ARTICLE

Predictors of Delays in EMS Transport in the City of Chicago

S. Okum[1, a b e f], E. Bussel[1, e f], K. Loetscher[1, a d f], S. Chang[1, a d f], A. Saparov[1, a b e]

Affiliations: [1]The College, University of Chicago

Roles: [a]Report Author, [b]Dataset Assembly, [c]Data Cleaning, [d]MCA Analysis, [e]Regression Analysis, [f]Data Interpretation

Received 19th March 2020,
Accepted 19th March 2020

Introduction

The City of Chicago, like other major American metropolitan areas, has long struggled with equity in the provision of acute care services to diverse patient populations.[1] In particular, previous research has paid close attention to demographic and geographic disparities in access to trauma care in urban settings.[1] America's three largest cities—New York, Los Angeles, and Chicago—remain observably racially segregated, with these entrenched patterns of racial segregation accounting for inferior access to trauma care for historically marginalized groups.[1] Of these cities, Chicago is the most extreme case study in the association between race and trauma care access. According to Tung et. al., majority-black Chicago census tracts are 8.5 times more likely than their white counterparts to be located in a trauma desert, defined as a geographic area lacking immediate access to surgical trauma care.[1]

To mitigate these documented geographic and racial disparities in trauma care access, cities can work to reconfigure their acute care infrastructure in a fashion that improves access and equity for all residents, irrespective of the neighborhood in which they live. One such strategy that has been successfully employed in Chicago is the re-opening of a high-volume level 1 trauma center at the University of Chicago Medical Center (UCMTC), which is situated in the city's black-majority South Side and amongst communities with a significant burden of firearm violence.[2] Prior to UCMTC's opening on 5/1/2018, the trauma centers that serviced South Side residents were distantly located in Chicago's urban core, contributing to disproportionately long emergency medical services (EMS) transit times for black trauma patients in the city on average.[2] The opening of UCMTC contributed to a significant reduction of transport times for black patients and those from economically-disadvantaged Chicago zip codes. Most importantly, however, the UCMTC opening preceded a near elimination of the previous, statistically-significant transport time disparity between white and black trauma patients.[2] This could be clinically significant, since EMS transport time is modestly correlated with patient mortality in Chicago based on previous research.[3]

However, geospatial reorganization of trauma systems is only one component of broader policy initiatives that can reduce disparities in access to critical care. Crucially, cities must also examine their EMS systems holistically to address determinants—demographic or otherwise—that underlie delays in ambulance response, as timely EMS transport has been shown to reduce mortality and morbidity for certain patient populations in need of emergency care. [4,5] Though EMS services on the whole remain plagued by a chronic lack of compliance with internal response time standards in the City of Chicago, incompliance is more pronounced for ambulance runs conducted in the majority-minority South and West sides of the city.[6] However, whether racial and geographic factors are statistically significant determinants of delays in ambulance response in Chicago remains an open question. The failure of previous research to robustly investigate these connections could pose a barrier to the equitable provision of EMS services in Chicago, as a thorough understanding of the role of social determinants in EMS response times is necessary to best implement policy to address disparities in care. This need has served as the impetus of this study, in which we conducted a multimodal exploratory analysis of an EMS database to investigate the existence of a correlation between demographic and temporal explanatory variables and delays in EMS response.

Methods

For our investigation, we explored the same database as was analyzed by Abassi et. al. (2020), which includes data for all ambulance runs related to traumatic injury originating in Chicago from 5/1/2015 to 5/1/2019. The data was filtered to only include adult patients (defined as individuals older than 15 years of age) and runs in which EMS transport to a hospital was completed. We adopted the following definitions in this study:

1. "Run time" includes the total duration between ambulance dispatch and patient arrival at the hospital
2. A "delay" in EMS transit occurred for any ambulance run with a duration greater than 1 standard deviation above the mean run time (2234 sec)

Our quantitative analysis involved two central components: multiple correspondence analysis (MCA), a dimensionality-reduction technique, as well as a series of regression analyses.

Multiple Correspondence Analysis

MCA was used in this study for dimensionality reduction of our complex dataset, which allowed us to better visualize associations between demographic, temporal, and geographic variables. The following eight variables were examined with MCA:

1. Patient race/ethnicity
2. Patient gender
3. Patient age
4. Income quintile of the injury scene zip code
5. Scene zip code classification as rural or urban
6. Time of day of the ambulance run
7. Month of the year during which the ambulance run took place
8. Injury classification (penetrative wound or not)

The patient age variable was binned in ten-year increments and the time of day variable was binned at 3 hour increments to streamline analysis. Additionally, since run duration was used as the response variable in later regression analyses performed in this study, it was excluded from MCA for two reasons: (1) MCA inherently does not allow for interpretation between explanatory and response variables in the first place and (2) we did not want the run time variable to affect derivation of the principal axes.

Upon determination of principal axes, we produced several plots—including scree plots, plots of projected categories, plots of projected individuals, and squared cosines plots—to visualize the data.

For these analyses, R packages FactoMineR and factoextra were used to run MCA and derive descriptive plots and statistics respectively.

Regression Analysis

We employed three principal regression schemes in this study: a multivariate logistic regression, a multiple linear regression, and a series of univariate linear regressions. All regressions were run using Project Jupyter for Python.

We initially framed this analysis as a binary classification problem, where runs with duration above the delay threshold were classified as "on time", while "delayed" runs had duration below the threshold. A multivariate logistic regression was used to investigate if and how seven explanatory variables of interest correlated with our binary delay outcome variable. The explanatory variables examined were the following:

1. Distance from the zip code of the scene of injury to the destination hospital
2. Patient ethnicity (Latinx or not)
3. Patient race
4. Patient gender
5. Time of day of the ambulance run
6. Month of the year during which the ambulance run took place
7. Injury classification (penetrative wound or not)

Unfortunately, we were unable to also incorporate another potentially intriguing variable—the scene of injury zip code—into this analysis. Doing so would necessitate the construction of dozens of additional dummy variables, which would lead to concerns of over fitting when running the logistic regression.

To prevent overlooking any underlying relationships in the data, we also performed a multiple linear regression in the event that our seven variables of interest could better explain trends in ambulance run durations if we were to treat our time variable as a continuous distribution rather than a binary state. The multiple linear regression was run with the same combination of explanatory variables as were used in the logit regression above.

Finally, univariate linear regressions—which separately compared each of the explanatory variables to the run time distribution—were performed to account for autocorrelative effects between explanatory variables.

Results

There were 151,692 EMS runs spanning over a 5-year period that were included in this study. Of these runs, 80,873 (53%) involved black patients, 33,803 (22%) involved white patients, and 22,407 (15%) involved Hispanic patients. In addition, 85,626 (56%) patients were male, while the remainder (44%) were female. Overall, 133,226 (88%) runs were below the delay threshold, while 18,466 (12%) runs were above it.

Multiple Correspondence Analysis

1. Scree Plots

Figure 1 includes the scree plots produced from MCA. **Panel A** of this figure shows the percent of variance captured by each dimension of the MCA when it was run with all eight variables of interest mentioned in the methods section. The percent of explained variance is notably low across the dimensions. We therefore removed temporal variables (run time of day and run month) to reduce dimensionality further, which successfully improved resolution of correlations across multiple variables and increased the percent of variance explained by the first two dimensions. The **panel B** of **Figure 1** shows the scree plot produced from this six-variable analysis.

Although the six-variable MCA accounted for a small segment of the total variance, our results are comparable to other papers that have performed similar MCA analyses on public health data. The small capture of variance could perhaps be explained by a correction MCA uses to deflate eigenvalues.[7] However, despite the marginally improved ability of the six-variable MCA to account for variance in our dataset, we ultimately chose to implement the eight-variable model to generate subsequent figures in order to explore relationships between a more expansive cohort of variables.

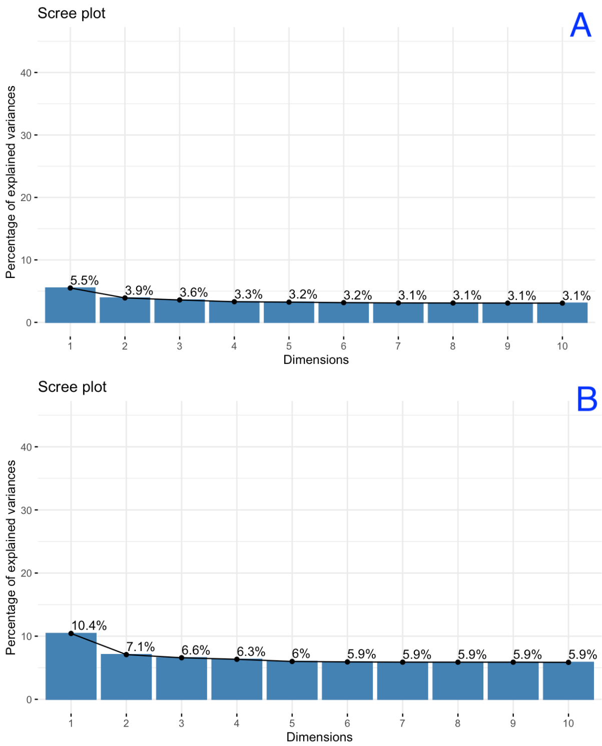

Figure 1: Scree plots for the eight-variable MCA model (panel A) and the six-variable MCA model (panel B)

2. MCA Coordinate Plot

The MCA Coordinate Plot in **Figure 3** was produced following our decision to use the eight-variable model. It represents clouds of our categorical variables of interest projected onto the first two dimensions returned by MCA. Oppositely correlated categories are found in opposing quadrants, while variables with similar profiles cluster together. For instance, the race categories “white” and “black” are located opposite quadrants in the dimensional plane (I and III respectively), while the gender categories “male” and “female” are also located in opposing quadrants (II and IV).

Interesting clustering patterns are also observed in the coordinate plot below. Notably, non-penetrating wounds seem to cluster with colder months, older age, female patients, and more active periods of the day in quadrant IV, while penetrating wounds appear to cluster with warmer months, younger people, male patients, and less active hours of the day. These patterns in clustering could help us understand overlapping patient and temporal characteristics for different injury classifications.

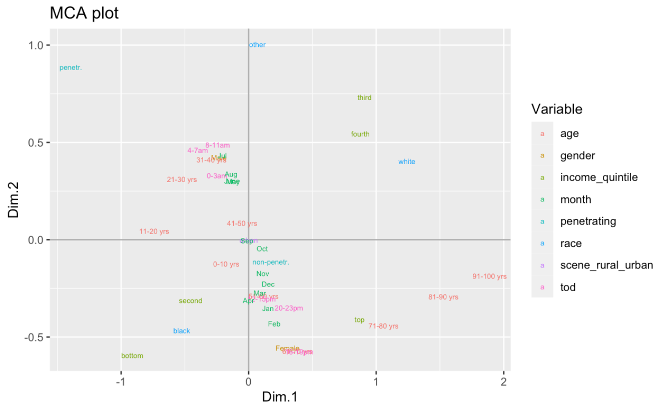

Figure 2: MCA coordinate plot for the eight-variable model

3. Squared Cosine Plot

The Squared Cosine Plot showed in **Figure 4** represents the contributions of each categorical variable to a given dimension. The $\cos^2$ of the distance between a particular categorical variable and an axis provides a quantifiable degree of association between the two. Larger $\cos^2$ values indicate that the categorical variable of interest has a higher degree of association to a particular axis. Higher association values occur for variables that have strong influence on the determination of the principal axes.

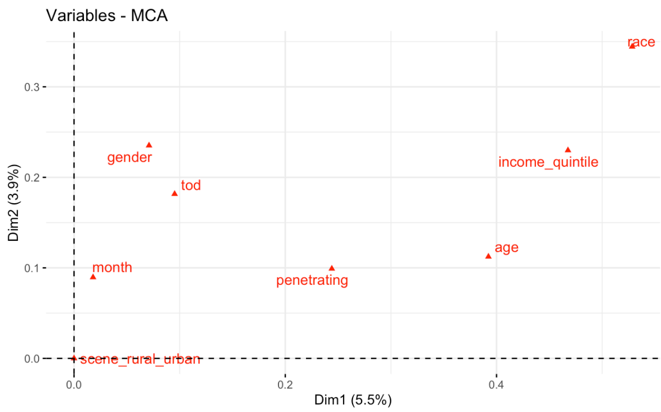

Figure 3: Squared cosine plot for the eight-variable MCA model

Looking at the Squared Cosine Plot above, the most discriminative variables for the first dimension are income quintile, age, and type of injury. The most discriminative for the second dimension are time of day, month, and gender. The race category is well described by both with the highest sum of association values (sum of x and y coordinates).

4. MCA Coordinate Plot Colored by Squared Cosines

Figure 4 combines analytical perspectives offered by Figures 1 and 3. By displaying variable categories coded by their $\cos^2$ values, we can adjust our level of confidence regarding how these variables are placed in low dimensional space. Warmer colors imply greater confidence in placement, while colder colors imply lower confidence in placement and may be better captured with a greater number of dimensions. Overall, the most salient variables are injury type, race, income quintile, and gender.

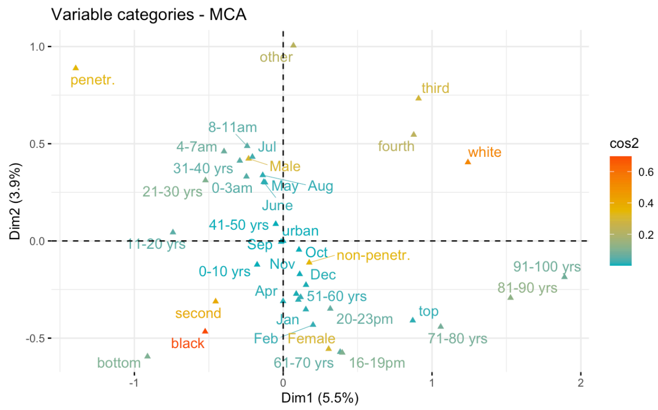

Figure 4: MCA coordinate plot colored by association/$\cos^2$ values for the eight-variable model

Regression Analyses

The results of the multivariate logistic regression are summarized in **Table 1** and point to an overall poorness of fit of the model. First, the LLR p-value rounds to 1, meaning that our fit does not show a statistical relationship between the seven explanatory variables we selected and the binary time distribution. In addition, the pseudo R^2 value is negative, which indicates that the logistic model is actually worse at accounting for the variance in the binary time variable than a horizontal line. Comparing the log likelihood value of the model with the null log likelihood value echoes this conclusion. Though both are negative with a large magnitude (-45176 and -44251 respectively), the null log likelihood value is actually greater than that of the logistic model, meaning that the null hypothesis (fitting the data with a horizontal line) better captures relationships in the data than the model itself.

Logistic Regression Results				
	Coefficient	Std. Error	z	P > |z|
Distance	2.70e-06	2.26e-07	11.939	0
Ethnicity	-1.3088	0.028	-46.926	0
Time of Day	-0.0307	0.001	-27.902	0
Month	-0.0801	0.002	-36.117	0
Penetrating	-0.3282	0.018	-18.189	0
Female	0.0507	0.029	1.765	0.078
Male	-0.948	0.023	-41.494	0
White	-1.3953	0.02	-69.062	0
Black	2.70e-06	2.26e-07	11.939	0
			Pseudo R^2	-0.021
			LLR p-value	1.000

Table 1: Logistic Regression Output

Table 2 summarizes the results of the multiple linear regression, which also did not suggest a collective association between our explanatory variables of interest and the continuous distribution of run times. Notably, the condition number for this fit ($1.7\bullet10^4$) was extremely large, which reflects a high degree of autocorrelation (or multicolinearity) between our chosen explanatory variables. Resultantly, this fit is not useful to delineate between explanatory variables that actually correlate with the continuous distribution of run times and those that do not.

Please do not adjust margins

Multiple Linear Regression Results				
	Coefficient	Std. Error	z	P > |z|
Distance	0.0018	7.32e-05	23.997	0
Ethnicity	925.889	6.896	134.274	0
Time of Day	17.6164	0.27	65.36	0
Month	48.4781	0.552	87.88	0
Penetrating	124.8846	4.204	29.708	0
Female	191.6907	6.901	27.777	0
Male	1020.9745	6.234	163.77	0
White	852.0339	5.352	159.197	0
Black	0.0018	7.32e-05	23.997	0
			Condition No.	1.7e+4

Table 2: Multiple Linear Regression Output

The output statistics of the univariate regressions are summarized in **Table 3**. Across the board, these fits have low p-values and low Pearson correlation coefficients, which suggest the lack of a linear relationship between the explanatory variable and distribution of run times. The only univariate regressions that are potentially of interest are those for ethnicity (Hispanic or non-Hispanic) and scene to hospital distance. These fits yielded low p-values, which possibly suggest a linear relationship to run time. This result unsurprising for the distance variable because run time should naturally increase for ambulance runs in which travel distance was longer. However, after accounting for multiple testing, it's possible that we would find the correlation observed for the ethnicity variable to be spurious.

Univariate Linear Regression Results		
Explanatory Variable	r	p-value
Gender	0.0168	0.9936
Distance	0.0020	4.725e-12
Race (White/Not White)	0.5921	0.8094
Penetrating	1.3217	0.6937
Hour of Day	1.3163	6.2160
Month of Year	0.0337	0.9236
Hispanic	6.8110	0.0165

Table 3: Univariate Linear Regression Outputs

Conclusions and Future Directions

Overall, the multiple correspondence analysis conducted in this study is more interesting than the regression analyses from the standpoint of trying to understand underlying relationships in the EMS dataset. Data visualization through MCA underscored the existence of major variable clusters. The first cluster was defined by penetrative wounds, less-active hours, warm months, younger patient ages, and male patients. The second cluster was defined by non-penetrating wounds, more active mid-day hours, colder months, and female patients. Although $\cos^2$ analysis revealed that certain variables—particularly those temporal in nature—have reduced salience in interpretations of data clustering, the observed relationships between injury class, gender, and race could have real implications in framing further analyses that go a step further to quantify the association between these factors.

Unfortunately, our regression analyses did not underscore similarly interesting underlying trends in the data. Both the multivariate logistic regression and multiple linear regression did not provide significant evidence to support the existence of a correlation between our explanatory variables of interest and ambulance run duration or propensity of delay respectively. Despite the identification of multicolinearity between explanatory variables from our multiple linear regression, univariate analyses were largely unhelpful in untangling this issue.

In the future, the dataset employed in this study should be subjected to further quantitative analysis to continue the broader effort of uncovering disparities in the provision of acute care. Though this analysis largely dispelled the notion that demographic and temporal factors may be deterministic in ambulance run delays, it remains possible for new equity issues to come to light if researchers ask more specific research questions in further analysis of this dataset.

Acknowledgements

As a group, we would of course like to thank Dr. Dmitry Kondrashov for his help in exposing us to some of the quantitative methods used in this study and for the general guidance he provided to us over the duration of this project.

Sam would also like to express gratitude to Dr. Peter Bendix and Dr. Ali Abbasi for serving as research mentors and exposing him to the broader considerations that drive the provision of trauma care from a systems perspective as well as quantitative methods used in public health research. This project could not have been brought to fruition without their help!

References

1. E.L. Tung, D.A. Hampton, M. Kolak, S.O. Rogers, J.P. Yang, M. Peek. *Race/Ethniciy and Geographic Acces to Urban Trauma Care*. JAMA Network Open, 2019, 2(3).
2. A.B. Abbasi, J. Dumanian, S.E. Okum, D. Nwaudo, D. Lee, P. Prakash, P. Bendix. *Association of a New Trauma Center with Racial, Ethnic, and Socioeconomic Disparities in Access to Trauma Care*. JAMA Surgery, 2020, 156(1): 97-99.
3. T. Karrison, P. Schumm, M. Kocherginsky, R. Thisted, D. Dirschl, S.O. Rogers. *Effects of Driving Distance and Transport Time on Mortality Among Level I and II Traumas Occurring in the Chicago Metropolitan Area.* Journal of Trauma and Acute Care Surgery, 2018, 85(4): 756-765.
4. C.E. Begley, Y. Chang, R.C. Wood, A. Weltge. *Emergency Department Diversion and Trauma Mortality: Evidence from Houston, Texas*. Journal of Trauma and Acute Care Surgery, 2004, 57(6): 1260-1265.
5. Y. Shen, R.Y. Hsia. *Ambulance Diversion Associate with Reduced Access to Cardiac Technology and Increased One-Year Mortality*. Health Affairs, 2015, 34(8).
6. *Chicago Fire Department Fire and Medical Incident Response Times Audit,* Office of the Inspector General for the City of Chicago, 2013.
7. P.S. Costa, N.C. Santos, P. Cunha, J. Cotter, N. Sousa. *The Use of Multiple Correspondence Analysis to Explore Associations between Categories of Qualitative Variables in Healthy Ageing*. Journal of Aging Research, 2013, 1-12.