	[0.751, 9.206]	1.689	0.534	1.693	0.548	1.675	0.665
Infant mortality rate, ca. 1910 (%)	[5.217, 34.293]	17.098	4.965	17.054	5.053	16.465	3.865
Higher educ. pupils p.c., 1911 (%)	[0, 30.747]	1.892	2.919	1.948	2.915	2.006	2.907
Listed firms p.c., 1911 (%)	[0, 0.299]	0.013	0.023	0.013	0.022	0.012	0.021
Filed patents p.c., 1912 (%)	[0, 0.684]	0.008	0.024	0.008	0.024	0.008	0.027

Note: This table shows summary statistics for the most important variables used in our analyses. See text for detailed information.

Table B.2: Spatio-temporal resolution and sources for the pre-epidemic CDR measure

Territory	Cities (%)	Years	Spatial units	Sources
Prussia (old provinces)	1167 (45.73)	1849 (1862/64)	683 cities, 348 counties (*Kreise*)	KSB (1851b) and KSB (1867)
Bavaria	344 (13.48)	1862-1864	33 cities, 187 counties (*Gerichtsämter*)	Mayr (1878)
Saxony (kingdom)	182 (6.31)	1862-1864	20 cities, 121 counties (*Gerichtsämter*)	KSMI (1865)
Württemberg	107 (6.43)	1862-1864	64 counties (*Oberämter*)	KMCW (1863/66)
Hanover	95 (4.27)	1862-1864	41 cities, 101 counties (*Ämter*)	KSBH (1865)
Baden	83 (4.74)	1861-1863	58 counties (*Oberämter*)	HMB (1865)
Alsace-Lorraine	71 (1.96)	1862-1864	11 cities, 3 regions (*départements*)	SGF (1870)
Thuringian states	69 (3.53)	1862-1864	15 cities, 6 counties, 2 states	Hildebrand (1867), Kraus (1980)
Hesse-Darmstadt	64 (2.27)	1863-1865	64 cities	GCLS (1877)
Holstein, Schleswig, Lauenburg	54 (2.19)	1857-1859	51 cities, 45 counties (*Ämter*)	SBD (1863)
Mecklenburg (Schwerin and Strelitz)	46 (1.72)	1862-1864	5 cities, 38 parishes (*Präposituren*)	GHMS (1863/65a) and GHMS (1863/65b)
Nassau	39 (1.65)	1857-1859	39 cities	Franque et al. (1863)
Hesse-Kassel	38 (2.47)	1859-1861	21 cities, 14 counties (*Ämter*)	KCSA (1867)
Oldenburg	31 (0.67)	1862-1864	31 cities	GHO (1864/66)
Anhalt	18 (0.71)	1862-1864	18 cities	HASB (1865)
Brunswick	14 (0.59)	1862-1864	13 cities, 23 counties (*Ämter*)	SBHS (1874)
Westphalian states	13 (0.94)	1862-1864	5 cities (Lippe-Detmold), 2 states	SL (1863/65) and Kraus (1980)
Hanse cities and Frankfurt/Main	10 (0.35)	1862-1864	10 cities	PBASB (1865), SBAL (1873), and Kraus (1980), SAFVGS (1866) and N.N. (1862)

Note: This table lists the different territories that make up our study area, where some smaller territories with similar properties have been aggregated. For each territory, the table lists the number of cities and the respective share in the total sample, as well as the reference years and spatial resolution of the data used to compute the pre-epidemic crude death rate.

C Epidemiological model

As part of our identification strategy, we employ a compartmental spatial epidemiological model to simulate the cholera epidemic across Germany in 1866. Following common practice in cholera modelling, we use the classical **S**usceptible-**I**nfectious-**R**emoved (SIR) framework with direct (infectious-to-susceptible) transmission, and add a waterborne bacterial reservoir (**W**) to account for indirect transmission via contaminated water (Torres Codeço 2001; Mukandavire et al. 2011; Tien and Earn 2010). We further model spatial transmission across discrete geographic units, each with their own compartment structure (Bertuzzo et al. 2010; Sattenspiel and Dietz 1995; Tuite et al. 2011). Our main innovation is to integrate mobile epidemiological units–the North German battalions–within a realistic spatial exposure model. In total, the model features 2,685 cities and 852 rural areas ("immobile units"), 492 army battalions ("mobile units"), and 356 external regions neighboring Germany ("external units").

Each unit's population is divided daily into susceptible, infectious, and removed (recovered or deceased) individuals.[51] The central purpose of the model is to predict daily cholera deaths, which we aggregate into unit-specific cumulative epidemic death counts. The model parameters are then calibrated to match the cumulative deaths across the 3,537 immobile units observed in our dataset. The model-predicted estimates of soldiers' contribution to the 1866 epidemic then enter our reduced-form regression framework.[52]

Model setup. Let $i = 1, \dots, n$ index immobile and mobile units and $e = n + 1, \dots, n + E$ index external units, whose epidemic dynamics we treat as exogenous. Each day d, we track currently infectious individuals $I_{i,d}$, susceptible (not-yet infected) individuals $S_{i,d}$, and removed individuals (recoveries $R_{i,d}$ and deaths $D_{i,d}$). Newly

[51] We abstract from several model extensions discussed in the cholera literature, such as vaccination, sanitation interventions, hyperinfectious bacterial stages, climate-driven seasonality, waning immunity, and cholera's short incubation period (Azman et al. 2013). Given the short time span studied, we also ignore demographic change and permanent migration, except for documented battle-related losses and reinforcements among mobile units. See Anteneh et al. (2024) and Wang (2002) for an overview of recent cholera compartmental models.

[52] The use of calibrated compartmental models to explore counterfactual scenarios is established practice, i.e. to analyse the effect of alternative vaccination strategies (Tuite et al. 2011).

infected individuals remain infectious for p days, after which they die with probability π.[53] Thus, each unit's fixed total population $N_i > 0$ consists of the mutually exclusive compartments $S_{i,d}$, $I_{i,d}$, $R_{i,d}$, and $D_{i,d}$. In addition, the immobile units posses a local waterborne bacterial load $W_{i,d}$.

We distinguish two infection pathways. *Direct transmission* arises as susceptibles $S_{i,d} = N_i - I_{i,d} - R_{i,d} - D_{i,d}$ are exposed to infectious individuals in their own and other units. Following standard cholera models, we assume frequency-dependent transmission, where the force of infection is proportional to the share of infectious people, not their absolute number (Tien and Earn 2010). In large cities, contact between susceptibles and infectious may be more likely, e.g. because of shared facilities, public transport, markets and dense housing. To approximate such density-amplification effects, we allow the transmission frequency to scale non-linearly with elasticity $\theta > 1$. We also take the vanishing probability of transmission with spatial distance into account and scale cross-unit transmissions by distance penalties $\omega_{i,j,d} \in [0, 1]$. These are defined as

$$\omega_{i,j,d} = \exp\left(-\lambda \cdot t_{i,j,d}\right)^{\eta} \text{ if } t_{i,j,d} \leq 15, \text{ else } \omega_{i,j,d} = 0, \tag{14}$$

where $t_{i,j,d}$ is the travel time (in hours) between units i and j on day d. λ is an exponential scaling parameter and η controls how quickly spatial transmission fades with distance. By construction, $\omega_{i,i,d} = 1$. To reduce computational complexity, we ignore direct spatial interactions between units further apart than 15 hours.[54]

Indirect transmission occurs as susceptibles are also exposed to contaminated water. We model waterborne transmission using a saturating Holling type II function (Mukandavire et al. 2011), where infection pressure increases with bacterial load $W_{i,d}$ but plateaus at high concentrations. The indirect transmission channel generates a

[53] As in standard compartmental models, we require some exogenous initial cases to model the spread, not the genesis, of the epidemic. We treat the spread of cholera until the start of the military campaign as exogenous, thus initializing $I_{i,0}$ empirically as of June 15, 1866, and restrict our simulation to the 300 days thereafter.

[54] Travel times are estimated using a complete representation of Germany's 1866 railway network and navigable waterways (see section B.3). When faster overland travel is available, we use travel time estimates based on the Human Mobility Cost layer approximating pre-railroad travel. Because mobile units change their location over time, some pairwise weights change over time, i.e. $\omega_{i,j,d} \neq \omega_{i,j,d'}$ for some (i, j).

"long wave" infection curve governed by a half-saturation constant that is endogenously estimated within the model.

Summing up both transmission routes, daily realized new infections for immobile units are

$$\Delta I_{i,d} = \underbrace{\beta^I \times \left(\frac{S_{i,d-1}}{N_i}\right)^\theta \times \sum_{j=1}^{n+E} \omega_{i,j,d-1} I_{j,d-1}}_{\text{direct infections}} + \underbrace{\delta_i S_{i,d-1} \times \frac{W_{i,d}}{\kappa + W_{i,d}}}_{\text{indirect infections}}, \tag{15}$$

where β^I is the direct transmission rate for immobile units, δ_i is unit-specific waterborne susceptibility, and κ is the half-saturation constant. For mobile units (soldiers), infections occur solely through direct transmission, reflecting their relative independence from local water sources.[55] Finally, external units ($e = n+1, \ldots, n+E$) are treated as exogenous: their daily new infections $\Delta I_{e,d}$ follow gamma-distributed epidemic curves fitted to the known cumulative deaths and epidemic durations of those regions.

Bacterial load dynamics in the local water reservoirs follow

$$W_{i,d} = \underbrace{(1-\phi)W_{i,d-1}}_{\text{bacterial decay}} + \underbrace{\gamma^I I_{i,d-1}}_{\text{civilian shedding}} + \underbrace{\gamma^M \left(\sum_{j \in \mathcal{M}_{i,d-1}} \omega_{i,j,d-1} I_{j,d-1}\right)}_{\text{military shedding}}, \tag{16}$$

where γ^I and γ^M are the bacterial shedding rates for infected civilian and military individuals, respectively, ϕ denotes the daily decay rate of cholera bacteria, and $\mathcal{M}_{i,d-1}$ is the set of mobile units (battalions) located close to immobile unit (city or county) i at day $d-1$.[56]

Waterborne susceptibility δ_i is endogenously determined as a logistic function of six standardized geo-hydrological characteristics $\mathbf{X}_i = (X_{i,1}, \ldots, X_{i,6})$, including elevation, slope, groundwater productivity, groundwater salinity, distance to surface waters, and precipitation:

[55] Thus, for mobile unit i, $\Delta I_{i,d} = \beta^M \times \left(\frac{S_{i,d-1}}{N_i}\right)^\theta \times \sum_{j=1}^{n+E} w_{i,j,d-1} I_{j,d-1}$, where β^M is the direct transmission rate for mobile units.

[56] We define a mobile unit j as being close to immobile unit i if $t_{i,j,\cdot} < 1$, allowing for a maximum travel time of an hour.

$$\delta_i = \left(1 + \exp\left[-\left(\alpha_0 + \sum_{p=1}^{6} \alpha_p X_{i,p}\right)\right]\right)^{-1} \in [0, 1]. \tag{17}$$

The coefficients $\alpha_0, \ldots, \alpha_6$ are jointly estimated during model calibration.[57]

Figure C.1 illustrates the model structure. Note that each unit's population N_i remains constant over time, while the distribution across compartments (S_i etc.) can change. The potentially different epidemic dynamics between immobile and mobile units are symbolized by subscripts for the direct and indirect transmission rates β_i and γ_i.

Model parameters and estimation. Given our historical setting, we avoid calibrating model parameters by borrowing values from other studies.[58] The only parameters we externally set are the infection period ($p = 3$ days, the mean of the typical range reported by Azman et al. (2013)), the case fatality rate ($\pi = 0.5$), and the parameters of the transport network (including speeds τ, as discussed in section B.3). All other structural parameters $\Gamma = (\beta^I, \beta^M, \gamma^I, \gamma^M, \phi, \kappa, \lambda, \eta, \theta, \alpha_0, \ldots, \alpha_6)$ are estimated by minimizing a loss function $\mathcal{P}$, i.e.

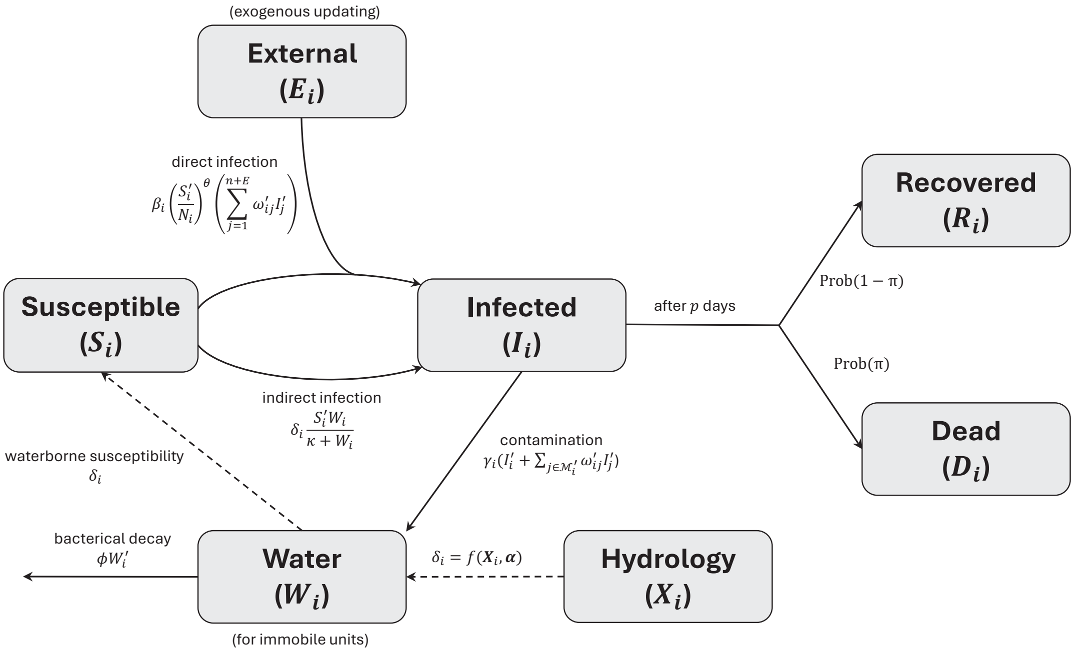
$$\hat{\Gamma} = \arg\min_{\Gamma} \left\{\mathcal{P}(D_i^o, \hat{D}_i) | N_i, \mathbf{X}_i, p, \pi, \mathbf{T}\right\}, \tag{18}$$

for observed and simulated cumulative end-of-outbreak deaths D_i^o, $\hat{D}_i$, and given population counts N_i, geo-hydrological characteristics $\mathbf{X}_i$, parameters p and π, and the matrix of bilateral travel times $\mathbf{T}$ (with elements $t_{i,j,d}$). We simulate the epidemic spread for 300 days, hence $\hat{D}_i \equiv \hat{D}_{i,300}$.

[57]Linking local transmission risks to fundamental geo-hydrological factors ensures plausibility and improves model identifiability, especially in remote areas with few observed cases. It also reduces the number of free parameters compared to unrestricted unit-specific estimates. We cannot condition on local demographics (besides the population count), such as the age structure, because we lack systematic information at this at the granular level.

[58]The most important reason is that the cholera bacterium has changed substantially: While 19$^{\text{th}}$-century outbreaks originated from the so-called "classical" *Vibrio cholerae* strain, post-1970s outbreaks (on which almost all empirical studies are based) overwhelmingly originate from the modern *El Tor* biotype. Furthermore, we think that socioeconomic, medical and sanitary conditions are too different between 19$^{\text{th}}$-century and modern scenarios to allow directly using established parameters.

Figure C.1: Epidemiological model flow chart



Note: This flow chart summarizes the compartment-to-compartment flow, as well as other central mechanisms in the epidemiological model. For brevity, temporal subscripts are substituted by prime marks ($'$), denoting values at $d-1$. The dashed line between the hydrology inputs, and the water and susceptible compartments represents the estimation of unit-specific waterborne susceptibility rates δ_i. $\beta_i \in \{\beta^I, \beta^M\}$ and $\gamma_i \in \{\gamma^I, \gamma^M\}$ differ among immobile and external (I) and mobile (M) units i.

The loss function $\mathcal{P}$ rewards five objectives: (1) observed and simulated log cumulative deaths rates vectors should be similar in *shape*–we minimize their cosine distance; (2) the relative *ranking* of units with respect to log cumulative death rates should be similar–we minimized Spearman's R for those vectors; (3) the total *scale* of the epidemic should be realistic–we minimize the difference between the total observed and simulated death counts across all units; (4) both direct (person-to-person) and indirect (waterborne) transmission should be powerful with each channel accounting for at least 30% of the total deaths; (5) mobile units (batallions) should be epidemically relevant but not dominant, contributing between 20 and 50% of total deaths.[59] The

[59]We implement all these criteria as "soft penalties", penalizing only deviations above 10%, and

focus on death counts in the estimation approach reflects the fact that our empirical data consistently reports reliable death counts, while case data (including non-fatal cases) and timing data is only available for a subset on observations and generally less reliable.

We use a differential evolution approach to search the parameter space for cost-minizing parameters, similar to Mari et al. (2012).[60] Table C.1 lists all estimated and calibrated parameters.

Table C.1: Estimated and calibrated model parameters

Symbol	Value	Description	Unit / Dimension
		Endogenous (estimated) parameters	
β^I	0.00148	Contribution to new infections per infectious civilian per day (direct transmission)	day^{-1}
β^M	0.00091	Contribution to new infections per infectious soldier per day (direct transmission)	day^{-1}
γ^I	0.04927	Contribution to bacterial load per infectious civilian per day (indirect transmission)	(water-stock units) $\times$ person^{-1} $\times$ day^{-1}
γ^M	0.08836	Contribution to bacterial load per infectious soldier per day (indirect transmission)	(water-stock units) $\times$ person^{-1} $\times$ day^{-1}
ϕ	0.01642	Bacterial decay fraction per day (water reservoir)	day^{-1}
κ	8,671,052	Half-saturation constant for waterborne transmission	water-stock units (model scale)
λ	0.34283	Travel-time decay rate in spatial transmission kernel	hour^{-1}
η	1.51294	Exponent shaping travel-time decay (spatial kernel)	dimensionless
θ	1.55343	Density elasticity of direct transmission	dimensionless
$(\alpha)_0^6 \to \delta_i$	[0.41682]	Average waterborne susceptibility across immobile units	dimensionless index $\in [0, 1]$
		Exogenous (calibrated) parameters	
p	3	Infectious period (fixed duration)	days
π	0.5	Case fatality rate	dimensionless probability $\in [0, 1]$

Note: This table lists the values of endogenous and exogenous model parameters. Endogenous parameters are estimated by minimizing the penalized loss function described in equation 18. Exogenous parameters are fixed based on historical or medical evidence (see text). Water-stock units denote an internal model scale, only interpretable relative to κ. For δ_i, the unweighted mean of unit-specific values is presented. The standard deviation is 0.28966.

How plausible are the estimated parameters in light of the historical evidence and existing cholera simulation studies? First, the direct-contact rates (β^I, β^M) are comparatively small: even with $S/N \approx 1$, one infectious person generates only $\beta^I \times p = 1.48 \times 10^{-3} \times 3 = 0.0044$ direct infections over their infectious period. By contrast, the waterborne parameters $(\phi, \delta_i, \gamma^I, \gamma^M, \kappa)$ imply a potent environmental pathway. The estimated decay $\phi = 0.0164\,\text{day}^{-1}$ implies a $\sim$ 42-day half-life—higher than

thus allowing for a corridor around perfect fits (e.g., a rank correlation of 1).

[60]Other studies with calibrated cholera models use simpler least-square fitting methods (Mukandavire et al. 2011; Tuite et al. 2011). Note that generally no closed-form solutions to the epidemic model classes described here exist, making regression-based approaches to parameter optimization less feasible. See Tuite et al. (2011) for a closed-form analysis of the final epidemic death count in a simpler, non-spatial SIRW model.

many modern calibrations, but credible in pre-modern settings with low sanitation and repeated re-seeding.[61] A back-of-the-envelope example illustrates how the water-related parameters shape infections: An infectious civilian sustains about $\gamma^I/\phi \approx 3$ units of W. For a large city with representative waterborne susceptibility ($\delta = 0.42$), $S = 100{,}000$ suceptibles and $I = 100$ infected (i.e. $W \approx 300$), the water term yields $\delta SW/\kappa \approx (0.42 \times 10^5 \times 300)/(8.67 \times 10^6) \approx 1.4$ new infections per day via indirect transmission. In short, the estimated parameters imply that direct transmission alone can rarely sustain an outbreak, while strong, environmentally mediated transmission sustains it—consistent with historical evidence and modern simulations.

Second, the civilian–soldier contrasts are historically plausible: direct infections from soldiers were less likely given the limited contact between soldiers and civilians during campaigns ($\beta^M/\beta^I \approx 0.61$). At the same time, local water sources were more endangered from concentrated troop movements ($\gamma^M/\gamma^I \approx 1.79$). This is consistent with armies as mobile seeders of contaminated water rather than efficient contact spreaders.

Third, the remaining parameters mediating the role of density and distance fit the 19th-century context. With $\theta = 1.553$, direct infection intensity falls faster than proportionally as susceptibles deplete ($S/N = 0.8 \Rightarrow (S/N)^\theta \approx 0.71$ vs 0.80; $0.5 \Rightarrow 0.34$ vs 0.50), curbing runaway growth in large cities. With $\lambda = 0.3428\,\mathrm{h}^{-1}$ and $\eta = 1.513$, the spatial weight halves every 1.34 hours of travel time, is down $\sim 90\%$ by 4.44 h, $\sim 99\%$ by 8.88 h, and $\sim 4.2 \times 10^{-4}$ at the 15-hour cutoff—implying primarily local, corridor-bound spread.

In figure C.2, we plot the distribution of simulated death rates next to the targeted deaths rate.

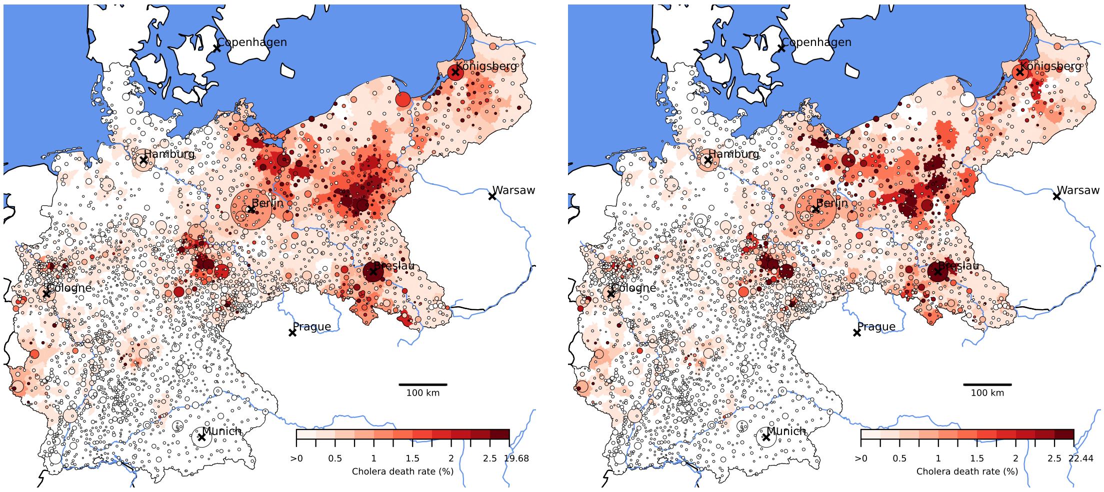
Counterfactual scenario estimation. To quantify soldiers' epidemic contribution, we use our calibrated model to conduct a counterfactual simulation with mobile units' populations set to zero:

$$\hat{D}_i^{(\text{due to soldiers})} = \hat{D}_i(\hat{\Gamma}^{(\text{baseline})}) - \hat{D}_i(\hat{\Gamma}^{(\text{no soldiers})}). \tag{19}$$

[61] The summer of 1866 was unusually warm, which also may have slowed bacterial decay.

Figure C.2: Observed and simulated spatial spread

(a) Observed 1866 deaths p.c. **(b)** Simulated 1866 deaths p.c.



Both maps shows cities and rural counties in Germany, with colour shading indicating cholera deaths per 100 in 1866. The map on the left shows observed cholera death rates; the map on the right shows those implied by our simulation. Circles are cities, with circle size proportional to their 1866 population. Cross markers place the cities of Cologne, Hamburg, Munich, Berlin, Breslau, Königsberg, and foreign capitals. Black lines show the German Empire in its 1871 borders and blue lines show major rivers. Note that panel a of this figure differs slightly from figure 2, which includes 1865 and 1867 deaths.

These differences provide city- and region-specific estimates of additional cholera deaths attributable directly or indirectly to soldiers' mobility patterns.

References

Albers, Thilo N. H., and Kalle Kappner. 2023. "Perks and pitfalls of city directories as a micro-geographic data source." *Explorations in Economic History* 87:101476.

Almon, Shirley. 1965. "The Distributed Lag Between Capital Appropriations and Expenditures." *Econometrica* 33 (1): 178–196.

Anderson, D. Mark, Kerwin Kofi Charles, and Daniel I. Rees. 2022. "Re-Examining the Contribution of Public Health Efforts to the Decline in Urban Mortality." *American Economic Journal: Applied Economics* 14 (2): 126–157.

Anteneh, Leul Mekonnen, Bruno Enagnon Lokonon, and Romain Glèlè Kakaï. 2024. "Modelling techniques in cholera epidemiology: A systematic and critical review." *Mathematical Biosciences* 373:109210.

Azman, Andrew S., Kara E. Rudolph, Derek A. T. Cummings, and Justin Lessler. 2013. "The incubation period of cholera: a systematic review." *Journal of Infection* 66 (5): 432–438.

Beach, Brian. 2022. "Water Infrastructure and Health in U.S. Cities." Urban Economics and History, *Regional Science and Urban Economics* 94:103674.

Becker, Sascha O., and Francesco Cinnirella. 2020. "Prussia disaggregated: the demography of its universe of localities in 1871." *Journal of Demographic Economics* 86 (3): 259–290.

Belzyt, Leszek. 1998. *Sprachliche Minderheiten im preußischen Staat 1815-1914.* Herder-Institut.

Bertuzzo, E., R. Casagrandi, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo. 2010. "On spatially explicit models of cholera epidemics." *Journal of the Royal Society Interface* 7 (43): 321–333.

Brix, Josef, Karl Imhoff, and Robert Weldert. 1934. *Die Stadtentwässerung in Deutschland.* Jena: Gustav Fischer.

Brown, John C. 1988. "Coping with Crisis? The Diffusion of Waterworks in Late Nineteenth-Century German Towns." *The Journal of Economic History* 48 (2): 307–318.

Cantoni, Davide, Cathrin Mohr, and Matthias Weigand. 2020. *Princes and Townspeople: A Collection of Historical Statistics on German Territories and Cities.* Data set (v.3). Harvard Dataverse.

Colella, F., R. Lalive, S. O. Sakalli, and M. Thoenig. 2023. "acreg: Arbitrary correlation regression." *The Stata Journal* 23 (1): 119–147.

Conley, Timothy G. 1999. "GMM estimation with cross sectional dependence." *Journal of Econometrics* 92 (1): 1–45.

Conley, Timothy G., and Morgan Kelly. 2025. “The standard errors of persistence.” *Journal of International Economics* 153:104027.

Cutler, David, and Grant Miller. 2005. “The Role of Public Health Improvements in Health Advances: The Twentieth-Century United States.” *Demography* 42 (1): 1–22.

Deutscher Verein von Gas- und Wasserfachmännern. 1913. *Statitische Zusammenstellung der Betriebs-Ergebnisse von Wasserwerken.* Vol. 20. München: R. Oldenbourg.

———. various years. *Journal für Gasbeleuchtung und verwandte Beleuchtungsarten, sowie für Wasserversorgung.* München: R. Oldenbourg.

Donaldson, Dave, and Richard Hornbeck. 2016. “Railroads and American Economic Growth: A “Market Access” Approach.” *The Quarterly Journal of Economics* 131 (2): 799–858.

Duscher, Klaus, Andreas Günther, Andrea Richts, Patrick Clos, Uta Philipp, and Wilhelm Struckmeier. 2015. “The GIS Layers of the “International Hydrogeological Map of Europe 1:1,500,000”' in a Vector Format.” *Hydrogeology Journal* 23:1867–1875.

Engel, Ernst. 1867. “Die wahren Verluste der königlich preussischen Armee im Krieg des Jahres 1866.” *Zeitschrift des Königlich Preussischen Statistischen Bureaus* 7:157–166.

———. 1869. “Die Cholera-Epidemie des Jahres 1866 mit einem Rückblick auf die früheren Epidemien.” *Zeitschrift des Königlich Preussischen Statistischen Bureaus* 9 (1-3): 70–98.

Farbmacher, Helmut, and Harald Tauchmann. 2023. “Linear fixed-effects estimation with nonrepeated outcomes.” *Econometric Reviews* 42 (8): 635–654.

Fick, S. E., and R. J. Hijmans. 2017. “WorldClim 2: New 1km Spatial Resolution Climate Surfaces for Global Land Areas.” *International Journal of Climatology* 37 (12): 4302–4315.

Franque, J. B., W. Fritze, and C. Vogler. 1863. *Medicinische Jahrbücher für das Herzogthum Nassau.* Vol. 19 & 20. Wiesbaden: Christian Wilhelm Kreidel.

Freiherr von Schweiger-Lerchenfeld, Amand. 1894. *Vom rollenden Flügelrad. Darstellung der Technik des heutigen Eisenbahnwesens.* Wien: Hartleben.

Gallardo-Albarrán, Daniel, and Kalle Kappner. 2025. “Epidemics and urban growth: the impact of cholera in Nineteenth-century Prussia.” *The History of the Family* 30 (2): 468–499.

Geisenheimer, Paul. 1913. *Hauptbericht des Arbeitsausschusses für die Wasserversorgung des oberschlesischen Industriebezirks.* Kattowitz: Gebrüder Böhm.

Grahn, Ernst. 1878. *Die städtische Wasserversorgung.* Vol. 1. Statistik der städtischen Wasserversorgungen mit einer geschichtlichen Einleitung. Munich: R. Oldenbourg.

———. 1883. *Die Art der Wasserversorgung der Städte des Deutschen Reiches mit mehr als 5000 Einwohnern.* Munich and Leipzig: Druck und Verlag von R. Oldenbourg.

———. 1904. "Die städtischen Wasserwerke." In *Die deutschen Städte. Geschildert nach dem Ergebnissen der ersten deutschen Städteausstellung zu Dresden 1903,* edited by R. Wuttke, 301–344. Leipzig: Friedrich Brandstetter.

———. 1898-1902. *Die städtische Wasserversorgung im Deutschen Reiche sowie in einigen Nachbarländern.* Munich and Leipzig: R. Oldenbourg.

Großh. Hessische Zentralstelle für Landesstatistik. 1911. *Die Bevölkerungsbewegung in den einzelnen Gemeinden des Großh. Hessen während des Zeitraums 1906 bis 1910.* Vol. 62/1. Beiträge zur Statistik des Großherzogtums Hessen. Darmstadt: Großherzoglicher Staatsverlag.

Grossherzogliche Centralstelle für die Landes-Statitik. 1877. *Beiträge zur Statistik des Grossherzogtums Hessen.* Vol. 15/2. Darmstadt: G. Jonghaus.

Großherzogtum Mecklenburg-Schwerin. 1863/65a. *Großherzoglich Mecklenburg-Schwerinscher Staats-Kalender.* Vol. 1863, 1864, 1865. Schwerin: Bärensprung.

Großherzogtum Mecklenburg-Strelitz. 1863/65b. *Großherzoglich Mecklenburg-Strelitzscher Staats-Kalender.* Vol. 1863, 1864, 1865. Neustrleitz: Spalding & Sohn.

Grossherzogtum Oldenburg. 1864/66. *Hof- und Staatshandbuch des Großherzogtums Oldenburg.* Vol. 1864, 1865, 1866. Oldenburg: W. Berndt.

Guttstadt, Albert. 1873. "Die Pocken-Epidemie in Preussen, insbesondere in Berlin 1870/72." In *Zeitschrift des Königlich Preussischen Statistischen Bureaus,* edited by Ernst Engel, 13:116–158. Berlin: Verlag des Königlichen Statistischen Bureaus.

———. 1900. *Krankenhaus-Lexikon für das Deutsche Reich. Die Anstaltsfürsorge für Kranke und Gebrechliche und die hygienischen Einrichtungen der Städte im Deutschen Reich am Anfang des zwanzigsten Jahrhunderts.* Berlin: Georg Reimer.

Handels-Ministerium. 1865. *Beiträge zur Statistik der inneren Verwaltung des Großherzogtums Baden.* Vol. 18: Bewegung der Bevölkerung in den Jahren 1856 bis mit 1863 und Medicinische Statistik. Carlsruhe: Chr. Fr. Müller.

Hansen, Bruce E. 2000. "Sample Splitting and Threshold Estimation." *Econometrica* 68 (3): 575–603.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1): 153–161.

Herrfurth, L. 1878. "Beiträge zur Finanzstatistik der Gemeinden in Preussen." In *Zeitschrift des Königlich Preussischen Statistischen Bureaus,* vol. Ergänzungsheft VI. Berlin: W. Koebke.

Herrfurth, L., and W. von Tzschoppe. 1884. "Beiträge zur Finanzstatistik der Gemeinden in Preussen." In *Zeitschrift des Königlich Preussischen Statistischen Bureaus,* vol. Ergänzungsheft XVI. Berlin: Verlag des Königlichen Statistischen Bureaus.

Herzoglich Anhaltisches statistisches Bureau. 1865. *Mittheilungen des Herzoglich Anhaltischen statistischen Bureaus.* Vol. 1: Statistische Uebersicht der in den Jahren 1861 bis 1864 in Anhalt Geborenen und Gestorbenen. Dessau: A. Lange.

Hildebrand, Bruno. 1867. *Statistik Thüringens. Mittheilungen des Statistischen Büreaus Vereinigter Thüringischer Staaten.* Vol. 1. Jena: Friedrich Frommann.

Kaiserliches Gesundheitsamt. 1907-1911. *Veröffentlichungen des Kaiserlichen Gesundheitsamtes.* Vol. 31-35. Berlin: Julius Springer.

Kappner, Kalle. 2024. "Did Cholera 'Force' the Reform of Urban Water Infrastructure? Lessons from Nineteenth-Century Berlin." *Journal of Urban History* 51 (3): 545–566.

———. 2025. "The Mortality Impact of Cholera in Germany." *Jahrbuch für Wirtschaftsgeschichte [Economic History Yearbook]* 66 (2): 365–403.

Kersting, Felix, Iris Wohnsiedler, and Nikolaus Wolf. 2020. "Weber Revisited: The Protestant Ethic and the Spirit of Nationalism." *The Journal of Economic History* 80 (3): 710–745.

Königlich Preußisches Ministerium der Unterrichts-Angelegenheiten. 1867. *Statistische Nachichten über das Elementar-Schulwesen in Preußen für die Jahre 1862 bis 1864.* Berlin: Wilhelm Hertz.

Königlich Preussisches Statistisches Landesamt in Berlin. 1914-1920. *Preussische Statistik.* Vol. 243: Finanzstatistik der preusischen Städte und Landgemeinden für das Rechnungsjahr 1911. Berlin: Verlag des Königlichen Statistischen Landesamts.

Königlich Sächsischen Ministeriums des Innern. 1865. "Die Bewegung der Bevölkerung im Königreiche Sachsen während der Jahre 1862, 1863 und 1864." In *Zeitschrift des Statistischen Bureaus des Königlich Sächsischen Ministeriums des Innern,* 11:125–148. Dresden: B. G. Teubner.

Königlich Sächsisches Statistisches Landesamt. 1912. *Die Geburten und Sterbefälle in den einzelnen Jahren und im ganzen Jahrfünft.* Vol. 58. Zeitschrift des K. Sächsischen Statistischen Landesamtes. Dresden: R. von Zahn.

Königliche Commission für statistische Angelegenheiten. 1867. *Beiträge zur Statistik des vormaligen Kurfürstenthums Hessen.* Vol. 2: Die Bevölkerung Kurhessens. Hauptergebnisse der Volkszählung von 1861 in Kurhessen. Kassel: U. Freischmidt.

Königliches Finanzministerium. 1870. *Die Ergebnisse der Grund- und Gebäudesteuerveranlagung im Regierungsbezirk Aachen.* Berlin: Königliche Staatsdruckerei.

Königliches Medicinalkollegium. 1863/66. "Jahresbericht über die Geburts- und Sterblichkeitsstatistik Württembergs, mit Einschluss der Statistik des Heilpersonals." In *Medicinisches Correspondenz-Blatt des würtembergischen ärztlichen Vereins,* vol. XXXIII/39, XXXIV/40, XXXVI/15. Stuttgart: Schweizerbart.

Königliches Statistisches Bureau. 1873. *Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871. 11 volumes.* Berlin: Verlag des Königlichen Statistischen Bureaus.

Königliches Statistisches Bureau in Berlin. 1867. *Preussische Statistik.* Vol. 10: Die Ergebnisse der Volkszählung und Volksbeschreibung, der Gebäude- und Viehzählung, nach den Aufnahmen vom 3. December 1864, resp. Anfang 1865 und die Statstik der Bewegung der Bevölkerung in den Jahren 1862, 186 und 1864. Berlin: Ernst Kuehn.

———. 1871. *Die Gebäude im Preussischen Staate nach den Aufnahmen der Gebäudesteuer-Veranlagung auf Grund des Gesetzes vom 21. Mai 1861.* Vol. XVIII. Preussische Statistik. Berlin: Verlag des Königlichen Statistischen Bureaus (Dr. Engel).

———. 1877. *Die Ergebnisse der Volkszählung und Volksbeschribung vom 1. December 1875 im Preussischen Staate.* Vol. XXXIX. Preussische Statistik. Amtliches QuellenWerk. Berlin: Ernst Kuehn.

———. 1886. "Preußische Statistik. Amtliches Quellenwerk." In *Die Gewerbebetriebe im preussischen Staate nach der Aufnahme vom 5. Juni 1882: Zweiter Theil. Uebersichten für die einzelnen Kreise,* vol. 83/2. Berlin: Verlag des Königlichen Statistischen Bureaus.

Königliches Statistischs Bureau zu Hannover. 1865. *Beiträge zur Statistik des Königreichs Hannover.* Vol. 1, II: Uebersicht der Geborenen, Getraueten und Gestorbenen im Königreich Hannover für die Jahre 1862, 1863, 1864. Hannover: Fr. Culemann.

Kopsidis, Michael, and Nikolaus Wolf. 2012. "Agricultural Productivity Across Prussia During the Industrial Revolution: A Thünen Perspective." *The Journal of Economic History* 72 (3): 634–670.

Kraus, Antje. 1980. *Quellen zur Bevölkerungsstatistik Deutschlands, 1815-1875.* Edited by Wolfgang Köllmann. Quellen zur Bevölkerungs-, Sozial- und Wirtschaftsstatistik Deutschlands 1815-1875. Boppard am Rhein: Harald Boldt.

Lehmann, Sibylle H. 2010. "The German Elections in the 1870s: Why Germany Turned from Liberalism to Protectionism." *The Journal of Economic History* 70 (1): 146–178.

Mari, L., E. Bertuzzo, L. Righetto, R. Casagrandi, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo. 2012. "Modelling cholera epidemics: the role of waterways, human mobility and sanitation." *Journal of the Royal Society Interface* 9 (67): 376–388.

Martí-Henneberg, Jordi. 2023. *Creative Ways to apply Historical GIS.* Cham: Springer Nature.

Mayr, Georg. 1878. *Beiträge zur Statistik des Königreichs Bayern.* Vol. XXXIII: Die Bewegung der bayerischen Bevölkerung in den Jahren 1862/63 bis 1875 mit vergleichenden Rückblicken. I. Theil. Tabellenwerk. München: J. Gotteswinter & Mössl.

Mukandavire, Zindoga, Shu Liao, Jin Wang, Holly Gaff, David L. Smith, and J. Glenn Morris. 2011. "Estimating the reproductive numbers for the 2008–2009 cholera outbreaks in Zimbabwe." *Proceedings of the National Academy of Sciences* 108 (21): 8767–8772.

Mushacke, Eduard. 1866. *Schul-Kalender auf die Zeit vom 1. Januar 1866 bis 31. März 1867.* Vol. XV. Berlin: Wilhelm Schultze.

N.N. 1862. *Statistische Mittheilungen über den Civilstand der freien Stadt Frankfurt und ihrer Landgemeinden im Jahre 1862.* Frankfurt am Main: J. P. Streng.

———. 1911. *Statistisches Jahrbuch der höheren Schulen Deutschlands, Luxemburgs und der Schweiz und der höheren deutschen Schulen im Ausland.* Vol. 32. Leipzig: B. G. Teubner.

———. 1912. *Handbuch der Deutschen Aktien-Gesellschaften. Jahrbuch der deutschen Börsen.* Vol. 16/2. Berlin/Leipzig/Hamburg: Verlag für Börsen- und Finanzliteratur.

Nasse, Walther. 1905. "Der Rhein als Wasserstraße." In *Die Schiffahrt der deutschen Ströme. Untersuchungen über deren Abgabenwesen, Regulierungskosten und Verkehrsverhältnisse. Dritter Band,* edited by Verein für Socialpolitik, 1–300. Duncker & Humblot.

Neefe, Moritz. 1920. *Statistisches Jahrbuch Deutscher Städte.* Vol. 20. Breslau: Wilh. Gottl. Korn.

Niederstetter, J. 1867. *Staats-Almanach für das Königreich Preußen.* Berlin: Carl Heymann.

Özak, Ömer. 2018. "Distance to the pre-industrial technological frontier and economic development." *Journal of Economic Growth* 23:175–221.

Papke, Leslie E., and Jeffrey M. Wooldridge. 1996. "Econometric methods for fractional response variables with an application to 401(k) plan participation rates." *Journal of Applied Econometrics* 11 (6): 619–632.

Papke, Leslie E., and Jeffrey M. Wooldridge. 2008. "Panel data methods for fractional response variables with an application to test pass rates." *Journal of Econometrics* 145 (1-2): 121–133.

Petersilie, A. 1882. *Die Oeffentlichen Volksschulen in Preussen (einschliesslich Mittelschulen und höhere Mädchenschulen) und die zur Unterhaltung derselben erforderlichen persönlichen und sächlichen Gesammtaufwendungen im Jahre 1878.* X. Ergänzungsheft zur Zeitschrift des königlich preussischen statistischen Bureau's. Berlin: K. Preuss. Statist. Bureau.

Pfister, Ulrich. 2018. "Real wages in Germany during the first phase of industrialization, 1850–1889." *Jahrbuch für Wirtschaftsgeschichte* 59 (2): 567–596.

Provisorisches Bureau für allgemeine Statistik. 1865. *Zur Statistik des Bremischen Staats.* Vol. Die Gestorbenen in den Jahren 1826 bis 1864 nach Geschlechtern und Altersclassen. Bremen: Heinrich Strack.

Salomon, Hermann. 1906-1911. *Die städtische Abwässerbeseitigung in Deutschland. Wörterbuchartig angeordnete Nachrichten und Beschreibungen städtischer Kanalisations- und Kläranlagen in deutschen Wohnplätzen (Abwässer-Lexikon).* Jena: Gustav Fischer.

Sattenspiel, L., and K. Dietz. 1995. "A structured epidemic model incorporating geographic mobility among regions." *Mathematical Biosciences* 128 (1-2): 71–91.

Selter, Hugo. 1911. *Die Trinkwasserversorgung der Rheinprovinz.* Bonn: Martin Hager.

Silbergleit, Heinrich. 1908. *Preussens Städte. Denkschrift zum 100 jährigen Jubiläum der Städteordnung von 19. November 1808.* Berlin: Carl Heymanns Verlag.

Singer, Judith D., and John B. Willett. 1993. "It's about Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events." *Journal of Educational Statistics* 18 (2): 155–195.

Staat Lippe. 1863/65. *Fürstlich Lippisches Regierungs- und Anzeige-Blatt.* Vol. 1863, 1864, 1865. Detmold: Klingenberg.

Statistique Generale de la France. 1870. *Statistique de la France.* Vol. XVIII, 2e série: Mouvement de la Population, 1861-1865. Strasbourg: Imprimerie administrative de Veuve Berger-Levrault.

Statistische Abtheilung des Frankfurter Vereins für Geographie und Statistik. 1866. *Beiträge zur Statistik der Freien Stadt Frankfurt.* Vol. 2, 1. Frankfurt am Main: J. D. Sauerländer.

Statistisches Büreau des Herzoglichen Staatsministeriums. 1874. *Beiträge zur Statistik des Herzogthums Braunschweig.* Vol. 1: Die Bewegung der Bevölkerung in den 20 Jahren 1863 bis 1872. Braunschweig: Herzogliche Waisenhaus-Buchdruckerei.

Statistisches Bureau des Stadt- und Landamtes Lübeck. 1873. *Statistik des Lübeckischen Staats.* Vol. II: Die natürliche Bewegung der Bevölkerung in den Jahren 1846-70. Lübeck: Frerdinand Grautoff.

Statistisches Bureau zu Berlin. 1851a. *Tabellen und amtliche Nachrichten über den Preussischen Staat für das Jahr 1849.* Vol. 1. Berlin: A. W. Hayn.

———. 1851b. *Tabellen und amtliche Nachrichten über den Preussischen Staat für das Jahr 1849.* Vol. 2. Berlin: A. W. Hayn.

Statistisches Landesamt der Großherzogthums Baden. 1907-1913. *Statistisches Jahrbuch für das Großherzogthum Baden.* Vol. 36-39. Karlsruhe: Macklot.

Statistiske Bureau. 1863. *Statistisk Tabelværk.* Vol. 3, 3rd series: Tabeller over Vielser, Födsler og Dödsfals i Konigeriget Danmark, Hertugdömmet Slesvig, Hertugdömmet Holsteen og Hertugdömmet Lauenborg for Aarene 1855-1859. Kjöbenhavn: Bianco Lunos.

Suesse, Marvin, and Nikolaus Wolf. 2020. "Rural transformation, inequality, and the origins of microfinance." *Journal of Development Economics* 143 (102429).

Tien, Joseph H., and David J. D. Earn. 2010. "Multiple Transmission Pathways and Disease Dynamics in a Waterborne Pathogen Model." *Bulletin of Mathematical Biology* 72:1506–1533.

Tilly, Richard H. 1990. *Sozialer Protest in Deutschland im 19. Jahrhundert.* GESIS Datenarchiv, Köln. ZA8040 Data file version 1.0.0. https://doi.org/10.4232/1.8040.

Torres Codeço, Cláudia. 2001. "Endemic and epidemic dynamics of cholera: the role of the aquatic reservoir." *BMC Infectious Diseases* 1 (1).

Tuite, Ashleigh R., Joseph Tien, Marisa Eisenberg, David J.D. Earn, Junling Ma, and David N. Fisman. 2011. "Cholera Epidemic in Haiti, 2010: Using a Transmission Model to Explain Spatial Spread of Disease and Identify Optimal Control Interventions." *Annals of Internal Medicine* 154 (9): 593–601.

Wang, Jin. 2002. "Mathematical Models for Cholera Dynamics—A Review." *Microorganisms* 10 (12): 2358.