The contribution of infrastructure investment to Britain's urban mortality decline 1861–1900*

Jonathan Chapman†

January 13, 2017

Version 3.3

Abstract

It is well-recognized that both improved nutrition and sanitation infrastructure are important contributors to mortality decline. However the relative importance of the two factors is difficult to quantify, since most studies are limited to testing the effects of specific sanitary improvements. This paper uses new historical data regarding total investment in urban infrastructure, measured using the outstanding loan stock, to estimate the extent to which the mortality decline in England and Wales between 1861 and 1900 can be attributed to government investment. Fixed effects regressions indicate that infrastructure investment explains approximately 22–25% of the decline in mortality between 1861 and 1900, once time trends are accounted for. Since these specifications may not fully account for the endogoneity between investment and mortality, I perform additional specifications using lagged investment as an instrument for current investment. These estimates suggest that government investment was the major contributor to mortality decline, explaining up to 60% of the reduction in total urban mortality between 1861 and 1900 and 88% between 1861 and 1890. Additional results indicate that investment in urban infrastructure led to declines in mortality from both waterborne and airborne diseases.

Keywords: sanitation, mortality, public investment, Britain, urban infrastructure.

*I thank Philip Hoffman, Jean-Laurent Rosenthal and Erik Snowberg for their helpful advice. I am also grateful for comments from Matthew Chao, Samantha Myers, Matthew Shum, Frank Trentmann, and audiences at Caltech, All-UC Economic History workshops and the Economic History Society. I am grateful for financial support from the History Project of the Institute of New Economic Thinking, and from NSF grant 1357995. All errors remain my own.

†Division of Social Science, New York University Abu Dhabi.
e-mail: jchapman@nyu.edu.

1 Introduction

Between 1851 and 1900 mortality rates in Britain declined by almost 20%. Over the same period, local government expenditure on urban infrastructure increased rapidly, so that by 1890 spending by local authorities accounted for over 41% of total public expenditure, with much of the money used for water supply and sewers (Lizzeri and Persico, 2004). This simple pattern leads to the natural conclusion that government sanitation expenditure was the driving force behind the improvement in life expectancy. This belief is also supported by evidence from other countries showing that investment in sanitary infrastructure, such as clean water supply, can have positive effects on mortality both in the present day (e.g. Günther and Fink, 2011; Zwane and Kremer, 2007; Deaton, 2006) and historically (e.g. Cain and Rotella, 2008; Troesken, 1999).

Yet the role of public health in explaining British mortality decline in the nineteenth century remains disputed. The classic explanation of the dramatic fall in mortality rates after 1850—due to McKeown (1976)—has emphasized the importance of better nutrition rather than improvements in the sanitary environment. This conclusion followed from estimates showing that the greatest contribution to the decrease in mortality rates during this period came from reductions in airborne, rather than waterborne or foodborne, diseases. More recent studies (Williamson, 2002; Szreter, 2005), however, have questioned his conclusion without ending the debate or pinning down the precise quantitative impact that sanitary investment had on mortality. In particular, this later work has argued that McKeown's thesis overlooks the potential contribution of sanitary reform in reducing overcrowding (and hence deaths from airborne diseases) and does not account for differences in the death rates from different airborne diseases (e.g Woods, 1984; Szreter, 2005). After accounting for the latter factor Szreter (2005) argues that "the classic sanitation diseases come to the fore" in explaining the mortality decline after 1850 (p115).

The importance of government public health interventions in the early twentieth-century is supported by evidence from other countries. Cain and Rotella (2001), for example, estimate that a 1% increase in sanitation expenditures would have led to close to a 3% decline in the annual death rate in 48 American cities between 1899 and 1929. Cutler and Miller (2005) find that clean water technologies had a social rate of return that was 23 to 1 in major US cities in the early twentieth-century (see also Troesken, 2002; Kesztenbaum and Rosenthal, 2013). Ferrie and Troesken (2008) find that improvements to Chicago's water supply led to reduced mortality not only from waterborne disease, but also from several other causes of death including tuberculosis, pneumonia and kidney failure. Several studies within the development literature also show significant effects of water improvements and sanitation access on health outcomes, particularly amongst infants (e.g. Zwane and Kremer, 2007; Ahuja et al., 2010; Fink et al., 2011; Zhang, 2012). However, the relative importance of infrastructure and better nutrition in increasing life expectancy remains unresolved (Fogel, 2004). Few studies assess several types of infrastructure spending together (although see Alsan and Goldin (2015)), and as a result cannot measure the overall importance of government's ability and willingness to invest in public infrastructure to achieving mortality decline.

In this paper I analyze Britain's mortality decline through constructing and putting to use a new panel dataset identifying town-level infrastructure investment across England and Wales between 1861 and 1900. This dataset combines information on town–level infrastructure investment—which I measure by the extent of town council loans outstanding—with information on local mortality rates. During this period decisions over investment in public goods were made by local town councils, leading to great variation in the extent of investment across the country—variation which can be exploited for empirical analysis. Most of the investment that town councils undertook was focused on goods that improved the sanitary environment, including items such as the street paving, public parks and sewer systems alongside clean water. In contrast to previous studies, I use data from a large number of districts, rather than relying on particular case studies (e.g. Woods, 1984) or using small samples of towns (Millward and Sheard, 1995; Millward and Bell, 1998). By combining this expenditure data with mortality information drawn from Registration reports I am able to estimate the relative importance of spending by town councils in reducing mortality, accounting for changes in town wealth.

Several features of the particular historical setting facilitate identifying the overall impact of infrastructure investment. First, this period marked the very beginning of the public health movement, meaning that the counterfactual—of essentially no public investment—is very clear. Second, at this time responsibility for infrastructure investment fell almost exclusively on local governments, removing concerns that the data excludes spending by other authorities (such as different levels of government, or non-governmental organizations). Third, I can capture the combined impact of a broad range of infrastructure, rather than focusing on one particular type of investment (e.g. sewage systems).

The analysis proceeds in two steps. First, I use ordinary least squares specifications to establish the fact that infrastructure expenditure had a negative impact on overall mortality. Once demographic control variables or town fixed effects are included in the specifications, there is a clear evidence that infrastructure investment led to significant declines in mortality rates. In particular, infrastructure investment is estimated to explain between 22% and over 100% of the decline in mortality between 1861 and 1900, with the range determined largely by whether time trends are accounted for.

While these results indicate the effectiveness of infrastructure investment, the breadth of the range provides limited insight into the relative contribution of infrastructure as opposed to other causes of mortality decline. While the bottom of the range provides a useful lower bound, it is likely to underestimate the effects since it does not account for the fact that expenditure on public goods was not random: towns would be more likely to invest where health problems were greatest. This conclusion is supported by the fact that a simple regression shows a positive correlation between mortality and public infrastructure stock.

To address this issue, I instrument for expenditure on infrastructure using the level of infrastructure in the previous decade. This approach requires that lagged investment affects mortality only through contemporary investment; an approach justified by the fact that previous research has found that reductions in mortality are associated with future mortality declines only over periods that are very short (1–2 year) in comparison to the data aggregated by decade used in this analysis.

The results of the instrumental variables regressions show that infrastructure investment was the major contributor to urban mortality decline in the second half of the nineteenth century. The main results indicate that between 54% and 60% of the decline in total mortality is explained by infrastructure investment. Only considering the period before 1890, before new infrastructure such as tramways and electricity supply appeared in town accounts, 88% of the total urban mortality decline is explained by infrastructure investment.

I undertake a number of tests to ensure the validity of the instrumental variables estimates. First, I check that the effect is not driven by particular sub-periods of the analysis. Second, as a placebo test I estimate the same specifications utilizing mortality from childbirth and (separately) violence as dependent variables. Change in childbirth mortality during this period was driven largely by improved understanding of hygiene and so this specification serves as a test of whether the results are capturing behavioral change (e.g. hand-washing) rather than infrastructure investment. Similarly, mortality from violence is a useful placebo since it declined significantly during this period, is likely to have been associated with increases in town wealth, but again should not be directly related to spending on urban infrastructure. The results show no evidence of any statistically significant relationship between infrastructure investment and either variable. Finally, using separate data for the 1871–1890 period I control for mortality trends in the rural areas surrounding towns.

By so doing, I treat these rural areas as a counterfactual for urban areas, and account for any district-specific time trends such as weather shocks or improved medical understanding. The estimated effect of the sanitation infrastructure is robust to this test and remains large and statistically significant. Further, as expected, there is no evidence of any relationship between infrastructure investment and mortality in the parts of districts not containing rural areas.

Building on these results, I analyze the contribution of infrastructure investment to the decline in mortality from different types of disease. The largest effects are on waterborne diseases (cholera, diarrhea and typhoid), with infrastructure investment accounting for approximately 100% of the decline in mortality from these diseases between 1871 and 1900. However, I also find significant evidence that infrastructure investment accounted for up to 30% of the decline in mortality from airborne diseases. This shows that public health investment had effects beyond diseases most directly effected by sanitation either directly through reducing transmission of disease (for instance as a result of reducing overcrowding) or indirectly through strengthening immune systems.

Together, these estimates indicate that government infrastructure investment was the major contributor to the mortality decline in England and Wales between 1861 and 1900. Government engagement in public health was crucial to overcoming the mortality penalty associated with urbanization. These findings are particularly striking if we consider that the benefits to public health investments were by no means exhausted at this point in time. Even in 1914 not all urban households had access to piped water. It was not until the very end of the century that the benefits of water chlorination were recognized. Similarly programs of social housing and slum clearance were by no means fully developed until after 1900. Once these investments are properly accounted for, the longer run contribution of public works to urban mortality decline may have been even greater.

2 Historical background and data

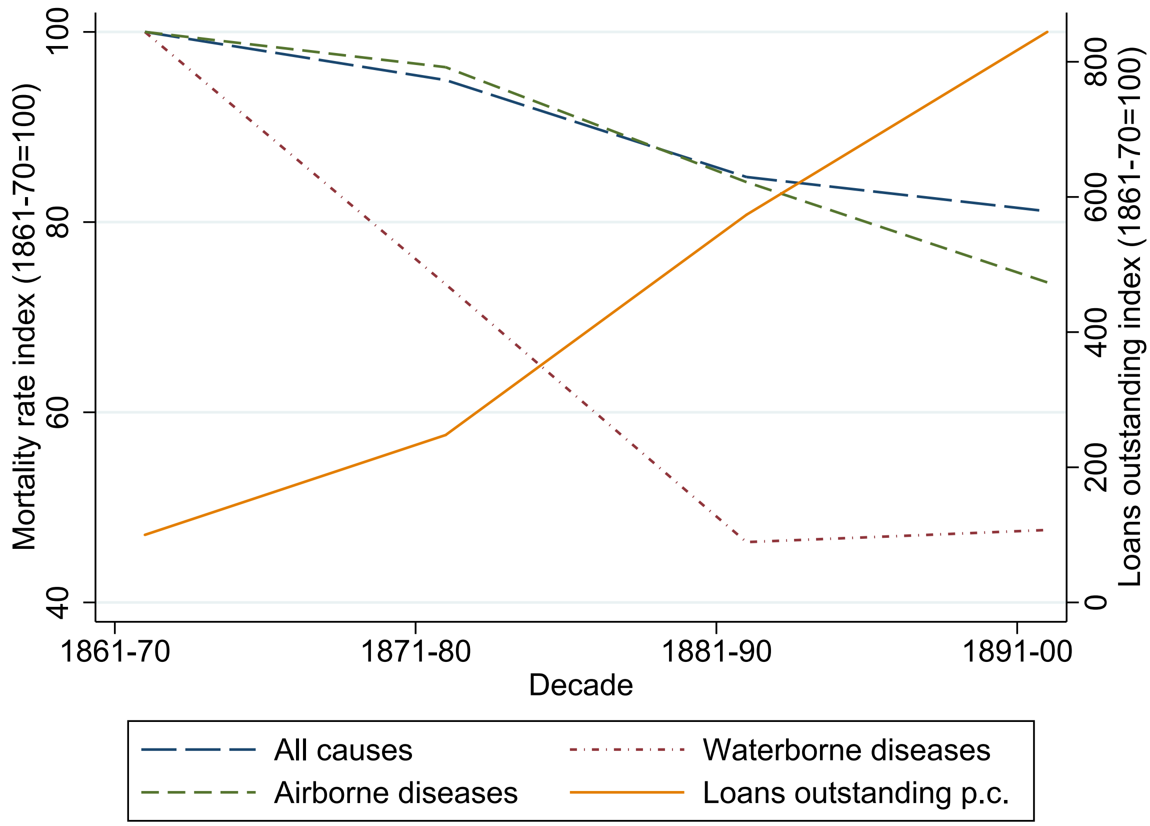
Britain became a much healthier place in the second half of the nineteenth century, with crude total mortality rates falling from 22 to 18 per 1,000 living between 1851 and 1900. Deaths from waterborne diseases such as cholera and diarrhea fell at an even faster rate, as shown in Figure 1. At the same time—as shown on the right hand axis of the figure—the level of spending on urban infrastructure increased dramatically, with the level of loans outstanding used to finance that investment increasing more than eight fold over the same period.

However, this overall picture of mortality decline and urban investment masks significant variation in the experience across different localities. While life expectancy increased across all major cities during the second half of the century, the extent of the increase differed considerably across different towns. This is illustrated by the two towns, Hull and Sunderland, highlighted in Figure 2. While both towns had similar life expectancy at birth in the decade 1861-1870—if anything slightly lower in Hull—by 1891-1900, life expectancy in Hull was three years higher than Sunderland.

The question for this paper is whether, and to what extent, these differences in mortality between towns were caused by different levels of sanitation investment. As suggestive evidence, in 1891-1900 Hull—where life expectancy rose sharply—spent an average of £6.6 per capita each year on sanitation public goods, while Sunderland—where life expectancy stagnated—spent only £3.2 per capita.[1] To answer this question comprehensively I construct a dataset that measures mortality and infrastructure investment across England and Wales in the second half of the nineteenth century.

[1] These figures relate to the Sunderland and Hull Registration Districts respectively, and are based on the dataset discussed in detail in the following subsection.

Figure 1: Rapid growth in infrastructure loans between 1861 and 1900 coincided with decline in overall mortality of 20% and decline in mortality from waterborne diseases of over 50%.



Each variable is displayed as an index with 1861-70 as the base period. Source: Author's calculations using mortality data from Decennial Reports of the Registrar General and loans outstanding data from Local Taxation Returns. Estimates based on approximately 400 registration districts containing an urban area in 1881. See text for further details.

2.1 Data sources

Financial data are drawn from the *Local Taxation Returns* reported to Parliament and collected in the *Parliamentary Papers* collection. These reports detail the annual accounts of every town council—the bodies responsible for the vast majority of infrastructure investment. Data was collected for all "urban sanitary authorities" for each year from 1867 to 1900.[2] This includes approximately 900 towns in total, which were all granted standardized

[2]Prior to 1873 the accounts are reported under the titles of Local Boards of Health and Improvement Commissions— the bodies which were renamed Urban Sanitary Authorities in the 1872 Public Health Act.

Figure 2: General increase in life expectancy across English cities, but with significant variation in experiences between towns.

Source: Data from Szreter (2005).

expenditure powers under the terms of the 1872 Public Health Act. The accounts report the value of loans outstanding in each year, with the values disaggregated from 1884 onwards. They also report the value of the rateable value of property in each district, which formed the tax base available to councils. I translate these nominal values into real values using the Rousseaux Price Index (Mitchell, 1971, pp. 723-4) following Millward and Sheard (1995).

Data on cause of death in different districts are drawn from official statistics reported by the Registrar General for the period 1861-1900. The geographic unit of analysis is the Registration District, of which there were approximately 630 across England and Wales during this period. The primary source is a series of decennial reports digitized by Woods (1997). These reports are well known to both economic historians and demographers since they provide a wealth of data on both cause and age of death (in five or ten year intervals)

The towns included in the analysis are those designated as Urban Sanitary Authorities in 1881.

in each district averaged by decade.[3]

I supplement the decennial data with information from the *Quarterly Returns of the Registrar General* regarding annual third quarter mortality data for the period 1871-1890.[4] This source reports mortality statistics at a more disaggregated geographical level than the decennial data, and allow me to distinguish between mortality in urban and rural parts of Registration Districts (although at a lesser level of detail than in the decennial reports).[5] Unfortunately, collecting this data is complicated by the fact that sub-district information was only reported in quarterly, rather than annual, reports. To create a consistent time series, the third quarter was chosen for transcription since waterborne diseases—such as diarrhea—were particularly likely to strike during the summer months. As such, this period provides the best test of whether infrastructure had an effect—if it had no impact in the third quarter, it seems unlikely it would have made a substantial contribution in the remainder of the year.

Unfortunately, town boundaries during this period did not match the boundaries of the Registration Districts (or Subdistricts) for which mortality data was reported. Large towns comprised whole (and sometimes multiple) Registration Districts, while some Registration Districts included multiple smaller sanitary authorities. Given this issue, I link the financial and mortality data by first linking each town to the Registration Subdistrict(s) in which it was situated using information reported in the 1881 census. Where town boundaries crossed multiple Registration Districts (a relatively rare occurrence), town spending was allocated to each Registration District according to the population residing in each district at the time of the census. Where multiple Registration Districts were combined in a single town (such

[3]Examples of works using these sources include Szreter (2005), Woods and Shelton (1997) and Hanlon (2015).

[4]The reports for the years 1880 and 1882 were not available. In addition, data was missing for some districts in other years as the reports were illegible.

[5]Specifically this information was reported at the level of Registration Subdistrict which were the smaller administrative units underlying Registration Districts. There were approximately 2000 subdistricts in England and Wales.

as Liverpool and Manchester), I combine them into a single district for the purposes of the analysis.

This approach has the advantage of matching directly to the mortality information reported in the Registrar General's Dicennial reports—and it is those reports that provide the most detailed disaggregation of mortality. Further, the boundaries of these districts were, in general, relatively stable between 1860 and 1900, allowing me to construct a panel dataset. Major boundary changes were largely limited to mergers or splits of sanitary districts; where this occurred I construct "synthetic" districts consisting of the larger, merged, district.

More difficulties arise when using the data on Registration Subdistricts to construct urban and rural mortality series, since there were frequent reallocations of boundaries within Registration Districts. To address this issue, when analyzing the rural and urban mortality patterns, I adjust the mortality data for each year to consistent 1881 district boundaries. To do this, I first identified all subdistrict boundary changes between 1871 and 1891 and then re-weighted the data to the 1881 district boundaries based on population weight. A fuller explanation of this procedure is provided in the Appendix. As a result, analysis utilizing the subdistrict data only covers the two decades between 1871 and 1890.

Finally, additional demographic data was collected from census reports.

2.2 Variable definition

Crude mortality rates The key dependent variables in the analysis are crude mortality rates, disaggregated by cause of death and age group. Because each cross-section in our panel covers a decade—following the information reported in the decennial reports—the appropriate measure is the average death rate over the decade:

$$death\_rate\_ageI = \frac{\text{Number of deaths at ageI}}{\text{(Average population at ageI/ 100,000)}}$$

Infrastructure investment The key measure of urban infrastructure investment is the average level of loans outstanding per capita in each district over a decade. This variable is an accurate measure of the level of investment since nearly all town investments in infrastructure needed to be funded by borrowing. In 1902 on average over 95% of the capital invested in trading entities (such as water and gas supply bodies) had been borrowed.[6] Furthermore, the stock of loans outstanding was seen as the single best measure of urban progress by contemporaries (Wohl, 1983, p.112).

Further, and importantly for this paper, most of the loans that were taken out were dedicated to infrastructure that had a clear sanitary component. On average, after 1884 around one-quarter of towns' loans were devoted to each of water and sewer systems, with approximately a further 12% used for spending on streets—all of which would improve the quality of the urban environment (Millward and Sheard, 1995).[7] Other items of infrastructure which were not disaggregated included public parks, public baths and public housing, which could also have had an impact on reducing mortality.[8] There were also some spending items which would not have contributed to mortality declines including gas supply and, in larger towns after 1890, tram systems and electricity supply. Any concern that the measure partially captures infrastructure which does not have a clear sanitary impact should be balanced against the fact that not all urban spending on sanitation would be included in a measure of infrastructure. For instance, neither "scavenging"—the process by which privy middens were emptied—nor cleaning of streets are included.

One issue with the loans data is that it is significantly right skewed, since a few towns

[6]Author's calculation based on figures in *Report from the Joint Select Committee of the House of Lords and the House of Commons on Municipal Trading*, **1903** (270)VII.1.

[7]Unfortunately, detailed disaggregated information is not available before 1884, and so I use the total stock of loans outstanding throughout the analysis. More detailed information on the percentage of loans devoted to different purposes is presented in the descriptive statistics in the appendix.

[8]Public housing in particular would be likely to have a significant impact but was only a very small part of spending until the end of the nineteenth century.

spent an extremely high amount.[9] As a result, some observations have very high leverage in some specifications. These high leverage points are a concern since understanding the size of the effect (rather than just its direction) is an important goal of this paper. As such, I transform the loan stock per capita data using a square root transformation, and use the resulting variable as the main independent variable in the remainder of the paper.

Other variables The second major independent variable used is the urban "rateable value per capita" in each Registration District. This variable captures the size of the urban tax base in each district since the taxes raised by local authorities—and used to repay loans—were property taxes (rather than, for instance, income taxes). As such, I use this variable as a proxy for urban wealth. As with the level of the loan stock per capita, I apply a square root transformation to this variable.

I also include several demographic control variables for each district, including district population, population density, and the percentage of the population aged over 5. In additional specifications presented in the appendix, I include measures of the population of the largest town in each district, population growth, and the percentage of population in different age groups. Fuller details of the construction of these variables are presented in the Appendix.

3 Empirical specification and identification

The data is used to construct a four-period panel dataset, where each cross-section relates to a decade reported in the decennial reports of the Registrar General: 1861-1870, 1871-1880, 1881-1890 and 1891-1900.[10] I then estimate the effect of infrastructure investment on deaths

[9] Figure A.III in the appendix presents a density plot of the loans outstanding variable.

[10] Descriptive statistics of the variables used in the regressions are presented in the Appendix.

using the specifications of the following form:

$$death\_rate_{i,t} = \alpha + \beta InfrastructureInvestmentPC_{i,t} + \gamma X_{i,t} + \delta_0 Z_i + \delta_1 T + \epsilon_{i,t}$$

where i indexes Registration Districts and and t indexes each decade. The variable $death\_rate$ measures the number of deaths per capita, and $InfrastructureInvestmentPC$ is the per capita level of urban infrastructure investment in each district—measured by the square-rooted per capita stock of loans outstanding. X is a vector of control variables, Z_i includes district fixed effects, T is a vector of decade fixed effects, and ϵ is an error term. The basic set of control variables includes the decadal average population in the district, district population density, and the percentage of population aged over 65.[11] As a proxy for district wealth, I also control for the urban tax base per capita in the district (also square rooted). I run an additional set of tests using as a dependent variable third quarter mortality in the urban portion of districts for the period 1871–1890. These tests allow me to include rural mortality as an additional control variable in each district. By so doing, I can account for time varying factors that are common across a whole district—for instance, weather, or improved hygiene—and check whether the point estimates are substantially affected.

Identifying the effects of mortality change is complicated by the endogoneity issues in the location of infrastructure. Towns did not spend their resources at random, and were likely to increase infrastructure investment in response to the disease environment. The effect of this reverse causality could be to mask any beneficial results of infrastructure expenditure on mortality—a hypothesis supported by results below showing a positive correlation between higher spending and higher mortality.

I take two approaches to isolate the causal effect of infrastructure investment on mortality.

[11] In additional specifications presented in the appendix I also include controls for population under 5, and age 5 to 19, the population in the largest town in the district, the district population squared, and population growth. I also include mortality from childbirth and violence (causes unrelated to infrastructure) as control variables.

First, I estimate specifications including district fixed effects. By doing so I account for time-invariant factors, such as location, that affect both the level of mortality and the level of spending. While this approach accounts for many potential sources of endogoneity, it does not address any endogoneity resulting from reverse causality within a decade—for instance, if towns respond to high mortality by building more infrastructure. Even more problematic, it cannot account for the fact that towns may have acted to forestall *expected* increases in mortality through building additional infrastructure.

To account for this potential simultaneity bias I also estimate two stage least squared specifications using the *lagged* level of infrastructure as an instrument for the actual level of infrastructure in the decade. The exclusion restriction here is that the level of infrastructure (measured by the loan stock) in the previous period does not affect the level of mortality in the current period, except through its effect on the infrastructure in the current period. This is a plausible assumption in that the main effect of improved infrastructure would have been to prevent individuals catching the diseases that would eventually kill them. The assumption could be violated, on the other hand, if infrastructure prevents the weakening of the immune system that prevents deaths from a *different* disease in the future. If this were the case, the estimates from the two stage least squares regressions may be biased upwards—that is towards over-estimating the contribution of urban infrastructure. Is this likely to have a major contaminating effect? Ferrie and Troesken (2008) find that in Chicago survivors from typhoid were more likely to die from other diseases in the following years—suggesting that there could be follow on effects. However, they find no "evidence that lagged typhoid rates of greater than 1 or 2 years had any systematic effect" (footnote 11). Further, they suggest that typhoid was unique in having these strong knock-on effects on other diseases. As such, there is little reason to think that these effects will lead to considerable instrumental variables bias. To the extent there may be bias, however, it would be expected to be in the direction of exaggerating the effect. As such these estimates form an upper bound on the contribution of infrastructure investment to the mortality decline.

3.1 Magnitude of effects

To measure the *relative* importance of infrastructure compared to other causes of mortality decline, I compare the estimated effects to the overall mortality decline across the period of study. Specifically, I use the regression results to estimate the reduction in the mortality rate in each district explained by town spending on infrastructure in 1891-1900. I then take an average of this effect, weighted by district size, and compare it to the (weighted) average actual decline that occurred in these districts.

Denoting loans per capita outstanding in district i in the decade ending in year t as $l_{i,t}$, the estimated regression coefficient $\hat{\beta}$, the mortality rate as $d_{i,t}$, and the district population as $p_{i,t}$ the magnitude of the effect is then estimated as:

$$\text{Magnitude} = \frac{\Sigma_i p_{i,t}\hat{\beta}\left(l_{i,t}\right)}{\Sigma_i p_{i,t}\left(d_{i,t} - d_{i,1870}\right)}$$

where t refers to the decade at the end point of the period under consideration (either 1881–1890 or 1891–1900, depending on the specification), and 1870 refers to the decade 1861–70.

4 Results

4.1 Ordinary least squares estimates

Table 1 displays the results of seven specifications analyzing the relationship between infrastructure investment (measured by the per capita stock of loans outstanding) and the total mortality rate. All variables are standardized and so the coefficients should be interpreted as the effect of a one standard deviation increase in the independent variable in terms of standard deviations of the dependent variable.

The first specification includes only infrastructure investment per capita as an independent variable. There is evidence of a statistically significant *positive* relationship between infrastructure investment per capita and the mortality rate. The likely explanation for this is that towns with higher mortality invested more as a reaction to the disease environment.

[Table I here]

However the remaining specifications indicate that this relationship changes sign once other town characteristics likely to be associated with higher mortality are accounted for. Once either the set of demographic control variables (specification (2)) or district fixed effects (specification (4)) are included there is statistically significant evidence that higher investment led to lower mortality.

The remaining specifications explore the robustness of this finding to the inclusion of control variables and different time periods. Specifications (3) and (5) add decade fixed effects to capture changes over time that may be correlated both with mortality and with infrastructure investment. We can see that the coefficient on infrastructure investment is now no longer negative (or statistically significant) in the absence of district fixed effects. Where district fixed effects are included, the effect shrinks in magnitude by approximately 75%. Specification (6) reintroduces the control variables; we can see that there is a further reduction in the estimated effect but that it is still strongly statistically significant.[12] Interestingly, there is no evidence that mortality was affected by changes in the town tax base per capita (a proxy for wealth): specification 6 shows a statistically insignificant and close to zero effect of this variable. In specification (7) I exclude the final decade (1891-1900), since this was the period when new infrastructure that was less associated with sanitary im-

[12]The appendix contains the results of additional specifications including subsets of these control variables, as well as additional control variables including district population squared, percentage of population aged under 5 and aged 5 to 19, the population of the largest town in the district, and population growth. I also estimate one specification with mortality from childbirth and mortality as control variables. The effect sizes and statistical significance are robust to the inclusion of these different groups of control variables.

provements began to be built (electricity supply, for instance). The results are similar when this decade is excluded, suggesting that any difference in the composition of infrastructure in this decade is not affecting the results significantly.

Together these results provide clear evidence that greater spending on urban infrastructure led to lower mortality. They also show that the effect was large. However, the range of the estimated effects leaves the *relative* importance of infrastructure investment open to question. If we do not account for time trends, then the effect is extremely high—over 100% of the total decline—implying that mortality would actually have increased in the absence of infrastructure spending. However, we cannot then be sure that these effects are causal since we may be capturing other factors, such as higher incomes and improved nutrition that occurred during the same period.

The lower estimate on the other hand, is likely to underestimate the effect on mortality, by failing to address the fact that infrastructure investment is endogenous to the sanitary environment: investment would be higher where towns faced greater pressure on existing sanitary systems due to overcrowding or high population growth. Thus this estimate should be seen as providing a lower bound on the overall size of the effect.

4.2 Instrumental variables estimates

The inclusion of fixed effects removes a large degree of the potential endogeneity in investment decisions, by accounting for time-invariant factors that could affect both mortality and the decision to invest. However, as argued above, there may be other forms of bias—such as pre-emptive investment in infrastructure—meaning that these specifications do not capture the full contribution of infrastructure to mortality decline.

To estimate the contribution of infrastructure investment more precisely, I thus use the lagged level of loans outstanding per capita as an instrument for the contemporaneous loan stock. As discussed in more detail previously, this approach is valid so long as the lagged level of infrastructure only affects mortality through the contemporary level. I have argued that this is a valid assumption on the basis that "carry-over" effects, whereby one disease was associated with deaths from other diseases, were limited both in terms of the range of diseases that had this effect and the time span of those effects. Further, if there is bias, it should exaggerate the effects of investment in reducing mortality. As such, these instrumental variables specifications can provide an upper bound on the overall effect of infrastructure investment.

Table 2 presents the results of the two stage least squares estimations. Panel A displays the second stage results, while Panel B presents an abbreviated version of the first stage results, with coefficients on the control variables excluded.[13] Since we rely on the lagged value of the loans outstanding, these specifications consist of three periods only (in contrast to four in Table 1); thus for comparison I also include fixed effects regressions for the same periods. The estimated effects from these specifications indicate infrastructure investment can explain between 23% and 29% of the mortality decline—similar in magnitude to the corresponding specifications estimated over the entire period in Table 1.

[Table II here]

As expected, there is a strongly significant positive relationship between the instrument and the current level of infrastructure investment. Further, the large Kleibergen-Papp statistics (at the bottom of the table) indicate that there are no concerns of weak instruments. Additional tests, presented at the bottom of Table 2 confirm the validity of the instrumental variables approach. The C-statistic, which tests for endogoneity of infrastructure investment, is strongly statistically significant in all specifications.[14]

Columns (1) and (2) presents the estimates from the fixed effects and two stage least squares regressions with no control variables, while (3) and (4) present the results including

[13] Full regression results are presented in the Appendix.

[14] The C-statistic is the difference of the Hansen statistics from the unrestricted equation and the restricted equation.

the control variables. Mirroring the last specification in Table 1, in columns (5) and (6) I exclude the final decade (1891-1900) from the analysis. In all cases, the estimated effect size is significantly higher in the two stage least squares regressions. In all three specifications, the estimated coefficient on the measure of infrastructure investment is negative and strongly statistically significant. The estimated effect sizes indicate that infrastructure investment can account for between 54% and 60% of the decline in urban mortality between 1861 and 1900—or even higher if the effect is estimated using the period before 1890. After accounting for the endogoneity, therefore, spending on infrastructure thus appears to be the major force behind Britain's urban mortality decline.

4.3 Effects on mortality by age and by cause of death

Having established the importance of sanitation investment in reducing overall mortality, we can examine in more detail the specific causes of deaths that were affected. Since the infrastructure investments in question were largely associated with sanitary improvements, the primary causes of death affected are likely to be waterborne diseases such as cholera (although this accounted for only a few deaths during this period), diarrhea and typhoid. However, there is reason to suspect that by reducing the rate at which individuals caught these waterborne diseases, sanitation could also reduce the likelihood that individuals would die of other diseases at a later date. Ferrie and Troesken (2008) find that in Chicago reductions in mortality from typhoid were associated with declines in mortality from kidney failure, tuberculosis and pneumonia. Szreter (2005) argues that infrastructure investment may have reduced mortality from airborne disease through reducing overcrowding and hence the spread of disease.

Table 3 explores these hypotheses through re-estimating the fixed effects and instrumental variables specifications, but using mortality by various different causes of death as the dependent variable. The first two specifications examine mortality from three major waterborne diseases, cholera, diarrhea and typhoid. These diseases would be directly affected by sanitation investment, and so are likely to have particularly sizable effects. Unfortunately, however, typhoid was not separately distinguished in the Registrar Generals reports until the decade 18711880, and so the estimated decline in mortality is measured over the period after 1871.[15]

Specifications (3) and (4) analyze a group of airborne diseases—including "diseases of the respiratory system" (such as bronchitis), pulmonary tuberculosis, smallpox, scarlet fever, whooping cough, measles and diphtheria. Specifications (5)–(8) then carry out placebo tests using as the dependent variable mortality from two causes, violence and childbirth, that would not be expected to be affected by investments in urban infrastructure.

[Table III here]

The results show strong and statistically significant effects of infrastructure spending on mortality from both waterborne and airborne diseases. The instrumental variables specifications indicate that mortality from waterborne diseases was reduced by approximately 100% by infrastructure investment—that is, it would have increased in the absence of public health expenditure.

There is also evidence that urban infrastructure contributed significantly to the reduction in airborne disease, accounting for between 16% and 30% of the decline between 1861 and 1900. It is interesting—and reassuring—that, in comparison to the waterborne disease estimates, the instrumental variables estimate is closer to the OLS specification (and the endogeneity test indicates they cannot be statistically distinguished at conventional significance levels). Such a finding is consistent with the argument that the fixed effects estimates are biased downwards because a reaction to those (waterborne) diseases directly affected by sanitation infrastructure—it is intuitive that the effect would be smaller in the case of

[15] Additional specifications using alternative measures for waterborne mortality as the dependent variable are included in the Appendix.

airborne diseases.

Disaggregating mortality by cause of death also provides a valuable placebo test to check that the instrumental variables estimates are not incorrectly capturing improvements in hygiene or medical understanding. Specifications (5) and (6) use mortality in childbirth (including puerperal fever) as the dependent variable. This provides a good test of whether the effects we find are causal, since the major contributor to declines in maternal mortality was most likely improved medical knowledge, rather than an improved sanitary environment (Loudon, 2000). As such, finding a negative coefficient in these specifications would cause concern that the measure of infrastructure investment is still capturing the effects of broader improvements in medical understanding. Similarly, in specifications (7) and (8) the dependent variable is mortality from violence (including suicide), as a check that the results for infrastructure investment are not capturing increasing wealth (on the basis that richer towns would have lower crime). There is no evidence of such an effect in either the fixed effects or the instrumental variables regressions in these specifications.[16]

4.4 Comparing urban and rural areas

One remaining concern could be that the measure of infrastructure investment are picking up something more general about "urban" versus "rural" areas. The registration districts we have analyzed were, in many cases, comprised of both rural and urban portions. An increase in spending per capita could therefore result from the spread of urban areas within a registration district. Thus the estimated effect of infrastructure investment could be capturing other factors associated with urbanization such as better education or understanding of disease transmission.

To address this concern, in this subsection I use data at the registration subdistrict level

[16] Additional results presented in the appendix indicate that these tests are also passed when looking only at the period 1871–1890.

to distinguish the urban and rural parts of registration districts for the period 1871–1890 (as discussed in Section 2 difficulties in creating consistent district boundaries at the subdistrict level precluded carrying out this analysis over a longer time period). The urban parts of subdistricts are those that contained an urban area in 1881, while rural parts are those subdistricts that contained no urban area at all in 1881. I then use data regarding mortality in the third quarter from cholera, diarhea and fever (which would include both typhoid and other forms of fever).[17]

I carry out three tests of the effects of urban infrastructure using this data. First, I check that there is evidence that infrastructure led to a decline in mortality when focusing only on mortality in urban sub-districts. Second, I include rural mortality as a control variable in these specifications, as a check that the results are not spuriously capturing other time-varying factors such as local weather patterns that affect mortality. Third, I carry out a placebo test with mortality in rural areas as the dependent variable—checking whether greater infrastructure spending is capturing any effects that affected the broader area of a district. This latter test rules out a situation where everyone in a district becomes better informed about disease, leading to lower mortality and higher spending in urban areas.

The results, shown in Table 4 again show consistent evidence of infrastructure investment on waterborne mortality in urban areas. Specification (1) displays the estimates for all registration districts, while specification (2) includes only districts with both urban and rural areas. Specification (3) shows that the estimated effect is essentially unchanged when controlling for mortality in rural areas. Finally, specifications (4) and (5) use rural mortality as the dependent variable— and show that there is no evidence that infrastructure spending

[17]The appendix presents similar specifications using total mortality as the dependent variable. The results are similar, except that the relationship with infrastructure investment is statistically insignificant for districts that were partly rural when the linear control for district population is included. This is likely to reflect two factors. First, the third quarter data is noisier than the annual data since it relates to a smaller sample and because of the need to adjust for boundary changes. Second, many of the forms of mortality that had the largest decline, such as tuberculosis, had a lower "base" mortality in the summer months Fares et al. (2011).

affected mortality in those areas whether control variables are included (specification (4)) or not (specification 5)).

[Table IV here]

Together these results provide strong evidence that the effects we are capturing for urban infrastructure investment relate directly to mortality in urban areas of the registration districts. There is no evidence that we are capturing any effect that would affect mortality in both towns and their surrounding areas.

5 Discussion

This paper has tested the effects of government spending on sanitation infrastructure on mortality rates from waterborne disease in England between 1861 and 1900. During this period local government took responsibility for improving urban environments, leading to rapid growth in expenditure on public goods such as clean water supply, sewer systems and street paving and cleaning. Using a new panel dataset, I estimate that this investment was responsible for between 22% and over 100% of the mortality decline during this period. Using instrumental variables regressions to estimate this effect more precisely identifies an effect of approximately 60%: indicating that public investment was the single most important factor in reducing urban mortality at this period. Nor were these benefits limited to classic "sanitation diseases", with investment in infrastructure also associated with a decline in mortality from airborne diseases. Further estimates suggest that between 1871 and 1890, when investment was most focused on sanitation infrastructure, infrastructure investment accounted for 88% of the mortality decline.

Together, these results support an explanation of Britain's mortality decline based predominantly around the provision of public infrastructure, rather than nutrition. Further, they suggest that in estimating the potential benefits of public investment, we should be careful to properly account for the wide range of investments that can improve health outcomes. Some of these investments may offer less clear cut causal mechanisms than, for instance, a new water filtration plant. However they may nevertheless offer important cumulative benefits that have significantly improve urban environments and hence life expectancy—for instance through reducing overcrowding and hence the spread of airborne disease. Future research will look at disaggregating the role of these different types of infrastructure in greater detail.

References

Ahuja, A., M. Kremer, and A. P. Zwane (2010). Providing safe water: evidence from randomized evaluations. *Annual Review of Resource Economics 2*(1), 237–256.

Alsan, M. and C. Goldin (2015). Watersheds in infant mortality: The role of effective water and sewerage infrastructure, 1880 to 1915. Technical report, National Bureau of Economic Research.

Cain, L. and E. Rotella (2001). Death and spending: Urban mortality and municipal expenditure on sanitation. *Annales de démographie historique* (1), 139–154.

Cain, L. and E. Rotella (2008). Epidemics, demonstration effects, and investment in sanitation capital by us cities in the early twentieth century. *Quantitative History: The Good of Counting*, 34–53.

Cutler, D. and G. Miller (2005). The role of public health improvements in health advances: the twentieth-century United States. *Demography 42*(1), 1–22.

Deaton, A. (2006). The great escape: A review of robert fogel's the escape from hunger and premature death, 1700–2100. *Journal of Economic Literature 44*(1), 106–114.

Fares, A. et al. (2011). Seasonality of tuberculosis. *Journal of global infectious diseases 3*(1), 46.

Ferrie, J. P. and W. Troesken (2008). Water and Chicago's mortality transition, 1850–1925. *Explorations in Economic History 45*(1), 1–16.

Fink, G., I. Günther, and K. Hill (2011). The effect of water and sanitation on child health: evidence from the demographic and health surveys 1986–2007. *International Journal of Epidemiology 40*(5), 1196–1204.

Fogel, R. W. (2004). *The escape from hunger and premature death, 1700-2100: Europe, America, and the Third World*, Volume 38. Cambridge University Press.

Günther, I. and G. Fink (2011). Water and sanitation to reduce child mortality: The impact and cost of water and sanitation infrastructure. *World Bank Policy Research Working Paper Series, 5618*.

Hanlon, W. W. (2015). Pollution and mortality in the 19th century. Technical report, National Bureau of Economic Research.

Kesztenbaum, L. and J.-L. Rosenthal (2013). Public goods and health inequality: lessons from Paris, 1880–1914. *Unpublished Manuscript*.

Lizzeri, A. and N. Persico (2004). Why did the elites extend the suffrage? Democracy and the scope of government, with an application to Britain's Age of Reform. *Quarterly Journal of Economics 119*(2), 707–765.

Loudon, I. (2000). Maternal mortality in the past and its relevance to developing countries today. *The American Journal of Clinical Nutrition 72*(1), 241s–246s.

McKeown, T. (1976). *The modern rise of population*. Edward Arnold (London).

Millward, R. and F. N. Bell (1998). Economic factors in the decline of mortality in late nineteenth-century Britain. *European Review of Economic History 2*(3), 263–288.

Millward, R. and S. Sheard (1995). The urban fiscal problem, 1870–1914: Government expenditure and finance in England and Wales. *Economic History Review 48*(3), 501–535.

Mitchell, B. R. (1971). *Abstract of British historical statistics*. CUP Archive.

Registrar General (1895). *Supplement to the fifty-fifth annual report of the Registrar General of Births, Deaths and Marriages for England.* London: HMSO.

Schürer, K. and E. Higgs (2014). *Integrated Census Microdata (I-CeM), 1851-1911. [computer file].* Colchester, Essex: History Data Service, UK Data Archive [distributor].

Southall, H.R; Gilbert, D. G. I. (January 1998). *Great Britain Historical Database: Vital Statistics, Mortality Statistics, 1851–1920 [computer file].* Colchester, Essex: UK Data Archive [distributor], SN: 3708.

Southall, H., P. Ell, D. Gatley, and I. Gregory (August 2004). *Great Britain Historical Database: Census Data: Parish-Level Population Statistics, 1801–1951 [computer file].* Colchester, Essex: UK Data Archive [distributor] SN: 4560.

Szreter, S. (2005). *Health and Wealth: Studies in History and Policy.* Rochester University Press.

Troesken, W. (1999). Typhoid rates and the public acquisition of private waterworks, 1880-1920. *Journal of Economic History 59*(4), 927–948.

Troesken, W. (2002). The limits of Jim Crow: race and the provision of water and sewerage services in American cities, 1880–1925. *Journal of Economic History 62*(3), 734–772.

Williamson, J. (2002). *Coping with city growth during the British industrial revolution.* Cambridge University Press.

Wohl, A. (1983). *Endangered lives: public health in Victorian Britain.* JM Dent and Sons Ltd.

Woods, R. (1984). Mortality patterns in nineteenth-century England. In R. Woods and J. Woodward (Eds.), *Urban disease and mortality in nineteenth-century England.* London and New York: Batsford Academic and Educational Ltd.

Woods, R. (March 1997). *Causes of death in England and Wales, 1851–60 to 1891–1900: The Decennial Supplements [computer file]*. Colchester, Essex: UK Data Archive [distributor] SN: 3552, http://dx.doi.org/10.5255/UKDA-SN-3552-1.

Woods, R. and N. Shelton (1997). *An atlas of Victorian mortality*. Liverpool University Press.

Zhang, J. (2012). The impact of water quality on health: evidence from the drinking water infrastructure program in rural China. *Journal of Health Economics 31*(1), 122–134.

Zwane, A. P. and M. Kremer (2007). What works in fighting diarrheal diseases in developing countries? A critical review. *World Bank Research Observer 22*(1), 1–24.

Tables

Table 1: Infrastructure investment associated with large declines in total mortality between 1861 and 1900 after accounting for time invariant town characteristics.

	DV=Total mortality rate (all ages, standardized coefficients)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Infrastructure investment p.c.	0.073**	-0.165***	0.078	-0.772***	-0.208***	-0.147***	-0.133***
	(0.036)	(0.056)	(0.053)	(0.029)	(0.027)	(0.030)	(0.035)
Population		0.267***	0.266***			-0.149*	-0.214**
		(0.049)	(0.047)			(0.087)	(0.104)
Population density		0.179***	0.113***			-0.043	-0.058
		(0.038)	(0.036)			(0.073)	(0.118)
Percent aged over 65		-0.427***	-0.209***			0.193***	0.218***
		(0.040)	(0.042)			(0.052)	(0.057)
Tax base p.c.		-0.234***	-0.081			0.017	0.002
		(0.059)	(0.058)			(0.040)	(0.042)
Reg Dist FE	N	N	N	Y	Y	Y	Y
Decade FE	N	N	Y	N	Y	Y	Y
Period	1861-1900	1861-1900	1861-1900	1861-1900	1861-1900	1861-1900	1871-1890
% decline explained	0.00	0.28	0.00	1.29	0.35	0.25	0.22
Obs.	1520	1520	1520	1520	1520	1520	1140
No. Districts	380	380	380	380	380	380	380

All coefficients are standardized. Observations are "Registration District–decades", between 1861-1870 and 1891-1900. *Infrastructure investment p.c.* is the average stock of loans outstanding over the decade divided by average district population. *Tax base p.c.* is the average per capita rateable value of property. Population density is the population per acres. "% decline explained" is the estimated reduction in mortality explained by the level of infrastructure investment as a percentage of the total decline in mortality from 1861–1900 (specifications (1)–(6)) or 1861–1890 (specification (7)). Standard errors are clustered by Registration District, and are displayed in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Instrumental variable regressions show that infrastructure investment explained more than half of the urban mortality decline between 1861 and 1900.

	(1) FE	(2) IV	(3) FE	(4) IV	(5) FE	(6) IV
Panel A: Fixed effects and 2SLS specifications for mortality at all ages						
Infrastructure investment p.c.	-0.174***	-0.324***	-0.152***	-0.361***	-0.135**	-0.519***
	(0.030)	(0.047)	(0.037)	(0.068)	(0.059)	(0.140)
Tax base p.c.			0.138**	0.270***	0.210***	0.410***
			(0.055)	(0.067)	(0.075)	(0.105)
Panel B: Abbreviated first stage regressions for infrastructure investment per capita						
Lag Infrastructure investment p.c.		0.557***		0.419***		0.399***
		(0.033)		(0.038)		(0.066)
Tax base p.c.				0.444***		0.440***
				(0.049)		(0.068)
Controls	Y	Y	Y	Y	Y	Y
Reg Dist FE	Y	Y	Y	Y	Y	Y
Decade FE	Y	Y	Y	Y	Y	Y
Period	1871-1900	1871-1900	1871-1900	1871-1900	1871-1890	1871-1890
% decline explained	0.29	0.54	0.25	0.60	0.23	0.88
Hansen C statistic		15.57		10.51		9.03
p-value		0.00		0.00		0.00
Kleibergen-Papp Stat		290		124		37
Obs.	1140	1140	1140	1140	760	760
No. Districts	380	380	380	380	380	380

All coefficients are standardized. Observations are "Registration District–decades", between 1861-1870 and 1891-1900. *Infrastructure investment p.c.* is the square root of the average stock of loans outstanding over the decade divided by average district population. *Tax base p.c.* is the square root of the average per capita rateable value of property. Control variables include the population density, district population, the percentage of population aged over 65 and the level of tax base—see the Appendix for full results for these variables. "% decline explained" is the estimated reduction in mortality explained by the level of infrastructure investment as a percentage of the total decline in mortality from 1861–1900 (specifications (1)–(4)) or 1861–1890 (specification (5)–(6)).

Two stage least squares regressions instrument for infrastructure investment using the lagged level of infrastructure investment.

Standard errors are clustered by Registration District, and are displayed in parentheses.

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

Table 3: Infrastructure spending caused significant decline in mortality from both waterborne and airborne diseases, but does not have any effect in a placebo tests with mortality from childbirth or violence as a dependent variable.

	DV=Mortality rate at all ages by cause							
	Waterborne		Airborne		Childbirth		Violence	
	FE	IV	FE	IV	FE	IV	FE	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Infrastructure investment p.c.	-0.167***	-0.372***	-0.092**	-0.170**	-0.084	-0.047	-0.055	-0.160
	(0.045)	(0.087)	(0.038)	(0.077)	(0.069)	(0.125)	(0.070)	(0.113)
Reg Dist FE	Y	Y	Y	Y	Y	Y	Y	Y
Decade FE	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y
Period	1871-1900	1871-1900	1871-1900	1871-1900	1871-1900	1871-1900	1871-1900	1871-1900
% decline explained	0.50	1.11	0.16	0.30	1.34	0.76	0.19	0.54
Hansen C statistic		7.42		1.53		0.12		1.09
p-value		0.01		0.22		0.73		0.30
Kleibergen-Papp Stat		124		124		124		124
Obs.	1140	1140	1140	1140	1140	1140	1140	1140
No. Districts	380	380	380	380	380	380	380	380

Observations are "Registration District–decades", between 1861-1870 and 1891-1900. *Infrastructure investment p.c.* is the square root of the average stock of loans outstanding over the decade divided by average district population. *Tax base p.c.* is the square root of the average per capita rateable value of property. Control variables include the population density, district population, the percentage of population aged over 65 and the level of tax base—see the Appendix for full results for these variables.
"% decline explained" is the estimated reduction in mortality explained by the level of infrastructure investment as a percentage of the total decline in mortality from 1871–1900 (specifications (1)–(2)) or 1861–1900 (specification (2)–(8)). The shorter period is used in the first two specifications since typhoid was not distinguished in the Registrar General's reports before 1871,

Instrumental variable regressions instrument for infrastructure investment using the lagged level of infrastructure investment. First stage results are the same as Table 2.

Standard errors are clustered by Registration District, and are displayed in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Estimated effect on waterborne mortality in urban areas is similar after controlling for neighbouring rural mortality.

	All districts	Districts with rural portions			
	DV = urban mortality	DV = urban mortality		DV = rural mortality	
	(1)	(2)	(3)	(4)	(5)
Infrastructure investment p.c.	-0.287***	-0.214**	-0.209**	-0.025	-0.032
	(0.078)	(0.088)	(0.088)	(0.108)	(0.104)
Rural Waterborne morality			0.136***		
			(0.039)		
Reg Dist FE	Y	Y	Y	Y	Y
Decade FE	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	N
Period	1871-1890	1871-1890	1871-1890	1871-1890	1871-1890
Obs.	757	550	550	550	550
No. Districts	380	275	275	275	275

All variables are standardized. Observations are "Registration District–decades", for the two decades 1871-1880 and 1881-1890, and using synthetic district boundaries to account for sub-district boundary changes over this period. *Infrastructure investment p.c.* is the square root of the average stock of loans outstanding over the decade divided by average district population. *Tax base p.c.* is the square root of the average per capita rateable value of property. Control variables include the population density, district population, the percentage of population aged over 65 and the level of tax base—see the Appendix for full results for these variables.
Only Registration Districts with both rural and urban subdistricts are included in specifications (2)–(5). The dependent variable in specifications (1)–(3) is the waterborne mortality rate in the urban subdistricts of each registration district—those subdistricts containing part of a town in 1881. The dependent variable in specifications (4)–(5) is the mortality rate in the rural subdistricts of each registration district—those not containing part of a town in 1881. Waterborne mortality in these specifications refers to mortality from cholera, diarrhea and fever.
Standard errors are clustered by Registration District, and are displayed in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Online appendix - not intended for publication

A Data

A.1 Data sources and variable construction

A.1.1 Mortality data

The most basic reporting unit was the Registration Subdistrict, of which there were approximately 2,000 in England and Wales. Each subdistrict then formed part of a larger Registration District—which in turn formed part of a Registration County and, finally, a Registration Division. There were approximately 600 Registration Districts, 50 Registration Counties and 9 Registration Divisions.

Mortality data reported at Registration District level are drawn from the decennial reports of the Registrar General for 1851-1891, which report the annual average number of deaths by cause and by age group split by Registration District. This information is obtained from Woods (1997).

Mortality data at Registration subdistrict level were collected from the Quarterly Returns of the Registrar General for the third quarter of each year between 1871 and 1890, with the exception of 1880 and 1882. This information was supplemented with the equivalent data for the years 1871, 1881 and 1891 which had been digitized previously by Southall (1998).

The Quarterly Reports during this period contain information on the total number of deaths, and deaths from nine causes: smallpox, measles, scarlet fever, cholera, diarrhea, violence, whooping cough, diptheria and fever. The latter category covered a range of maladies including typhoid (or enteric fever), simple continued fever and puerperal fever. The reports, however, do not detail a set of important causes of death during this period—

particularly airborne diseases such as tuberculosis. Nor do they disaggregate the cause of death by age group, precluding us from identifying the effect of sanitary intervention on specific age groups.

A.1.2 Financial data

Information regarding expenditures on infrastructure are drawn from the *Local Taxation Returns* reported to Parliament and collected in the *Parliamentary Papers* collection. Data is collected for all "urban sanitary authorities" for each year from 1867 to 1900. Prior to 1872 the accounts are reported under the titles of Local Boards of Health and Improvement Commissions— the bodies which were renamed Urban Sanitary Authorities in the 1872 Public Health Act. This includes approximately 900 towns, granted standardized expenditure powers under the terms of the 1872 Public Health Act.

A.1.3 Census data

Information on town and district population, number of houses and area was collected from decennial census reports. The collected data was supplemented by additional information for Registration Districts using the parish-level census information for 1871 and 1891 digitized by Southall et al. (2004). 1901 census population was downloaded from the Integrated Census Microdata project at the UK data archive (Schürer and Higgs, 2014).

A.2 Variable definition

A.2.1 Mortality rates

Mortality data for deaths in Registration Subdistricts, on the other hand, is reported by year (precisely, the third quarter in each year). The numerator of the measure is the average number of deaths in the subdistrict for years for which data is available.

The Decennial Registration District reports also detail average district population across the decade. Average populations are not available for the Registration Subdistricts, so I estimate an average population by assuming that subdistrict population grew at an compound average growth rate between decennial censuses and interpolating. The average of this interpolated population then serves as the denominator of the measure.

The Registrar General attempted to enforce a consistent nosology on registration officials around the country, and in general we can consider the categories of individual diseases as reasonably accurate. However, there are some exceptions to this general rule. The most major relates to typhoid which was not distinguished at all from typhus—a disease with similar symptoms but that is not waterborne—until 1869, and not in the decennial reports until 1871-1880. A second potential issue is that relatively substantial revisions were made to the nosology used in 1881. Fortunately for the purposes of this paper, most of the changes were to relatively minor categories or were later reported as separate categories allowing the original classification to be reconstructed (See Registrar General, 1895, Table H).

A.2.2 Financial variables

Financial measures, including loans outstanding and rateable value are calculated as decennial averages. The average annual total over the decade is estimated by summing over all years for which data is available and dividing by the number of years for which data is available. Amounts were then allocated to each registration district (as explained in the following section), and per capita variables were calculated using the average district population in the decade.

In most cases towns had data for loans outstanding available in all years, with the exception of the years before 1867. Rateable value on the other hand was missing for some years between 1866 and 1870, and for 1871. Missing values were linearly interpolated. In some cases, no data was available for the 1860s—in this case I use the first year in which data was available.

Data on interest rates (the adjusted yield on British consols) and prices (the Rousseax price index) are taken from Mitchell (1971).

A.2.3 Demographic variables

Demographic variables were defined in the following way:

- Average district population: directly taken from the Registrar General's decennial reports.

- Percentage of population in different age groups:calculated using information from the Registrar General's decennial reports.

- Average urban and rural population (used in the calculation of urban/rural mortality rates): Annual intercensal population was estimated using a 10 year geometric growth data, and then averaged across the 10 year period.

- Population density: Total district population divided by district area. Estimated using population data for the census year at the end of each decade: for instance, the population for 1861-1870 is estimated using the 1871 census. 1861 and 1881 census population using the reports of the previous census. Registration district area for 1861–1880 was taken from the Registrar General's Quarterly Report in 1871, and for 1881-1900 was taken from the Registrar General's Quarterly Report in 1891 (using the data digitized by Southall et al. (2004)).

A.3 Boundary changes and linking towns to Registration Districts

A.3.1 Linking towns to Registration Districts

Each town is linked to a Registration Subdistrict using information reported in the 1881 census, Vol II. This report splits the population of each town according to Registration Subdistrict. For example, of a total town population of 10,000 it identifies that 4,000 lived in subdistrict A, and 3,500 in subdistrict B, and 2,500 in subdistrict C. To aggregate expenditure data at the level of Registration District, expenditure is allocated to each Registration District allocate proportionally to the portion of the town population that falls in each district. That is, if 85% of the town live in district X, and 15% in district Y, then 85% of town expenditure is assigned to district X and 15% to district Y. Registration district level expenditure is then calculated through summing the spending amounts for all towns (whole or part) within the district.

A.3.2 Boundary changes

A further complication is that the boundaries of Registration Districts changed over time, with some added and others removed. To account for this, in analyses focusing on registration subdistricts I adjust all subdistrict mortality data to the 1881 boundaries by first identifying all subdistrict boundary changes (using the reports of the Registrar General) and then created a synthetic district based on population weight. That is, deaths in each year were reassigned to the 1881 district based on the population of the actual district reporting that lived in the 1881 district boundary in 1881. For instance, if two equally-sized districts merged in 1885, mortality data from the new district after this point would be split evenly between the two synthetic districts.

A.4 Descriptive statistics

Table A.4 presents summary statistics of the main variables used in the paper. Figure A.III displays the density of the loans outstanding per capita variable before and after the square root transformation. Figure A.IV displays the trends in mortality from the causes of death analyzed in the specification in Table 3.

Table A.V: Descriptive statistics of main variables included in regressions

Variable	Obs	Mean	Std. Dev.	Min	Max
Total average loans outstanding per capita (£p.c.)	1520	1.68	2.92	0	34.81
Square root of average loans outstanding per capita	1520	.99	.84	0	5.9
Tax base per capita (£p.c.)	1520	1.75	1.36	0	8.33
Square root of tax base per capita (£p.c.)	1520	1.22	.5	.03	2.89
Deaths from all causes per 100,000 popn; all ages	1520	18882.11	2671.77	12315.97	32126.56
Deaths from waterborne diseases per 100,000 popn; all ages	1140	771.68	436.28	84.16	3006.49
Deaths from airborne diseases per 100,000 popn; all ages	1520	6143.7	1450.91	2970.19	12873.54
Deaths from violence per 100,000 popn; all ages	1520	639.44	178.73	320.07	1815.36
Deaths from violence per 100,000 popn; all ages	1520	152.61	45.87	14.86	330.17
Deaths from all causes per 100,000 popn; age under 5	1520	50347.47	14500.67	24761.9	110954.2
Deaths from all causes per 100,000 popn; age 5-19	1520	4502.45	1307.65	2006.05	8910.69
Deaths from all causes per 100,000 popn; age over 20	1520	20010.09	2306.69	12545.55	28018.31
avgRDpopnAllAgesTemp	1520	4.86	6.32	.58	76.78
Population density (population per acre)	1520	2.5	6.96	.06	69.18
pctRDpopnAgeUnder5	1520	.13	.01	.09	.16
pctRDpopnAgeOver65	1520	.06	.02	.02	.1
% of loans outstanding in Water 1884-1890	369	.24	.28	0	1
% of loans outstanding in Sewers 1884-1890	369	.25	.27	0	1
% of loans outstanding in Street 1884-1890	369	.14	.19	0	1
% of loans outstanding in Gas 1884-1890	369	.08	.17	0	.85
% of loans outstanding in other category 1884-1890	369	.28	.28	0	1
% of loans outstanding in Water 1891-1900	374	.28	.27	0	1
% of loans outstanding in Sewers 1891-1900	374	.25	.24	0	1
% of loans outstanding in Street 1891-1900	374	.12	.14	0	1
% of loans outstanding in Gas 1891-1900	374	.09	.17	0	.85
% of loans outstanding in other category 1891-1900	374	.27	.24	0	1

Figure A.III: Square root transformation reduces positive skewness in average outstanding loans per capita

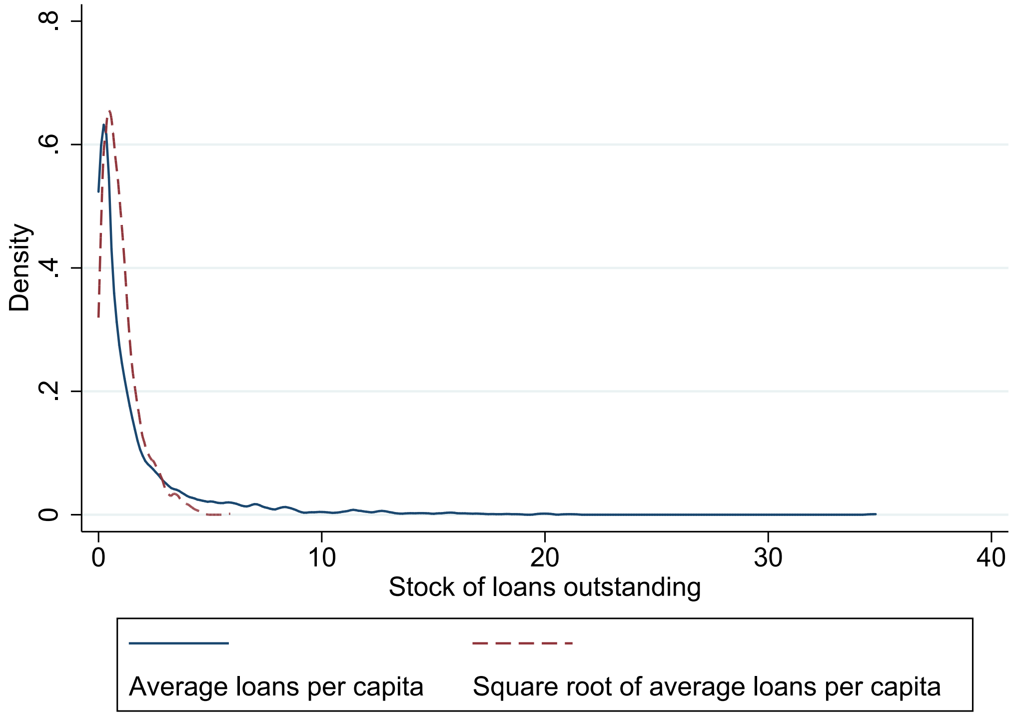
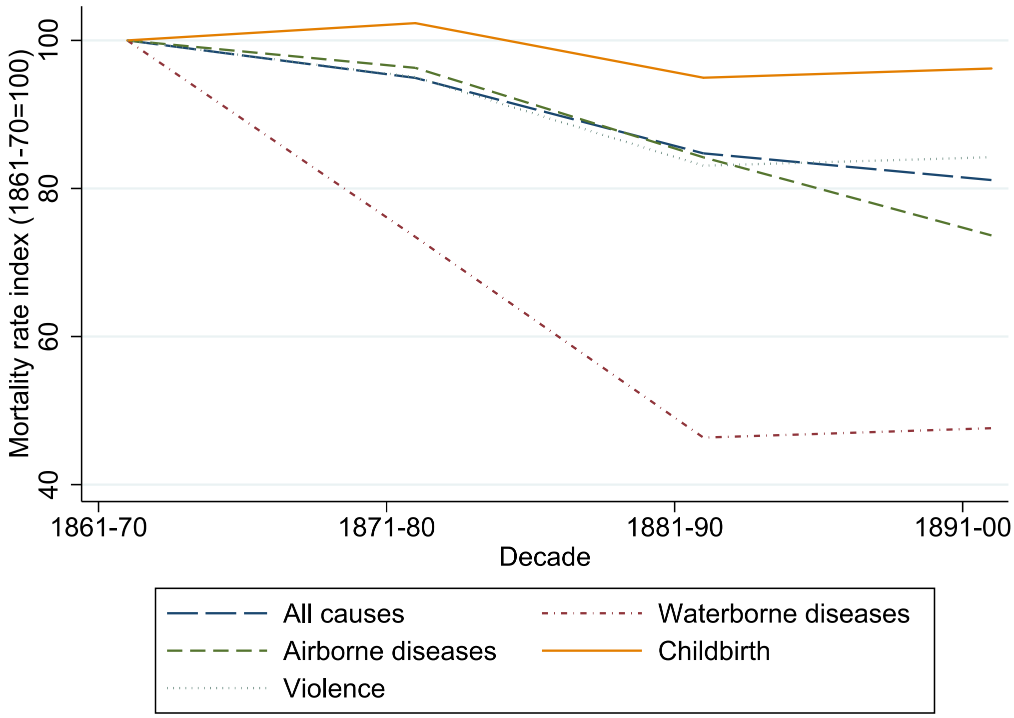


Figure displays the kernel density of the decadal average outstanding loans per capita for each registration district and decade included in the main regressions presented in the text.

Figure A.IV: Trends in mortality from different causes 1861 to 1900



Source: Author's calculations using data from Decennial Reports of the Registrar General. Estimates based on approximately 400 registration districts containing an urban area in 1881. See text for further details.

B Additional specifications and extended versions of tables in main paper

This section presents some additional results indicating the robustness of the results in the main paper, and displays the full regression results for the abbreviated specifications presented in the paper.

Table A.VI presents results of additional fixed effects estimates, including various subsets of control variables—including both those in the main paper and some additional variables. Neither the statistical significance nor the size of the effect varies considerably across specifications.

Table A.VII and Table A.VIII present the full results of the main instrumental variable regressions and the analysis by cause of death in the main paper.

Table A.IX presents specifications estimating mortality from different definitions of waterborne disease which are available for the entire period 1861–1900 (unlike the category used in the paper). First, specifications are estimated using only mortality from cholera and diarrhea as the dependent variable. A second set of specifications then estimate a "broad" mortality category—this includes the three causes of death listed in the main paper (cholera, diarrhea and typhoid) but also typhus and continued fever—which were combined with typhoid until 1869 and 1873 respectively—and other causes of death which Szreter (2005) suggests may have been affected by waterborne diseases: diseases of the nervous system and non-pulmonary tuberculosis. The results show that the estimated effects are still high (albeit lower) when the broad category is analyzed. The effect sizes when analyzing mortality from just cholera and diarrhea, on the other hand, are higher than when typhoid is also included. This may reflect the ongoing uncertainty in diagnosing typhoid.[18]

[18]There were considerable reporting inaccuracies even after deaths from typhoid were distinguished from typhus in the annual reports of the Registrar General after 1869, with typhoid often incorrectly diagnosed as either typhus or continued fever. For discussion of these problems, see *The Lancet*, September 21 1878

Table A.X presents the analysis by cause of death limited to the period 1871–1890. As in the estimates for total mortality in Table A.VII, the estimated effects for both waterborne and airborne diseases are much higher when analyzing this period alone. However, there remains no statistically significant effect for mortality from either childbirth or violence.

Finally, Table A.XI presents analysis of third quarter total mortality and Table A.XII does the same for waterborne mortality (including full results for the specifications in Table 4 in the paper.)

and *Supplement to the Fifty-Fifth Report of the Annual Report of the Registrar-General*, p.xxvii.

Table A.VI: Estimated effect on total mortality in fixed effects regressions robust to inclusion of different sets of control variables

	DV=Mortality rate at all ages 1861–1900							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Infrastructure investment p.c.	-0.162***	-0.161***	-0.142***	-0.167***	-0.153***	-0.129***	-0.135***	-0.126***
	(0.031)	(0.030)	(0.030)	(0.031)	(0.029)	(0.029)	(0.029)	(0.028)
Population	-0.204**	-0.193**	-0.146*	-0.207**	-0.146	0.034	0.036	0.118
	(0.080)	(0.085)	(0.086)	(0.081)	(0.120)	(0.101)	(0.101)	(0.092)
Population density		-0.025	-0.042					
		(0.075)	(0.072)					
Percent aged over 65			0.193***			0.339***	0.340***	0.307***
			(0.052)			(0.061)	(0.061)	(0.058)
Tax base p.c.				0.017			0.023	0.013
				(0.040)			(0.037)	(0.036)
District popn sq					0.008	-0.072	-0.075	-0.090
					(0.094)	(0.073)	(0.073)	(0.064)
Largest town pop					-0.096	-0.116	-0.121*	-0.167**
					(0.075)	(0.071)	(0.072)	(0.066)
District population growth					0.010	-0.001	-0.001	-0.009
					(0.013)	(0.012)	(0.012)	(0.013)
Percent aged under 5						0.226***	0.226***	0.179***
						(0.049)	(0.049)	(0.048)
Percent aged 20 to 64						0.250***	0.249***	0.202***
						(0.043)	(0.043)	(0.040)
Mortality from violence								0.135***
								(0.020)
Mortality from childbirth								0.094***
								(0.014)
Reg Dist FE	Y	Y	Y	Y	Y	Y	Y	Y
Decade FE	Y	Y	Y	Y	Y	Y	Y	Y
Period	1861-1900	1861-1900	1861-1900	1861-1900	1861-1900	1861-1900	1861-1900	1861-1900
% decline explained	0.27	0.27	0.24	0.28	0.25	0.22	0.23	0.21
Obs.	1520	1520	1520	1520	1520	1520	1520	1520
No. Districts	380	380	380	380	380	380	380	380

Table illustrates the robustness of the results of the fixed effects specifications presented in Table 1 to inclusion of different sets of control variables. "District popn sq" is the square of the population variable. Largest town population is the population of the largest town in the registration district, identified according to 1881 population. Mortality from violence and childbirth are the mortality rates at all ages from the respective causes. All variables are standardized. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.VII: Full results of instrumental variables regressions (extension of Table 2).

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Two stage least squares for mortality at all ages						
Infrastructure investment p.c.	-0.174***	-0.324***	-0.152***	-0.361***	-0.135**	-0.519***
	(0.030)	(0.047)	(0.037)	(0.068)	(0.059)	(0.140)
Population			-0.092	0.015	-0.213	0.096
			(0.103)	(0.078)	(0.195)	(0.175)
Population density			-0.065	-0.046	-0.104	-0.118
			(0.069)	(0.065)	(0.188)	(0.157)
Percent aged over 65			0.307***	0.262***	0.519***	0.389***
			(0.063)	(0.065)	(0.097)	(0.112)
Tax base p.c.			0.138**	0.270***	0.210***	0.410***
			(0.055)	(0.067)	(0.075)	(0.105)
Panel B: Abbreviated first stage regressions for infrastructure investment per capita						
Lag Infrastructure investment p.c.		0.557***		0.419***		0.399***
		(0.033)		(0.038)		(0.066)
Population				0.249**		0.527**
				(0.114)		(0.222)
Population density				0.078		0.033
				(0.066)		(0.273)
Percent aged over 65				-0.138**		-0.254**
				(0.062)		(0.101)
Tax base p.c.				0.444***		0.440***
				(0.049)		(0.068)
Controls	Y	Y	Y	Y	Y	Y
Reg Dist FE	Y	Y	Y	Y	Y	Y
Decade FE	Y	Y	Y	Y	Y	Y
Period	1871-1900	1871-1900	1871-1900	1871-1900	1871-1890	1871-1890
% decline explained	0.29	0.54	0.25	0.60	0.23	0.88
Hansen C statistic		15.57		10.51		9.03
p-value		0.00		0.00		0.00
Kleibergen-Papp Stat		290		124		37
Obs.	1140	1140	1140	1140	760	760
No. Districts	380	380	380	380	380	380

All coefficients are standardized. Observations are "Registration District–decades", between 1861-1870 and 1891-1900. See Table 2 for further details. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.VIII: Extended results for effects on mortality by cause 1871-1900 (following from Table 3)

	DV=Mortality rate at all ages by cause							
	Waterborne		Airborne		Childbirth		Violence	
	FE	IV	FE	IV	FE	IV	FE	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Infrastructure investment p.c.	-0.167***	-0.372***	-0.092**	-0.170**	-0.084	-0.047	-0.055	-0.160
	(0.045)	(0.087)	(0.038)	(0.077)	(0.069)	(0.125)	(0.070)	(0.113)
Population	0.101	0.206**	-0.113	-0.073	0.100	0.082	-0.361**	-0.307*
	(0.106)	(0.096)	(0.096)	(0.094)	(0.101)	(0.109)	(0.160)	(0.158)
Population density	-0.147	-0.128	-0.082	-0.075	-0.001	-0.005	0.285**	0.295***
	(0.122)	(0.117)	(0.061)	(0.062)	(0.069)	(0.069)	(0.113)	(0.110)
Percent aged over 65	-0.173*	-0.218**	0.097*	0.080	-0.151	-0.143	0.395***	0.373***
	(0.090)	(0.092)	(0.055)	(0.055)	(0.140)	(0.140)	(0.095)	(0.097)
Tax base p.c.	0.185**	0.314***	-0.017	0.033	-0.017	-0.040	0.099	0.165*
	(0.077)	(0.093)	(0.053)	(0.068)	(0.107)	(0.125)	(0.081)	(0.096)
Reg Dist FE	Y	Y	Y	Y	Y	Y	Y	Y
Decade FE	Y	Y	Y	Y	Y	Y	Y	Y
Period	1871-1900	1871-1900	1871-1900	1871-1900	1871-1900	1871-1900	1871-1900	1871-1900
Obs.	1140	1140	1140	1140	1140	1140	1140	1140
No. Districts	380	380	380	380	380	380	380	380

Observations are "Registration District–decades", between 1871-1880 and 1891-1900. See notes to Table 3 for further details.
Standard errors are clustered by Registration District, and are displayed in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.IX: Estimated results for effects on different forms of waterborne mortality.

	DV=Mortality from cholera and diarhea			DV=Broad waterborne category		
	FE	FE	IV	FE	FE	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Infrastructure investment p.c.	-0.123***	-0.152***	-0.304***	-0.158***	-0.161***	-0.361***
	(0.034)	(0.040)	(0.082)	(0.034)	(0.037)	(0.074)
Population	-0.080	0.111	0.189*	-0.199**	-0.093	0.009
	(0.104)	(0.111)	(0.109)	(0.088)	(0.086)	(0.073)
Population density	-0.244**	-0.177	-0.163	-0.034	-0.038	-0.019
	(0.095)	(0.139)	(0.136)	(0.065)	(0.078)	(0.074)
Percent aged over 65	-0.192***	-0.202**	-0.235***	0.110	0.134*	0.090
	(0.069)	(0.084)	(0.084)	(0.067)	(0.077)	(0.080)
Tax base p.c.	-0.059	0.043	0.139*	-0.068	0.142**	0.268***
	(0.051)	(0.068)	(0.083)	(0.047)	(0.058)	(0.071)
Reg Dist FE	Y	Y	Y	Y	Y	Y
Decade FE	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y
Period	1861-1900	1871-1900	1871-1900	1861-1900	1871-1900	1871-1900
% 1871-1900 decline explained	0.54	0.68	1.35	0.37	0.38	0.85
% 1861-1900 decline explained	0.33	0.41	0.81	0.25	0.25	0.57
Hansen C statistic			4.57			8.95
p-value			0.03			0.00
Kleibergen-Papp Stat			124			124
Obs.	1520	1140	1140	1520	1140	1140
No. Districts	380	380	380	380	380	380

Observations are "Registration District–decades", between 1871-1880 and 1881-1900. See Table 3 for further details of control variables and specifications.

Standard errors are clustered by Registration District, and are displayed in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.X: Estimated results for effects on mortality by cause deaths for the period 1871–1890.

	DV=Mortality rate at all ages by cause							
	Waterborne		Airborne		Childbirth		Violence	
	FE	IV	FE	IV	FE	IV	FE	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Infrastructure investment p.c.	-0.210***	-0.702***	-0.105*	-0.284**	-0.053	-0.384	-0.027	-0.011
	(0.074)	(0.164)	(0.056)	(0.139)	(0.108)	(0.276)	(0.091)	(0.197)
Population	-0.225	0.171	-0.019	0.125	0.037	0.304	-0.785***	-0.798***
	(0.264)	(0.224)	(0.173)	(0.182)	(0.199)	(0.326)	(0.250)	(0.306)
Population density	-0.379	-0.398**	-0.328**	-0.335**	0.161	0.148	0.550*	0.551*
	(0.244)	(0.200)	(0.156)	(0.146)	(0.234)	(0.297)	(0.314)	(0.312)
Percent aged over 65	0.176	0.010	0.218**	0.157	0.256	0.145	0.492***	0.498***
	(0.109)	(0.129)	(0.095)	(0.107)	(0.231)	(0.247)	(0.151)	(0.164)
Tax base p.c.	0.192**	0.447***	0.022	0.115	0.035	0.206	0.157	0.149
	(0.082)	(0.121)	(0.075)	(0.102)	(0.157)	(0.197)	(0.119)	(0.149)
Reg Dist FE	Y	Y	Y	Y	Y	Y	Y	Y
Decade FE	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y
Period	1871-1890	1871-1890	1871-1890	1871-1890	1871-1890	1871-1890	1871-1890	1871-1890
% decline explained	0.46	1.55	0.25	0.68	0.59	4.28	0.07	0.03
Hansen C statistic		9.76		2.06		1.79		0.01
p-value		0.00		0.15		0.18		0.93
Kleibergen-Papp Stat		37		37		37		37
Obs.	760	760	760	760	760	760	760	760
No. Districts	380	380	380	380	380	380	380	380

Observations are "Registration District–decades", between 1871-1880 and 1881-1890. See Table 3 for further details of control variables and specifications.

Standard errors are clustered by Registration District, and are displayed in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.XI: Extended results from regressions for comparison of total urban and rural mortality in the third quarter.

	All districts		Districts with no rural portions		Districts with rural portions				
	DV=urban mortality		DV=urban mortality		DV=urban mortality			DV=rural mortality	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Infrastructure investment p.c.	-0.232***	-0.241***	-0.245*	-0.358**	-0.219***	-0.144	-0.146	-0.164	0.035
	(0.074)	(0.090)	(0.146)	(0.158)	(0.084)	(0.106)	(0.106)	(0.140)	(0.141)
Population		-0.149		0.062		-1.704***	-1.548**		-3.479***
		(0.264)		(0.271)		(0.628)	(0.608)		(0.904)
Urban population density		-0.085		-0.168		4.069**	3.739**		7.386**
		(0.193)		(0.186)		(1.685)	(1.702)		(2.884)
Percent aged over 65		0.230		0.328		0.142	0.137		0.119
		(0.153)		(0.300)		(0.171)	(0.174)		(0.266)
Tax base p.c.		0.280***		0.485***		0.130	0.126		0.082
		(0.100)		(0.139)		(0.137)	(0.138)		(0.169)
Rural total morality							0.045		
							(0.053)		
Reg Dist FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Decade FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Period	1871-1890	1871-1890	1871-1890	1871-1890	1871-1890	1871-1890	1871-1890	1871-1890	1871-1890
Obs.	757	757	206	206	550	550	550	550	550
No. Districts	380	380	104	104	275	275	275	275	275

Extended results from specifications presented in Table 4, but with dependent variable of total (rather than waterborne) mortality. See notes to that table for further details.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.XII: Extended results from regressions for comparison of urban and rural waterborne mortality in the third quarter 1871–1890.

	All districts		Districts with no rural portions		Districts with rural portions				
	DV=urban mortality		DV=urban mortality		DV=urban mortality			DV=rural mortality	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Infrastructure investment p.c.	-0.323***	-0.287***	-0.299***	-0.325**	-0.314***	-0.214**	-0.209**	-0.025	-0.032
	(0.063)	(0.078)	(0.110)	(0.127)	(0.081)	(0.088)	(0.088)	(0.108)	(0.104)
Population		-0.202		0.003		-0.783	-0.650		-0.974
		(0.285)		(0.306)		(0.706)	(0.641)		(1.178)
Urban population density		-0.201		-0.218		0.067	-0.122		1.386
		(0.138)		(0.137)		(2.382)	(2.326)		(2.160)
Percent aged over 65		0.009		-0.013		-0.055	-0.019		-0.261
		(0.129)		(0.254)		(0.151)	(0.151)		(0.225)
Tax base p.c.		0.045		0.127		0.004	-0.037		0.299*
		(0.101)		(0.196)		(0.102)	(0.103)		(0.158)
Rural Waterborne morality							0.136***		
							(0.039)		
Reg Dist FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Decade FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Period	1871-1890	1871-1890	1871-1890	1871-1890	1871-1890	1871-1890	1871-1890	1871-1890	1871-1890
Obs.	757	757	206	206	550	550	550	550	550
No. Districts	380	380	104	104	275	275	275	275	275

Extended results from specifications presented in Table 4: see notes to that table for further details.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.