Data Sources and Analysis for "How Deadly Were Gotham's Tenements? Infectious Disease in the 19th Century" Parts I and II

Troy Tassier and Jason M. Barr

November 27, 2023

I. Data Sources for Paris I and II

Data on mortality was obtained from the NBER Union Army Data - Historical Urban Ecological (HUE) 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, 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 and Density: 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: $r_t = \frac{lnPop_t - lnPop_{t-n}}{n}$, where n is number of years between two censuses. Next for inter-census years, we estimated population values via:

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

Land areas for each ward are given the HUE files each year. We used land area ca. 1900. Density was calculated as population per acre.

Furthermore, several years included personal property and real property values by the ward. Earlier years tended to contain personal property values while later years contained real property. For the analysis below we include the latest years for which data are available in all wards (1856 personal property, 1893 real property).

Historical Ecology: Data was collected from the Welikia Mannahatta Project maps, which estimate the location of various ecosystems, including the presence of wetlands, rivers and streams, lakes, and main tree types. The data was geocoded in the following way. If a Manhattan block contained one of these features, the feature was given a 1 for the block, and 0 otherwise. For example, if there were oak tulip trees the block the Oak Tulip variable for the block was assigned a one. Then for each block, we added up the number of blocks that had each feature and divided by the total number of blocks to get the percentage with the feature. The variables of

interest here were the percentage of blocks with a ward that had an oak tulip tree, and the percentage that were below sea level (indicating the presence of wetlands).

Demographics: Using data from the 1890 census we calculated for each ward the percentage of Russian or Polich people living in each ward from *Vital Statistics of New York City and Brooklyn (1894)*.

Data on permits for water/ sewer and ventilation collected from HUE data set summaries of New York City Department of Health reports. Data on rear tenements and total tenements taken from HUE data set tenement house census reports.

Tenement Population Data: For 1870 and 1900 are part of the HUE data above in their respective years.

II. Part I Data and Results

1. 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

2. Regressions

To compare death rates across wards, we first computed the death rate for each ward by calculating *DR = deaths from diseases/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 $\leq$*-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***	0.00500***
	(7.11)	(6.04)
ward2	-0.00173**	-0.00172**
	(-2.24)	(-2.28)
ward3	-0.000813	-0.000813
	(-1.07)	(-1.14)
ward4	0.00370***	0.00370***
	(4.86)	(4.77)
ward5	0.00162**	0.00162***
	(2.12)	(4.22)
ward6	0.00232***	0.00232***
	(3.04)	(6.64)
ward7	-0.00147*	-0.00147***
	(-1.93)	(-5.36)
ward8	0.00116	0.00116***
	(1.52)	(2.75)
ward9	-0.000664	-0.000664**
	(-0.87)	(-2.23)
ward10	-0.00180**	-0.00180***
	(-2.37)	(-5.42)
ward11	-0.00234***	-0.00234***
	(-3.07)	(-5.44)
ward12	0.000949	-0.0000462
	(1.25)	(-0.05)
ward13	-0.00203***	-0.00203***
	(-2.66)	(-5.16)
ward14	0.00254***	0.00254***
	(3.34)	(3.97)
ward15	-0.00167**	-0.00167**
	(-2.19)	(-2.41)
ward16	-0.00172**	-0.00172***
	(-2.25)	(-6.80)
ward17	-0.00176**	-0.00176***
	(-2.31)	(-5.79)
ward18	-0.000761	-0.000761**
	(-1.00)	(-2.38)
ward19	0.000545	0.000545
	(0.72)	(0.92)
ward21	-0.000495	-0.000495**

	(-0.65)	(-2.11)
ward22	-0.00115	-0.00115***
	(-1.51)	(-3.52)
_cons	0.00965***	0.00965***
	(17.92)	(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.

III. Part II Results

1. Anecdotal Evidence on Risk Differences Across Wards

i. The History of Paradise Square: The Fall of Paradise

We can get a better sense of what the older tenements were like by considering the early history of one of New York City's most notorious tenement neighborhoods, *Five Points,* which sat in the 6th Ward of Manhattan. Much of this neighborhood sat upon or in the vicinity of what was once Collect Pond. Collect Pond was a 48-acre body of fresh water fed by an underground spring. Until the early 1800s, it was a primary source of drinking water for lower Manhattan.

However, it came to be surrounded by slaughterhouses, tanneries, breweries, and other industries that required water or a place to dump debris and refuse. Over time the pond became polluted. Eventually, the stench and contamination led to the pond being backfilled and drained by a set of canals that led to the Hudson River. (The main canal is the present-day location of Canal Street in Manhattan.)

After the pond was filled, real estate developers took advantage of the new land atop Collect Pond to build a residential neighborhood which they named *Paradise Square*. Originally Paradise Square attracted the wealthy who were steadily settling farther and farther north on the island. However, the presence of the underground spring and the marshy land surrounding it led to the neighborhood being anything but a paradise. Continually, the soil and streets were soggy and wet.

Many of the newly built apartment dwellings and homes began to sink and to settle unevenly. The affluent people for whom Paradise was built began to leave as the neighborhood fell into disrepair. The owners then began to subdivide the now undesirable buildings, as well as the old factories and tanneries, into small rooms that were lent primarily to recent immigrants seeking the most inexpensive housing. Because many of the newly partitioned apartments and rooms were made by subdividing larger buildings, they often lacked ventilation and exterior sources of light. In addition, most lacked connections to sewers and some did not even have access to

outhouses. In this case, human waste was simply discarded into alleys or holes were dug into the soggy soil. It was within this unsanitary and unventilated building stock that many lived in the Five Points/ Sixth Ward tenements. The original tenement laws were written to improve these abhorrent living conditions.

ii. Report of the Tenement House Department

Beginning with the *Report of The Tenement House Department of the City of New York 1902/1903* areas of New York City were divided into *tenement districts*. All of Manhattan contained 11 tenement districts. Two of these are of interest to the present question. The area south of 14th Street contained two districts, east and west. Unfortunately, these districts did not conform precisely to the ward boundaries of the city so we cannot match the data exactly to information about wards. But the report does provide information on the number of new tenements built throughout the city.

Between January 1902 and July 1903, the area south of 14th Street on the east side of the city erected 260 "new law" tenement buildings while the west side erected only 57. Further, the total number of new tenements on the Lower East Side represented 45 percent of all new tenements in the entire city. This is a huge fraction of such a tiny portion of the city's land area. Additionally, one can see from a map provided within the report that most of the 260 new tenements were built in the safer wards. Subsequent years of the Tenement House Department Reports also list much larger shares of tenement construction on the east side of lower Manhattan than on the west.

2. Descriptive Statistics

Risky Wards	Infectious Disease Deaths relative to City AVG	1870 % Tenement	1900 % Tenement	1900 % Rear Tenements	1868-1910 PopGrowth	1900 Density	1900 Irish+Italalian % Tenement	1900 Eastern European % Tenement	1860 Avg Personal Property	1893 Avg Real Property	OakTulips
1	52%	57.79	0.93814485	4.63	57%	54.75	61%	20%	3738.662292	8597.308259	20%
4	39%	51.39	0.90364075	11.04	97%	234.74	78%	9%	82.21992362	858.9126854	68%
6	24%	80.97	0.85960845	23.88	74%	197.86	68%	10%	55.63447708	1222.723734	31%
14	26%	72.28	0.76973992	23.05	136%	315.14	95%	3%	92.82115385	907.4860167	80%
Mean		65.6075	0.86778349	15.64999935	91%	200.62	75%	11%	992.3344616	2896.607674	50%

Safer Wards	Infectious Disease Deaths relative to City AVG	1870 % Tenement	1900 % Tenement	1900 % Rear Tenements	1868-1910 PopGrowth	1900 Density	1900 Irish+Italalian % Tenement	1900 Eastern European % Tenement	1860 Avg Personal Property	1893 Avg Real Property	OakTulips
7	-16%	50.01	0.83286496	4.93	255%	433.19	20%	75%	97.43549597	361.1170717	61%
10	-19%	55.59	0.81006201	12.04	229%	659.44	5%	85%	37.43097504	368.9858451	100%
11	-24%	88.6	0.7882364	8.96	229%	465.46	7%	88%	10.23392255	267.609214	43%
13	-21%	54.35	0.77156595	10.15	196%	588.23	7%	88%	16.24084819	290.2698512	53%
17	-19%	75.82	0.7149822	9.32	236%	491.71	14%	74%	91.47174208	390.2661509	91%
Mean		64.874	0.7835423	9.081036119	229%	527.61	10%	82%	50.56259677	335.6496266	70%
City Mean		46.9454545	0.75849231	7.408181155	304%	226.78	44%	35%	428.2309243	3631.142091	67%
City Median		50.7	0.78588417	5.874136188	126%	165.07	45%	23%	82.28214394	947.716021	69%

This table shows that density, tenement percentage, and population growth, and wealth were unlikely to account for the infectious disease risk difference across wards. Lack of rear tenements, new construction, and access to running water and sewage likely explained a large fraction of the difference. The wards with Eastern Europeans likely had newer building stocks and those with higher relative social mobility than those who remained in Five Points, for example.

3. Regression Results

Here, we use the 22 ward fixed effects from the Regression Table I above, using Equation (2) (i.e., without extreme outliers). We regress the fixed effects on several ward characteristics, including the presence of oak tulip trees (which indicate land with good drainage), percentage of blocks below sea level, log of density, percentage of residents from Eastern Europe, percentage of rear tenements, log of the land area of the ward, the log of number of structures that applied for plumbing/ sanitation permits and the log of number of structures that applied for additional light or ventilation permits.

As the table below suggests, more rear tenements were associated with a higher death rate, while more access to plumbing was associated with a lower death rate. Oak Tulip trees and the Eastern European population appear to have had lower death rates, though statistical significance varies.

Given that there are only 22 observations, the results can best be interpreted as suggestive, rather than causal.

Table II: Regression Results

	(1)	(2)	(3)	(4)	(5)
Russia_pct	-0.0000347	-0.0000133	-0.0000176	0.00000289	-0.00000437
	(-1.13)	(-0.42)	(-0.65)	(0.11)	(-0.18)
lnDensity	-0.000255	0.000186			
	(-0.33)	(0.21)			
Oaktulip_pct	-0.0000334		-0.0000321		-0.0000287
	(-1.51)		(-1.66)		(-1.62)
Rear_pct	0.000139	0.000177*	0.000154**	0.000194**	0.000162**
	(1.43)	(1.83)	(2.20)	(2.77)	(2.63)
lnArea	0.0000678	0.00121	0.00108	0.00170**	0.00159*
	(0.14)	(1.23)	(1.22)	(2.14)	(2.07)
belowsea_pct		0.0000502		0.0000422	
		(1.25)		(1.29)	
lnLightVent		-0.000795	-0.000753		
		(-1.18)	(-1.39)		
lnPlumb_drain				-0.00108**	-0.00107**
				(-2.33)	(-2.39)
_cons	0.00168	-0.0203	-0.0145	-0.0270**	-0.0225*
	(0.23)	(-1.22)	(-1.06)	(-2.16)	(-1.88)
N	22	22	22	22	22
R-sq	0.402	0.432	0.462	0.531	0.556
adj. R-sq	0.215	0.204	0.294	0.384	0.417
AIC	-211.1	-210.3	-213.5	-216.5	-217.7
BIC	-204.6	-202.6	-206.9	-209.9	-211.1

Dep. Var.: Ward infectious disease mortality fixed effect from Table 1, eq. (2).
t statistics in parentheses * p<0.10, ** p<0.05, *** p<0.01