Spatial Clustering of Hospital Readmissions and Academic Hospital Proximity in Chicago

Michelangelo Pagan

March 10th, 2025

1 Introduction & Motivation

Hospital readmissions have long served as a benchmark for evaluating the quality, efficacy, and equity of healthcare among patients. In the past, studies have used metrics such as staffing numbers, number of beds, rates of hospital acquired injuries, and others alongside insurance coverage and demographic data from patients to better understand the limitations and current state of our healthcare networks. However, the spatial dimension of healthcare access remains an overlooked area in the literature, and specifically how the type of medical center (academic or non-academic) can impact this relationship has not been explored.

Past work has examined the effects of hospital proximity and academic status on health outcomes separately. In particular, increases in the likelihood of mortality have been associated with longer distances between patients and the nearest hospital (Nicholl et al., 2007). When considering academic status, previous work has found that academic hospitals are rated higher in patient satisfaction surveys and roughly equal in disease-level readmission rates (Chen et al., 2019). Academic medical centers have been shown to improve patient outcomes when compared to non-academic centers, probably due to their emphasis on research and access to state-of-the-art medical resources. In fact, it has even been suggested that non-academic hospitals located closer to academic centers may benefit from knowledge spillover and better care coordination (Mitchell, 2023).

This study leverages spatial data analysis and clustering techniques to better elucidate the spatial trends related to hospital readmissions, with a focus on how academic status can differentially impact this relationship when accounting for socioeconomic factors. The analysis for this project was conducted with data sets in the Chicago area, as Chicago is a nexus of healthcare fragmentation and contains a high density of academic and non-academic hospitals.

2 Research Question

How do spatial clusters of readmission hot spots in Chicago community areas relate to the density and proximity of academic versus non-academic hospitals?

3 Hypotheses

Two hypotheses are investigated in this study:

1. Community areas that are closer to academic hospitals will have lower readmission rates compared to other community areas.

2. The impact of hospital type on readmission rates is more pronounced in socioeconomically disadvantaged community areas.

4 Data Methods

4.1 Data Sources

This research project leverages the Chicago community area shapefile from the GeoDa Center website and an additional shapefile containing hospital locations across the United States. The community area shapefile contains community area numbers, names, polygons, and a variety of socioeconomic data for each community area. In the analyses for this project, the socioeconomic data used from this file includes employment (count), bachelor degrees (count), and median income.

Readmission rate data was extracted from the Illinois Public Health Community Map. This data was downloadable as a CSV and included readmission count per 100,000 cases. A simple query was made to only include regions in Cook County and include counts from all races and ages. Specifically, readmission counts were extracted for heart attack, heart failure, and pneumonia hospitalizations. For each of these, readmission data was reported yearly in the time range of 2019 to 2021. It is worth noting that these data points were initially grouped by zip code in the Illinois Public Health Community Map, and therefore needed to be mapped to the community area level as part of the data processing. Other data sources included files that mapped zip codes to census tracts, and census tracts to community areas.

4.2 Data Processing

Before conducting the analysis using GeoDa and GeoPandas, the data had to be pre-processed to achieve the goal of creating a final dataset with community areas, socioeconomic information, hospital readmission data, distance, and density metrics. The pre-processing steps included:

1. Loading the US Hospitals shapefile and selecting for Chicago hospitals.
2. Manually assigning academic status and community area number to each hospital in the dataset.
3. Saving updated shapefiles, with separate files for academic hospitals and non-academic hospitals in Chicago.
4. Deriving exploratory measures from the separated hospital shapefiles.
5. Mapping community areas to zip codes using census tract walk files and coverage ratios.
    - Census tract to community area and census tract to ZIP code files (for Chicago) were read.
    - Census tracts were filtered to only include those in Cook County, Illinois (Chicago area)
6. Integrating Readmission Data for Heart Failure, Pneumonia, and Heart Attacks
    - Three-year data for readmissions was averaged yearly from the reported period in 2019 to 2021.
    - An outer join was performed with the census tract to zip code file and census tract to community area file, providing a dataframe with the tract number, zip code it falls within, ratio of the coverage of the tract in the zip code (from 0 to 1, 1 meaning the tract is entirely encompassed in the zip code and 0 meaning it is not within the zip code), and community area name that covers the census tract (community areas encompass multiple census tracts entirely).
    - Admissions data was merged to this dataframe, with each readmission rate multiplied by the coverage ratio to produce a dataframe with readmissions by census tract and the community areas the census tracts are in.
    - The dataframe was grouped by community area, with the weighted census tract readmission data summed. Dropping unnecessary columns, this resulted in a dataframe with community area number, community area name, heart attack readmissions per 100k (scaled), heart failure readmissions (scaled), and pneumonia readmissions (scaled).
7. The final dataframe with community area readmissions data was merged to the existing community area shapefile from GeoDa, producing an aggregated dataset (**final_CA_df**) with all sociodemographic data, community area polygons, and readmissions data.
8. Creating distance metrics for each community area.
    - Using GeoPandas, the final_CA_df was reloaded alongside the Chicago academic and Chicago non-academic hospital point layer data.

- A new column with centroid locations was created, and the distance from each community area centroid to the nearest academic and non-academic hospital (in meters) was derived and added to the dataframe.

9. The dataframe was turned into a shapefile through GeoPandas for analysis in GeoDa. A new column was created directly in GeoDa that summed all the readmissions counts for each of heart attack, heart failure, and pneumonia readmissions.

For the full code used to process the data, this GitHub Repository contains a Jupyter Notebook with all the pre-processing conducted and steps labeled.

5 Analysis

5.1 Exploratory Data Analysis

The point-layer data in this analysis of includes 45 total hospitals. Out of these, 14 are academic hospitals (31%) and the remaining 31 are non-academic hospitals (69%). Despite the disparity in the amount of each hospital type, each carry a roughly equal burden when it comes to number of patients since academic hospitals are generally a lot larger than their non-academic counterparts. The 14 academic hospitals have a combined count of 5,895, or 48%, of total hospital beds in Chicago, with the remaining 52% split over the non-academic centers.

To explore underlying trends in Chicago and how hospital location might play a role depending on hospital type, community area quartile maps were generated based on readmission rates for each of the three conditions tested. From lightest to darkest, the quartiles go from 0-25% up to 75-100% quartiles. In addition, the marked hospitals point layer was added on top of these quartile maps to observe any potential interactions that may be present. In the figures, blue points represent non-academic hospitals while red points represent academic hospitals.

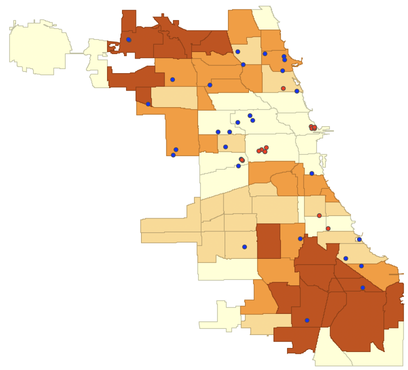

Figure 1: Quartile Heart Attack Readmissions

Figure 1 shows a clustering of heart attack readmissions in the Northern-most and Southern-most community areas of Chicago, regions in which academic hospitals are not significantly present. It is clear from the point-layer hospital data that there are more academic centers in central Chicago compared to non-academic centers, which appear more in Northern Chicago yet seemingly have less of an impact when it comes to limiting heart attack readmissions. In Northern Chicago, there are 8 non-academic hospitals within community areas that still are in the third quartile bucket for heart attack readmissions. On the other hand, nearly all of the community areas (except two) that contain academic hospitals are in the lowest quartile for heart attack readmissions. It is worth noting that many of these academic centers are in very close proximity to other academic hospitals, which could contribute to the amelioration in readmission data in these regions.

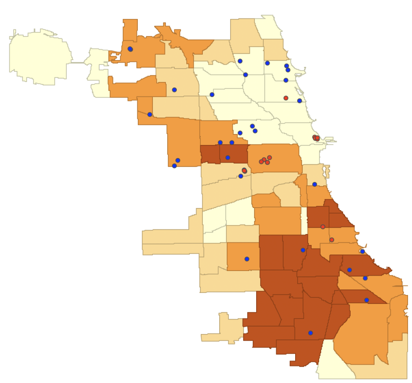

Figure 2: Quartile Heart Failure Readmissions

Figure 2 shows heart failure readmissions quartile data, which follows a different pattern than heart attack readmissions. In this map, Northern Chicago as a whole contains community areas with lower readmissions rates, and community areas containing academic centers do not seem to benefit differentially when compared to those with non-academic hospitals. In addition, there is a lot of overlap in readmissions in the south-side of Chicago, a trend that is expected due to the known reduction in health outcomes, insurance coverage, and access to care in these regions. For the entire south-side, the closest academic center is in Hyde Park (University of Chicago Medical Center), and non-academic centers are not as evenly distributed as they are in northern Chicago.

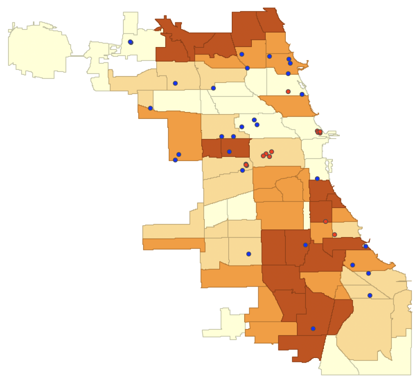

Figure 3: Pneumonia Readmissions

Figure 3 shows the pneumonia readmissions quartile map, indicating overall similar trends to the heart attack readmissions. Once again, it visually appears that community areas with academic hospitals tend to fare better than those without, but this can be further tested through analysis of actual distance metrics to uncover how significant these relationships are.

5.2 Selected Methods

In order to find underlying spatial clusters in the data while accounting for distance and spatially-constrained relationships, the clustering methods **Bi-variate Local Moran's I** and **KMeans Clustering** were used.

Bi-variate Local Moran's I analysis lends itself to the identification of spatial clusters based on two input variables. The incorporation of this method allows for the visualization of hotspots and coldspots in a spatial context, which can be compared to see whether and how much selection of different variables impact the clustering output. For this project, both hypotheses are addressed through this analysis by clustering Chicago areas based on readmissions and distance to academic hospitals, then on readmissions and distance to non-academic hospitals, and comparing the outcomes.

KMeans analysis is also employed in this study to identify potential spatial relationships in hospital readmissions data. Since KMeans is not specifically spatially constrained, comparison of the KMeans clustering results between clusters and their spatial configuration will lend itself to insights into the robustness of the spatial analysis done with LISA. In addition, KMeans is apt for handling continuous numerical data, as is the case with each of the indicators used in the model (employment number, bachelor's degree number, readmissions data, distance to academic and non-academic hospitals). Given the output metrics provided with KMeans, we will be able to directly relate the distances from academic or non-academic hospitals to readmissions rates, helping to address Hypothesis 1.

6 Results

6.1 Bi-variate Local Moran's I

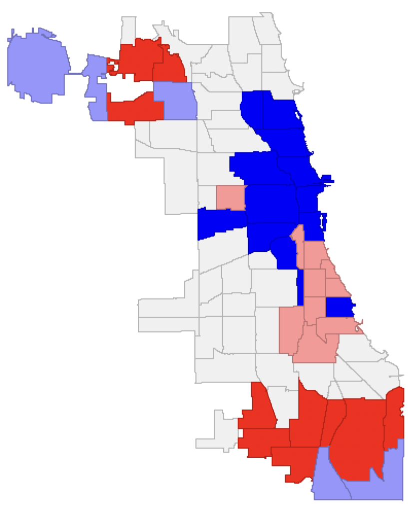

Figure 4: Total Readmissions & Academic Hospital Distance (meters)

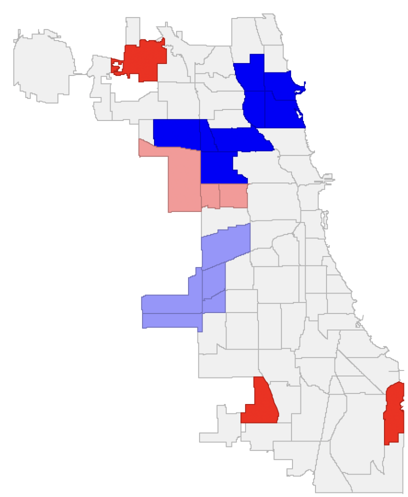

Figure 5: Total Readmissions & Non-Academic Hospital Distance (meters)

Figure 4 shows the result of Bi-variate Local Moran's I analysis when the input variables are total readmissions and minimum distance to an academic center. The low-low areas (blue) signify community areas with low total readmissions and close proximity to academic hospitals, high-low areas (light red) represent community areas with high readmissions but close proximity to academic centers, low-high areas (light blue) represent areas with low readmissions rates but high distance from academic hospitals, and high-high (red) areas represent community areas with high readmission rates that are farthest from academic centers. There are two major high-high clusters in the West and South of Chicago, where health disparities are most pronounced. These clusters suggest that distance to an academic center may be a contributing factor to high readmissions. In addition, this figure reveals that being closer to an academic center improves readmission rates and is most prominent in the central and Northern parts of Chicago. However, the presence of high-low clusters weaken this finding, suggesting that there are other factors that may be responsible for readmission rates such as income, distance to non-academic centers, and other factors.

Figure 5 uses the same methodology, clustering total readmissions with minimum distance to a non-academic center. This figure shows less low-low clusters in which there are lower readmissions and close proximity to centers than figure 4, suggesting that the protective effect of being near a non-academic center is not as pronounced as being near an academic hospital. Interestingly, the two high-high clusters near South Chicago in this figure are also high-high clusters in Figure 4, indicating that these community areas in particular (Beverly and East Side) are particularly impacted by their distance from both kinds medical centers, as reflected in their readmission rates. Excluding community areas such as O'Hare and the Southern-most community areas of Chicago, Figure 5 also has larger low-high clusters, meaning readmission rates are low despite increased distance from non-academic hospitals. This strengthens the finding that academic hospital distance may have a stronger protective effect, though this finding is further explored with KMeans.

There is also a marked increase in grey community areas in Figure 5, indicating that these areas are not statistically significant in the bivariate local Moran's I. This further suggests that the spatial relationship between readmissions and non-academic hospital distance is weaker when compared to academic hospital distance (figure 4).

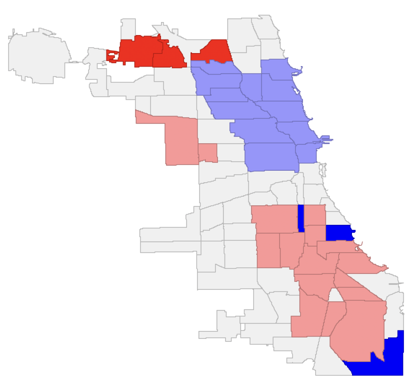

Figure 6: Total Readmissions & Median Income

Bivariate analysis has the limitation of not being able to consider other factors that may complicate the relationship between hospital distance and readmission rates. Figure 6 shows one of these potential confounders, clustering community areas by total readmission and median income. Northern Chicago has a large cluster of low-high areas in which readmissions are low and median income is high, suggesting better insurance coverage and access to higher quality care. In addition, there is a large high-low cluster in South Chicago, indicating that poorer areas have high readmission rates and by extension worse outcomes. This reinforces previous findings in literature and underscores the significant health fragmentation that is present in Chicago, which is exacerbated by a lower density of hospitals and greater distance from academic centers for south Chicago community areas.

6.2 KMeans Clustering

Through the use of the KMeans algorithm, the biggest decrease in total sum of squares occurred with four clusters. In this model, employment number, bachelor's degree number, median income, readmissions, and distance data were used to create clusters without a specified spatial constraint. The results of the KMeans algorithm can be found in Table 1 and Figure 7.

	EMP	BACH	MEDINC	HA_Re	HF_Re	PN_Re	dist_ac	dist_non
C1	19052.1	5300.62	61240.4	35.9271	92.685	28.4546	7512.9	2349.07
C2	8827.13	1992.3	34770.3	41.0728	172.459	45.5043	4850.53	2334.64
C3	7218.25	1741.25	54720	11.0016	46.1928	10.9184	11480.4	5929.65
C4	50093.6	24195.4	100220	7.39578	55.2433	16.6131	1861.24	2258.89

Table 1: Cluster Characteristics for Community Areas

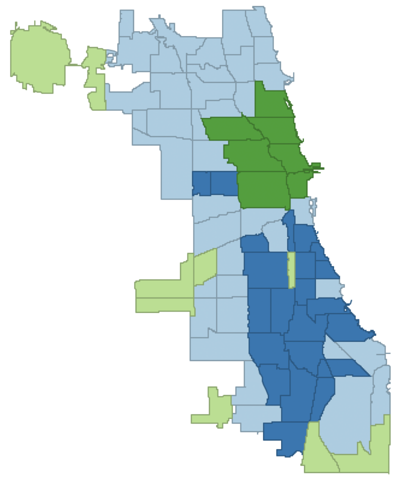

Figure 7: KMeans Clustering

Table 1 indicates the creation of four unique clusters with distinctive characteristics:

1. Cluster 1 (Light Blue): Middle and low-class areas with moderate hospital access. Include high employment (19,052 people) due to cluster size, slightly above-average income (over $60,000), and moderate readmission rates. Hospitals are not too far but not the closest when compared to other clusters.

2. Cluster 2 (Blue): Low-income areas with high readmission rates. This areas also have low employment (8,827 people), lower education rates, and significantly lower income (around $34,000). Moderate hospital access but suffer disproportionately from heart failure and pneumonia readmissions.

3. Cluster 3 (Light Green): Community areas on the outer areas of Chicago, with the least hospital access. Farthest distance to academic and non-academic centers yet still have low readmissions, indicating a lack of hospital dependent residents or alternative means of healthcare services outside of Chicago.

4. Cluster 4 (Green): Wealthy areas with best access to healthcare. These community areas have the highest employment, education levels, and median income. Closest to both types of hospitals but academic centers in particular (1,861m) and contain the lowest readmission rates on average over the three types of conditions.

The KMeans clustering highlights distinct patterns in socioeconomic variables, hospital distance, and accessibility that provide context for the bivariate Moran's I analyses (Figures 4 and 5). Notably, several low income areas in

Clusters C1 and C2 remain in high-readmission clusters despite having non-academic hospitals nearby. This aligns with prior research suggesting that non-academic hospitals located farther from academic centers may lack the resources to achieve comparable patient outcomes and are outside of the range to benefit from "spillover" effects of an academic network.

7 Discussion

The analyses in this project provide insight into the clustering of readmission hotspots in Chicago based on proximity of hospitals, with consideration of hospital academic status. This project substantiates the inequities of healthcare in Chicago, specifically the fragmentation between Northern and Southern community areas. Although it was found that being in closer proximity to academic centers generally improved (that is, lessened) readmissions, this is actually just one part of the larger access issue at hand. Socioeconomic factors including education level, income inequalities in community areas, and employment fragmentation on their own have a strong impact on hospital readmissions (Figures 6 and 7). Combined with the fact that most academic centers are concentrated in Northern/central Chicago, the fragmentation of health outcomes is exacerbated. Non-academic centers, although present in Southern Chicago (Figures 1-3), do not have that much of a mitigating effect in reducing heart attack and heart failure readmissions.

There are certainly limitations to this analysis and factors that are not considered that may provide a better picture of the specific benefits, drawbacks, and relationship between academic and non-academic hospitals. Performance metrics of physicians, patient satisfaction surveys, length-of stay, and presence of other centers for care outside of hospitals may all play a role in the readmissions data and are not considered in the current project.

Overall, the findings of this spatial cluster analysis indicate that proximity to academic hospitals matters. Community areas closer to academic hospitals tend to have lower readmissions rates, and they also tend to be found in wealthier areas of Chicago than those farther from academic centers. On the other hand, the relationship between non-academic hospital distance and readmissions is more variable, with more areas showing non-significant associations. This is likely due to the more pronounced effect of socioeconomic factors and smaller overall scope of care for non-academic facilities, which often don't deal with more complex cases or transfer patients to academic centers in many instances. These findings highlight the need to address the healthcare disparities in Chicago and targeted interventions in high-readmissions areas, particularly those far from academic hospitals. Creating a less clustered network of medical knowledge may improve the outcomes of nearby hospitals without necessitating the construction of many new academic hospitals.

References

Chen, A. S., Revere, L., Ratanatawan, A., Beck, C. L., and Allo, J. A. (2019). A comparative analysis of academic and nonacademic hospitals on outcome measures and patient satisfaction. *American Journal of Medical Quality*, 34(4):367–375.

Mitchell, J. (2023). Presence of teaching hospitals improves outcomes. *Harvard Medical School News*.

Nicholl, J., West, J., Goodacre, S., and Turner, J. (2007). The relationship between distance to hospital and patient mortality in emergencies: an observational study. *Emergency Medicine Journal*, 24(9):665–668.