



UNIVERSITY OF MINNESOTA

## **Did the Urban Mortality Penalty Disappear? Revisiting the Early Twentieth Century's Urban-Rural Mortality Convergence**

James Feigenbaum  
Boston University

Lauren Hoehn-Velasco  
Georgia State University

Elizabeth Wrigley-Field<sup>†</sup>  
University of Minnesota

December 2020

Working Paper No. 2020-09

DOI: <https://doi.org/10.18128/MPC2020-09>

<sup>†</sup>Address correspondence to Elizabeth Wrigley-Field, University of Minnesota, Department of Sociology, 1156 Social Science Building, 267 19<sup>th</sup> Ave S., Minneapolis, MN 55455 (email: [ewf@umn.edu](mailto:ewf@umn.edu)). Support for this work was provided by the Minnesota Population Center at the University of Minnesota (P2C HD041023) and the Fesler- Lampert Chair in Aging Studies at the University of Minnesota.

# Revising the canonical view of urban/rural mortality reversal in the United States

James Feigenbaum<sup>1</sup>, Lauren Hoehn-Velasco<sup>2</sup>, and Elizabeth Wrigley-Field<sup>\*3</sup>

<sup>1</sup>*Boston University and NBER*

<sup>2</sup>*Georgia State University*

<sup>3</sup>*University of Minnesota*

May 2021

*Preliminary Draft*

This paper reexamines the predominant narrative suggesting that urban areas became cleaner, safer, and healthier during the early twentieth century, eliminating the “urban penalty” in mortality by 1920. While canonical narrative holds for crude mortality in select states, we show that in the national panel of counties and cities, or for age-standardized mortality, the urban penalty persists past 1936. We attribute these divergent results to the differences in population composition by DRA entry and for urban versus rural areas. We then consider why the urban penalty failed to disappear and document two intriguing patterns in respiratory and waterborne-infectious mortality. First, both large and small cities had persistently high waterborne/gastroenteric mortality. Second, large cities struggled with airborne/respiratory deaths, while these respiratory illnesses killed fewer in small cities and rural areas. An interesting caveat to these findings is that tuberculosis mortality quickly converged between urban and rural areas after 1918, suggesting that the 1918 influenza pandemic may have played a role in narrowing the urban penalty.

**JEL codes:** I15, J11, N32, O18.

**Keywords:** Mortality rates, urban mortality penalty, rural health, demographic transition, population health.

---

\*Authors are listed alphabetically to reflect equal authorship.

We thank Carl Kitchens and seminar participants from Florida State University for their feedback on this project.

# 1 Introduction

The predominant narrative in the historical demography and economic history literature maintains that from 1800 to the early 20th century, urban areas became cleaner, safer, and healthier than their rural counterparts. The central studies documenting this transition suggest that urban-rural mortality had nearly converged by 1900 (Condran and Crimmins, 1980) and the urban penalty had mostly disappeared by the 1920s (Haines, 2001).<sup>1</sup> Much of the dramatic improvement in urban living conditions has been attributed to public health infrastructure investments in cities (Troesken, 1999, 2001; Haines, 2001; Cutler and Miller, 2005; Beach et al., 2016; Alsan and Goldin, 2019).

In this study, we reconsider the urban-rural convergence using newly digitized, geographically disaggregated measures of mortality. Our study is the first to document the urban penalty using place-specific city-level and county-level mortality data for all areas in the Death Registration Area (DRA). Using this novel data, we provide evidence suggesting that the urban penalty persists later into the 20th century than previously recognized.

In the full (national) sample of counties and cities, the urban penalty not only fails to disappear by 1936 but *increases* for overall crude mortality. In 1900, the crude mortality penalty for urban areas was 1.06 and grew to 1.2 by 1936.<sup>2</sup> This ratio suggests that urban mortality was 6% higher than rural areas in 1900, but 28% higher in 1936 (at the median). The growth in the urban penalty mainly occurs over 1910 to 1936, with, a small and inconsistent penalty over 1900-1910. The persistence (or growth) of the urban penalty represents a divergence from the traditional narrative of a disappearing urban mortality penalty (Haines, 2001; Condran and Crimmins, 1980).

Why do our findings diverge from the traditional narrative? As we show, the canonical story is true for crude mortality in the select states that were already in the Death Registration Area. Our conclusions are similar to Haines (2001) when we subset to the balanced panel of ten states in the DRA by 1900; indeed, urban mortality shrinks so much and so fast that these ten states show a *rural mortality penalty* beginning by 1910. However, when mortality is age-standardized, in any sample, or when the sample is broadened beyond the first ten DRA states, the urban penalty persists at least until 1936 (when our essential data series ends). We attribute the importance of the age-standardized findings to two features of the sample. First, the differences are due to the large shares of (low mortality) young adults in cities and (high

---

<sup>1</sup> Haines (2001) finds that the urban penalty disappeared by 1920 for whites and the 1930s for the remainder of the population

<sup>2</sup> We quantify the penalty as a ratio of urban to rural mortality for the median rural area versus the median urban area. We measure the size of the urban penalty for each year as  $\frac{\text{Median Urban Mortality}}{\text{Median Rural Mortality}}$ . A ratio over one will indicate a higher urban penalty, with higher ratios indicating higher penalties.

mortality) children in rural areas. Second, the changing composition of rural and urban areas entering the sample of DRA states. Early-entry rural areas had higher crude mortality rates while later-entry rural areas lower overall mortality. Conversely, mortality rates are higher in later-entry urban areas.

Our results for total mortality are replicated in infectious causes of death, as well as a specific focus on respiratory-infectious and waterborne-infectious mortality through 1915. 1915 is the last year until the 1940s where cause-specific data is available in rural counties. Our primary grouping of infectious illnesses, respiratory and waterborne mortality, collectively account for about 40% of deaths.<sup>3</sup> However, we find notably different relationships to city size in respiratory and waterborne causes of death, reflecting prior work in [Cain and Hong \(2009\)](#). Both large and small cities had high waterborne-infectious mortality. By contrast, respiratory-causes of death were persistently higher in large cities than in smaller cities and rural areas.

Finally, we note a convergence in respiratory mortality between large and small cities after 1918. We find suggestive evidence that the 1918 influenza sped urban-rural mortality convergence, at least in respiratory causes, through selective mortality ([Noymer, 2009, 2011](#)). This post-1918 pattern is apparent in a few ways. First, the convergence between large and small cities post-1918. Second, we show that the urban penalty is substantially lower in the Northeast after 1918. The Northeastern cities had the highest tuberculosis mortality pre-1918. Third, we formalize this analysis using a difference-in-differences strategy comparing large and small cities pre and post-1918. Using this methodology, we find similar convergence. Fourth, we show that national urban-rural tuberculosis mortality substantially converged after 1918. All findings align with the selective mortality story ([Noymer, 2009, 2011](#)). Though selective mortality is the most plausible explanation, we acknowledge that other explanations include regulatory changes or behavioral adjustments in the wake of the pandemic, but we leave further examination into these factors for future work.

Overall, our findings suggest several revisions to the previous narrative are warranted. First, the urban-rural mortality penalty failed to disappear by 1936 for place-specific mortality. In 1940, the mortality penalty was still clear and present for urban areas in aggregate mortality. Second, our results raise questions about the predominant public health narrative. Cities have persistently higher waterborne causes of death (at least through 1915), and those large cities continued to struggle with respiratory illnesses relative to towns and rural areas. Together, these findings suggest that public health infrastructure, in the form of water and sewage treatments, may not account for a substantial convergence in urban-rural mortality.

---

<sup>3</sup>The remainder are a mixture of known non-infectious causes (e.g., cancer, accidents), ambiguous causes ("diseases of the heart"), and unknown causes (e.g., "all other causes"). Among those attributable to airborne or waterborne sources, roughly 4 in 5 were airborne.

Our findings align with [Anderson et al. \(2019a\)](#), which revises the stylized narrative that major infrastructure advancements were the predominant driver of urban mortality declines. Third, our results do suggest that respiratory causes of death in large cities converged to levels of small cities after 1918, aligning with the selective mortality documented in [Noymer \(2009, 2011\)](#).

In addition to the works specifically examining the urban mortality penalty, ([Higgs, 1973](#); [Condran and Crimmins, 1980](#); [Haines, 2001](#); [Cain and Hong, 2009](#)), this study adds to a broad literature considering the mortality declines of the early twentieth century. Much of that literature has demonstrated the importance of infrastructure investments for typhoid-related mortality and mortality in infancy ([Troesken, 1999, 2001](#); [Haines, 2001](#); [Cutler and Miller, 2005](#); [Beach et al., 2016](#); [Alsan and Goldin, 2019](#); [Anderson et al., 2019a](#)). Another portion of the literature have documented the importance of particular public health campaigns ([Olmstead and Rhode, 2004](#); [Bleakley, 2010](#); [Kitchens, 2013b](#); [Moehling and Thomasson, 2014](#); [Komisarow, 2017](#); [Hoehn-Velasco, 2018](#)) as well as public health regulations through occupational licensing ([Anderson et al., 2016](#)). A caveat to this literature is that public health campaigns for illnesses such as tuberculosis appear less effective than public health efforts targeting infant or typhoid mortality ([Clay et al., 2020](#); [Anderson et al., 2019b](#)).

## 2 Data

In this section, we present our data sources and document differences across states and counties that correlate with their entrance into the DRA.

### 2.1 Data Sources

For this project, we combine digitized city-level and rural county-level mortality from annual volumes of the *US Vital Statistics (1890-1938)* with population data from the complete count of the Federal Census.

The vital statistics report the number of deaths and the population totals for urban and rural areas over an unbalanced panel of states, counties, and cities.<sup>4</sup> For our primary analysis, we rely on overall mortality for rural and urban areas in the DRA over 1900-1936. In addition to the overall mortality rates, we also use by-cause and by-age counts for specific years. Over the period 1900-1915, the vital statistics report mortality by cause. By-age deaths are available

---

<sup>4</sup>The data used in this project were digitized for previous work and described in these studies ([Hoehn-Velasco, 2018](#); [Feigenbaum et al., 2019](#)).

from 1900 to 1910. After 1915, only state-level estimates for rural areas are available and are not accessible by county.

We combine these death counts with population data from the Full Count Census, where we calculate rural and urban population totals and the population's age structure. We compute death rates for all cities and rural counties in the DRA by combining the full count populations and the deaths counts from the vital statistics. Overall, our data series includes (1) total death rates over 1900-1936, (2) by-cause death rates over 1900-1915 for individual cities and counties, (3) by-cause death rates over 1916-1944 for rural and urban areas at the state level, (4) by-age death rates over 1900-1910, (5) the population totals and age structure from the complete-count Censuses (Ruggles et al., 2021).

To account for compositional differences, we indirectly age-standardize the city-level and county-level mortality rates. Mortality is highest at the youngest and oldest ages in the early twentieth century. Thus, differences in crude mortality attributed to urban or rural status (or across states or census regions) may, in fact, be driven by variation in age structure. In our data, age-specific deaths are reported only from 1900 to 1910. This prevents us from directly age-standardizing mortality rates. However, indirectly age-standardized mortality rates allow us to compare geographies with different age structures.

The coverage of the mortality data series is limited, as states enter the Death Registration Area (DRA) from 1900 to 1933; U.S. states were not required to report vital statistics until 1933 (Haines, 2001). Rural areas were especially late to publish vital statistics, and some states did not report these statistics for rural areas until the 1930s. The available mortality statistics are consequently an unbalanced panel of counties and cities. In 1900 there were ten states and 319 counties that provided mortality statistics. By 1910, this number had grown to 20 states and 832 counties. By the end of the period, 1933-1936, the entire country had entered into the data with 2,450 counties in total. Our focus on the DRA is a limitation of this study because areas inside and outside the growing DRA may have differed in their mortality (as, indeed, our analyses suggest they did). In some cases, such as child mortality in 1900, the DRA would be expected to have higher mortality than non-DRA areas (Haines, 2001). However, in other instances, such as urban mortality in the South, non-DRA areas likely experienced a higher death rate than DRA areas (Feigenbaum et al., 2019).

## 2.2 Compositional Differences Across Entry into the Death Registration Area

The DRA expanded from just 10 states in 1900 to cover the entire country by 1933 but states and counties did not enter the DRA (and thus our sample) at random. In Table 1 we display the demographic (Panel A) and mortality (Panel B) conditions by DRA entry year.

We group entry years into the original 1900 states, entry between 1901-1910, 1911-1920, and 1921-1933.

Rural areas vary substantially in demographic characteristics by when they entered the DRA. For the earliest-entry DRA rural areas, population growth is declining, 99% are white, 10% are over 60, and 41% are under age 20. For the later-entry rural areas, only 5% are over 60, but over 50% are under age 20. These later-entry counties are also experiencing substantial population growth, and 19% of the residents are non-white. Much of the high population growth can be attributed to entry by western states, especially Texas. The large share of the population under 20 can be attributed to the Midwest and South. We show the regional urban and rural compositions in Table [A.1](#)

Urban areas are more homogeneous across entry years than rural areas, except for the non-white share. The 1900 population shares by age are relatively stable across entry years, and population growth is around 4% annually. The population under 20 is stable between 37-40%. The share white is the most noticeable change from year to year, with 98% of urban areas being white in early-entry states and only 79% of urban areas being white for later-entry states.

When it comes to mortality, rural and urban areas had opposite DRA entry patterns, as we show in Table [1](#) Panel B. For urban areas, the lowest mortality cities are in the DRA at the start of the twentieth century, and the highest mortality cities enter between 1921 and 1940. For rural areas, the pattern is reversed. Low mortality areas gradually enter the DRA, with mortality declining over each decade of entry. These differential mortality conditions between areas intuitively makes sense from what we know about the later-entry states. After 1910, many of the states that enrolled in the DRA were western and southern states. Many of the western rural counties were low-density and had lower crude mortality rates. The southern cities, by contrast, had some of the highest urban mortality in the nation ([Feigenbaum et al., 2019](#)). The mortality figures in the table reflect the conclusions of both [Haines \(2001\)](#); [Feigenbaum et al. \(2019\)](#).

We also show the age-standardized mortality rates to address the age composition differences illustrated in Panel A. The age adjustment accounts for the difference between mortality in newly-entering vs. longstanding rural areas, but not urban areas. The age-standardized rates suggest that the urban areas that enter the panel during the twentieth century have higher mortality that is not fully accounted for by age differences. On the other hand, after accounting for the population's age structure, rural areas have similar mortality conditions, no matter when they entered.

### 3 Did the Urban Penalty Disappear?

In this section, we trace out trends in the urban penalty for total mortality. We start with the full (unbalanced) sample of all counties and cities in the DRA in Section 3.1, focusing on crude mortality rates. Though we estimate a minimal urban penalty in 1900, we document a growing and persistent urban mortality penalty from 1910-1936, results that do not align with Haines (2001); Condran and Crimmins (1980). In Section 3.2, we turn a balanced panel and reconcile our results with Haines (2001) for the select set of states already in the DRA as of 1900. Results suggest that the canonical view of the disappearing urban penalty is heavily affected by examining the small set of states that entered the DRA earliest. Then, in Section 3.3, we replace the crude mortality rate with the age-standardized mortality rate. Properly accounting for age distribution differences across areas, the urban penalty is substantial and persistent for the duration of our period.

Throughout this section, we refer to results presented in Figure I, Figure II, and Figure III. Figure I shows 25th, 50th, and 75th percentiles of urban (blue) and rural (red) mortality from 1900 to 1936. (We also show the full distribution of county and city-level mortality in the Appendix as Figure B.1, which includes individual annual histograms.)

Figure II and Figure III present alternative summaries of the urban penalty. We calculate the ratio of urban mortality to rural mortality at time  $t$  as the ratio of median mortality values across geographic units:

$$\text{Urban Penalty}_t = \frac{\text{Urban Mortality}_t}{\text{Rural Mortality}_t} \quad (1)$$

A ratio over one will indicate an urban penalty, with a higher number suggesting worse relative conditions in urban areas. A number below one will indicate a *rural mortality penalty*.

Figure II summarizes the median urban penalty for key years across a larger set of samples, capturing differential selection into the DRA over time, and Figure III shows the median urban penalty for the main samples over all years.

#### 3.1 The Urban Mortality Penalty is Negligible in 1900, but Re-emerges in 1910 in the Unbalanced Panel

To begin, we explore the urban mortality panel in our full dataset, including all counties and cities for which we have data from 1900 to 1936, in Panel A of Figure I. The plotted distribution shows a narrow mortality penalty in 1900 for cities. Then between 1904 and



1909, any urban-rural mortality penalty that exists at the median is negligible. By 1906 rural areas have a higher variance than urban areas, but the medians are comparable. In 1910, the distribution of urban mortality shifts upward such that the median urban mortality is higher than in rural areas; indeed, in 1910, the 25th percentile of urban areas is comparable to the 50th percentile in rural areas. As emphasized in Figure II and Figure III, this stark urban mortality penalty continues through the 1920s, and—contrary to the conventional view of its disappearance—the urban-rural gap appears higher in 1936 than it did in 1900.

As Figure II shows, for the full DRA (in purple), there is a small penalty in 1900 (1.06), which grows over time and reaches 1.20 in 1936. The urban mortality penalty not only persists but grows over our series. We plot the full urban mortality penalty over time in Figure III. The dashed dark purple line shows the crude mortality penalty steadily growing over time. By 1936, the urban penalty in crude mortality has reached its highest point, in stark contrast with the canonical narrative.

### 3.2 The Urban Mortality Penalty Disappears in 1910 in the Balanced Panel

One simple explanation for why our findings diverge from the canonical narrative is simply the sample. As shown in the data section, the national panel has shrinking rural mortality by entry and growing urban mortality. This non-random entry of cities and counties makes the urban penalty dependent on the sample.

To illustrate the importance of the sample, Panel B of Figure I shows the balanced panel of the 10 DRA states. Here the mortality trend is the opposite of the full panel of counties and cities (the first graph of Figure I). Beginning in 1908, there appears to be a *rural mortality penalty*, and this rural mortality penalty grows over time. The appearance of a rural mortality penalty in the balanced panel starkly contrasts with the reappearance of the urban mortality penalty in the national sample. A similar picture is shown in the top set of histograms in Appendix Figure B.2, where the full distribution of mortality for rural areas gradually increases past urban areas over time. Figure III shows that, in the 1900-balanced panel, the urban penalty disappeared and became a persistent *rural penalty* as early as 1908.

The clear convergence in the balanced panel is illustrated clearly in point estimates in Figure II. The balanced panel, as of 1900, shows a negligible mortality penalty in 1900 of 1.06. By 1910 the urban penalty disappears and turns into a rural penalty (0.93). The rural penalty persists until 1936 (0.94). This result aligns with the Haines (2001) findings and suggests that for the original DRA states, urban areas may have been quicker to improve their urban conditions relative to rural areas and the remainder of the country. These findings only hold for the 1900 panel. If we consider the 1910 or the 1920 panel of states, there is a clear urban

penalty which persists until 1936.

Overall, our findings in the balanced 1900 panel indicate that the reappearance of the urban mortality penalty in the full DRA could be due to the mortality conditions in entering cities and counties. Later-entry rural areas have lower mortality, while later-entry urban areas have high mortality (Table 1). In Figure II, we also consider alternative balanced panels, reflecting all and only cities and counties present in the data in 1920 and 1930, respectively. These alternative panels confirm that the canonical story of the disappearing urban penalty is distinctive to the geographic areas that were already in the DRA by 1900.

### **3.3 The Urban Penalty Persists—and Grows—when Accounting for the Age Structure of the Population**

In the previous two subsections, we found that the trends in the urban mortality penalty depend on the sample: the disappearing urban penalty is limited to the small set of areas present in the DRA in 1900. Here we show that, even in that sample, the urban penalty's disappearance is driven by differences in the age composition of urban and rural areas.

Panels C and D of Figure I show the distribution of mortality using indirectly age-standardized mortality rates. These adjusted distributions are quite distinct from those shown in Panels A and B. In Panel C the age-standardized rates suggest a clear urban mortality penalty in 1900 that persists at least through 1936. The age-standardized penalties are also notably larger than in non-age-standardized mortality. Further, there is no disappearance and re-emergence of the urban mortality penalty. Urban mortality appears consistently higher than rural mortality, but the gap does decline slightly over the period, especially over 1918. Overall, urban mortality appears to be suppressed by an over-representation of young adults, while an over-representation of children inflates rural mortality.

The difference between balanced and unbalanced panels also appears smaller in the age-adjusted results than in crude mortality. Panel D of Figure I shows the age-standardized rates over the balanced panel. While the balanced panel does show a smaller mortality penalty than the unbalanced panel, the age correction largely addresses the concerns with the unbalanced panel. Accounting for differences in composition, mortality across rural and urban areas appear much more similar in the age-standardized balanced panel. The main difference between the age-standardized balanced and the unbalanced panels is before and after 1918. In the balanced panel, after 1918, the urban penalty completely disappears, whereas there is a clear urban penalty in the unbalanced panel until the end of the series.

Our results using age-adjusted mortality rates suggest that the canonical story of a disappearing urban penalty reflects the different age compositions of rural versus urban areas,

along with the distinctive trends in the areas with the earliest data. Illustrating this fact in a point estimate, Figure II shows the urban penalty persists in all measures of age-adjusted mortality—even in the 1900 balanced panel, where it declines from 1.28 (in 1900) to 1.09 (in 1936). By contrast, in the full panel, the age-adjusted mortality penalty declines from 1.28 to 1.15. The quantified urban penalty highlights the different conclusions depending on our chosen samples and preferred mortality rates. However, our findings show that the urban penalty persisted longer than has been understood, when accounting for the compositional differences between areas.

## **4 Was there an Urban Penalty in Infectious, Waterborne, and Respiratory Mortality?**

Does the trend in the urban mortality penalty extend to specific causes of death? In this section, we explore infectious mortality and, within infectious, waterborne and airborne mortality. Here, we are limited somewhat by our data as we lack cause specific rural mortality rates for individual counties after 1915. Instead, we supplement the geographically detailed 1900-1915 data (with individual rural counties) with a separate series of cause-specific mortality for urban and rural areas summarized within each state, for 1916-1944.

### **4.1 The Urban Penalty Persists in Infectious Disease Mortality**

Beginning with overall infectious mortality, the results suggest no narrowing of the urban penalty between urban and rural areas from 1900 through 1915. Figure IV (and Appendix Figure B.3) displays the distribution of mortality rates from infectious causes over the available period (1900-1915). Figure I shows that for infectious disease deaths, the urban penalty persists through 1915. In fact, the urban mortality in infectious disease at the 50th percentile is closer to the 75th percentile in rural areas.

One conclusion from overall mortality that does not extend to infectious mortality is the importance of age-standardization. In Panel B, the age-adjusted infectious mortality rates show similar findings to the crude infectious rates in Panel A. The compositional differences between rural areas and cities do not affect the interpretation of the urban penalty from the infectious mortality rates. This is a surprising fact, as there was a clear age gradient to infectious mortality in general mortality (with young children at the greatest risk).

Overall, these results suggest that the urban mortality penalty in crude mortality and infectious disease mortality continues past 1915. Unfortunately, due to the limited cause-

specific data in rural areas, we cannot track infectious disease mortality after 1915. Until 1915, the infectious-disease mortality tells a similar story as all-cause mortality, where the urban penalty clearly persists.

## 4.2 The Urban Penalty Persists in Waterborne Mortality, Despite Public Health Investment

To understand why the urban mortality penalty did not fade, we then consider disaggregated illness-specific causes of deaths in digestive mortality (largely waterborne) and respiratory mortality (largely airborne). These categories broadly following [Haines \(2001\)](#). This division is not a partition of the infectious mortality evaluated in the last section. Here we include causes of death that are ambiguous as to an infectious or other environmental etiology (e.g., chronic bronchitis, "other respiratory causes") and exclude infectious causes that were vector-borne (typhus, malaria), sexually transmitted (syphilis), or ambiguous ("other epidemic causes").<sup>5</sup>

Of infectious causes of death we can properly categorize, only one in five were waterborne. However, past scholarship on the improvement in urban mortality has singled out waterborne deaths as the mostly likely to be eliminated by public health infrastructure, interventions, and advancements in the early twentieth century ([Troesken, 1999, 2001](#); [Haines, 2001](#); [Cutler and Miller, 2005](#); [Beach et al., 2016](#); [Alsan and Goldin, 2019](#)). Even [Anderson et al. \(2019a\)](#), which downplays the role of clean water and sewerage interventions in broader declines in mortality, estimates large effects of infrastructure on typhoid deaths. If public health plays a role in the reduction of any urban-rural gaps, we would expect to see it in waterborne deaths. Despite these prior findings, we find that the gap between mortality from waterborne causes of death between urban and rural areas persisted until at least 1915 (when our data stop), as we show in Panels A and B of Figure [V](#).

Next, due to important mortality differences by city size ([Cain and Hong, 2009](#)), we break out urban groups by large and small city. For rural areas, we break counties into urban-adjacent rural areas versus rural counties completely surrounded by other rural counties. Using this classification of mortality, we estimate a relatively stable urban-rural gap whether or not we age-adjust our data (Panel A versus Panel B of Figure [V](#)). Urban areas seem to persistently struggle with waterborne causes of death, whether large or small.

What do our results for infectious waterborne deaths suggest about the effects of public health investments and infrastructure? We know that larger cities were more likely to im-

---

<sup>5</sup>While illnesses such as malaria also declined during the early 20th century, [Bleakley \(2010\)](#); [Kitchens \(2013b,a\)](#), these illnesses have not been attributed in the literature as major drivers of the relative urban penalty during the early-20th century [Haines \(2001\)](#).

plement clean water and sewage projects, both relatively to smaller cities and rural areas. In fact, rural areas saw essentially *no* public health investments during the period (Higgs, 1973; Hoehn-Velasco, 2018).<sup>6</sup> Still, rural waterborne death rates fall at similar rates as urban waterborne death rates and from a lower base. The patterns in Figure V suggest that major public health investments likely did not lead to urban-rural convergence or the erasure of the urban mortality penalty.

### 4.3 Respiratory Causes of Death Suggest that the 1918 Influenza Pandemic may have Hastened Urban-Rural Convergence

Next, we consider respiratory causes of death, including airborne-infectious illness, which accounted for the vast majority of infectious deaths during our period. We show respiratory causes of death in Panels C and D of Figure V as well as Appendix Figure B.5. These figures show a new pattern by city size: the essential divide in respiratory mortality is between large cities and everywhere else (urban or rural). These patterns are strengthened when we control for variation in population age structures in the age-adjusted data (Figure V Panel D). By 1915, while there is no penalty for small cities, there is still a clear and consistent penalty for large cities in respiratory causes of death.

Noymer (2009, 2011) documents mortality selection during the 1918 influenza pandemic, where the 1918 influenza increases the likelihood that someone dies from a respiratory cause of death (particularly TB). The argument, in short, is that many of the victims of the pandemic would have been at high risk for respiratory deaths in the years after 1918. However, the pandemic killed these people “early,” reducing later rates of TB mortality. We investigate what role, if any, this culling could have played in our story, we have to push our data beyond 1915. While we cannot test whether the 1918 pandemic *caused* the gap to close, we can consider whether this explanation is plausible in several ways.

First, we show the full results for the series of cities in Figure VI. While the gap in waterborne causes of death between large and small cities is never very large and closes before the pandemic (Panel A), for respiratory mortality, large cities only approach small cities’ levels after 1918 (Panel B). While large cities still have higher respiratory causes of death, the gap is smaller post-1918.

Second, to formalize the post-1918 decline in mortality, we consider whether large cities experienced significant decline in mortality relative to smaller cities, especially for airborne causes. To accomplish this, we group the post-1918 period and test whether larger cities had significantly lower mortality post-1918 relative to smaller cities. More specifically, we

---

<sup>6</sup>The county health department movement is one notable exception (Hoehn-Velasco, 2018). However, fewer than ten rural counties had a health department by 1915.

estimate the following:

$$M_{jt} = \alpha + \beta \text{Post-1918} \times \text{Large City}_{jt} + \alpha_j + \eta_t + \epsilon_{jt} \quad (2)$$

Where  $M_{jst}$  reflects the mortality rate for city  $j$  in year  $t$ .  $\text{Post-1918} \times \text{Large City}_{jt}$  is a dummy that equals one beginning in the post-1918 period for large cities only.  $\alpha_j$  accounts for the city fixed effects..  $\eta_t$  captures the year fixed effects.  $\epsilon_{jt}$  is the regression error. We cluster the standard errors at the city level.

Table 2 shows the results from Equation 2 across the mortality rates. Across all age-adjusted mortality, airborne mortality, and other non-infectious, there is a clear decline in mortality after 1918 for large cities relative to smaller cities. There is a less clear impact for waterborne and other-infectious, with waterborne showing no difference, and other infectious increasing post-1918. Overall, these coefficients suggest that there is a large relative decline in age-adjusted and airborne mortality causes after 1918.

To test whether the post-1918 period is particularly important, we next plot alternative post-periods 1901-1929 (29 different regressions) in Figure VII. There is a clear jump in the post-1918 through post-1929 coefficients. This pattern appears between 1917 and 1918, where mortality is lower in large cities relative to small cities. There does appear to be a closure in relative mortality between large and small cities post-1918.

In Panel B, we show the same set of regressions over airborne mortality. In the post-1918 period onward, there is a clear jump in the relative mortality from airborne causes in large v. smaller cities. This decline in the post-1918 period indicates that the 1918 influenza pandemic may have produced some change in mortality from airborne causes. The same pattern does not appear in other causes of death in Figure A.8. Overall, these results suggest that mortality did narrow between big cities and smaller cities after 1918.

#### 4.4 Tuberculosis Converges between Urban and Rural Areas After 1918

The shrinking respiratory gap between large and small cities is compelling, but does not speak to our central question about urban and rural mortality. To shed light on this, we turn to the aggregate data. When we narrow in on TB deaths in particular, the picture is even clearer (Figure VIII). After 1918, rural and urban areas almost completely converged in tuberculosis deaths. These aggregate numbers suggest that a decline in tuberculosis deaths may have sped the convergence in rural and urban areas, where urban areas with previously high tuberculosis deaths, had lower mortality after the 1918 influenza pandemic.

In particular, the age-standardized rates show a strong convergence in the Northeastern after the 1918 influenza pandemic. This fact makes intuitive sense as the Northeastern cities had the highest tuberculosis mortality pre-1918 and would face the highest influenza-based selection from the pandemic.

## 5 Conclusion

In this study, we reexamine the canonical narrative surrounding urban-rural mortality convergence. Did urban areas become cleaner, safer, and healthier during the early twentieth century? Did the urban penalty disappear? We replicate the traditional answer of yes when we look only at crude mortality in the ten states in the DRA from 1900 forward. However, the narrative breaks down for the United States as a whole. Using the full sample, or age-adjusted rates the results reveal a persistent urban penalty in overall mortality past 1936.

When we consider age-standardized mortality rates, a similar pattern suggests that the urban mortality penalty declines but persists until 1936. The age-standardized rates are less sensitive to the compositional changes as states enter the DRA, and the age-standardized series show similar urban penalties across the balanced and unbalanced panels. The importance of using age-standardized rates suggests that compositional differences between urban and rural areas (as well as differences by DRA entry) played a more critical role in the mortality declines than previously known. Urban areas had an over-representation of young adults, suppressing mortality, while rural areas had an abundance of children, inflating local mortality. Further, states that entered the DRA later in the 20th century were systematically different than early entry states. Most notably, these states had larger nonwhite populations with worse mortality in cities ([Feigenbaum et al. \(2019\)](#) and Panel B of Table 1). These findings suggest that accounting for age composition is essential when quantifying mortality declines.

The urban penalty that does appear is distinct across samples and mortality measures, highlighting three crucial points. First, we show that when the penalty disappears is sample-dependent, but in three out of four samples, urban mortality is still higher than in rural areas in 1936. Second, we highlight the importance of accounting for compositional differences. When we use the balanced panel relative to the full sample of DRA states, we come to different conclusions about when the mortality penalty disappeared. Our analysis demonstrates the importance of accounting for demographic differences in comparing mortality rates across areas. Our conclusions are mixed and depend on which measure of mortality (and sample) we choose. While a portion of this point is discussed with infant mortality in [Haines \(2001\)](#), our study illustrates that the age-standardized rates can help quantify differences between areas in a single measure of mortality.



After quantifying the urban penalty, we explore the role of infectious causes of death in the urban-rural convergence, focusing on waterborne-infectious and airborne-infectious. We show that both large and small cities experienced persistently elevated waterborne causes of death relative to rural areas (until 1915). Together, these findings suggest that public health infrastructure did not account for a substantial convergence in urban-rural mortality over the period considered in this study. Our findings fit well with recent work, [Anderson et al. \(2019a\)](#), which critiques the stylized narrative that major infrastructure advancement produced the urban mortality decline.

We conclude by examining respiratory causes of death. The results show that large cities persistently struggled with high respiratory causes of death, while small cities had similar respiratory mortality to rural areas. This gap somewhat converges after 1918 between large and small cities, but our rural series drops off after 1915. Based on the respiratory findings, we explore whether mortality selection during 1918 may have driven down the urban penalty. We note three findings that align with [Noymer, 2009, 2011](#)). First, we show that there was substantial urban-rural convergence in aggregate tuberculosis mortality after 1918. Second, we show that the urban penalty is substantially lower in the Northeast after 1918, the region with the highest urban tuberculosis mortality. Third, we show that the penalty between large and small cities closes substantially after 1918. Put together, our findings suggest that the 1918 influenza may have lowered tuberculosis mortality, through both mortality selection and lower transmission, and led to urban-rural convergence after 1918 ([Noymer, 2009, 2011](#)).

Our findings open many avenues for future research. First, when precisely does the urban penalty disappear? We conclude the primary analysis in 1936, but future work could extend this pattern throughout the twentieth century. Second, since waterborne findings reveal that public health may have played a minor role ([Anderson et al., 2019a](#)), which factors lead to the ultimate convergence in mortality? We show that age structure played a role, but there were likely other factors. Third, did the 1918 influenza pandemic play a key role in reducing respiratory mortality in large cities in the years following the pandemic? Fourth, which other factors may have lead to rural-urban convergence? While there are prime candidates for consideration, including immigration [Eriksson and Niemesh \(2016\)](#); [Ager et al. \(2020\)](#), modern sulfa drugs and antibiotics [Jayachandran et al. \(2010\)](#), and vaccinations, we leave questions of the causes of convergence for future work.



## 6 Tables

Table 1: County Characteristics by DRA Entry Year

Panel A: 1900 Demographic Characteristics				
	Pre-1900	1901-1910	1911-1920	1921-1933
	1900 Mean	1900 Mean	1900 Mean	1900 Mean
<b>Rural</b>				
Share White	0.99	0.96	0.80	0.81
Population <5	0.11	0.12	0.14	0.15
Share 5-19	0.30	0.32	0.37	0.36
Share 20-29	0.16	0.17	0.17	0.18
Share Over 60	0.10	0.07	0.06	0.05
Annual Pop. Growth %	-0.27	2.98	1.16	10.16
<b>Urban</b>				
Share White	0.98	0.97	0.78	0.79
Population <5	0.09	0.10	0.10	0.10
Share 5-19	0.28	0.30	0.31	0.31
Share 20-29	0.19	0.19	0.20	0.20
Share Over 60	0.08	0.07	0.06	0.07
Annual Pop. Growth %	3.52	4.56	4.10	4.05
Panel B: 1936 Mortality Conditions				
	Pre-1900	1901-1910	1911-1920	1921-1933
	1936 Mean	1936 Mean	1936 Mean	1936 Mean
<b>Rural</b>				
Overall Mortality - Rural	1,532.53	1,332.69	1,198.14	1,133.16
Age-St. Mortality - Rural	1.09	1.08	1.05	1.09
<b>Urban</b>				
Overall Mortality - Urban	1,334.09	1,586.86	1,614.96	1,814.76
Age-St. Mortality - Urban	1.15	1.35	1.59	1.71

NOTES: Entry year reflects entry into the vital statistics within the years specified in the column header. The mortality years are for 1936 to have a complete mortality picture of urban and rural areas. The demographic composition is presented for the 1900 census.

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county.

Table 2: Mortality by Large City x Post-1918

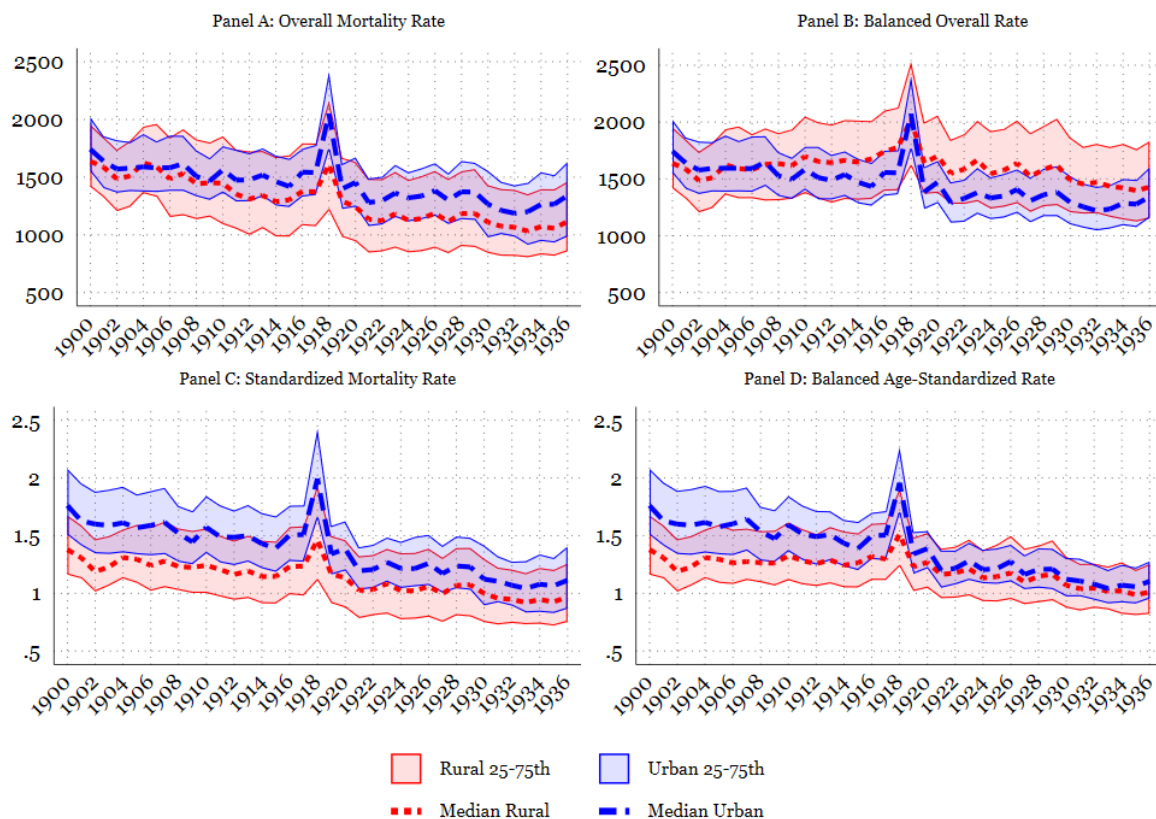
	Age-Adj. All	Age-Adj. Cities	Air- borne	Water- borne	Other- Infect	Other- Non-Infect
	(1)	(2)	(3)	(4)	(5)	(6)
Big City x Post	-0.2149*** (0.0270)	-0.0709** (0.0292)	-0.0667** (0.0267)	0.0006 (0.0378)	0.0392*** (0.0128)	-0.0407*** (0.0151)
N	82,657	20,836	32,419	32,419	32,419	32,419
Adjusted R-squared	0.692	0.799	0.824	0.727	0.561	0.763
Mean Dependent	1.24	1.40	0.48	0.32	0.39	0.96
State and Year FE	X	X	X	X	X	X

NOTES: Estimates reflect results from estimating Equation 2. The estimates reflect the of Post-1918 x Big City. Each regression includes year and city fixed effects. Only Column (1) includes rural areas. Columns (2)-(6) include big cities relative to smaller cities. Standard errors are clustered at the city level. \*, \*\*, \* represent statistical significance at 1, 5 and 10 percent levels.

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county.

## 7 Figures

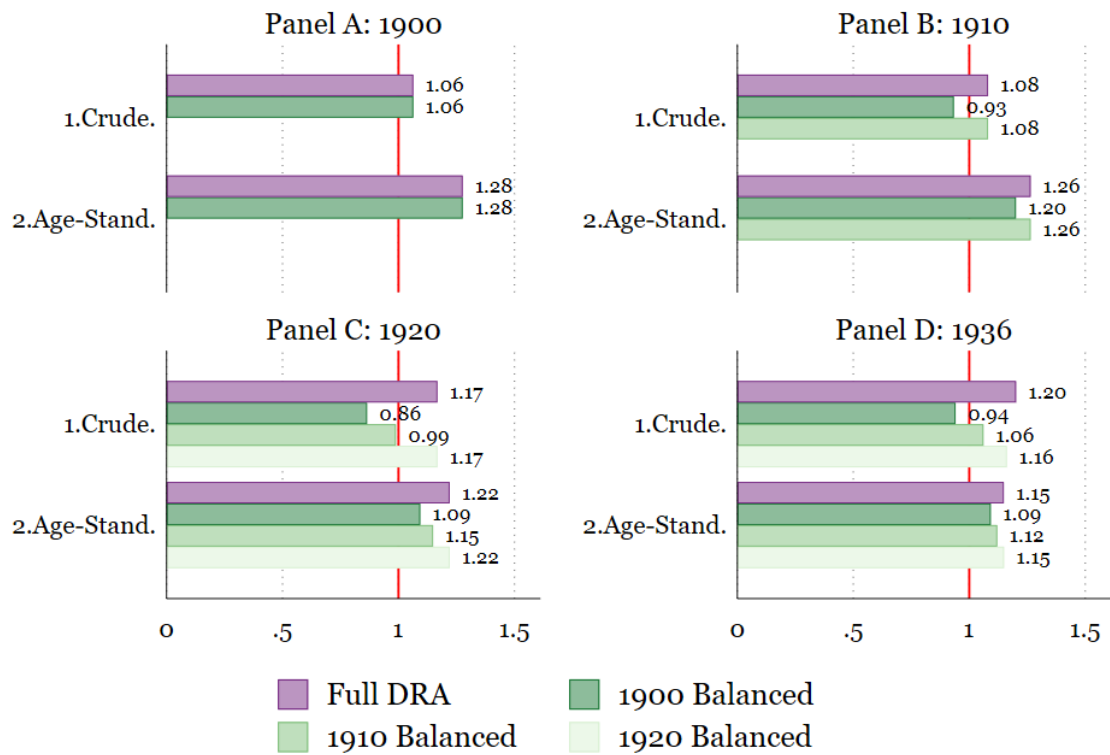
Figure I: Urban v. Rural Mortality, 1900-1936



NOTES: The plotted points reflect the 25th, 50th, and 75th percentiles for each year over 1900-1936. The first panel shows the overall mortality rates for the unbalanced panel. The second panel shows the balanced panel for overall mortality. The third panel shows the age-standardized rates for the unbalanced panel. The final panel shows the balanced panel for the age-standardized rates. The blue indicates urban areas, and the red shows rural areas. Mortality rates are reported per 100,000 persons.

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county. County-level population counts and demographic characteristics are calculated from the IPUMs Restricted Complete Count Census data.

Figure II: Urban Mortality Comparison -  $\frac{\text{Median Urban Mortality}}{\text{Median Rural Mortality}}$

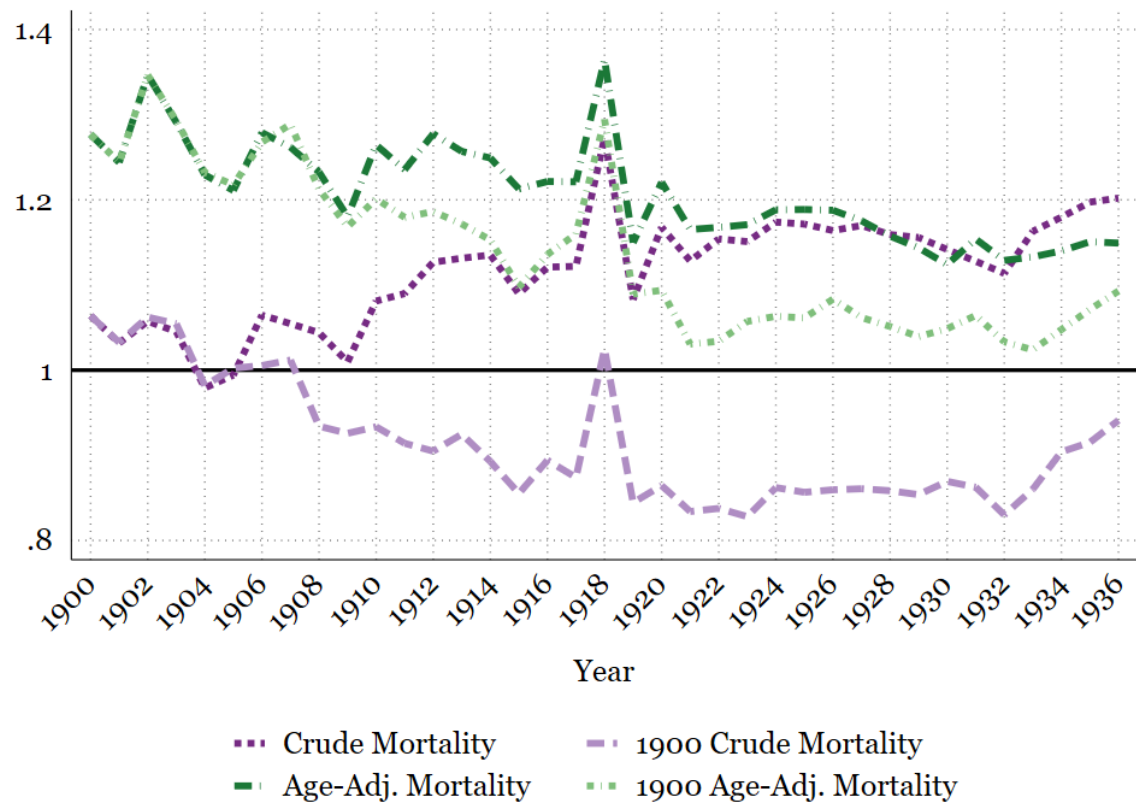


Graphs by Year

NOTES: Points reflect the urban mortality divided by the rural mortality at the 50th percentile (median) for each year. Ratios above one indicate an urban penalty.

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county. County-level population counts and demographic characteristics are calculated from the IPUMs Restricted Complete Count Census data.

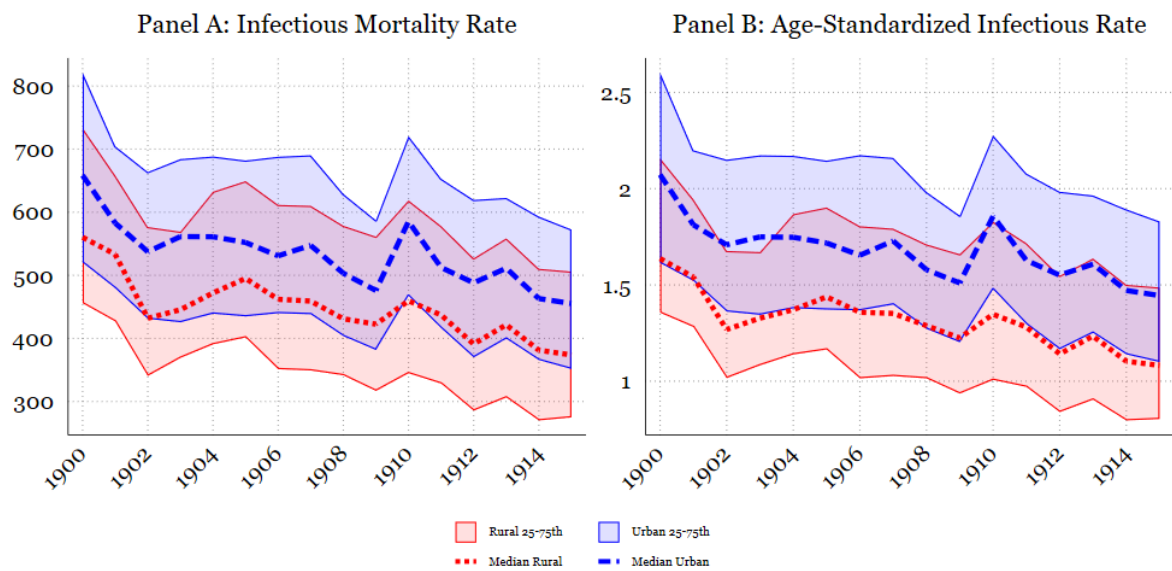
Figure III: Urban Mortality Penalty -  $\frac{\text{Median Urban Mortality}}{\text{Median Rural Mortality}}$



NOTES: Points reflect the urban mortality divided by the rural mortality at the 50th percentile (median) for each year. Ratios above one indicate an urban penalty.

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county. County-level population counts and demographic characteristics are calculated from the IPUMs Restricted Complete Count Census data.

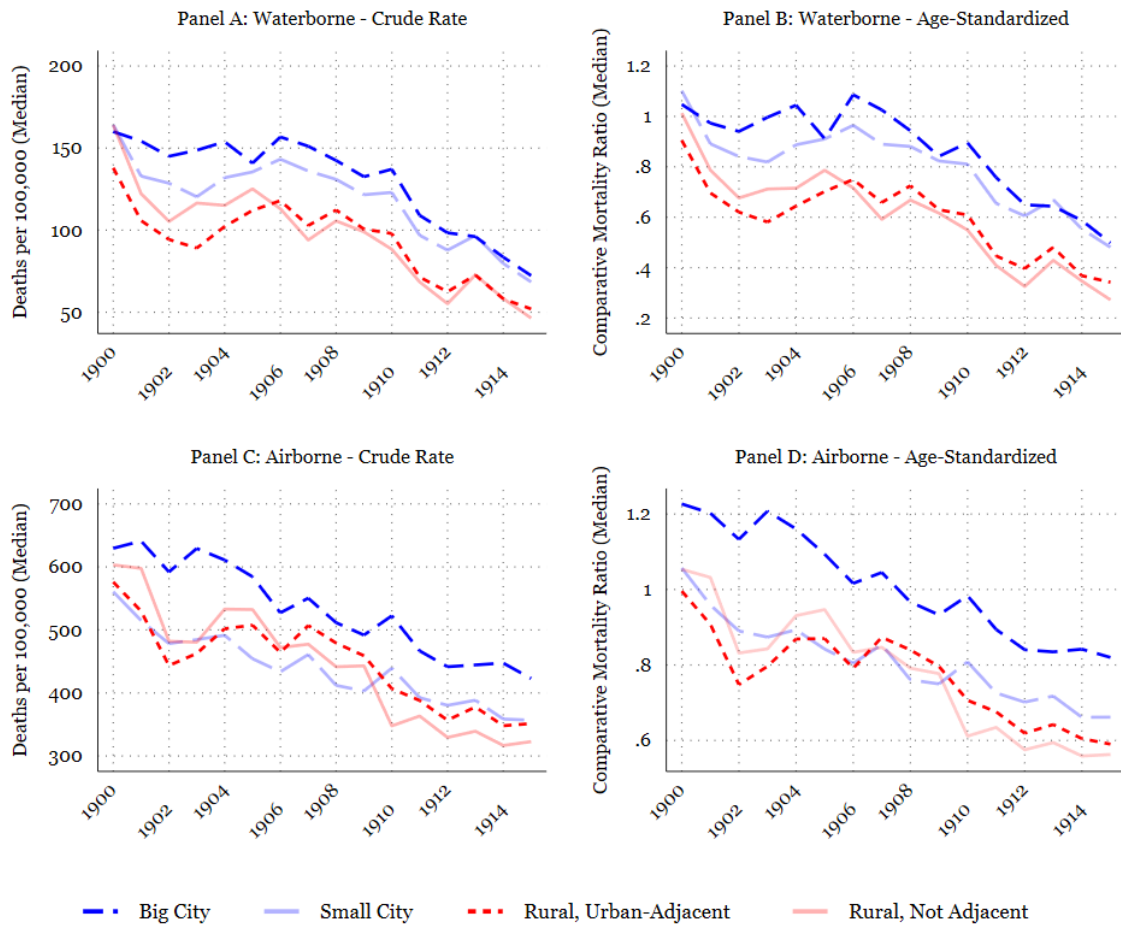
Figure IV: Urban v. Rural Infectious Disease Mortality, 1900-1915



NOTES: The plotted points reflect the 25th, 50th, and 75th percentiles for each year over 1900-1915. Mortality rates are reported per 100,000 persons. The blue indicates urban areas, and the red shows rural areas.

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county. County-level population counts and demographic characteristics are calculated from the IPUMs Restricted Complete Count Census data.

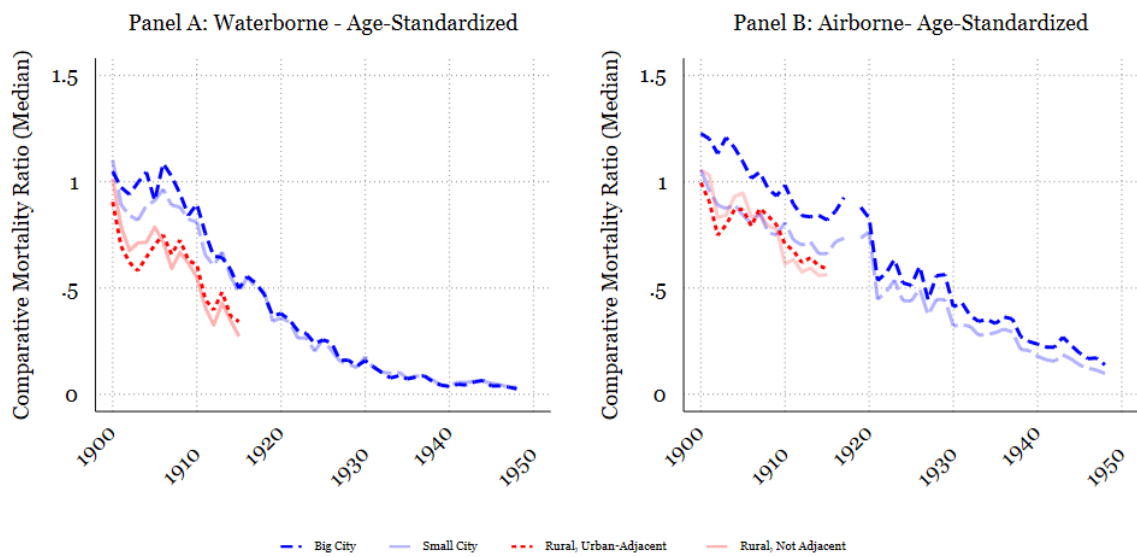
Figure V: Waterborne and Airborne, 1900-1915



NOTES: The blue indicates urban areas, and the red shows rural areas. Mortality rates are reported per 100,000 persons.

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county. County-level population counts and demographic characteristics are calculated from the IPUMs Restricted Complete Count Census data.

Figure VI: Waterborne and Airborne, 1900-1950



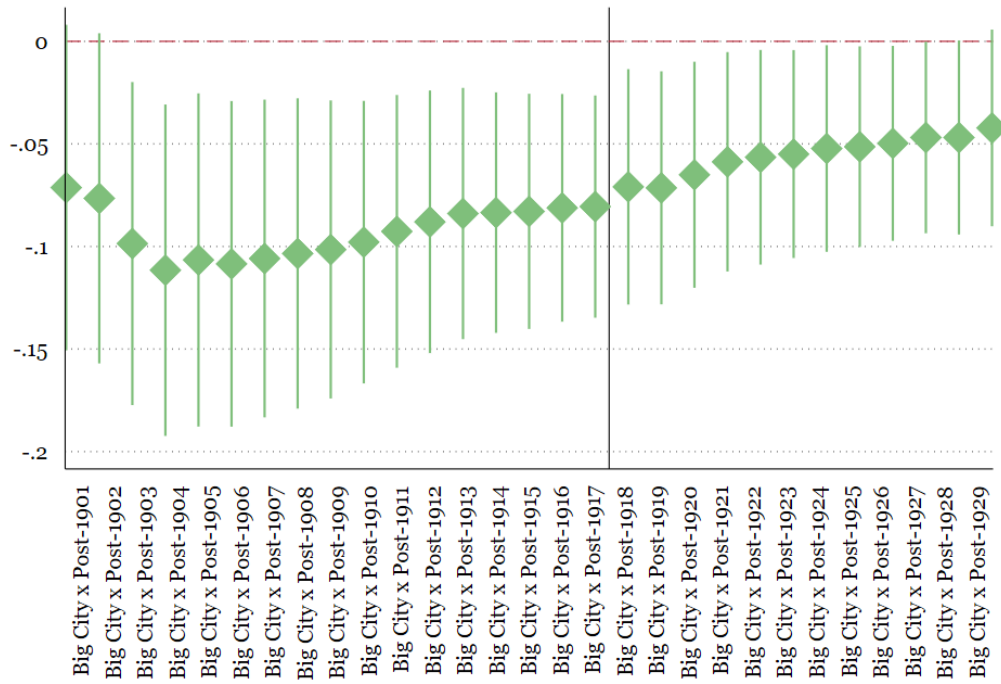
NOTES: The blue indicates urban areas, and the red shows rural areas. Mortality rates are reported per 100,000 persons.

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county. County-level population counts and demographic characteristics are calculated from the IPUMs Restricted Complete Count Census data.

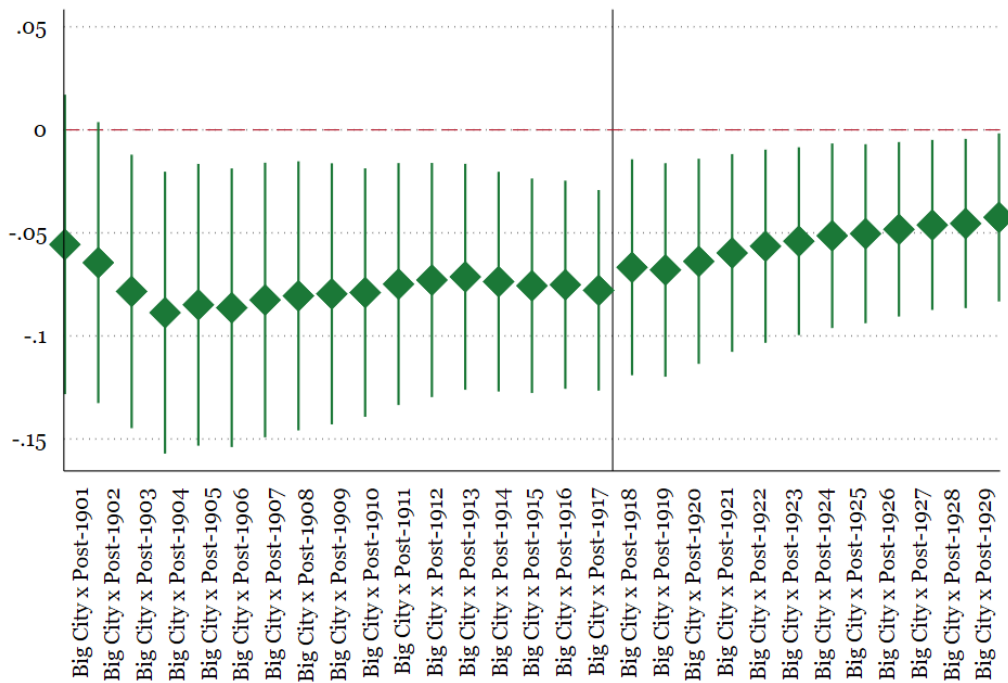


Figure VII: Mortality by Large Cities x Post-Year

Panel A: Age-Standardized Rate



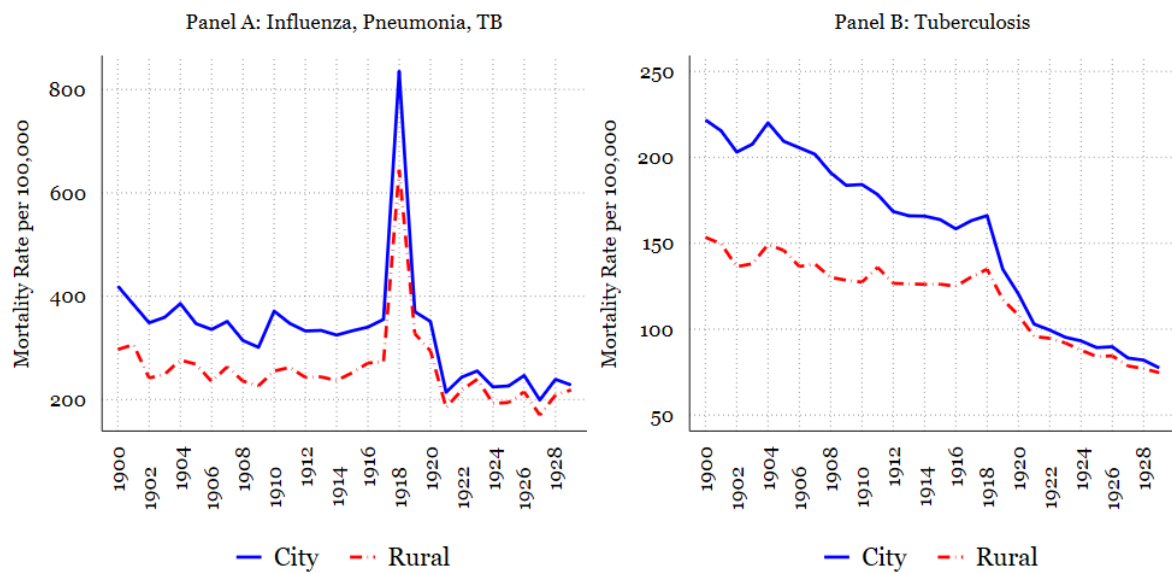
Panel B: Airborne



NOTES: Results are shown for separate regressions estimating the effect of Post-year x Big City. Each regression includes year and city fixed effects. Standard errors are clustered at the city level. Lines represent 95% confidence intervals.

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county.

Figure VIII: Tuberculosis, Influenza, and Pneumonia Aggregate Mortality, 1900-1930



NOTES: The blue indicates urban areas, and the red shows rural areas. Mortality rates are reported per 100,000 persons.

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county. County-level population counts and demographic characteristics are calculated from the IPUMs Restricted Complete Count Census data.

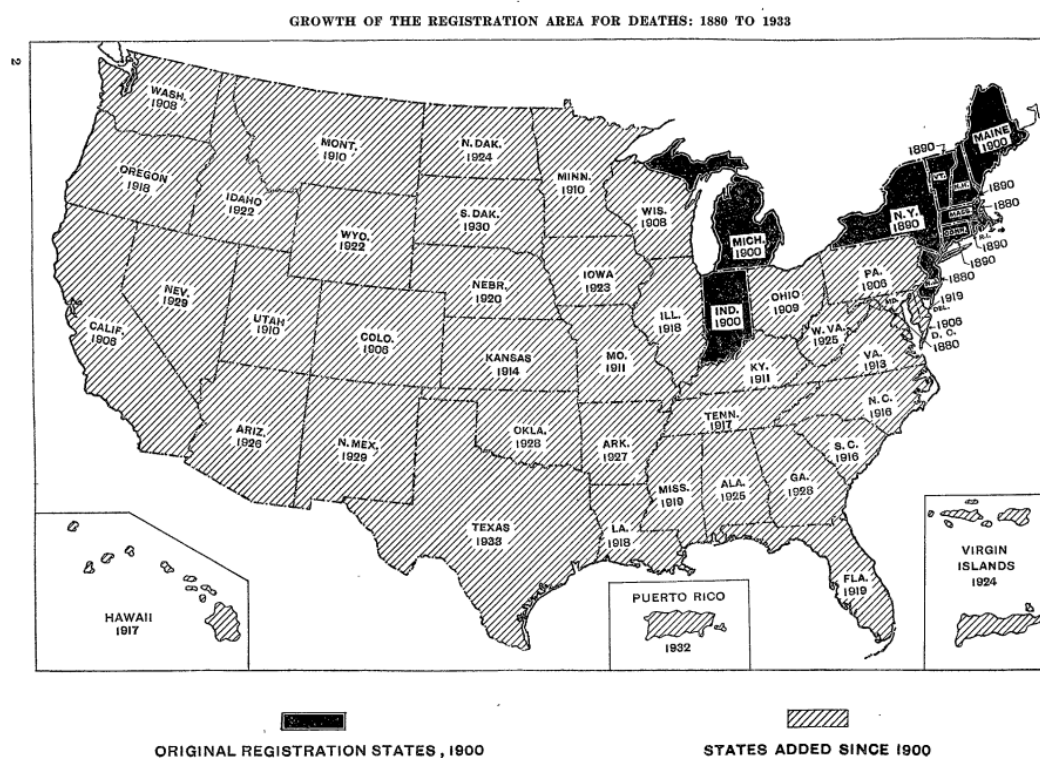
## References

- Ager, Philipp, James J Feigenbaum, Casper Worm Hansen, and Hui Ren Tan**, “How the Other Half Died: Immigration and Mortality in US Cities,” Technical Report, National Bureau of Economic Research 2020.
- Alsan, Marcella and Claudia Goldin**, “Watersheds in Child Mortality: The Role of Effective Water and Sewerage Infrastructure, 1880 to 1920,” *Journal of Political Economy*, 2019, 127 (2), 586–638.
- Anderson, D Mark, Kerwin Kofi Charles, and Daniel I Rees**, “Re-Examining the Contribution of Public Health Efforts to the Decline in Urban Mortality,” *American Economic Journal: Applied Economics*, 2019.
- , —, **Claudio Las Heras Olivares, and Daniel I Rees**, “Was the First Public Health Campaign Successful?” *American Economic Journal: Applied Economics*, 2019, 11 (2), 143–75.
- , **Ryan Brown, Kerwin Kofi Charles, and Daniel I Rees**, “The effect of occupational licensing on consumer welfare: Early midwifery laws and maternal mortality,” Technical Report, National Bureau of Economic Research 2016.
- Beach, Brian, Joseph Ferrie, Martin Saavedra, and Werner Troesken**, “Typhoid fever, water quality, and human capital formation,” *The Journal of Economic History*, 2016, 76 (1), 41–75.
- Bleakley, Hoyt**, “Malaria eradication in the Americas: A retrospective analysis of childhood exposure,” *American Economic Journal: Applied Economics*, 2010, 2 (2), 1–45.
- Cain, Louis and Sok Chul Hong**, “Survival in 19th century cities: The larger the city, the smaller your chances,” *Explorations in Economic History*, 2009, 46 (4), 450–463.
- Clay, Karen, Peter Juul Egedesø, Casper Worm Hansen, Peter Sandholt Jensen, and Avery Calkins**, “Controlling tuberculosis? Evidence from the first community-wide health experiment,” *Journal of Development Economics*, 2020, 146, 102510.
- Condran, Gretchen A and Eileen Crimmins**, “Mortality differentials between rural and urban areas of states in the northeastern United States 1890–1900,” *Journal of Historical Geography*, 1980, 6 (2), 179–202.
- Cutler, David and Grant Miller**, “The role of public health improvements in health advances: The twentieth-century United States,” *Demography*, 2005, 42 (1), 1–22.
- Eriksson, Katherine and Gregory Niemesh**, “Death in the promised land: The Great Migration and black infant mortality,” *Available at SSRN 3071053*, 2016.
- Feigenbaum, James J, Christopher Muller, and Elizabeth Wrigley-Field**, “Regional and Racial Inequality in Infectious Disease Mortality in US Cities, 1900–1948,” *Demography*, 2019, 56 (4), 1371–1388.
- Haines, Michael R**, “The urban mortality transition in the United States, 1800–1940,” *Annales de démographie historique*, 2001, 101 (1), 33–64.
- Higgs, Robert**, “Mortality in rural America, 1870–1920: Estimates and conjectures,” *Explorations in Economic History*, 1973, 10 (2), 177.
- Hoehn-Velasco, Lauren**, “Explaining declines in US rural mortality, 1910–1933: The role of county health departments,” *Explorations in Economic History*, 2018, 70, 42–72.

- Jayachandran, Seema, Adriana Lleras-Muney, and Kimberly V Smith**, “Modern medicine and the twentieth century decline in mortality: Evidence on the impact of sulfa drugs,” *American Economic Journal: Applied Economics*, 2010, 2 (2), 118–46.
- Kitchens, Carl**, “A dam problem: TVA’s fight against malaria, 1926–1951,” *The Journal of Economic History*, 2013, pp. 694–724.
- , “The effects of the Works Progress Administration’s anti-malaria programs in Georgia 1932–1947,” *Explorations in Economic History*, 2013, 50 (4), 567–581.
- Komisarow, Sarah**, “Public health regulation and mortality: Evidence from early 20th century milk laws,” *Journal of health economics*, 2017, 56, 126–144.
- Moehling, Carolyn M and Melissa A Thomasson**, “Saving Babies: The Impact of Public Education Programs on Infant Mortality,” *Demography*, 2014, 51 (2), 367–386.
- Noymer, Andrew**, “Testing the influenza–tuberculosis selective mortality hypothesis with Union Army data,” *Social Science & Medicine*, 2009, 68 (9), 1599–1608.
- , “The 1918 influenza pandemic hastened the decline of tuberculosis in the United States: an age, period, cohort analysis,” *Vaccine*, 2011, 29, B38–B41.
- Olmstead, Alan L and Paul W Rhode**, “An impossible undertaking: the eradication of bovine tuberculosis in the United States,” *The Journal of Economic History*, 2004, 64 (03), 734–772.
- Ruggles, Steven, Sarah Flood, Sophia Foster, Ronald Goeken Goeken, Jose Pacas, Megan Schouweiler, and Matthew Sobek**, “IPUMS USA: Version 11.0 [dataset],” 2021.
- Statistics, US Vital**, *Mortality Statistics: Vital Statistics of the United States: 1890-1938*, United States Bureau of the Census, 1890-1938.
- Troesken, Werner**, “Typhoid Rates and the Public Acquisition of Private Waterworks, 1880–1920,” *The Journal of Economic History*, 1999, 59 (04), 927–948.
- , “Race, Disease, and the Provision of Water in American Cities, 1889–1921,” *The Journal of Economic History*, 2001, 61 (03), 750–776.

## A Additional Tables and Figures

Figure A.1: Map of DRA Entry Year by State



NOTES: SOURCES: Page 6 of 1933 Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county.

Table A.1: County Characteristics, Alternative by Panel and Region

	Northeast	Midwest	South	West
	1900 Mean	1900 Mean	1900 Mean	1900 Mean
<b>Rural</b>				
Share White	0.99	0.98	0.71	0.92
Population <5	0.10	0.13	0.15	0.12
Share 5-19	0.28	0.34	0.38	0.30
Share 20-29	0.16	0.17	0.18	0.18
Share Over 60	0.11	0.07	0.05	0.06
Annual Pop. Growth %	-0.35	2.26	5.36	6.93
<b>Urban</b>				
Share White	0.98	0.97	0.63	0.95
Population <5	0.10	0.10	0.11	0.10
Share 5-19	0.27	0.30	0.32	0.28
Share 20-29	0.20	0.19	0.21	0.19
Share Over 60	0.08	0.07	0.06	0.06
Annual Pop. Growth %	4.45	2.85	4.00	9.43

NOTES: Regions are reflected in the column header. The census years are all for 1900.

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county.

Table A.2: Summary Statistics  
Panel A: Overall Mortality, 1900-1936

	1900	1936	Diff.
	Mean	Mean	Est.
<b>Full-Panel</b>			
Overall Mortality - Rural	1,777.87	1,237.60	-540.27***
Overall Mortality - Urban	1,817.46	1,354.65	-462.81***
Age-St. Mortality - Rural	1.54	1.07	-0.47***
Age-St. Mortality - Urban	1.88	1.18	-0.70***
<b>Balanced-Panel</b>			
Overall Mortality- Rural	1,777.87	1,532.53	-245.34***
Overall Mortality - Urban	1,817.46	1,392.50	-424.96***
Age-St. Mortality- Rural	1.54	1.09	-0.45***
Age-St. Mortality - Urban	1.88	1.14	-0.74***

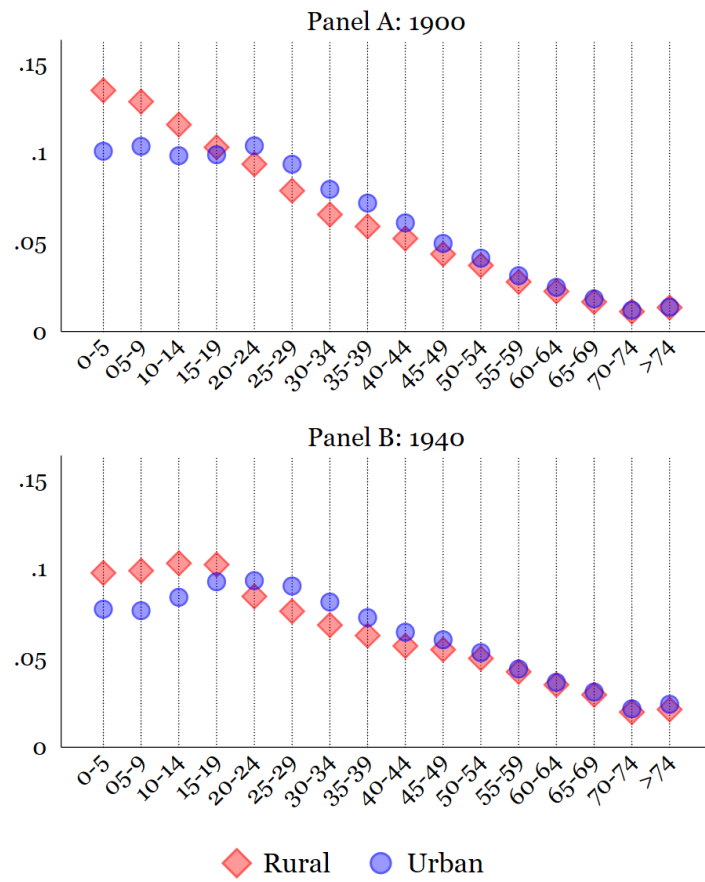
Panel B: Infectious Mortality, 1900-1915

	1900	1915	Diff.
	Mean	Mean	Est.
<b>Full-Panel</b>			
Infectious Rate - Rural	632.62	418.80	-213.82***
Infectious Rate - Urban	692.85	485.32	-207.54***
Age-St. Infectious - Urban	2.19	1.54	-0.65***
Age-St. Infectious - Rural	1.87	1.23	-0.63***
<b>Balanced-Panel</b>			
Infectious Rate - Rural	624.02	471.93	-152.09***
Infectious Rate - Urban	700.64	479.09	-221.55***
Age-Adj. Infectious - Rural	1.84	1.36	-0.48***
Age-Adj. Infectious - Urban	2.21	1.51	-0.70***

NOTES: Mortality rates are reported per 100,000 persons.

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county.

Figure A.2: Age Composition of Rural and Urban Areas



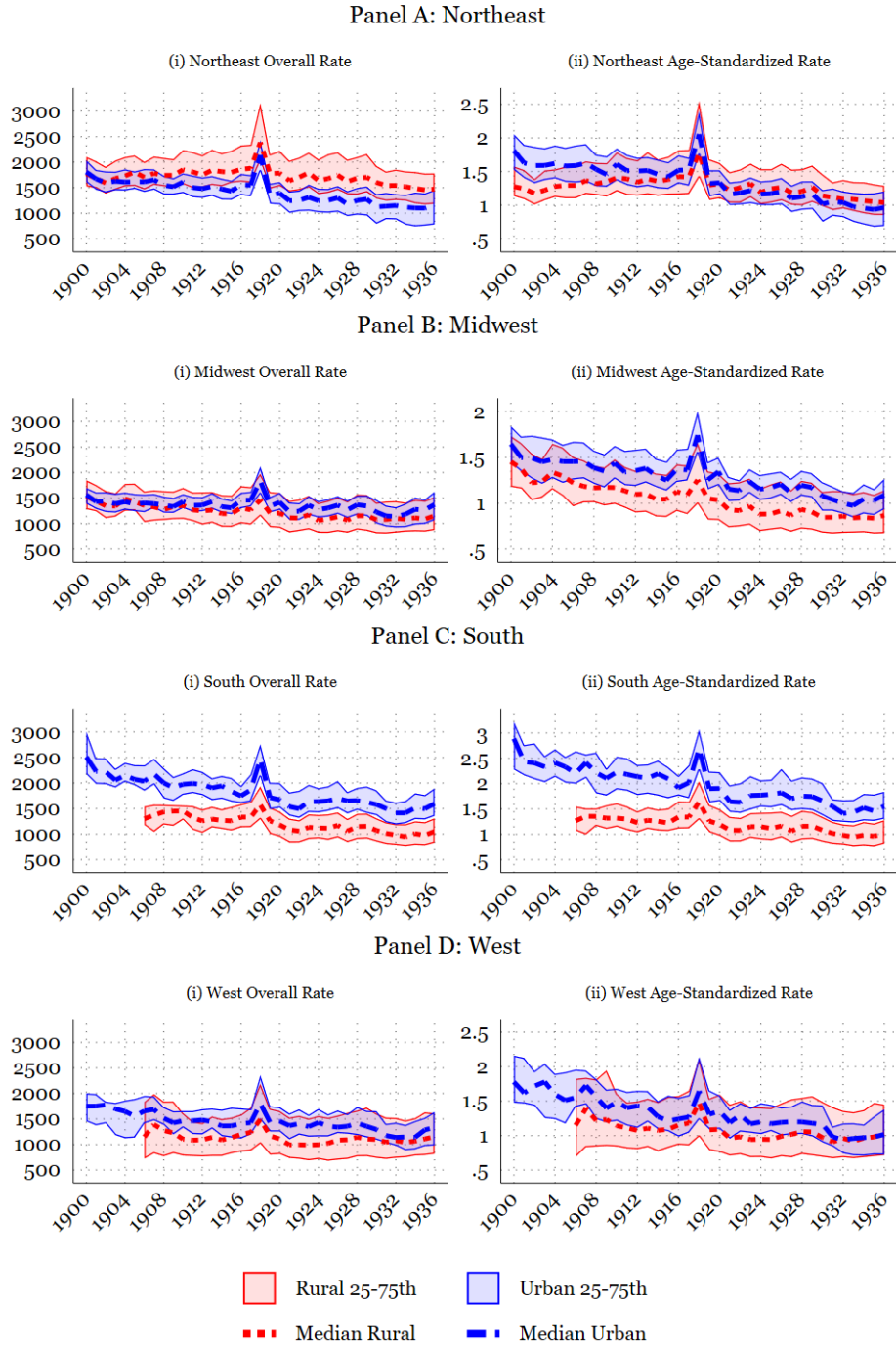
Graphs by Year

NOTES: Graph shows the average share over each age group in rural versus urban areas in 1900 (middle panel) and 1940 (bottom panel).

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county. County-level population counts and demographic characteristics are calculated from the IPUMs Restricted Complete Count Census data.



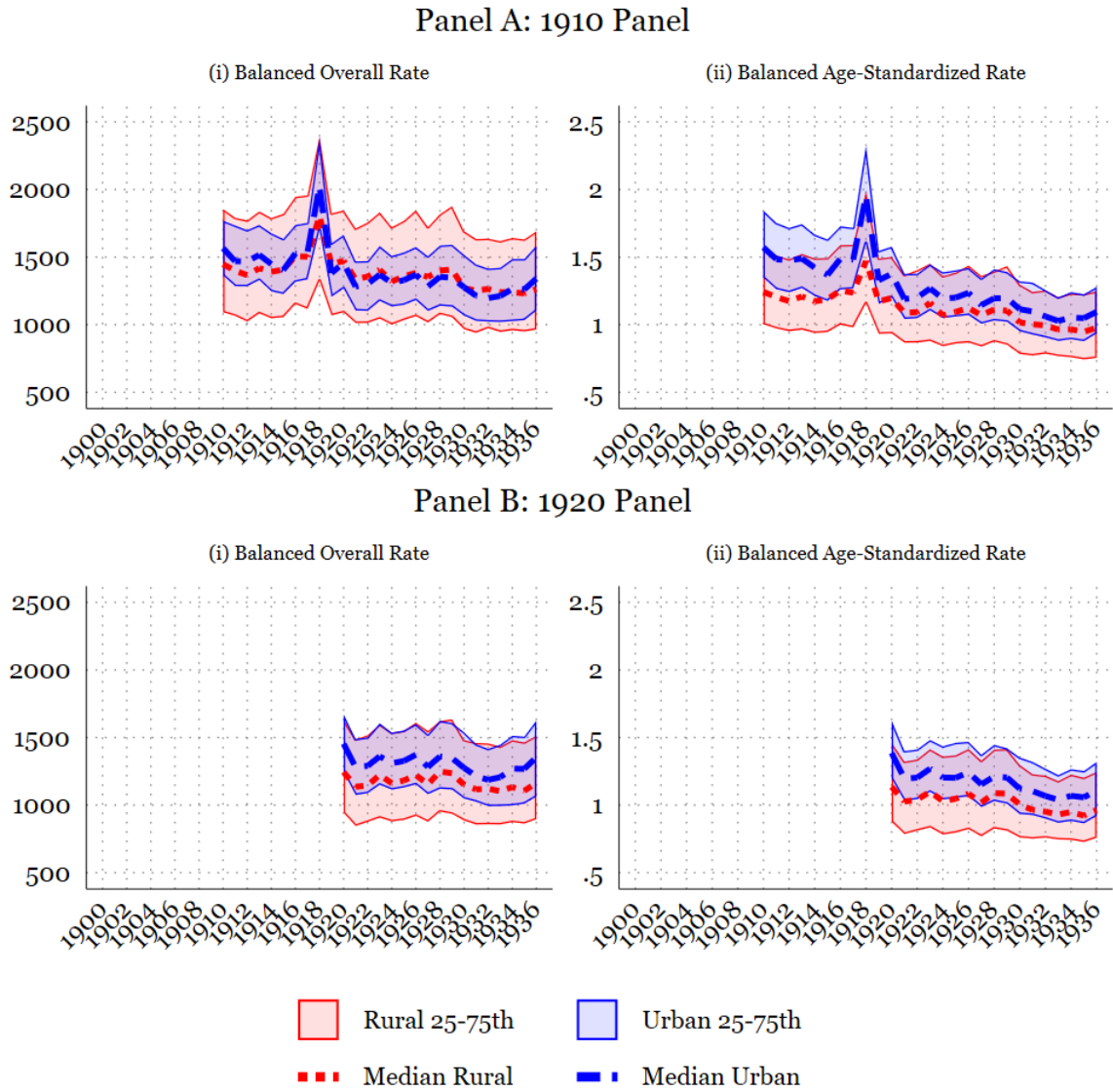
Figure A.3: Urban v. Rural Mortality by Region, 1900-1936



NOTES: The plotted points reflect the 25th, 50th, and 75th percentiles for each year over 1900-1936. The first panel shows the overall mortality rates for the unbalanced panel. The second panel shows the balanced panel for overall mortality. The third panel shows the age-standardized rates for the unbalanced panel. The final panel shows the balanced panel for the age-standardized rates. The blue indicates urban areas, and the red shows rural areas. Mortality rates are reported per 100,000 persons.

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county. County-level population counts and demographic characteristics are calculated from the IPUMs Restricted Complete Count Census data.

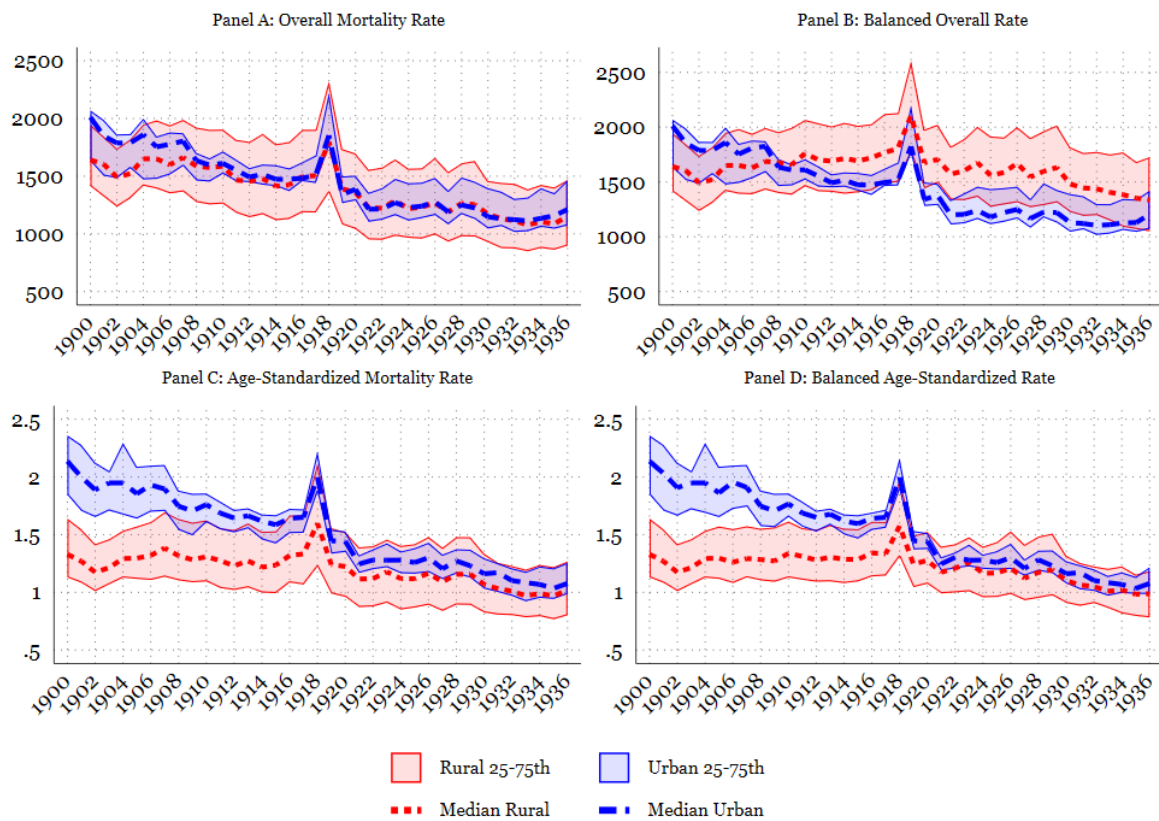
Figure A.4: Near-Balanced Panel, Alternative Years



NOTES: The plotted points reflect the 25th, 50th, and 75th percentiles for each year. The top panel shows that states that had entered by 1910 and the bottom two panels show the states that had entered by 1920. Mortality rates are reported per 100,000 persons. The blue indicates urban areas, and the red shows rural areas.

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county. County-level population counts and demographic characteristics are calculated from the IPUMs Restricted Complete Count Census data.

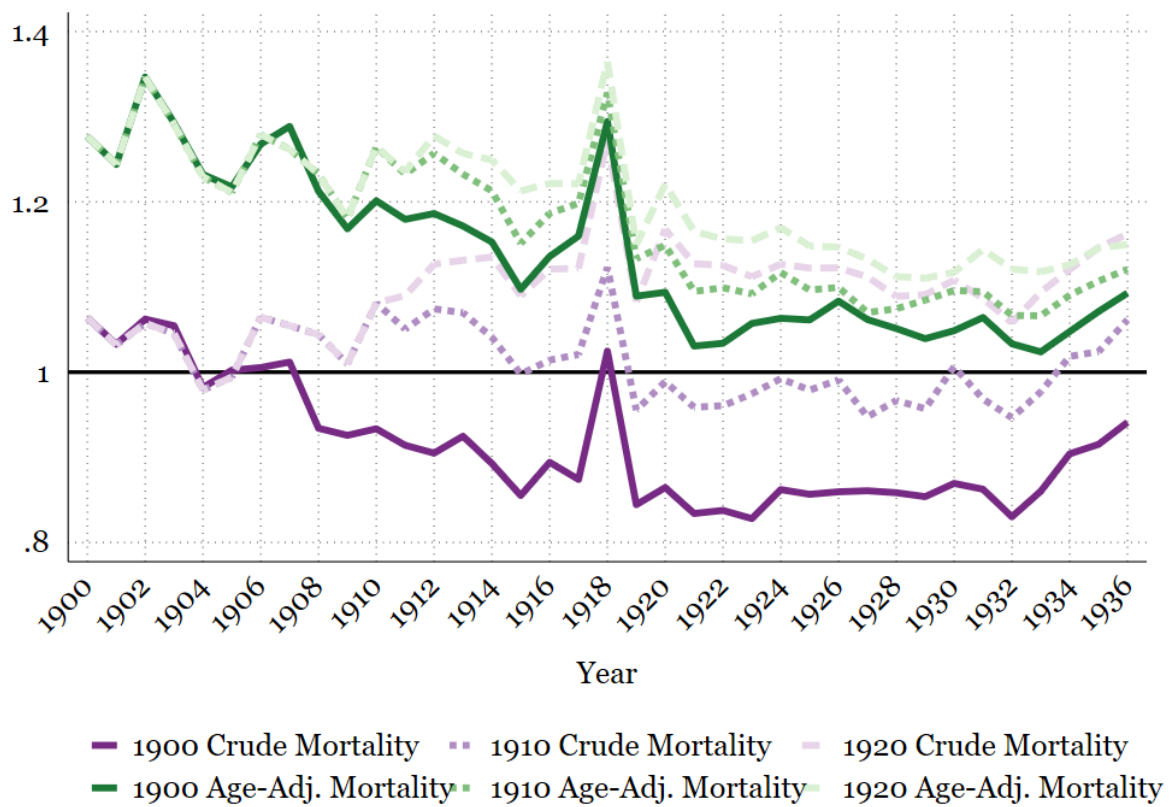
Figure A.5: Population-Weighted Estimates



NOTES: The plotted points reflect the 25th, 50th, and 75th percentiles for each year. The top panel shows that states that had entered by 1910 and the bottom two panels show the states that had entered by 1920. Mortality rates are reported per 100,000 persons. The blue indicates urban areas, and the red shows rural areas.

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county. County-level population counts and demographic characteristics are calculated from the IPUMs Restricted Complete Count Census data.

Figure A.6: Urban Mortality Penalty -  $\frac{\text{Median Urban Mortality}}{\text{Median Rural Mortality}}$  - Alternative Balanced Panel



NOTES: Points reflect the urban mortality divided by the rural mortality at the 50th percentile (median) for each year. Ratios above one indicate an urban penalty.

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county. County-level population counts and demographic characteristics are calculated from the IPUMs Restricted Complete Count Census data.

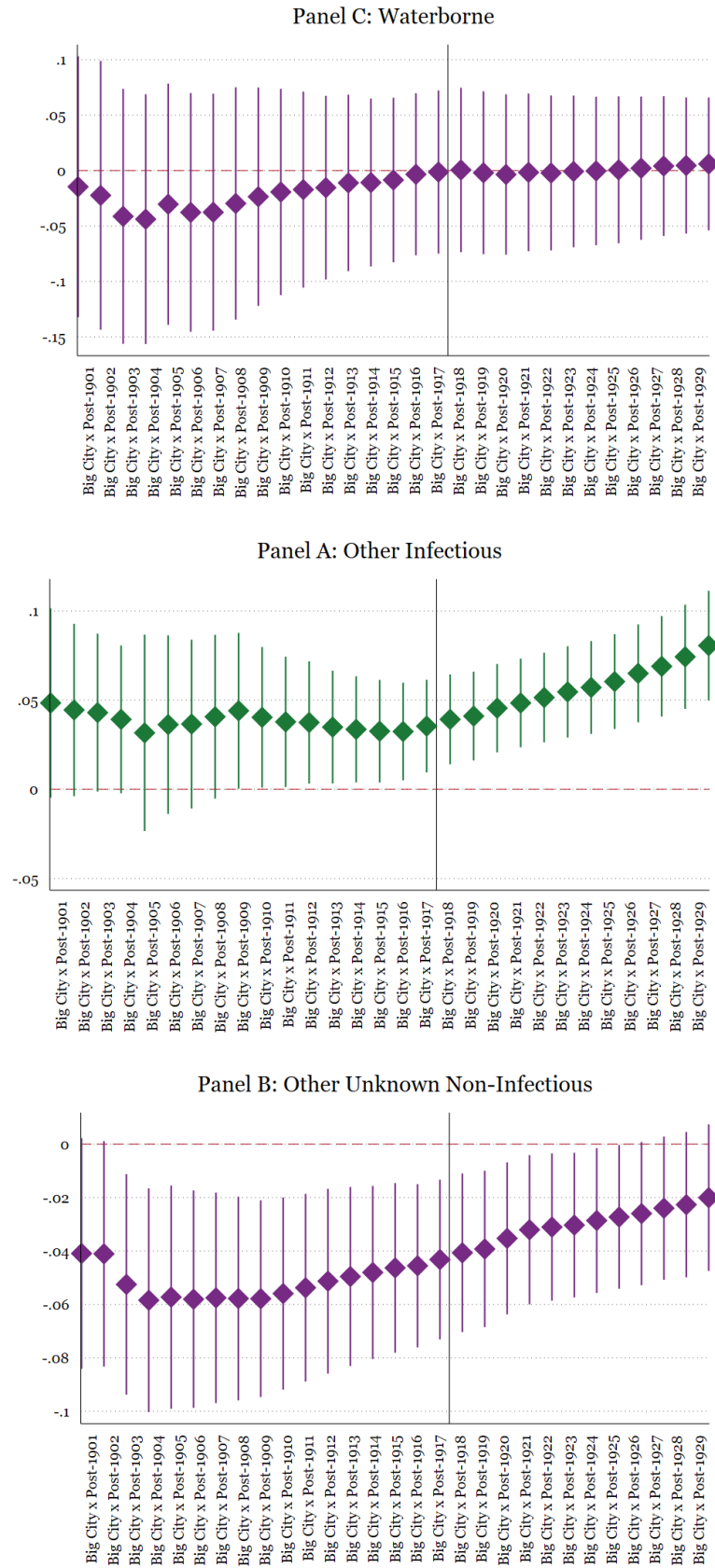
Figure A.7: Aggregate Death Rates from State-level Urban v. Rural Deaths



NOTES: Mortality rates are reported per 100,000 persons. The blue indicates urban areas, and the red shows rural areas.

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county. County-level population counts and demographic characteristics are calculated from the IPUMS Restricted Complete Count Census data.

Figure A.8: Mortality by Large Cities x Post-Year, Other/Unknown

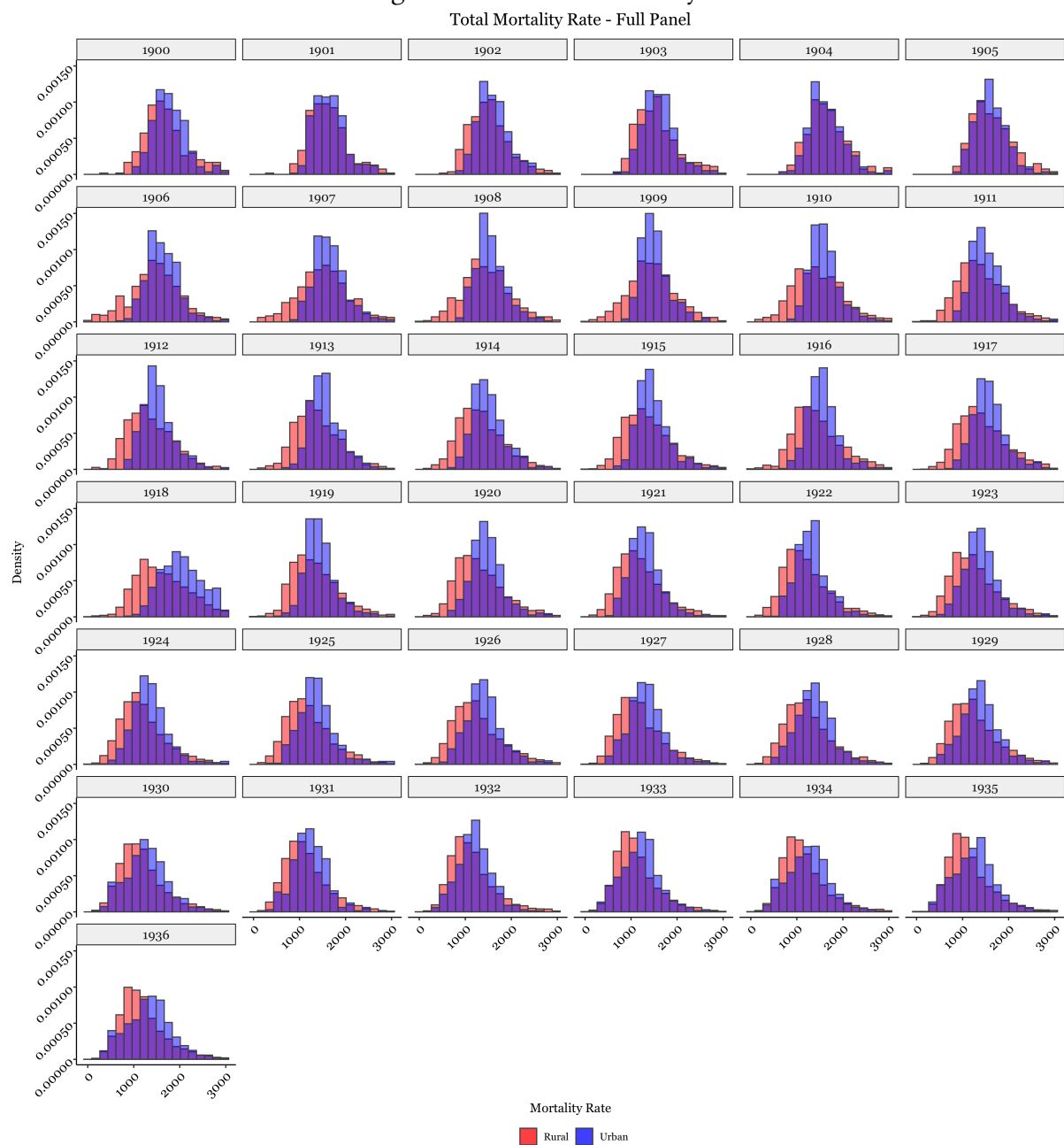


NOTES: Results are shown for separate regressions estimating the effect of Post-year x Big City. Each regression includes year and city fixed effects. Standard errors are clustered at the city level. Lines represent 95% confidence intervals.

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county.

## B Histograms Illustrating the Transition

Figure B.1: Overall Mortality

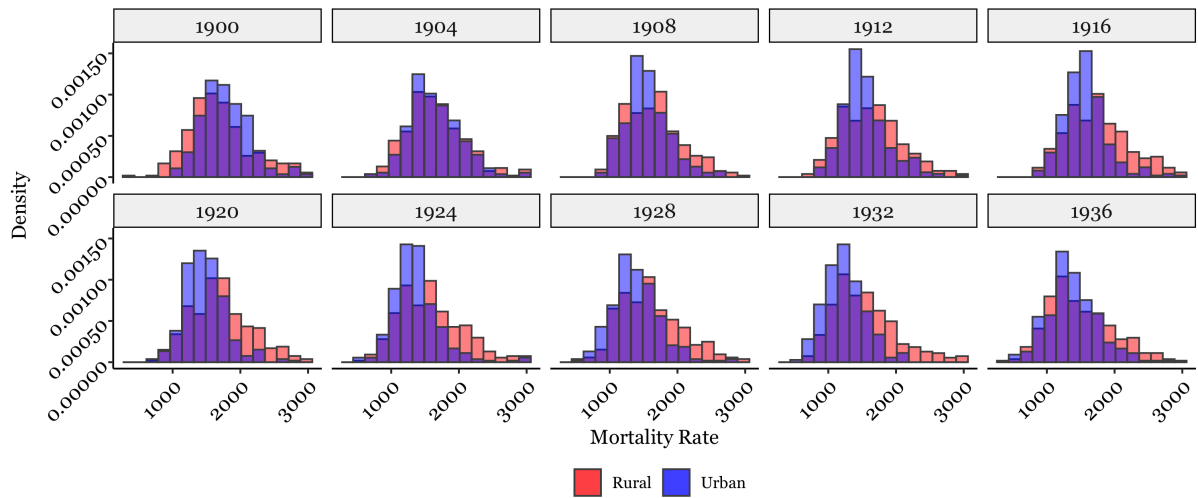


NOTES: Results are shown for separate regression estimating the effect of Post-year x Big City. Each regression includes year and city fixed effects. Standard errors are clustered at the city level. Lines represent 95% confidence intervals.

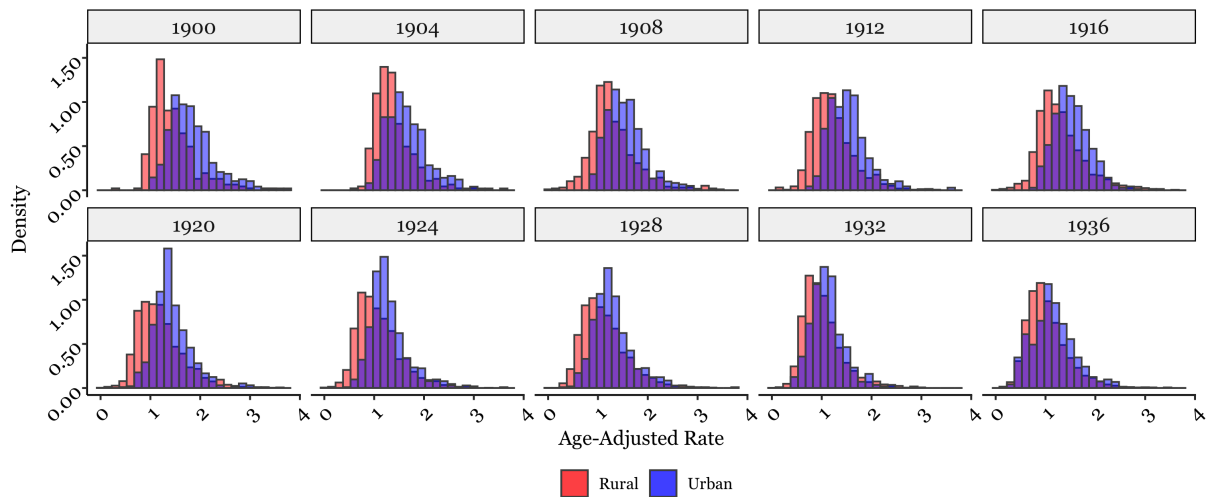
SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county. County-level population counts and demographic characteristics are calculated from the IPUMs Restricted Complete Count Census data.



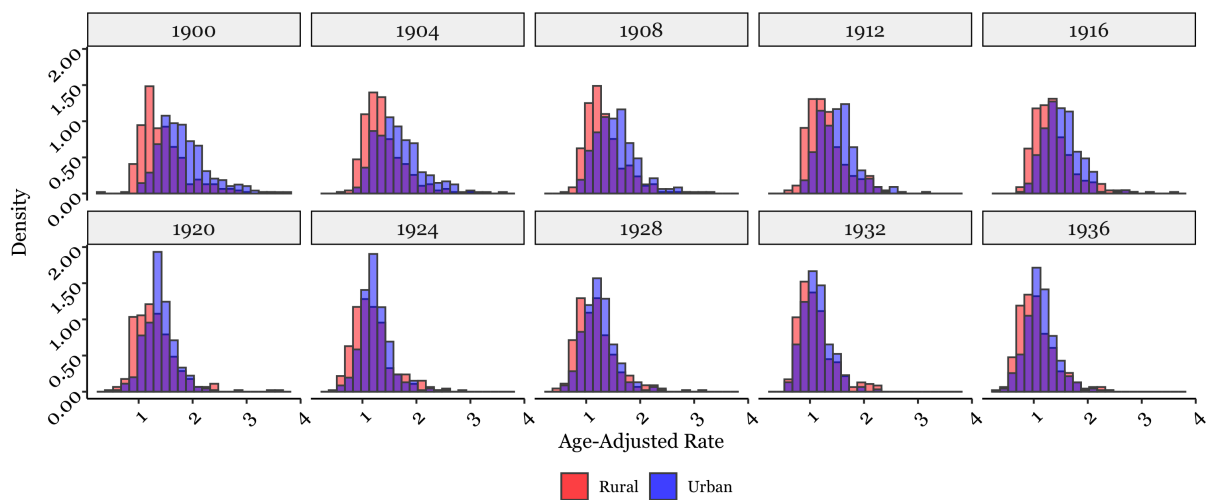
Figure B.2: Overall Mortality - Balanced and Age-Standardized  
Total Mortality Rate - Balanced Panel



Age-Standardized Total Rate - Full Panel



Age-Standardized Total Rate - Balanced Panel

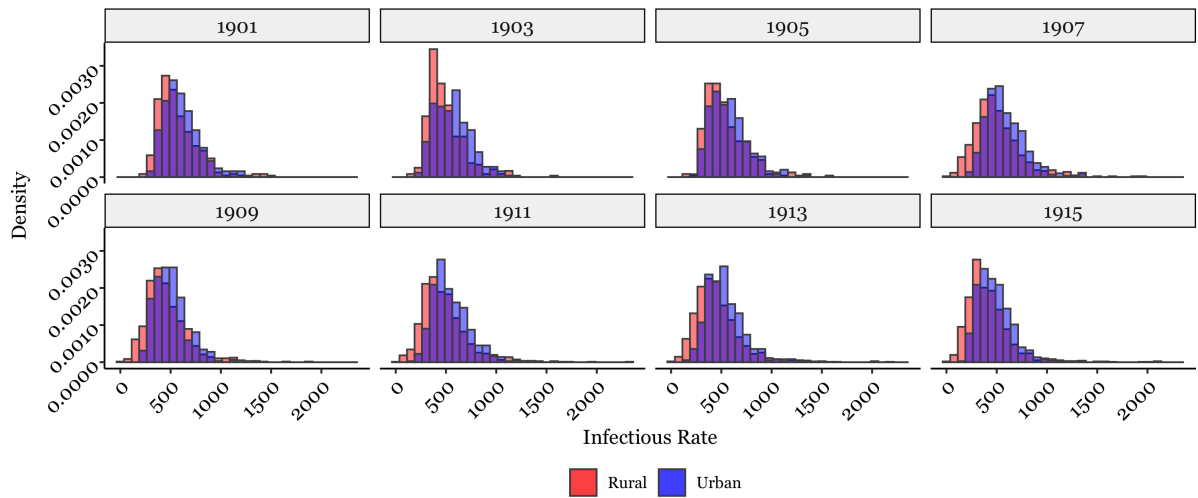


NOTES: Mortality rates are reported per 100,000 persons. The blue indicates urban areas, and the red shows rural areas.

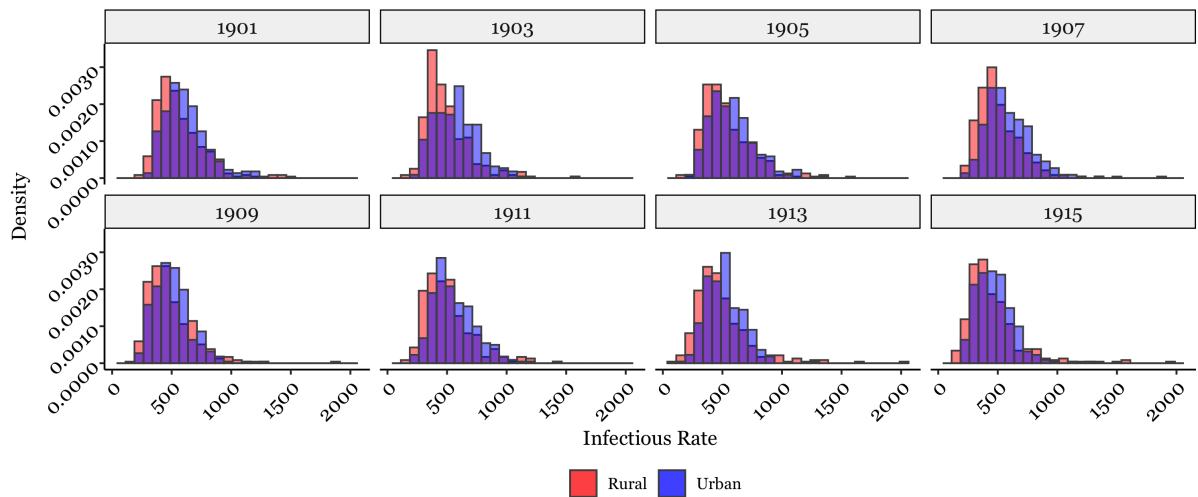
SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county. County-level population counts and demographic characteristics are calculated from the IPUMs Restricted Complete Count Census data.



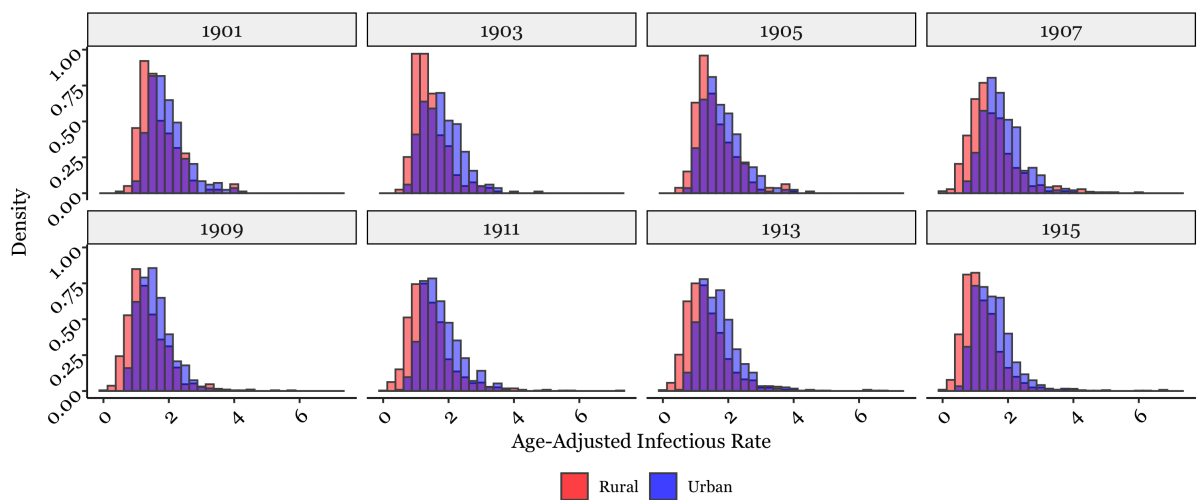
Figure B.3: Infectious Disease Mortality  
Infectious Rate - Full Panel



Infectious Rate - Balanced Panel



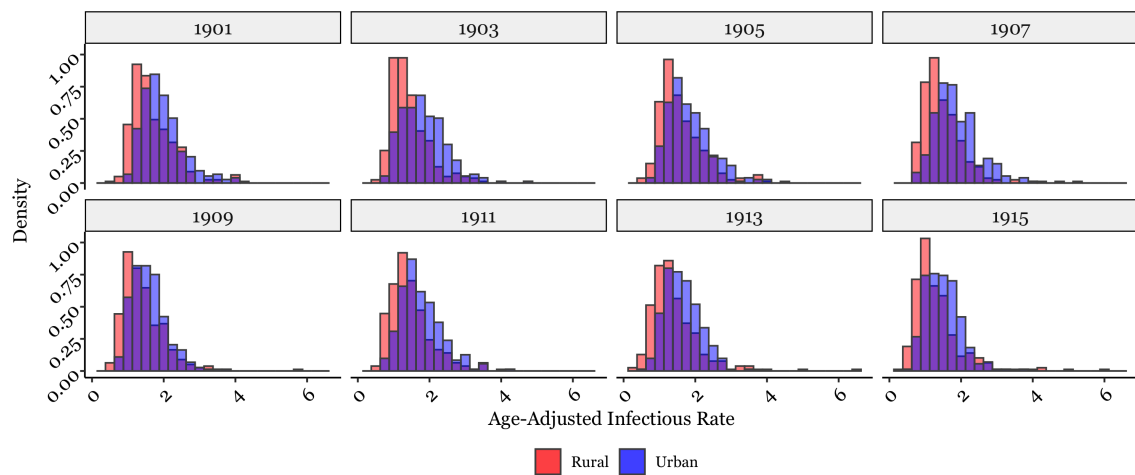
Age-Standardized Infectious Rate - Full Panel



NOTES: Mortality rates are reported per 100,000 persons. The blue indicates urban areas, and the red shows rural areas.

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county. County-level population counts and demographic characteristics are calculated from the IPUMs Restricted Complete Count Census data.

Figure B.4: Balanced Panel of Age-Standardized Mortality  
Age-Standardized Infectious Rate - Balanced Panel

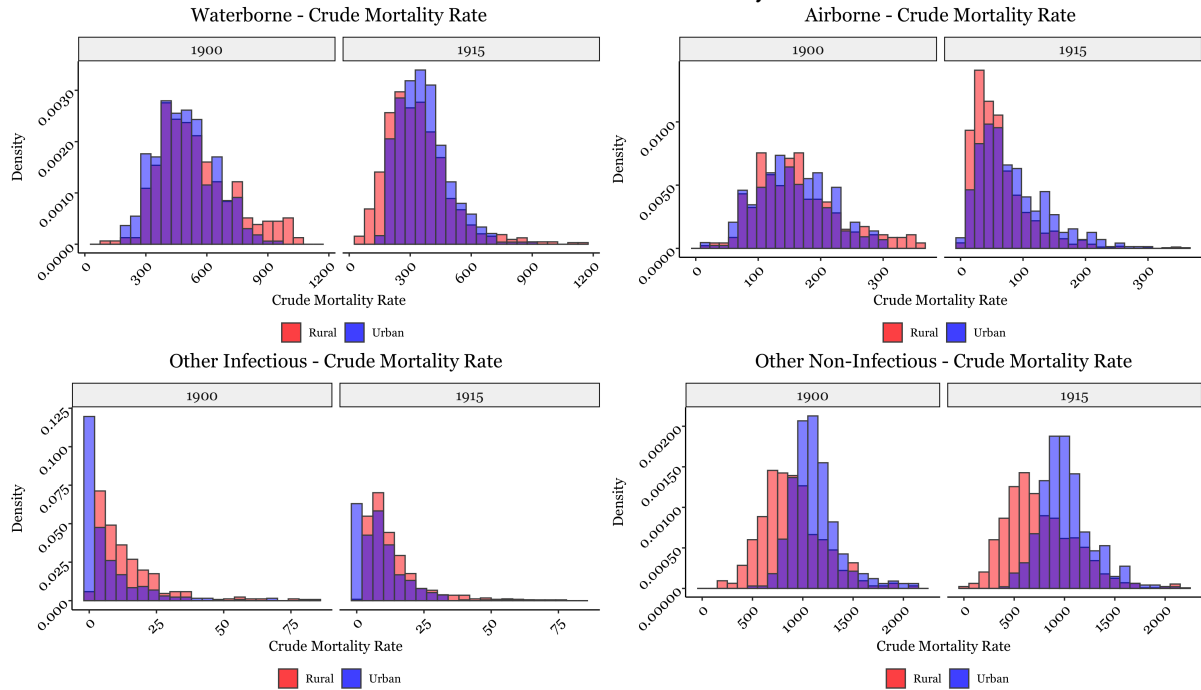


NOTES: Mortality rates are reported per 100,000 persons. The blue indicates urban areas, and the red shows rural areas.

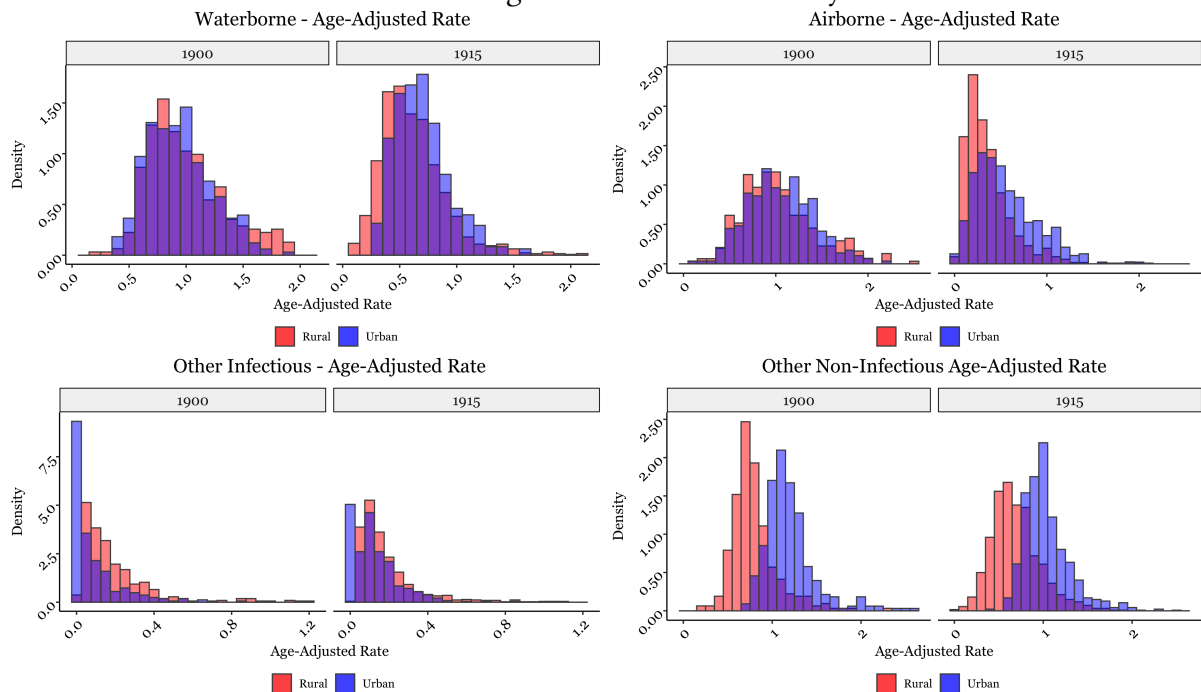
SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county. County-level population counts and demographic characteristics are calculated from the IPUMs Restricted Complete Count Census data.

Figure B.5: Mortality by Cause of Death

Panel A: Crude Mortality



Panel B: Age Standardized Mortality



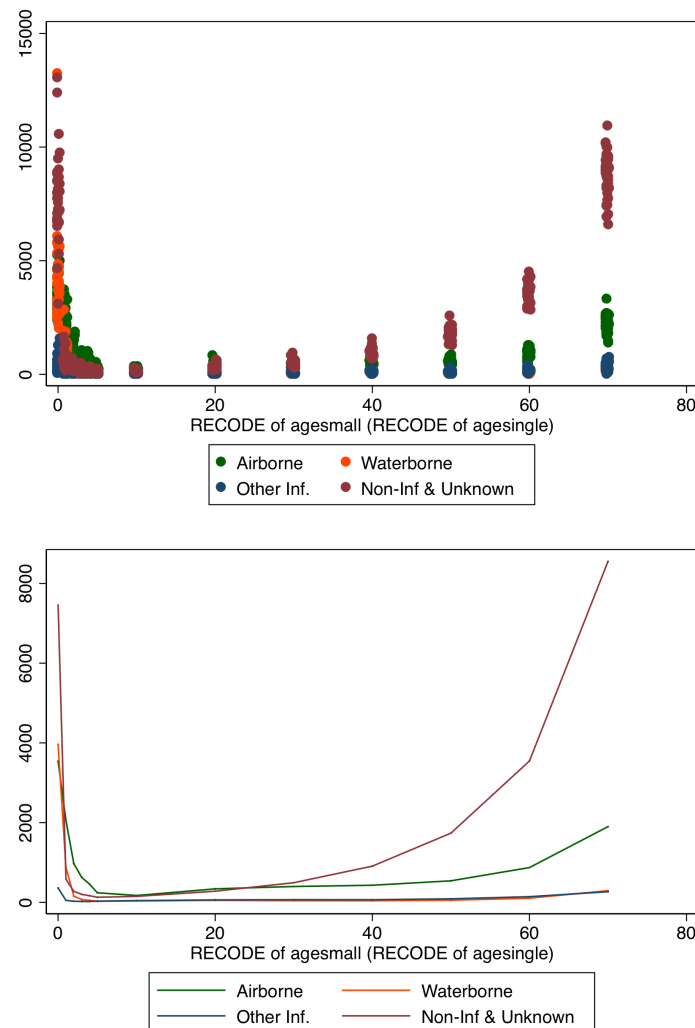
NOTES: Mortality rates are reported per 100,000 persons. The blue indicates urban areas, and the red shows rural areas.

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county. County-level population counts and demographic characteristics are calculated from the IPUMs Restricted Complete Count Census data.

## C Age schedules used to Standardize Mortality

Due to the fact that the age pattern of mortality changes during the 1900-1936 period, our choice of the age schedule used for standardizing may matter for the findings. We currently test three separate age schedules, based in three years. These mortality schedules include (1) the total mortality uses a 1935-based national schedule, (2) infectious mortality uses a 1925-based urban schedule, and (3) the cause-specific (e.g., airborne, waterborne) use a 1907-based urban schedules. These are shown in Figure C.1

Figure C.1: 1907 Age Schedule for Mortality Adjustments



NOTES: Mortality rates are reported per 100,000 persons.

SOURCES: Vital statistics are from the U.S. Vital Statistics for cities and rural portions of each county. County-level population counts and demographic characteristics are calculated from the IPUMs Restricted Complete Count Census data.