A Spatial Assessment of Lead Poisoning in Chicago
Jada Potter
3/4/2024

Introduction

Chicago has the largest number of lead service lines out of any city in the United States, which poses a major health hazard, particularly to young children, when these pipes corrode and contaminate drinking water with lead. Despite efforts to replace lead service lines through Chicago's Lead Service Line Replacement (LSLR) program, at current rates it will likely take 40 years or more for the city to be completely lead-free.[1]

The presence of lead in water can not only have devastating health effects, however; scholars have posited that lead may be partly to blame for elevated crime rates in the areas with the greatest levels of contamination, as childhood lead poisoning may impair the judgment of young adults (though the effect is often overstated).[2]

This paper will therefore assess the spatial distributions of early childhood lead poisoning in the early 2000s, searching for clusters of lead screening and testing in order to determine whether certain regions of the city have been disproportionately affected by lead contamination. Furthermore, it will assess the effect of the distribution of lead poisoning in that era on crime in the early 2020s, leaving a gap of 20 years in order to account for the fact that homicides tend to be committed by adults in their early 20s most frequently.[3]

This paper will thus address the three following questions:

In the early 2000s, were there spatial clusters of areas with high levels of lead blood screening, or of areas with high rates of elevated blood lead levels in children? Do these clusters overlap, or is there a significant difference in where they are found? Finally, does the spatial distribution of lead poisoning throughout the early 2000s have a strong latent effect (both spatially and temporally) on the distribution of violent crime 20 years later?

Null hypotheses

This paper will respond to each of these three questions with the following hypotheses:

1) There is no clustering of areas in which high rates of blood lead level screening took place, or of high rates of lead poisoning in children. In other words, the rates in one area tell nothing about the rates in another area.
2) There is no association between the locations of lead screening clusters and the locations of lead poisoning clusters. Areas within high-high or low-low clusters of screening are no more or less likely to also be within high-high or low-low clusters of lead poisoning.

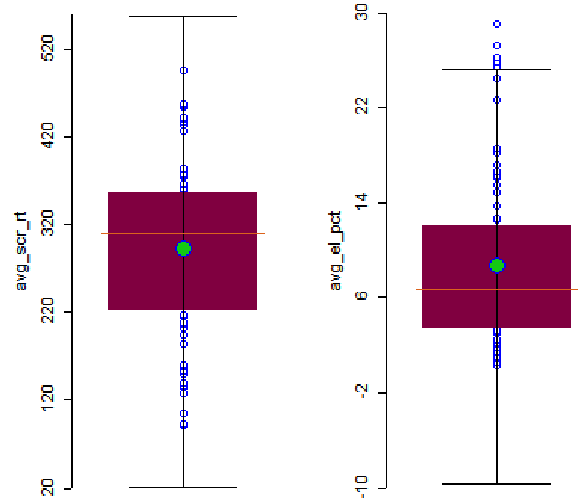
---

[1] https://blockclubchicago.org/2023/12/01/new-epa-proposal-to-eliminate-lead-pipes-nationwide-will-take-longer-to-implement-in-chicago/
[2] https://doi.org/10.1016/j.regsciurbeco.2022.103826
[3] https://www.statista.com/statistics/251884/murder-offenders-in-the-us-by-age/

3) There is no latent spatial effect of childhood lead poisoning in the early 2000s on the distribution of violent crime in the early 2020s.

Data

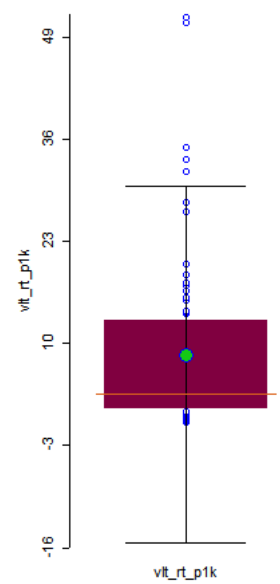
To address the first two hypotheses, this paper used historical data from the Chicago Data Portal on rates of lead screening and poisoning among children ages 0–6 by Community Area between 1999 and 2013,[4] which were joined to the spatial locations of each Community Area.[5] Though this dataset contains a large number of variables spanning over a long time period, only the "Lead Screening Rate" and "Percent Elevated" variables were used, and only for years between 1999 and 2003 (inclusive) in order to calculate a five-year average for the early 2000s. Summaries of these five year averages are shown to the right, with "avg_scr_rt" corresponding to the lead screening rate per 1,000 children ages 0–6 from 1999–2003, and "avg_el_pct" corresponding to the percentage of those children who were found to have an elevated level of lead in their blood (10 µg/dL). Each unit is a Community Area.



	avg_scr_rt		avg_el_pct
min	92.2000	min	0.3000
max	495.7400	max	29.0400
Q1	221.6000	Q1	3.3600
median	308.9600	median	6.6400
Q3	355.8000	Q3	12.0800
IQR	134.2000	IQR	8.7200
mean	293.3714	mean	8.7384
s.d.	97.3888	s.d.	7.2500

From the data summary, it is quite alarming to notice how high rates of lead poisoning were during this time period. However, it should be noted that these rates do not reflect current levels, which have greatly declined over the decades since.[6]

Finally, to address the third hypothesis, this paper used homicide and nonfatal shooting point data ranging from 1991 to the present, also from the Chicago Data Portal.[7] These data were filtered to only include entries from between 2019 and 2023 in order to correspond to the period 20 years after the lead poisoning data were collected. Once filtered, these data consisted of 17,917 points. They were then spatially joined to community areas and divided by population (2018–2022 five-year estimates[8]) to calculate a rate of homicides and nonfatal shootings per 1,000 population. A summary of this rate among Community Areas is shown to the right.



	vlt_rt_p1k
min	0
max	51.6270
Q1	1.6542
median	3.4857
Q3	13.0217
IQR	11.3674
mean	8.5391
s.d.	10.9867

One potential limitation of the use of this dataset is that the individuals involved in these incidents may not have actually lived in the same areas 20 years prior. However,

---

4 https://data.cityofchicago.org/Health-Human-Services/Public-Health-Statistics-Screening-for-elev about_data

5 https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Community-Areas-

6 https://chicagohealthatlas.org/indicators/LDPPH?tab=chart

7 https://data.cityofchicago.org/Public-Safety/Violence-Reduction-Victims-of-Homicides-and-Non-Fa/gumc-mgzr/about_data

8 https://chicagohealthatlas.org/indicators/POP?topic=population

given that there are no longitudinal data available on the children who were tested, a spatial approximation is likely the best mode of analysis possible. Furthermore, by accounting for spatial lag in the third hypothesis, a degree of local movement of population is accounted for.

Methods

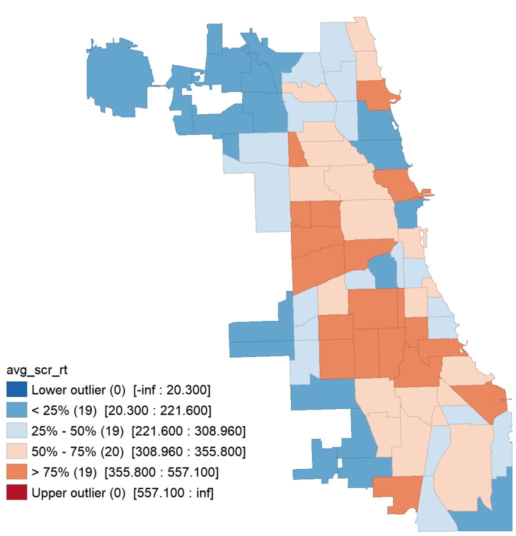
This analysis was conducted in GeoDa, using local and global spatial autocorrelation tools and cross-tabulation. Though additional exploratory methods were used, these are outlined in detail in the following section. The main methods used to address each hypothesis were the following:

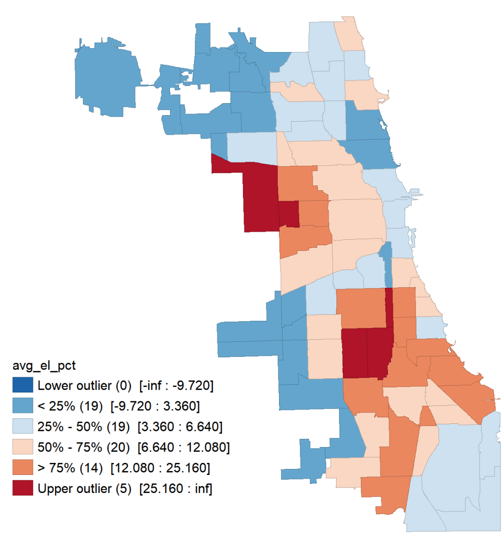
- **Hypothesis 1:** A global Moran's I test of lead screening and another of lead poisoning were used to address the hypothesis directly, while local Moran's I tests were used to delve deeper into the spatial distribution of any clusters that were found. The Moran's I test statistic was chosen due to its versatility (as it has both local and global variants) and ability to distinguish between types of positive and negative autocorrelation.
- **Hypothesis 2:** A cross-tabulation was used to determine the prevalence of overlap between clusters of lead screening and lead poisoning. A Chi-squared test of independence was then used to assess whether any overlap found was significant, which would allow Hypothesis 2 to be rejected. Furthermore, areas that deviate from the pattern found could also be of interest for further study, especially if they indicate a mismatch between screening and poisoning rates not found elsewhere in the city.
- **Hypothesis 3:** A bivariate Moran's I test was conducted in order to determine whether there was a strong spatially lagged effect of lead poisoning on rates of violent crime in the surrounding Community Areas.

Analysis, Results, and Discussion

Hypothesis 1

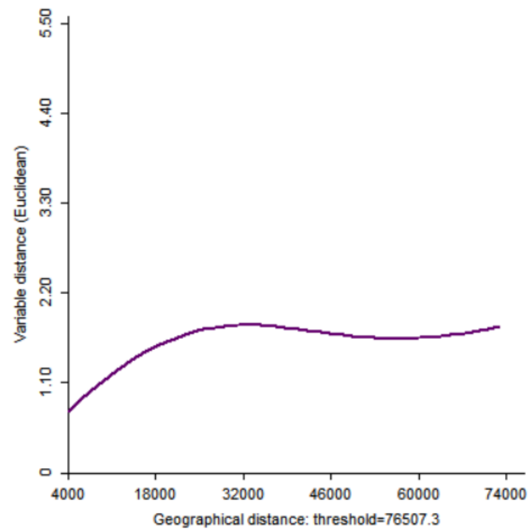
Exploratory data analysis: First, to gain a better understanding of the general distributions of lead poisoning and screening data, two box maps (hinge = 1.5) of these variables are shown below:



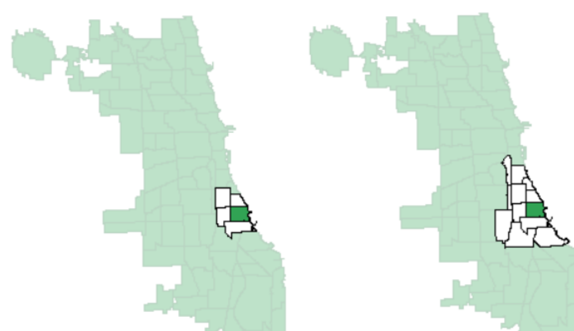


As the second map shows, lead poisoning is particularly prevalent in areas on the South and West Sides of Chicago. While these areas generally have higher rates of screening as well, the differences in screening rates between Areas are less extreme, as there are no outliers. This indicates that lead poisoning is potentially more clustered than lead screening, which may be an indication that screening is not distributed equitably according to need, a question which will be addressed further below.

Next, the impact of different spatial weights was assessed. This included 1st and 2nd order queen contiguity (Q1 and Q2), 5-nearest neighbors (5nn; this is approximately the median number of neighbors for Q1), and 25,000 ft distance band weights (25000d; roughly the shortest distance possible to avoid isolates). Issues with the 5nn weight were immediately apparent; it led to asymmetrical assignment and fairly extreme edge effects, assigning areas like Mount Greenwood and O'Hare neighbors much farther away than the neighbors of any other area simply due to their locations on the extremities of the map. It would not make sense for the spatial heterogeneity of lead poisoning and screening to follow such a pattern. The 25000d distance band weight and the two contiguity weights did not appear to create as many unreasonable artifacts. Because the Q2 and 25000d weights produced similar results (assessed through their histograms of neighbor counts and connectivity maps), only Q1 and Q2 were used for simplicity.



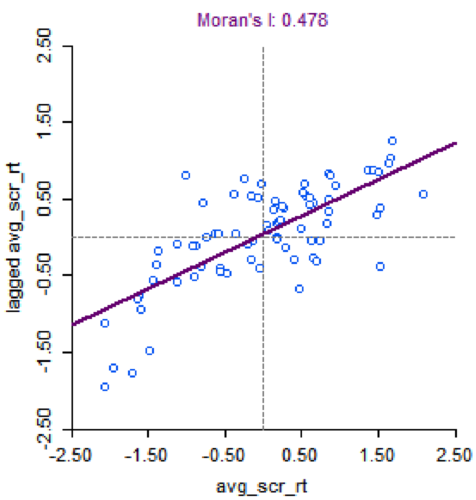
As the smoothed distance scatter plot on the right demonstrates, 25,000 ft is roughly around the distance past which geographic distance ceases to have a meaningful effect on variable distance (this accounts for both lead poisoning and screening variables); in other words, any weights with a larger distance band would be meaningless. Because Q2 was roughly equivalent to the 25000d weight, while the Q1 weight covered a smaller area, both will be included in this analysis as they are both reasonably within the bounds of the range of interaction.

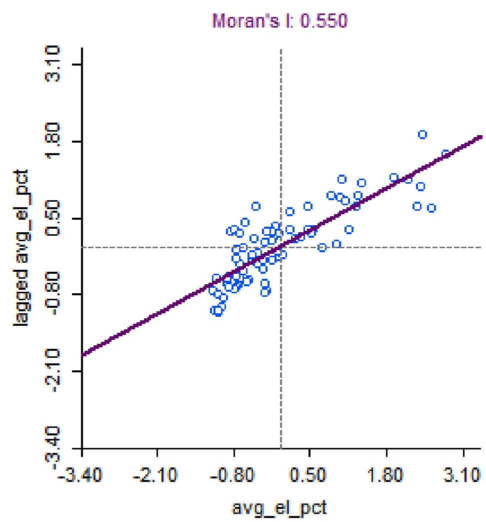


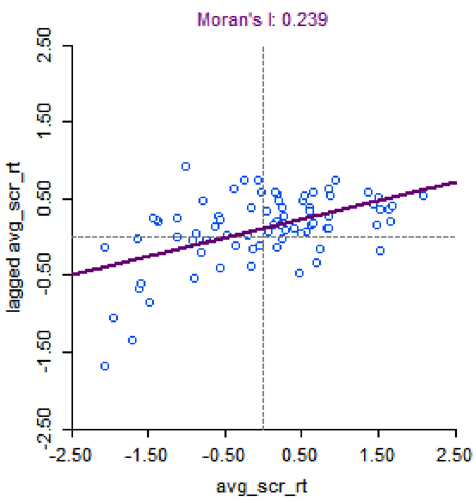
As an example, the connectivity maps to the right show the neighbors assigned to Hyde Park using Q1 and Q2 weights. Q1, shown on the left, only includes immediate neighbors and demonstrates more local neighborhood effects, while Q2, shown on the right, demonstrates the interactions between larger groups of community areas; in the case of Hyde Park, this larger group could be thought of as the "Near Southeast Side."

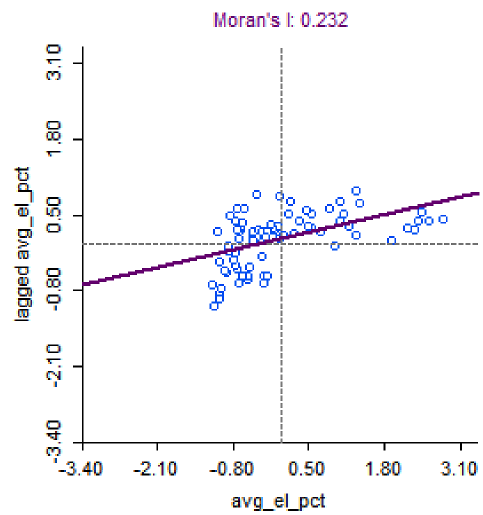
Assessing global spatial autocorrelation: The table below shows the Moran's I coefficients, z-scores, and pseudo p-values for both variables and for both spatial weights. To directly address Hypothesis 1, it is indeed clear that there is a global pattern of clustering for both variables regardless of spatial weight, given the high z-scores and low pseudo p-values of each iteration. Though clustering is slightly more apparent with a Q1 weight than a Q2 weight, it is still highly apparent in both cases.

As observed above, clustering of lead poisoning is more prevalent than clustering of lead testing, but this effect is greatly diminished when the reach of the spatial weight increases; in fact, the Moran's I values and z-scores of the two variables are nearly identical if a Q2 weight is used. This indicates that within the larger context of general areas of the city, lead screening and lead poisoning are clustered to roughly the same degree.

<table>
<tr><th>Weight</th><th>Lead Screening</th><th>Lead Poisoning</th></tr>
<tr><td>Q1</td><td>

z = 6.6420; pseudo p-value = 0.001</td><td>

z = 7.7343; pseudo p-value = 0.001</td></tr>
<tr><td>Q2</td><td>

z = 6.0728; pseudo p-value = 0.001</td><td>

z = 5.9380; pseudo p-value = 0.001</td></tr>
</table>

Table 1: Global Moran's I test results for both variables using Q1 and Q2 weights

Assessing local spatial autocorrelation: The table below shows the local Moran's I cluster maps for both variables using both spatial weights (significance threshold 0.05):

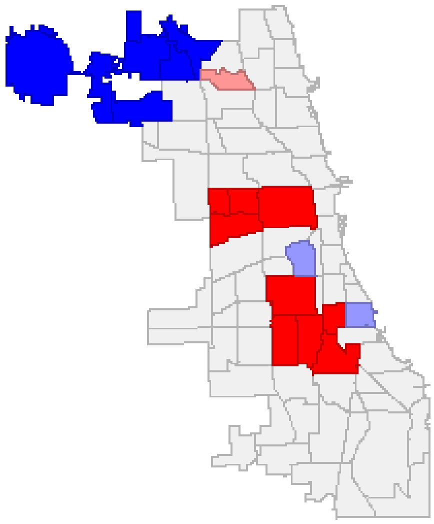
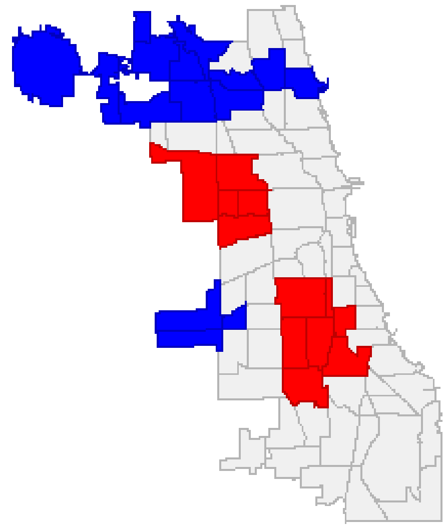
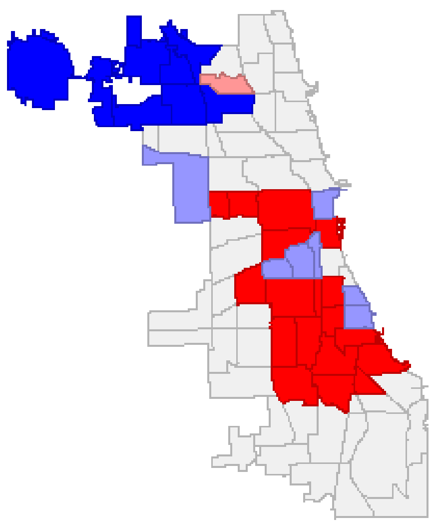
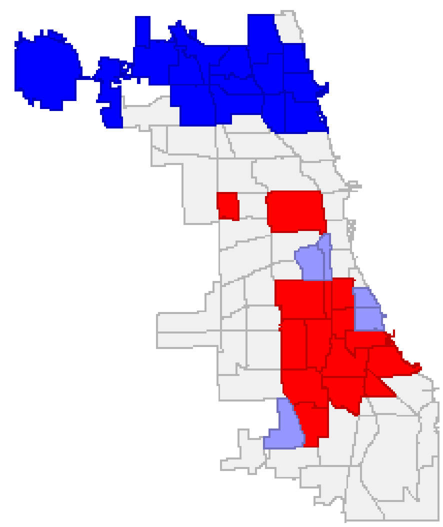
Weight	Lead Screening	Lead Poisoning
Q1		
Q2		

Table 2: Local Moran's I test results for both variables using Q1 and Q2 weights

As Table 2 demonstrates, not only is clustering of both variables apparent throughout the city, the clusters appear in the same general areas for both variables; low-low clusters of lead screening and poisoning are more prevalent in the Northwest Side, while high-high clusters of both variables appear on the Near South Side, as well as on the West Side, though less so if a Q2 weight is used (indicating that clusters of lead screening, and especially poisoning, are more locally constrained on the West Side than the South). This high degree of overlap suggests that higher rates of lead screening did take place in areas with more pervasive lead poisoning, though this overlap will be explored in more depth in the next section.

While the clusters appear in the same general location regardless of the choice of weight, the Q2 weight is also able to reveal a greater degree of negative autocorrelation; particularly in areas with low values surrounded by areas with high values. This negative autocorrelation highlights various interesting areas that are greatly different characteristically from their neighbors. Hyde Park and Bridgeport (the two cases of low-high clustering on the top-left map; see Appendix for map with labels) have relatively low rates of lead screening compared to their immediate neighbors (Q1), while even more areas in their vicinity have relatively low rates of both lead screening and poisoning compared to their more distant neighbors (Q2); in other words, these are islands of low lead concern surrounded by areas of high concern. Meanwhile, Albany Park (the only case of high-low clustering on any of the maps) exhibits a much higher rate of lead screening than its neighbors, despite being part of low-low clusters of lead poisoning.

Sensitivity analysis—change of test statistic: To ensure that the clusters found above were not merely artifacts of the Moran's I methodology, local Geary C and Getis-Ord G* statistical tests were run as well (with the same significance threshold). For simplicity, only the results using a Q2 weight are shown below. Overall, the two other methodologies produced similar results to Moran's I.

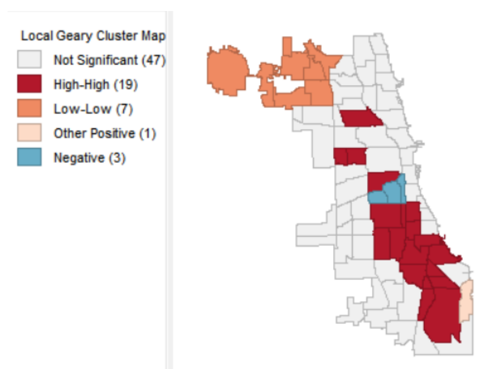
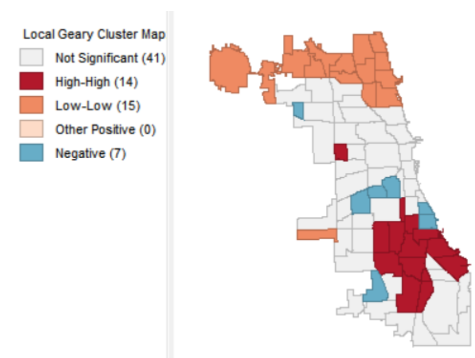
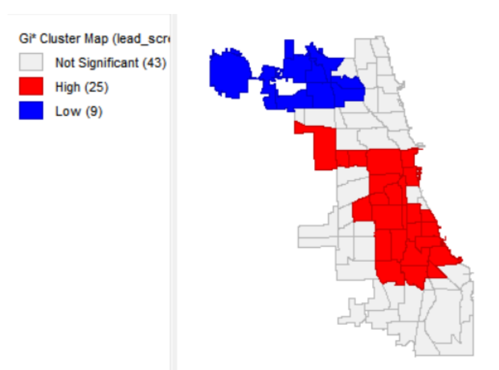
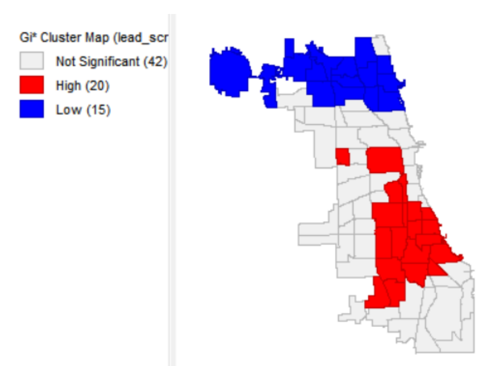
Statistic	Lead Screening	Lead Poisoning
Geary's C		
Getis-Ord G*		

Table 3: Geary's C and Getis-Ord G* tests for both variables using the Q2 weight

A few differences between these maps and those on the bottom row of Table 2 are apparent. Because the G* test does not distinguish areas with negative autocorrelation, areas like Bridgeport and Hyde Park are classified as being part of high-high clusters, despite their relatively low rates of lead screening and poisoning. Otherwise the results of the G* test and the Moran's I test are nearly identical. Geary's C does produce slightly different results, with slightly fewer high-high clusters on the West Side, and slightly more on the South Side, but the core locations of these clusters are the same. Although Geary's C does distinguish areas with negative autocorrelation, it does not differentiate low-high and high-low clusters, and thus the direction of the negative autocorrelation must be derived from context.

Hypothesis 2

Cross-tabulation: The following two tables show a cross-tabulation of lead screening and poisoning clusters, with one table for each spatial weight. The values correspond to numbers of Community Areas. "% Overlap" is the ratio of the highlighted cell to the total number of Areas in that row or column (ie. the percentage of areas in that row or column's cluster type that are of the same cluster type for the other variable). The "% Overlap" value in the bottom right is the total proportion of Areas where the lead screening and lead poisoning cluster types are the same.

		Lead Poisoning Clusters				
		Not Signif.	**High-High**	**Low-Low**	**Total**	**% Overlap**
Lead Screening Clusters	**Not Signif.**	**48**	4	7	59	81.4%
	High-High	1	**8**	0	9	88.9%
	Low-Low	0	0	**6**	6	100%
	Low-High	2	0	0	2	0%
	High-Low	0	0	1	1	0%
	Total	51	12	14	77	-
	% Overlap	94.1%	66.7%	42.9%	-	80.5%

Table 4a: Cluster cross-tabulation with the Q1 weight

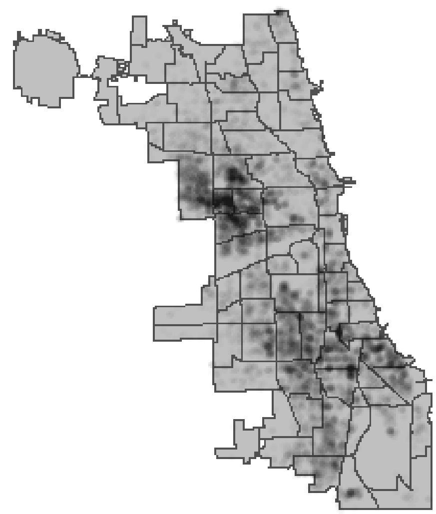
		Lead Poisoning Clusters					
		Not Signf.	**High-High**	**Low-Low**	**Low-High**	**Total**	**% Overlap**
Lead Screening Clusters	**Not Signf.**	**34**	1	7	1	43	79.1%
	High-High	4	**14**	0	0	18	77.8%
	Low-Low	1	0	**7**	0	8	87.5%
	Low-High	3	0	0	**4**	7	57.1%
	High-Low	0	0	1	0	1	0%
	Total	42	15	15	5	77	-
	% Overlap	80.1%	93.3%	46.7%	80.0%	-	76.6%

Table 4b: Cluster cross-tabulation with the Q2 weight

With total overlaps of 80.5% and 76.6%, both the Q1 and Q2 weights indicate that there is a fairly strong, but not perfect, overlap between lead screening and lead poisoning. In order to ensure that this overlap is not a product of chance, a chi-squared test of independence was conducted on both tables, which yielded p-values of $3.8*10^{-13}$ and $4.8*10^{-18}$ respectively. Thus, there is plenty of evidence to reject Null Hypothesis 2.

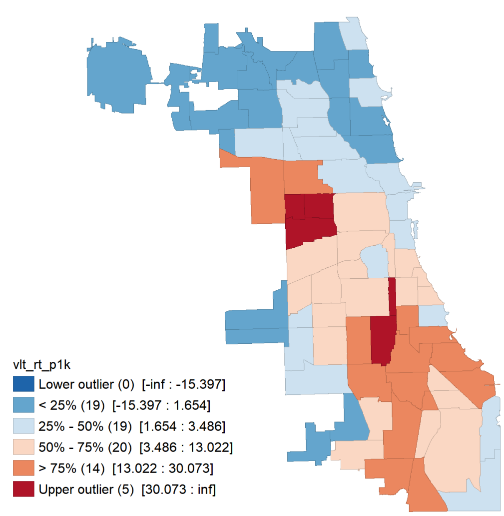
Exceptions to the pattern: 76 of the 77 Community Areas either fall along the diagonal, first row, or first column of the two matrices above, meaning that they either have the same type of clustering for both variables, or clustering for one variable but no significant clustering for the other. Only one Community Area, Albany Park (the only area with high-low clustering in Table 2; see Appendix for map with labels), deviates from this pattern. For both Q1 and Q2 weights, the Area is in a low-low cluster for lead poisoning but a high-low cluster for lead screening. In other words, it has a relatively high rate of lead screening compared to its neighbors, despite it being in a general area where lead poisoning rates are low. Although this does not seem to be indicative of a problem, as higher rates of lead screening are very unlikely to cause any harm, it does suggest that the Community Area could be a place of interest for future study in order to better understand the social forces behind the higher rates of lead screening.

One final individual Community Area worth discussing is Austin, on the far West Side of Chicago. Although the Area is in a high-high cluster of lead poisoning when a Q1 weight is used, it is also in a low-high cluster of lead screening with a Q2 weight, suggesting that the Area has a much lower rate of testing than its (more distant) neighbors, despite being in a (more local) cluster of Areas with high rates of lead poisoning. Though comparisons between results produced by different spatial weights should be taken with caution, this could suggest that the Area is being underserved by lead screening services compared to its neighbors, especially considering the high rate of lead poisoning within and nearby.

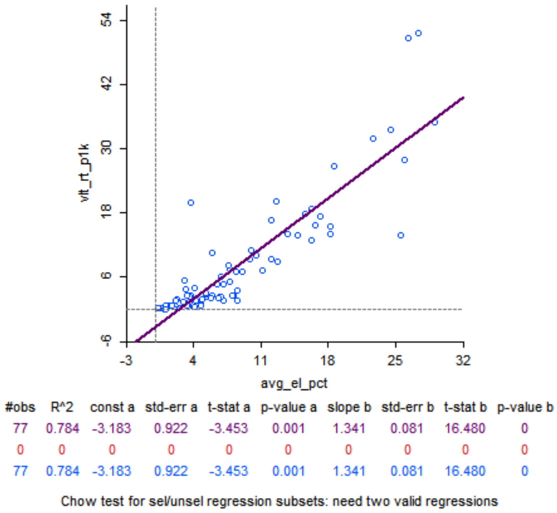


Hypothesis 3

Exploratory Data Analysis: The above-right heat map shows the distribution of homicides and nonfatal shootings in Chicago from 2019 to 2023. However, because it is difficult to make direct comparisons between point data and areal data without aggregation, the rate of homicides and nonfatal shootings per 1,000 population in each Community Area (the calculations for which are described in the Data section of this paper) was used in this analysis instead. A box-map of this variable (hinge 1.5) is shown to the right. Though a degree of granularity is lost in this abstraction, the general pattern of clustering throughout the city remains the same.

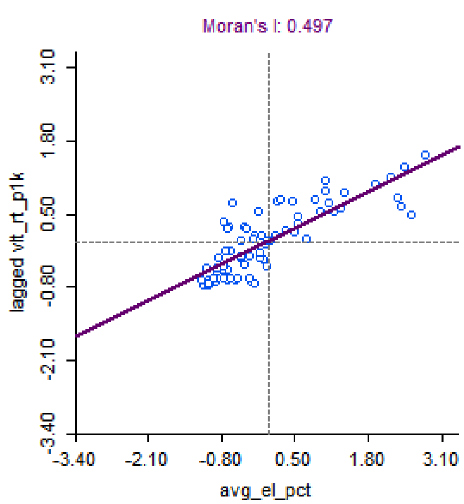
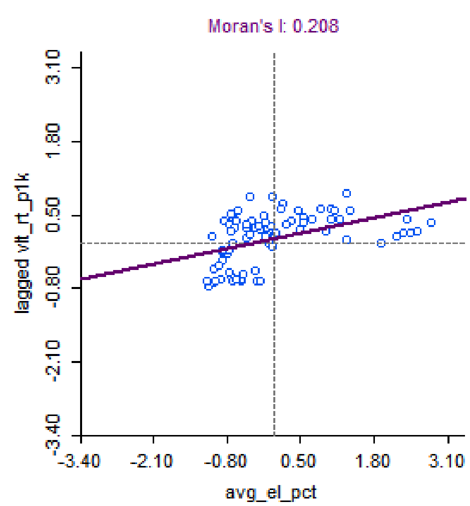


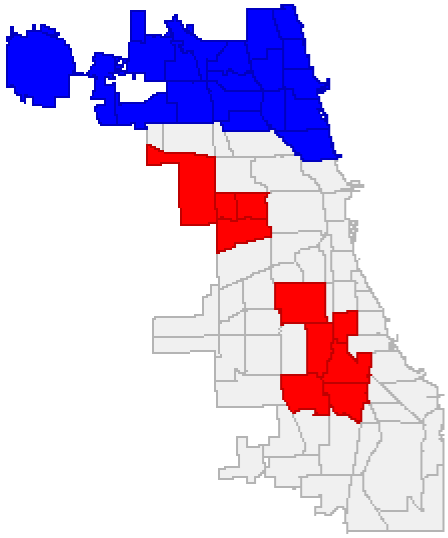
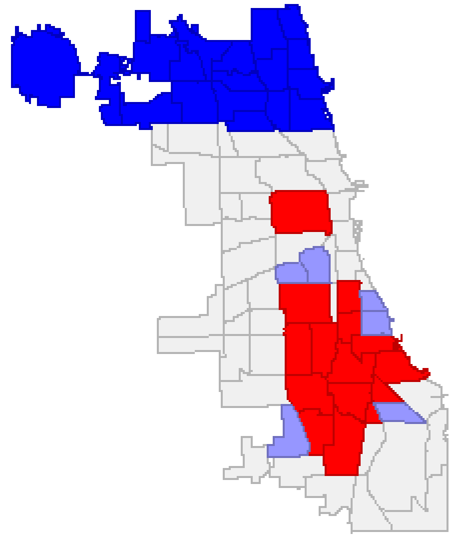
As both maps demonstrate, homicides and nonfatal shootings were most heavily concentrated on the West Side



around the Community Areas Austin, Humboldt Park, East and West Garfield Park, and North Lawndale, with an additional, more spread out cluster around the middle of the South Side, with Greater Grand Crossing around the center (see Appendix for Community Area name labels). Notably, the box map has a strong resemblance to the box map of lead poisoning on page 3, with positive outliers in almost all the same locations. This similarity is further shown in the scatter plot to the left (plotting the effect of percent childhood lead poisoning levels in the early 2000s on the rate of homicides and nonfatal shootings per 1,000 population in the early 2020s in each Community Area). These are all strong indications that rates of lead poisoning had a strong impact on rates of violent crime 20 years later, but they do not directly assess whether this effect remains true when spatial lag is accounted for.

Spatial lag is important. After all, it is unlikely that the effects of childhood lead poisoning would only remain in one Community Area as people move around throughout their lives. Though this movement is nearly impossible to comprehensively account for, one can assume that nearby Community Areas would be a more common destination than ones farther away. Thus, a global and local bivariate Moran's I test was used to calculate the effects of childhood lead poisoning in each Area on the rates of homicides and nonfatal shootings of its neighbors. The results of these tests are shown in the table below.

Type	Q1	Q2
Global	 z = 7.3420; pseudo p-value = 0.001	 z = 5.6320; pseudo p-value = 0.001

Local	 

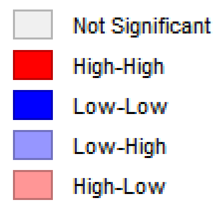


Table 5: Global and Local bivariate Moran's I tests for the effect of lead poisoning on neighboring rates of violent crime 20 years later.

Assessing global bivariate spatial autocorrelation: As the Moran's I scatter plots in the first row of Table 5 demonstrate, there is a strong direct correlation between lead poisoning and the rates of violent crime in neighboring areas, particularly for the Q1 weight (ie. when a more local area is considered), and thus Hypothesis 3 can likely be rejected. It is also worth noting that while cases of low-high autocorrelation are present in both scatter plots, there is virtually no high-low autocorrelation apparent in either one. In other words, while areas with low rates of lead poisoning are sometimes surrounded by areas with high rates of violent crime, areas with high rates of lead poisoning are almost never surrounded by areas with low rates of violent crime. This could indicate that while many factors may contribute to violent crime even when lead poisoning is not present, the presence of lead poisoning almost always leads to higher violent crime rates, though there are likely many other factors that contribute to this trend.

Assessing local bivariate spatial autocorrelation: The maps on the second row of Table 5 show the distribution of significant (p-value threshold 0.05) clusters from the Moran's I scatter plots above. These clusters are very similar to the clusters of lead poisoning found in Table 2, with more local high-high clusters on the West Side, more spread-out high-high clusters on the South Side, and a large section of low-low clustering across the far North Side. On a local level (Q1), there is no significant negative autocorrelation present, while at a broader scale (Q2) low-high negative autocorrelation appears again for Community Areas like Hyde Park and Bridgeport. These are Areas with low rates of lead poisoning, but surrounded by Areas with high rates of homicides and nonfatal shootings, indicating a major disconnect between them and their neighbors. Altogether, these findings suggest that the same areas that form the cores of lead poisoning clusters also are surrounded by neighbors with high rates of violent crime. Though this does not prove the lead-crime hypothesis, it suggests that actual observed patterns from the early 2000s to the early 2020s do align with its premise.

Conclusion

Outside of the realm of this spatial model, there are myriad other factors that could contribute to higher rates of lead screening, lead poisoning, and violent crime. This study could be improved by attempting to account for a broader range of those variables, and comparing the relative strength of their effects on public health outcomes. Without this type of comprehensive approach, this study cannot definitively ascertain whether the overlap between lead screening and lead poisoning was sufficient during the early 2000s, or whether the strong clustering of violent crime around areas with high lead poisoning rates 20 years prior is indicative of a causal relationship between the two.

However, this study has been able to provide strong evidence for the presence of various spatial patterns that may be of interest for future study through the high degree of clustering of its variables and the strong spatial associations between them. It has shown that childhood lead poisoning was strongly clustered in the early 2000s, with high values concentrated in particularly disadvantaged parts of the city, and that lead screening was also clustered with high values in these areas, though to a lesser extent. It has demonstrated that the overlap between clusters of lead poisoning and screening was both strong and significant. It has also highlighted a strong link between lead poisoning in this time period and rates of violent crime in neighboring areas 20 years later. Finally, this study has highlighted various areas of interest that stand out from these patterns, which could merit studies of their own, such as Albany Park with its uniquely high rate of lead screening relative to the Areas around it and Austin with a lower rate of lead screening than its lead poisoning clustering status might warrant.

Appendix: Community Area Labels