

Data Sources and Analysis for “How Deadly Were Gotham's Tenements? Infectious Disease in the 19th Century (Part I)”

Troy Tassier and Jason M. Barr

November 5, 2023

Data Sources

Data was obtained from the NBER Union Army Data - Historical Urban Ecological Data at <https://www.nber.org/research/data/union-army-data-hue>. For post Part I, for each year in each ward, we collected data from 1868 to 1910 that contained deaths from contagious and other types of disease that were collected annually from the New York City Department of Health. The files were generally entitled, “Deaths from Zymotic & Certain Other Preventable Diseases by Wards.”

Death by Disease Count: To analyze contagious disease deaths we included total deaths from the following categories: Cholera, Cerebro-Spinal Meningitis, Diphtheria & Croup, Typhoid or Typhus Fever, Malarial Fevers, Measles, Scarlet Fever, Small-Pox, Whooping Cough, Diarrheal Diseases, Tuberculous, and Pneumonia (over the years there were slight variations in categories, so we included counts of all contagious diseases for each year). We then added up all these deaths to get a total *deaths from disease*.

Population: For each ward, population counts were given in the Dept. of Health files and were collected from state or federal censuses. For intermediate years, we interpolated the population by first estimating the average growth rate via: $r_t = \frac{\ln Pop_t - \ln Pop_{t-n}}{n}$, where n is number of years between two censuses. Next for inter-census years, we estimated population values via:

$$Pop_t = Pop_{t-1}e^{r_t}$$

Descriptive Stats

Here are the ward-level descriptive stats for the per capita deaths by infectious disease from 1869 to 1910. In Equation (2) below, we removed the largest three, which were in around the 99.5th percentile.

Percentiles		Smallest		
1%	.00243	.0007042		
5%	.0033766	.0019472		
10%	.0038997	.0020411	Obs	659
25%	.0055212	.0021436	Sum of wgt.	659
50%	.0077923		Mean	.0096488
		Largest	Std. dev.	.0057459
75%	.0130044	.0307152		
90%	.0169765	.0315972	Variance	.000033
95%	.02155	.0342286	Skewness	1.352061
99%	.0297562	.0370832	Kurtosis	5.078971

Regressions

To compare death rates across wards, we first computed the death rate for each ward by calculating $DR = \text{deaths from diseases}/\text{population}$. We then regressed DR on ward and year dummies for the years we had data between 1866 and 1910. The omitted ward was 20, which had a death rate very close to the city average. Note we excluded the two largest outlier values from the regression, given that the outliers seemed like possible errors (results in Column (2)). In the table, we can see that the ward dummy coefficients with and without the outliers are very close. Only Ward 12—the largest northern most ward—changes in any fashion.

For the map in Post I, we set coeff. values ≤ -0.00115 as less than the average, >0.0055 above the average, and between $>-0.00115 - \leq 0.0055$ as about average

Table I: Regression Results

	(1)	(2)
ward1	0.00542*** (7.11)	0.00500*** (6.04)
ward2	-0.00173** (-2.24)	-0.00172** (-2.28)
ward3	-0.000813 (-1.07)	-0.000813 (-1.14)
ward4	0.00370*** (4.86)	0.00370*** (4.77)
ward5	0.00162** (2.12)	0.00162*** (4.22)
ward6	0.00232*** (3.04)	0.00232*** (6.64)
ward7	-0.00147* (-1.93)	-0.00147*** (-5.36)
ward8	0.00116 (1.52)	0.00116*** (2.75)
ward9	-0.000664 (-0.87)	-0.000664** (-2.23)
ward10	-0.00180** (-2.37)	-0.00180*** (-5.42)
ward11	-0.00234*** (-3.07)	-0.00234*** (-5.44)
ward12	0.000949 (1.25)	-0.0000462 (-0.05)
ward13	-0.00203*** (-2.66)	-0.00203*** (-5.16)
ward14	0.00254*** (3.34)	0.00254*** (3.97)

ward15	-0.00167** (-2.19)	-0.00167** (-2.41)
ward16	-0.00172** (-2.25)	-0.00172*** (-6.80)
ward17	-0.00176** (-2.31)	-0.00176*** (-5.79)
ward18	-0.000761 (-1.00)	-0.000761** (-2.38)
ward19	0.000545 (0.72)	0.000545 (0.92)
ward21	-0.000495 (-0.65)	-0.000495** (-2.11)
ward22	-0.00115 (-1.51)	-0.00115*** (-3.52)
_cons	0.00965*** (17.92)	0.00965*** (53.36)

N	659	657
Year Dummies	YES	YES
R-sq	0.756	0.788
adj. R-sq	0.736	0.770
AIC	-5817.6	-5930.4
BIC	-5718.8	-5831.6

t statistics in parentheses. * p<0.10, ** p<0.05, *** p<0.01. Dep. Var.: Contagious disease death per capita. Year dummy coefficients not shown. Column (1) is full sample; Column (2) has three outliers removed.