Examining Individual Health and Healthcare Utilization Patterns at the Intersection of Transportation, Environment and Communities

Center for Transportation, Environment, and Community Health
Final Report

by

Yiye Zhang PhD MS, Mohammad Tarayani PhD, Subhi AlAref MD FACC, Hannah Mitchell MS, Evan Sholle MS, Sajjad Abedian MS, Yongkang Zhang PhD, Arindam RoyChoudhury PhD, James K Min MD FACC, Jessica Ancker MPH PhD, Jyotishman Pathak PhD

February 13th, 2019

DISCLAIMER

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2.Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle
Examining Individual Health and Healthcare Utilization Patterns at the Intersection of Transportation, Environment and Communities

5. Report Date
February 12, 2019

6. Performing Organization Code

7. Author(s)
Yiye Zhang PhD MS, Mohammad Tarayani PhD, Subhi AlAref MD FACC, Hannah Mitchell MS, Evan Sholle MS, Sajjad Abedian MS, Yongkang Zhang PhD, Arindam RoyChoudhury PhD, James K Min MD FACC, Jessica Ancker MPH PhD, Jyotishman Pathak PhD

8. Performing Organization Report No.

9. Performing Organization Name and Address
Department Name: Healthcare Policy and Research Institution Name: Weill Cornell Medicine Institution City, State Zip Code: New York, 10065

10. Work Unit No.

11. Contract or Grant No.
69A3551747119

12. Sponsoring Agency Name and Address
U.S. Department of Transportation 1200 New Jersey Avenue, SE Washington, DC 20590

13. Type of Report and Period Covered
Final Report
1/1/2018 – 12/31/2018

14. Sponsoring Agency Code
US-DOT

15. Supplementary Notes

16. Abstract

A knowledge gap exists in our understanding of the association between individual health, healthcare, transportation, and the environment at a granular level. This study aimed to curate a longitudinal dataset that contains the history of individual demographics, health conditions, healthcare utilization, and community dwellings. Geo-coding residence information to communities in New York City, we matched over 1.5 million individuals with community-level information on travel behavior, active transportation, and built environment.
Three on-going studies are conducted as part of the study. Focusing on a group of heart failure patients, first two studies analyzed detailed patterns of health and healthcare utilization by varying travel behaviors, active transportation, and built environment in a cross-sectional and a longitudinal study, respectively. The third, on-going, study examines the role of social determinants of health in predicting hospital readmission.
Details about the database curated as part of the study is described in the Appendix.

17. Key Words
Electronic health records, heart failure, social determinants of health, built environment

18. Distribution Statement
Public access and for journal submissions

19. Security Classif (of this report)	20. Security Classif. (of this page)	21. No of Pages	22. Price
Unclassified	Unclassified		

Form DOT F 1700.7 (8-69)

Identifying the Association between Built Environment and Heart Failure Progression Using Electronic Health Records

Abstract

Electronic health record (EHR) data have emerged as a longitudinal data source to uncover the associations between the environment and health. In this paper, we aim to identify built environment factors that are associated with heart failure (HF) progression using EHR. A cohort of HF patients in New York City (NYC) and initially had normal ejection fraction were identified. Patients' EHR data were linked with public data on transportation, air quality, land use, and accessibility to identify built environment risk factors for HF progression across NYC and within NY Boroughs using mixed effects models. Increased distance to parks was found to have significant associations with HF progression in NYC, while controlling for demographics, comorbidities, and neighborhood poverty rates. Land use for retail and facility purposes were found to have significant associations within Brooklyn and Bronx. Insights from this study may help identify patients at higher risk for HF progression.

Introduction

Heart failure (HF) is a chronic, progressive condition where the heart cannot pump a sufficient volume of blood to satisfy the body's required blood and oxygen.(1) Risk factors of HF include male sex, high blood pressure, coronary artery disease, diabetes, valvular heart disease, tobacco use, obesity, education level, and socioeconomic deprivation.(2-4) It is estimated that more than one-third of HF patients suffer from comorbidities, including diabetes, obesity, chronic kidney disease, chronic obstructive pulmonary disease, anemia, and sleep apnea.(5, 6) Progression of HF such as declined cardiac function and calcification of the heart is known to be affected by cardiac- and non-cardiac comorbidities.(3, 5) HF is one of the leading causes of morbidity and mortality in the US, and is associated with substantial healthcare expenditure.(7) Global prevalence is estimated to be more than 26 million and increase further with an aging population.(1) The disease manifests with symptoms including a persistent coughing, breathlessness, lower extremity edema, and fatigue. An indicator for diagnosing and managing HF is ejection fraction, a ratio of the amount of blood pumped out to the amount of blood left in the left ventricle with each contraction.(8) Ejection fraction is commonly measured using an echocardiogram test.

HF is known to be associated with environmental risk factors through prior studies conducted using surveys, observational studies, and cohort follow-up studies.(9) Especially well-studied is the association between HF and air pollutants including particulate matter (PM) and nitrogen dioxide. HF incidence has been associated with particulate matter ≤ 2.5 μm in aerodynamic diameter (PM 2.5) in a 4-year prospective cohort study of women across US,(10) and an 11.5-year prospective cohort study in Europe. (11) HF mortality has been associated with PM 2.5 particulate matter ≤ 2.5 μm in aerodynamic diameter in the Cancer Prevention Study II of 1.2 million adults over a 16-year follow up.(12) HF-related hospitalization was associated with air pollutants including PM and nitrogen dioxide in a 12-year follow up study in Pittsburgh, Pennsylvania, as well as an observational study using national Medicare claims data from 1999 to 2002.(13, 14) Transportation-related risk factors for HF have also been reported in previous studies. HF mortality was associated with roadway proximity and noise volume in in 5-year follow up studies in Worcester, Massachusetts, a 9-year cohort study in the Netherlands, and a cross-sectional survey in Toronto, Canada, respectively.(15-17) Outcomes most frequently focused on these previous studies are disease incidence, mortality rates, and hospitalization rates. In comparison, how environmental factors affect the progression of HF is less studied.(18, 19)

Aside from the physical environmental risk factors mentioned above, there remains a knowledge gap on the association between HF and other environmental factors, notably, the built environment factors. Built environment refers to the *human-made* environment through urban planning, ranging from buildings and parks that provide space for human activities, including daily living, spending, occupation, and recreation.(20) The built environment significantly affects public health through its role in providing safe shelter, access to resources such as food and transportation, and space for maintaining a healthy lifestyle. In this study, we propose to study the associations between the built environment and HF, while adjusting for known clinical, social and physical environmental risk factors. We aim to contribute to the evidence that may inform urban planning strategies for healthy heart health.

Moreover, we propose to achieve our study goal using electronic health records (EHRs) as a source of longitudinal health data. By creating a linkage between EHRs and public data sources on the social, physical, and built environment, we aim to demonstrate that EHRs may facilitate an efficient extraction of detailed health information combined with information on the degree of exposure to non-clinical risk factors. In recent years, studies have started to adopt EHRs as a source of longitudinal data to study non-clinical risk factors.(21) Linked with environmental data sources through patients' address information, EHR data have shown promise to help study the associations between social deprivation and cardiovascular diseases,(22) air pollution and cardiovascular events during the labor and delivery,(23) air pollution and asthma,(24) and socioeconomic status and obesity.(25)

Leveraging the 10-year EHR data from a health system in New York City (NYC), in this study, we linked HF patients' longitudinal clinical information in EHRs with public data on air pollution, transportation, land use, and accessibility. The goal of the study is to identify built environment risk factors that are associated with HF progression, notably, deteriorated cardiac function as measured by reduced ejection fraction. Mixed effects models were constructed to identify risk factors for the study cohort across NYC and within NYC Boroughs, respectively. To the best of our knowledge, this study is among the first to use EHRs to study the association of HF progression and built environment risk factors.

This paper is organized as follows. We describe the data sources used in the study and modeling approaches in Methods and Data section. Descriptive statistics of the data and model results are described in Results section. We discuss findings, limitations and future work in Discussion section, and conclude the study in Conclusion section.

Methods and Data

Data were extracted from the EHR at Weill Cornell Medicine and NewYork-Presbyterian Hospital from 2007 to 2018. The data extraction was approved by Weill Cornell Medicine Internal Review Board. EHR data were stored using the (Observational Medical Outcomes Partnership) OMOP common data model.(26) A total of 12801 adult patients with at least one diagnosis of heart failure (ICD-9-CM: 428* or ICD-10-CM: I50*) were identified. From the unstructured notes of the patients' EHR, measurements for ejection fraction were extracted using a natural language processing pipeline at Weill Cornell Medicine.(27) In order to measure HF progression, we excluded 9907 patients who did not have at least two ejection fraction measurements. The average days between patients' first and last ejection fraction measurements is 1781 days, with a standard deviation of 1460 days. We excluded 1239 patients whose first and last EF measurements were taken fewer than one standard deviation above and below the mean difference in days. Furthermore, we removed 415 patients whose initial ejection fraction measurements were below normal as their already declined cardiac functions may lead them to have different lifestyles and affected by the built environment in a different fashion. Finally, 600 patients were excluded from the final study cohort as they did not have valid addresses required for geocoding in the 5 Boroughs of New York: New York, Queens, Bronx, Brooklyn, and Staten Island. A total of 840 patients are left in the study cohort.

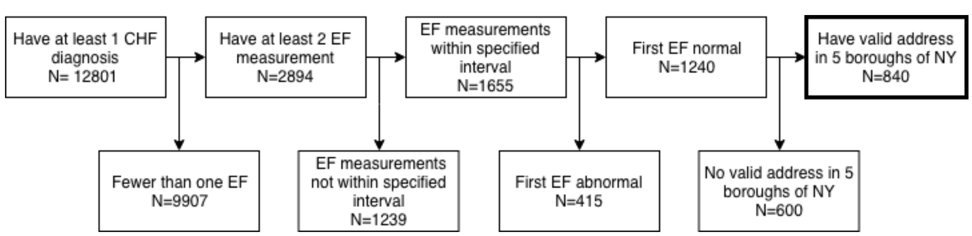


Figure 1. Patient inclusion criteria

Using definitions provided by the American Society of Echocardiography and the European Association of Cardiovascular Imaging (28), ejection fraction measurements were classified into 4 categories: normal (EF>51% in men and EF>53% in women), mildly abnormal (EF within 41% to 51% in men and EF within 41% to 53% in women), moderately abnormal (EF within 30% to 40%), and severely abnormal (EF<30%). Patient inclusion criteria are shown in Figure 1. Aside from ejection fraction, data elements extracted from patients' EHR data include age, sex, race, body mass index (BMI), smoking (yes/no), diabetes (yes/no), valvular heart disease (yes/no), coronary artery disease (yes/no), primary care location, patients' addresses, county of residence, and Federal Information Processing Standard (FIPS) code of residence.

Geospatial analyses were conducted using patients addresses. More than 22% of the patients in our study cohort were found to have multiple addresses. The most recent address for each patient was used for linkage. Four indicators were defined to measure accessibility to public and active transportation and green spaces: distance to the nearest bus stop, distance to the nearest subway station, distance to the nearest park space, and distance to the nearest bike facility. The spatial data were obtained in shapefile formats from the official website of New York state on city planning (https://www1.nyc.gov/site/planning/data-maps/open-data.page). The shapefiles were then intersected with the patients geocoded address to first find the nearest facility using the "nearest" function in ArcGIS and then calculate the distance for every patient. The traffic data were obtained from the New York activity-based travel demand model called the New York Best Practice Model (NYBPM) that includes traffic volume on highways, major arterials, and collector's links along several other transportation measures.(29) The model predicts daily traffic volume in each roadway link for different type of vehicles including passenger vehicles, bus, taxi, and trucks. We grouped the traffic volumes into two groups: Light vehicle duty such as passenger vehicles and taxies, and heavy-duty vehicles such as buses and trucks as their externalities are considered to be different. The vehicle kilometer traveled within the 250, and 500 meters buffers were then calculated.(30) Figure 2 displays displaces to nearest parks and distribution of floor area ratio for retail use mapped across NYC.

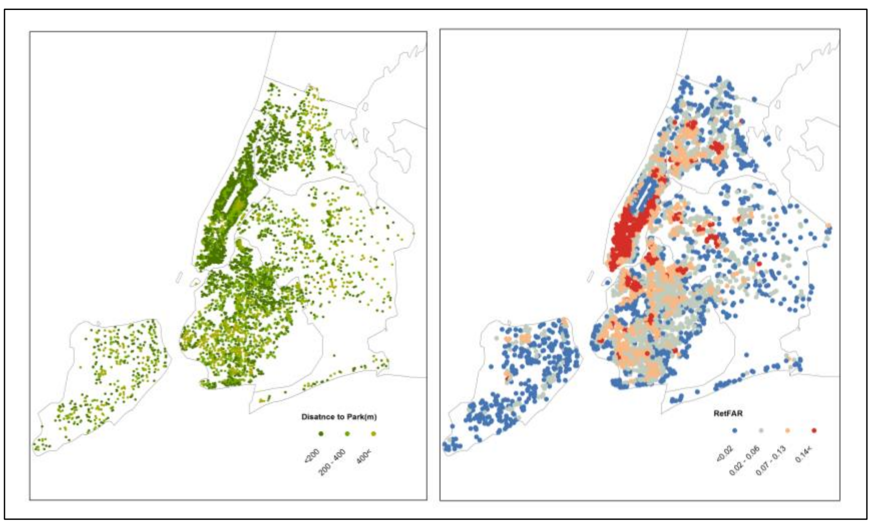


Figure 2. Distances to nearest park from patients' home locations (left) and distribution of floor area ratio for retail use (RetFAR) (right)

Three indicators were defined to measure the role of land use including Land Use Mix (LUM) index, floor area ratio, and street connectivity. Three indicators together measure walkability and availability and variety of resources within 500 meters of each patient's home location. The land use data were extracted from the parcel shapefile from the city planning section of the official website of New York state (https://www1.nyc.gov/site/planning/data-maps/open-data.page) which include information about land use type at the parcel level. The LUM index measures the heterogeneity of land uses around an area of interest and ranges between 0 to 1, where 0 represents homogeneity and 1 represents maximum heterogeneity.(31) Higher LUM values indicate higher walkability of the area and it is believed to have positive impacts on public health. Four types of floor area ratio were computed: retail floor area ratio, residential floor area ratio, commercial floor area ratio, and facility floor area ratio.(32) The floor area ratio is building floor area divided by land area. For example, the areas with a higher share of parking space have lower retail floor area ratio values while areas with smaller setbacks from the street has higher values. The areas with higher floor area ratio are believed to promote walkability. The number of intersections within 500 meters of patients' home location is the third land use indicator used to measure the walkability and connectivity of the neighborhood. As an indicator for street connectivity, the number of intersections was extracted from the transportation network developed for the NYBPM travel demand model.(29) Patient's exposure to two marker air pollutants, PM2.5 and nitrogen dioxide (NO2) were estimated using the Land Use Regression (LUR) model obtained from the Center for Air, Climate and Energy Solutions which estimated the pollutant concentration at the block group level using LUR models.(33) The two air pollutants together could cover both regional and local air pollution hotspots. Figures 1 displays the distances to the nearest park from patients' home locations, and distribution of floor area ratio for retail use in our study cohort.

Clinical and environmental variables above were used to construct mixed effects models with fixed and random effects to elicit factors associated with HF progression. Mixed effects models are extensions to the regression model but allow for hierarchies in the data that arise from data points occurring in groups. The intra- and inter-group variability can be accounted for by designating fixed and random effects.(34) In this paper, the response variable is a binary indicator for disease progression and declined cardiac function, as defined by the reduced EF measurement. Fixed effect variables considered include patient-level information: age, sex, race, body mass index (BMI), smoking (yes/no), diabetes (yes/no), valvular heart disease (yes/no), coronary artery disease (yes/no); and built environment information: floor area ratio for residential use, floor area ratio for facility use, floor area ratio for commercial use, floor area ratio area ratio for retail use, LUM index, number of intersections, daily PM 2.5 concentration (ug/m3), daily NO2 concentration (ug/m3), light-duty vehicle in 250m/500m buffer in kilometer, heavy-duty vehicle in 250m/500m buffer in kilometer, distance (ft) to nearest bus stops, distance (ft) to nearest parks, distance (ft) to nearest subway stops, distance (ft) to nearest bike paths. A random effect for clinics is included in the model to control for the possible care variations across patients' different primary care locations. The model was constructed for all study cohort, and also by NYC Boroughs as separate models. Tests for correlations and multicollinearity among variables were tested using the Variance Inflation Factor. Backward elimination was performed for variable selection. Models were constructed using Stata's generalized structural equation model.(35) All continuous data were standardized by subtracting the data points by the mean and divided by standard deviation. Missing data were imputed using multiple imputations.(36)

Results

Table 1 lists the variable categorized by outcome defined as HF progression. Majority of the patients were hypertensive so we omitted the variable. We tested univariate variable significance with respect to the outcome using Chi-Square test for categorical variables, and analysis of variance (ANOVA) for continuous variables.

Table 1. Descriptive patient characteristics. * indicates p-value <0.05 in Chi-square test or ANOVA

	Progression	
Variable (standard deviation)	No (n=646)	Yes (n=189)
Sex*		
Female	337	65
Male	309	124
Race		
Asian	41	10
Black or African American	112	32
White	233	72
Unknown	101	27
Other	159	48
Age*	74.1 (0.44)	72.2 (0.92)
BMI	36.4 (2.73)	39.5 (8.90)
Smoking		

No	633	186
At least one pack per day	16	5
Valvular heart disease*		
0	251	57
1	398	134
Coronary artery disease		
0	394	114
1	255	77
Diabetes		
0	319	80
1	327	109
Poverty rate*	18.5% (0.5%)	21.5% (1.1%)
floor area ratio for residential use	1.61 (0.060)	1.57 (0.104)
floor area ratio for commercial use	0.12 (0.006)	0.12 (0.011)
floor area ratio for retail use	0.12 (0.006)	0.12 (0.011)
floor area ratio for facility use	0.95 (0.060)	0.86 (0.106)
LUM index	8221463 (378104)	7842812 (677223)
Number of intersections	11.3 (0.31)	11.3 (0.56)
distance (ft) to nearest bus stops	346 (16.4)	324 (19.9)
distance (ft) to nearest subway stops	1869 (92.2)	1773 (130.2)
distance (ft) to nearest parks	704 (20.7)	742 (37.7)
distance (ft) to nearest bike paths	633 (40.0)	642 (64.6)
daily PM 2.5 concentration (ug/m3)	9.21 (0.019)	9.24 (0.036)
daily NO2 concentration (ug/m3)	19.3 (0.105)	19.5 (0.180)
light duty vehicle in 250m buffer in kilometer	26248 (1498)	24685 (2563)
heavy duty vehicle in 250m buffer in kilometer	3427 (179)	3508 (314)
light duty vehicle in 500m buffer in kilometer	258061 (14067)	246423 (23275)
heavy duty vehicle in 500m buffer in kilometer	32247 (1735)	33110 (3192)

Results from the mixed effects model for the entire study cohort are shown in Table 2. As in previous literature, we find that male sex, valvular heart disease, and poverty rate within census tract are adversely associated with HF progression. In terms of patient-level information, we also find that increased age, Asian race and other race are positively associated with HF progression. Increase in the distance (ft) to nearest parks was found to be adversely associated with HF progression.

Table 2. Mixed effects model for progression of ejection fraction. (N=840) *: p-value<0.05

	Odds ratio	P-value	[95% Conf.	Interval]
Male (vs. Female)	1.136	0.000*	1.083	1.191
Race (Base: White)				
Asian	0.914	0.031*	0.842	0.992
Black	0.983	0.590	0.925	1.045
Declined	0.958	0.276	0.887	1.035
Other	0.948	0.047*	0.900	0.999
Age	0.974	0.023*	0.953	0.996
BMI	1.000	0.333	1.000	1.000

Valvular heart disease	1.063	0.013*	1.013	1.115
Coronary artery disease	0.987	0.676	0.931	1.048
Diabetes	1.052	0.084	0.993	1.115
Census tract poverty rate	1.044	0.005*	1.013	1.076
floor area ratio for residential use	0.983	0.401	0.943	1.024
floor area ratio for retail use	1.000	0.988	0.958	1.044
floor area ratio for facility use	0.991	0.590	0.958	1.024
LUM index	1.011	0.628	0.967	1.057
distance (ft) to nearest bus stops	0.999	0.924	0.985	1.013
distance (ft) to nearest subway stops	0.998	0.925	0.955	1.043
distance (ft) to nearest parks	1.025	0.034*	1.002	1.048
distance (ft) to nearest bike paths	0.998	0.879	0.975	1.022
daily NO2 concentration (ug/m3)	1.023	0.237	0.985	1.063
light duty vehicle in 250m buffer in kilometer	0.984	0.341	0.951	1.017
heavy duty vehicle in 250m buffer in kilometer	1.022	0.373	0.975	1.071
_cons	1.118	0.000	1.057	1.183

We performed subgroup analyses for Manhattan, Brooklyn, and Bronx as these three boroughs are considered to have varying neighborhood characteristics and built environments. The number of variables in the mixed effects model was reduced to accommodate for reduced sample size in the subgroup analyses. Results for patients with home addresses in Manhattan are shown in Table 3. Same as the main model, we find that male sex and poverty rate within census tract are adversely associated with HF progression. Asian race remains positively associated with HF progression in addition to declined race information. No built environment factors were found to be associated with HF progression.

Table 3. Within Manhattan: Mixed effects model for progression of ejection fraction. (N=461) *: p-value<0.05

	Odds ratio	P-value	[95% Conf.	Interval]
Male (vs. Female)	1.107	0.004*	1.032	1.187
Race (Base: White)				
Asian	0.865	0.003*	0.786	0.953
Black	0.955	0.416	0.854	1.067
Declined	0.904	0.030*	0.825	0.990
Other	0.943	0.217	0.858	1.035
Age	0.985	0.474	0.945	1.026
BMI	1.000	0.791	0.999	1.001
Valvular heart disease	1.051	0.169	0.979	1.128
Coronary artery disease	1.042	0.232	0.974	1.115
Census tract poverty rate	1.124	0.000*	1.077	1.173
floor area ratio for retail use	0.988	0.551	0.949	1.028
floor area ratio for facility use	1.006	0.734	0.970	1.044
LUM index	1.015	0.515	0.970	1.062

distance (ft) to nearest parks	1.042	0.203	0.978	1.109
_cons	1.219	0.000	1.112	1.336

Results from the mixed effects model for patients with home addresses in Brooklyn are shown in Table 4. Male sex and increased neighborhood poverty rate are adversely associated with HF progression. Asian race and other race are positively associated with HF progression. Increase in the floor area ratio for retail use and LUM index were found to be adversely and positively associated with HF progression, respectively. Increased floor area ratio for retail use indicates a denser retail land use, while an increase in LUM index reflects better walkability in a neighborhood. While not statistically significant, having valvular disease was found to be adversely associated with HF progression with a p-value of 0.051.

Table 4. Within Brooklyn: Mixed effects model for progression of ejection fraction. (N=202) *: p-value<0.05

	Odds ratio	P-value	[95% Conf.	Interval]
Male (vs. Female)	1.149	0.000*	1.081	1.223
Race (Base: White)				
Asian	0.744	0.000*	0.659	0.840
Black	0.976	0.703	0.860	1.107
Declined	1.019	0.835	0.853	1.217
Other	0.859	0.034*	0.746	0.989
Age	0.985	0.731	0.903	1.074
BMI	0.999	0.543	0.995	1.002
Valvular heart disease	1.087	0.051	1.000	1.182
Coronary artery disease	0.954	0.413	0.853	1.068
Census tract poverty rate	1.008	0.769	0.957	1.061
floor area ratio for retail use	1.274	0.009*	1.061	1.529
floor area ratio for facility use	0.923	0.614	0.675	1.261
LUM index	0.510	0.000*	0.385	0.676
distance (ft) to nearest parks	0.989	0.552	0.952	1.027
_cons	0.802	0.132	0.602	1.069

Lastly, results from the mixed effects model for patients with home addresses in Bronx are shown in Table 5. Male sex and having valvular heart disease are adversely associated with HF progression. Increased age, Asian race and other race are positively associated with HF progression. Increase in the floor area ratio for facility use was found to be positively associated with HF progression, while increased LUM index and distance (ft) to nearest parks were found to be adversely associated with HF progression. Same as the main model, increased distance to the nearest parks is also found to be adversely associated with HF progression.

Table 5. Within Bronx: Mixed effects model for progression of ejection fraction. (N=94) *: p-value<0.05

	Odds ratio	P-value	[95% Conf.	Interval]
Male (vs. Female)	1.240	0.004*	1.073	1.433
Race (Base: White)				
Asian	0.831	0.041*	0.696	0.992
Black	1.162	0.036*	1.010	1.337

Declined	1.205	0.209	0.901	1.612
Other	1.231	0.012*	1.047	1.447
Age	0.932	0.033*	0.874	0.994
BMI	0.999	0.198	0.997	1.001
Valvular heart disease	1.271	0.013*	1.052	1.536
Coronary artery disease	0.920	0.115	0.829	1.021
Census tract poverty rate	0.965	0.206	0.914	1.020
floor area ratio for retail use	0.963	0.776	0.740	1.252
floor area ratio for facility use	0.424	0.000*	0.287	0.628
LUM index	4.800	0.000*	2.806	8.212
distance (ft) to nearest parks	1.125	0.003*	1.040	1.217
_cons	1.845	0.008	1.172	2.905

Discussion

We identified and confirmed factors reported from previous literature, including male sex, valvular disease, and poverty rate. In addition, our models consistently found Asian race to be positively associated with HF progression. Several built environment factors were found to be associated with the outcome, including distance to nearest parks, floor area ratio for retail use, floor area ratio for facility use, and LUM index. To the best of our knowledge, these associations have not been reported in previous literature. Interestingly, the patterns of the associations differ between NY Boroughs. In Manhattan, land use variables were not significant factors, possibly due to the similar nature of the land use across Manhattan. In Brooklyn, higher LUM index, an indicator for better walkability was positively associated with HF progression, but a higher floor area ratio for retail use was negatively associated with the outcome. In Bronx, higher LUM index was associated with negative outcome whereas higher floor area ratio for facility use was positively associated. Further investigations on the types of retail stores, such as grocery stores vs. convenience stores, in addition to facility types, may better explain these associations.

There is an opportunity that models created using EHR data and their findings may be integrated with the EHR's clinical decision support system. While healthcare service providers may not be familiar with the individual neighborhoods that patients reside in, information about the land use and availability of resources are readily available through governmental agencies. These insights may allow early identifications of HF patients who can benefit from more monitoring and support to improve healthcare delivery and patient outcomes

Limitations

A number of limitations exist in the study. First, the EHR data used for the study are limited to one health system, although patients may visit multiple outpatient clinics within the system. Thus, it is possible that crucial health information is missing if patients were treated outside the health system during the study period. In addition, while EHR and public data linkage were done using individual home addresses, we are not able to explain the environmental exposure for activities such as work and school outside the residence. In addition, since most recent addresses were used for geocoding, we are not able to account for changes due to moving or study the effect of the moving during the study period. Moreover, there are factors that were not considered in the study, such as patient-level income, family support, occupation, stress level, and living in high vs. low rise floors that may contribute to HF progression. Future work will aim to elicit this information by examining unstructured notes. Furthermore, because of the relatively narrow geographical area covered in this study (NYC), we were not able to identify significant differences in air quality and traffic volume. Future work may use larger national datasets which provide observations across areas with distinct geographical and neighborhood characteristics.

Conclusions

In this study, we demonstrate that EHR data can be linked with public data sources to study the associations of HF progression and built environment factors while adjusting for patient-level, social, and physical environmental factors. This approach may lead to future integration of public data sources with EHR as a form of clinical decision support.

Acknowledgment
We thank Dr. Andrew Danneberg and Ms. Renee Autumn Ray for their valuable feedback on the study.

References

1. Savarese G, Lund LH. Global Public Health Burden of Heart Failure. Card Fail Rev. 2017;3(1):7-11.
2. Ramsay SE, Whincup PH, Papacosta O, Morris RW, Lennon LT, Wannamethee SG. Inequalities in heart failure in older men: prospective associations between socioeconomic measures and heart failure incidence in a 10-year follow-up study. European heart journal. 2014;35(7):442-7.
3. He J, Ogden LG, Bazzano LA, Vupputuri S, Loria C, Whelton PK. Risk factors for congestive heart failure in US men and women: NHANES I epidemiologic follow-up study. Arch Intern Med. 2001;161(7):996-1002.
4. Dunlay SM, Weston SA, Jacobsen SJ, Roger VL. Risk factors for heart failure: a population-based case-control study. Am J Med. 2009;122(11):1023-8.
5. Lang CC, Mancini DM. Non-cardiac comorbidities in chronic heart failure. Heart. 2007;93(6):665-71.
6. Dahlstrom U. Frequent non-cardiac comorbidities in patients with chronic heart failure. Eur J Heart Fail. 2005;7(3):309-16.
7. Writing Group M, Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, et al. Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association. Circulation. 2016;133(4):e38-360.
8. Roger VL. Epidemiology of heart failure. Circ Res. 2013;113(6):646-59.
9. Cosselman KE, Navas-Acien A, Kaufman JD. Environmental factors in cardiovascular disease. Nat Rev Cardiol. 2015;12(11):627-42.
10. Miller KA, Siscovick DS, Sheppard L, Shepherd K, Sullivan JH, Anderson GL, et al. Long-term exposure to air pollution and incidence of cardiovascular events in women. New England Journal of Medicine. 2007;356(5):447-58.
11. Cesaroni G, Forastiere F, Stafoggia M, Andersen ZJ, Badaloni C, Beelen R, et al. Long term exposure to ambient air pollution and incidence of acute coronary events: prospective cohort study and meta-analysis in 11 European cohorts from the ESCAPE Project. BMJ. 2014;348:f7412.
12. Pope CA, Burnett RT, Thurston GD, Thun MJ, Calle EE, Krewski D, et al. Cardiovascular mortality and long-term exposure to particulate air pollution - Epidemiological evidence of general pathophysiological pathways of disease. Circulation. 2004;109(1):71-7.
13. Wellenius GA, Bateson TF, Mittleman MA, Schwartz J. Particulate air pollution and the rate of hospitalization for congestive heart failure among Medicare beneficiaries in Pittsburgh, Pennsylvania. American journal of epidemiology. 2005;161(11):1030-6.
14. Dominici F, Peng RD, Bell ML, Pham L, McDermott A, Zeger SL, et al. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. JAMA : the journal of the American Medical Association. 2006;295(10):1127-34.
15. Beelen R, Hoek G, Houthuijs D, van den Brandt PA, Goldbohm RA, Fischer P, et al. The joint association of air pollution and noise from road traffic with cardiovascular mortality in a cohort study. Occup Environ Med. 2009;66(4):243-50.
16. Medina-Ramon M, Goldberg R, Melly S, Mittleman MA, Schwartz J. Residential exposure to traffic-related air pollution and survival after heart failure. Environ Health Persp. 2008;116(4):481-5.
17. Chum A, O'Campo P. Cross-sectional associations between residential environmental exposures and cardiovascular diseases. Bmc Public Health. 2015;15.

18. Tibuakuu M, Jones MR, Navas-Acien A, Zhao D, Guallar E, Gassett AJ, et al. Exposure to ambient air pollution and calcification of the mitral annulus and aortic valve: the multi-ethnic study of atherosclerosis (MESA). Environ Health-Glob. 2017;16.
19. Roux AVD, Mujahid MS, Hirsch JA, Moore K, Moore LV. The Impact of Neighborhoods on CV Risk. Glob Heart. 2016;11(3):353-63.
20. Perdue WC, Stone LA, Gostin LO. The built environment and its relationship to the public's health: The legal framework. Am J Public Health. 2003;93(9):1390-4.
21. Schinasi LH, Auchincloss AH, Forrest CB, Roux AVD. Using electronic health record data for environmental and place based population health research: a systematic review. Ann Epidemiol. 2018;28(7):493-502.
22. Pujades-Rodriguez M, Timmis A, Stogiannis D, Rapsomaniki E, Denaxas S, Shah A, et al. Socioeconomic deprivation and the incidence of 12 cardiovascular diseases in 1.9 million women and men: implications for risk prediction and prevention. PloS one. 2014;9(8):e104671.
23. Mannisto T, Mendola P, Grantz KL, Leishear K, Sundaram R, Sherman S, et al. Acute and recent air pollution exposure and cardiovascular events at labour and delivery. Heart. 2015;101(18):1491-8.
24. Xie S, Greenblatt R, Levy MZ, Himes BE. Enhancing Electronic Health Record Data with Geospatial Information. AMIA Jt Summits Transl Sci Proc. 2017;2017:123-32.
25. Biro S, Williamson T, Leggett JA, Barber D, Morkem R, Moore K, et al. Utility of linking primary care electronic medical records with Canadian census data to study the determinants of chronic disease: an example based on socioeconomic status and obesity. BMC medical informatics and decision making. 2016;16:32.
26. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. Studies in health technology and informatics. 2015;216:574.
27. Johnson SB, Adekkanattu P, Campion TR, Jr., Flory J, Pathak J, Patterson OV, et al. From Sour Grapes to Low-Hanging Fruit: A Case Study Demonstrating a Practical Strategy for Natural Language Processing Portability. AMIA Jt Summits Transl Sci Proc. 2018;2017:104-12.
28. Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. Eur Heart J Cardiovasc Imaging. 2015;16(3):233-70.
29. Vovsha P, Petersen E, Donnelly R. Microsimulation in travel demand modeling: Lessons learned from the New York best practice model. Transportation Research Record: Journal of the Transportation Research Board. 2002(1805):68-77.
30. Karner AA, Eisinger DS, Niemeier DA. Near-Roadway Air Quality: Synthesizing the Findings from Real-World Data. Environ Sci Technol. 2010;44(14):5334-44.
31. Frank LD, Sallis JF, Conway TL, Chapman JE, Saelens BE, Bachman W. Many pathways from land use to health: associations between neighborhood walkability and active transportation, body mass index, and air quality. Journal of the American planning Association. 2006;72(1):75-87.
32. Barr J, Cohen JP. The floor area ratio gradient: New York City, 1890–2009. Regional Science and Urban Economics. 2014;48:110-9.
33. Muller NZ. ECONOMICS Boosting GDP growth by accounting for the environment. Science. 2014;345(6199):873-4.
34. Winter B. A very basic tutorial for performing linear mixed effects analyses. arXiv preprint arXiv:13085499. 2013.
35. Huber C. Introduction to Structural Equation Modeling Using Stata. California Association for Institutional Research. 2014.
36. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int J Meth Psych Res. 2011;20(1):40-9.

Study 2: Policy Lessons on Urban and Transportation Planning: New Evidences on the Impacts of Environmental and Social Determinants on Heart Failure in New York City

Abstract

Background: Health concerns are one of the main challenges in the world's agenda to tackle. Where people live could both, directly and indirectly, affect their well-being. The role of the built environment on the incidence of NCDs and in particular cardiovascular diseases, therefore, has drawn attention from researchers. However, there is no comprehensive investigation on the role of built environment factors on health outcomes that could help to highlight the importance of integrated health and urban planning.
Objectives: We aim to evaluate the role of built environment in public health in a more accurate and comprehensive way than the existing studies; and providing urban policy implications that aimed improving public health. The study also intends to find what built environment factors should be planned for in the favor of improving public health
Results: Among the built environment factors, land use, traffic, and air pollution measures significantly increase the risk of death in heart failure patients by 47.2%, 35% and 14%, respectively. Surprisingly proving access to public transit, green space, and active transportation do to significant attenuate odds of death in HF patients. The confiding factors such as household income play a vital role in finding the
Discussions: The findings imply that the smart growth strategies including compact development do not necessarily improve public health. the results have two major implications; first, combining smart growth strategies with technology solutions such as fleet electrification in order to balance out the benefits of placing people close to high amount of pollution; second, urban infrastructure can enhance public health if being blended at the early planning stages so the complex interactions being accounted for.

Key words: Public Health, Built Environment, Heart Failure, Urban Infrastructure Systems

Introduction

Health concerns are one of the main challenges in the world's agenda to tackle. The global population will reach to 10 billion by 2050 and with 22% of them expected to be above 65 years old, a significant increase from 8% in 2015 (United Nation), health challenges will draw more attention in the future world. Noncommunicable diseases (NCDs) are the leading cause of death around the world by causing about 40.5 millions out of the 56.9 million annual deaths in 2016, an 33% increase from 30.1 million in 2000 (World Health Organization, 20180). The cumulative costs of NCDs in the 20 years from 2011 to 2030 could reach to $47 trillion dollar (Bloom et al., 2012). Among NCDs, cardiovascular diseases are the leading contributor that count for 17.9 million annual deaths. Preventive measures, therefore, have drawn researchers' attention to improve public well-being and avoid tremendous costs on health system.

There exist two types of the risk factors: unmodifiable factors such as age, gender, race and ethnicity, and family history, and modifiable risk factors such as obesity, tobacco and alcohol use, environmental factors such as exposure to air and noise pollution, and access to healthy food. As an example of an unmodifiable factor, previous study found a higher 40-year cumulative incidence of hypertension as a cardiovascular risk factor among Black versus Asian people, 92.7% and 84.1%, respectively (Benjamin et al., 2017). However, it is reported that 70% of cardiovascular diseases are attributable to modifiable risk factors (Ezzati and Lopez, 2003). For example, Roux Av et al., 2001 (Diez Roux, 2001) found that low income people, regardless of their race, have 2 to 3 times greater risk of coronary heart disease.

Where people live could both, directly and indirectly, affect their well-being. Exposure to air, noise pollution and socioeconomic status of a neighborhood are among built environment factors that directly cause several negative health outcomes such as lung cancer, asthma, and birth defects (HEI, 2010; Hoffmann et al., 2007; Tonne et al., 2007). For instance, Chum and O'Campo (Chum and O'Campo, 2015) found that violent crimes, noise, and proximity to major roads increase the odds of cardiovascular diseases after controlling for smoking, drinking, age, gender, education, income, physical activity, and BMI. On the other hand, built environment can indirectly deteriorate public health. Physical inactivity, for example, is one of the factors believed to be responsible for range of NCDs such as obesity, diabetes, and cardiovascular diseases. Factors such as density of fast food restaurant, grocery stores, walkability of neighborhood, access to the green space areas, and the type of land use in the neighborhood are believed to affect physical activity. Casagrande et al., showed that people who are exposed to traffic, have access to proper sidewalks, and are safe from crime have more chance of being physically active. Other built environment factors are believed to affect the risk of cardiovascular diseases including street lighting at night, unattended dogs, places within walking distance, places to exercise, hills, enjoyable scenery, access to services, residential density, street connectivity, land use mix, characteristics of neighborhood, store density, population density, and access to public transit (Adams et al., 2012; Casagrande et al., 2009; Renalds et al., 2010; Witten et al., 2012).

The role of the built environment on the incidence of NCDs, therefore, has drawn attention from researchers (Malambo et al., 2016). The built environment policies are even believed to be more effective than individual factors in promoting public health (Diez Roux, 2003). Transportation and land use policies in urban areas are therefore among the levers that could be employed by policymakers to improve public health. Land use can both directly and indirectly affect public health. Residential and job locations can affect public health by reducing the need to travel and thus affect air quality and risk of vehicle accidents and also increasing active mode shares. For instance, Brown et al (Brownson et al., 2009) has shown that a quartile increases in land use mix index that indicates walkability of a neighborhood is associated with 12.2% reduction in the risk of obesity. Powell et al., (Powell et al., 2007) showed that increasing the chain supermarket outlet by one unit per 10,000 capita, can reduce BMI by 0.11 units, but one unit increase in convenience store per 10,000 capital can increase BMI by 0.03 units. Women who have access to fast food restaurants in their neighborhood have 13% higher stroke risk (Hamano et al., 2013).

The more the vehicle mode share dominates the transportation systems, the less people walk and bike for their daily trip. On the other hand, domination of single occupancy vehicles in urban transportation exacerbates the air and noise pollution and road injuries, it can decrease the physical activity by discouraging walkability (Srinivasan et al., 2003). Transportation is responsible for a significant part of ambient air pollution in urban areas and exposure to vehicle emission has been associated with several health outcomes such as asthma in children and preterm labor, cardiovascular diseases, respiratory disease and lung cancer. In particular, the report by Collaborative on Health and the Environment (CHE) finds strong evidence on exposure to particulate air pollution and cardiovascular disease. Kan et al., (Kan et al., 2008) find that high traffic density and distance to major roads are associated with coronary heart diseases and in particular those who live within 300m of major roads has 12% more chance of developing CHD. Previous research has shown that access to transit systems or access to active transportation infrastructure such as proper sidewalks or bicycle paths can damp the probability of using personal vehicle for daily commuting trips. Therefore, promoting active transportation not only can reduce obesity and diabetes by increasing daily activity, but also can play an important role in improving air quality. For instance, one hour increase in daily driving can increase the risk of obesity by 6% (Frank et al., 2004).

Despite the discussed evidences on the association between built environment and health and in particular cardiovascular diseases, still there is a need for more rigorous empirical evidence to support the need for policy changes including hypothetical testing on the association and measurement improvements (Diez Roux, 2003). Existing literature mostly focus on the association between physical activity, active transportation, and obesity, but not the direct impact of built environment on cardiovascular diseases. There is no comprehensive investigation on the role of built environment factors on cardiovascular diseases (Chum and O'Campo, 2015). Prior studies mostly focus on one aspect of built environment factors, such as traffic air pollution. The literature survey by Malambi et al., revealed that only 17% of existing studies devoted to the direct impacts of built environment on cardiovascular diseases (Malambo et al., 2016). Beside the lack of direct investigation on the role of built environment, the few available studies also face technical imperfection. They may overlook the impact of the built environment factors on public health by using low resolution data for both the dependent and independent variables. The combination of low-resolution data at both ends can propagate the errors and mask the true impacts. This is also important since some factors may affect the health condition through a direct-causation relation, others may play a confounder role. While the existing studies usually obtained patients' data at the coarse levels such as zip code level, they also estimate the independent variables' values such as access to sidewalks or exposure to vehicle emission at low resolution scale. For instance, Jerrette et al., (Jerrett et al., 2009) used postal code addresses in Toronto among with LUR model to predict air pollution; Finkelstein et al (Finkelstein et al., 2004) used postal codes to derive mean air pollution and proximity to road; and Beelen et al (Beelen et al., 2008) used home address for patients in Netherlands (1987-1996).

The contribution of our study is twofold: evaluating the role of built environment in public health in a more accurate and comprehensive way than the existing studies; and providing urban policy implications that are aimed at improving public health. We aim to find what built environment factors should be planned for in the favor of improving public health. For this purpose, we evaluate how built environment and social determents are associated with the death of heart failure (HF) patients in New York area. Finding the significant associations, we then propose urban policies that could help to boost public health.

MATERIAL and METHODS

We studied how built environment affect the risk of death in HF patients. two types of risk factors: individual and built environment factors.

Study Population

Study population were extracted from the electronic health record (EHR) at Weill Cornell Medicine and New York-Presbyterian Hospital from 2012 to 2017 on 12610 adult patients with at least one diagnosis of heart failure (ICD-9-CM: 428.*). The dataset initially included age, gender, smoking status, street address, ethnicity, and BMI. From the initial datasets 10,630 were geocoded by converting the street location address to longitude and latitude, and the rest where exclude from the dataset due to lack of a valid address. The dataset then was modified according to the covariate under study and thus number of observations then were excluded because of the wrong data entry for different covariates. The spatial analyses were conducted in ArcMap 10.5 and statistical analyses were performed with R-3.4.4.

Individual Covariates

Besides age, gender, ethnicity, smoking, and BMI covariates from EHR, the rest of socioeconomic data including median household income, and violent crime rate is obtained from TIGER products of US Census Bureau data at the block group level. To obtain data, the patients address layer is overlapped with the block group shapefile. The patients' median household income is assumed to be equal to the median household income of the block group in which the person resides. The TIGER data provides the percentage of people in each block with a specific level of education: no school, high school but no degree, high school or General Educational Development (GED), some college degree, college degree, master's degree and above, and others. To obtain patients' education level, two categories are defined: primary education and college or higher. The patient is assumed to be in one of the two categories if the resident's block has higher than 50% of each of the two categories. During our analysis, we removed those patients with BMI higher than 60 and those who death date was before 2012, which we believe was due to error in data entry. Death is obtained using the social security death index.

Built Environment Covariates

Accessibility Measures:

Four indicators where defined to measure accessibility to public and active transportation and green spaces: distance to the nearest bus stop, distance to the nearest subway station, distance to the nearest park space, and distance to the nearest bike facility. The spatial data were obtained in shapefile formats from the official website of New York state ("The Official Website of New York State," 2018). The shapefiles were then intersected with the patients geocoded address to first find the nearest facility using the "nearest" function in ArcMap and then calculate the distance to it for every patient.

Exposure to Traffic Measures

The traffic data were obtained from the New York activity-based travel demand model called the New York Best Practice Model (NYBPM) that includes traffic volume on highways, major arterials, and collectors' links along several other transportation measures. The model predicts daily traffic volume in each roadway link for different type of vehicles including passenger vehicles, bus, taxi, and trucks. The externalities from light and heavy-duty vehicles are unalike, and literature suggest separating them to study their negative health impacts. Thus, we grouped the traffic volumes into two groups: light vehicle duty (passenger vehicles and taxies), and heavy-duty vehicles (buses and trucks). The Vehicle Kilometer Traveled within the 250, and 500 meters buffers (Karner et al., 2010) were then calculated.

Walkability

Three indicators were defined to measure the role of land use on risk of death in heart failure patients including Land Use Mix (LUM) index, retail floor area ratio, and street connectivity. Three indicators together measure walkability and availability and variety of destinations within 500 meter of each patient's home location. The land use data were extracted from the parcel shapefile from the official website of New York state which include information about land use type at parcel level. The LUM index measures the heterogeneity of land uses around an area of interest and ranges between 0 to 1, where 0 represents homogeneity and 1 represents maximum heterogeneity. Higher LUM values indicates higher walkability of the area and it is believed to have positive impacts on public health.

$$LUM = \frac{-\sum_{i=1}^{n}(P_i \ln P_i)}{\ln N}$$

where i is the land use categories, and P is the proportion of the land area of each land use category, and N is the number of land use categories.

The Retail Floor Area Ratio (RetFAR) is retail building floor area divide by retail land area. The areas with higher share of parking space have lower RetFAR values while areas with smaller setbacks from the street has higher values. The areas with higher RetFAR are believed to promote walkability and, therefore, improve public health. The number of intersections is the third land use indicator used to measure the walkability of the neighborhood that could affect heart failure. The number of intersections where extracted from the transportation network developed for the NYBPM travel demand model.

Air pollution

Two methods were deployed to estimate patient's exposure to two marker air pollutants: Land Use Regression (LUR) model and air monitoring model; the PM2.5 which is believed to affect human health and NO2 as a marker for traffic pollution. The two air pollutants together could cover both regional and local air pollution hotspots. Both PM2.5 and NO2 estimates were obtained from the Center for Air, Climate and Energy Solutions (Kim et al., 2018) which estimated the pollutant concentration at the block group level using LUR models. We also use PM2.5 from air monitoring stations as the most common method used in epidemiological studies to estimate impact of air pollution on public health. Besides the differences that the two methods may cause in our analysis, we were interested to evaluate if air quality methods could alter the epidemiology outcomes.

Statistical analysis

The association between exposure to different kinds of built environment factors and risk of death in heart failure patients first is estimated using the unadjusted odds ratio for different ranges of exposure. Except for gender, BMI, ethnicity, education and accessibility measure the quartiles of exposures are used to calculate the odds ratios while the first quartile is assumed to be the base condition. While gender, education and ethnicity were modeled as binary exposure variables, the biomedical definition of obesity is used to estimate odd ratio for different BMI values. For accessibility measures, 200 and 400 meters were considered as exposure threshold since they are plausible walking distance to get access to those facilities. Considering that the patients' information has different sources, the controlling could reduce the size of the sample due to unavailability of data at different levels. Therefore, the second model calculates the adjusted odds ratio controlling for ages, BMI, gender, and smoking status. The fully adjusted model with 5,402 observation calculates the odds ratio controlled for age, gender, smoking, BMI, ethnicity, income, crime, education using the regression modeling.

Results

The patients profile reveals that more than the half of the patients, 55%, are men, only 10% of them are Latino, most of them have some level of higher education, and about 34% of them are smokers (Table 1). Expectedly, the HF patients are older than the average population, about 71 years old. The median patient's household income is higher than the average New York state population and also than the average US population as well, $70,009 versus $61,741, $60,052, respectively. On average, the patients are considered overweight with the BMI of 28.5. The majority of the patients live in a close proximity of the public and active transportation facilities, considering the 200 m walking threshold. Living in one the most congested areas in the world, make the patients' average air pollution close to the US EPA PM2.5 10ug/m3 thresholds.

Table 1. Descriptive Statistics

	Number	**%**
Gender(n=9908)		
Male	5750	55.03%
Female	4158	41.97%
Ethnicity(n=9908)		
Latino	1028	10.38%
Not Latino	8880	89.62%
Education(n=5402)		
Primary Education	791	14.64%
College and Higher	4611	85.36%
Smoking(n=9908)		
Smokers	3363	33.94%
Non-Smokers	6545	66.06%
	Mean	**95% CI**
Individual		
Age*(n=8951)*	71.03	(70.77, 71.30)
Annual Household Median Income ($) *(n=5402)*	70,009	(69,093, 70,926)
BMI*(n=9274)*	28.5	(28.4, 28.6)
Accessibility (Meter)*(n=7812)*		
Distance to the nearest Bike facility (ft)	229.4	(221.9 236.9)
Distance to the nearest Subway Station (ft)	634.7	(616.1, 653.2)
Distance to the nearest Bus Station (ft)	109.5	(106.8, 112.1)
Distance to the nearest Park Space (ft)	222.5	(218.9, 226.0)
Transportation*(n=10360)*		
Light Duty Vehicles VKT in 250 m buffer	1,8510	(17910, 19113)
Heavy Duty Vehicles VKT in 250 m buffer	2,378	(2299, 2457)
Light Duty Vehicles VKT in 500 m buffer	178,862	(172983, 184741)
Heavy Duty Vehicles VKT in 500 m buffer	22,140	(21375, 22906)
Land Use (Square Mile) *(n=7839)*		
Building Area	1.67	(1.62, 2.06)
Commercial Area	0.64	(0.62, 0.67)
Residential Area	0.99	(0.97, 1.02)
Office Area	0.28	(0.27, 0.30)
Retail Area	0.07	(0.07, 0.08)
Storage Area	0.011	(0.011, 0.012)
Factory Area	0.004	(0.004, 0.005)
Ret Far	0.120	(0.115, 0.123)

LUM	0.599	(0.596, 0.604)
Number of Intersections in 500 m	10.7	(10.5, 10.8)
Air pollution*(n=10360)*		
PM2.5-Air Quality Monitoring (ug/m3)	7.94	(7.92, 7.96)
PM2.5- LUR Model (ug/m3)	8.96	(8.94, 8.97)
NO2- LUR Model (ug/m3)	16.72	(16.63, 16.82)
Safety*(n=5402)*		
Total Housing Violation per 1,000 Population	0.56	(0.54, 0.58)
Felony per 1,000 Population	104.95	(103.25, 106.65)

Except for the patients in the highest income quartile, the rest of socioeconomic characteristics are not significantly associated with risk of death in HF patients. Interestingly, individuals with very low BMI had a 50% higher risk of death but those with higher than average BMI had a 26%-33% lower risk of death. This is similar to the findings by Lavie et al., (Lavie et al., 2009), where they reported higher survival rate for cardiovascular patients with higher BMI.

Surprisingly, accessibility to public transportation, green space, or bike facilities were not correlated with risk of death in HF patients. HF patients who live in areas with the highest light and heavy-duty vehicles activity within 500 meters buffer around their residence have faced significantly higher risk of death, between 11% to 16%.

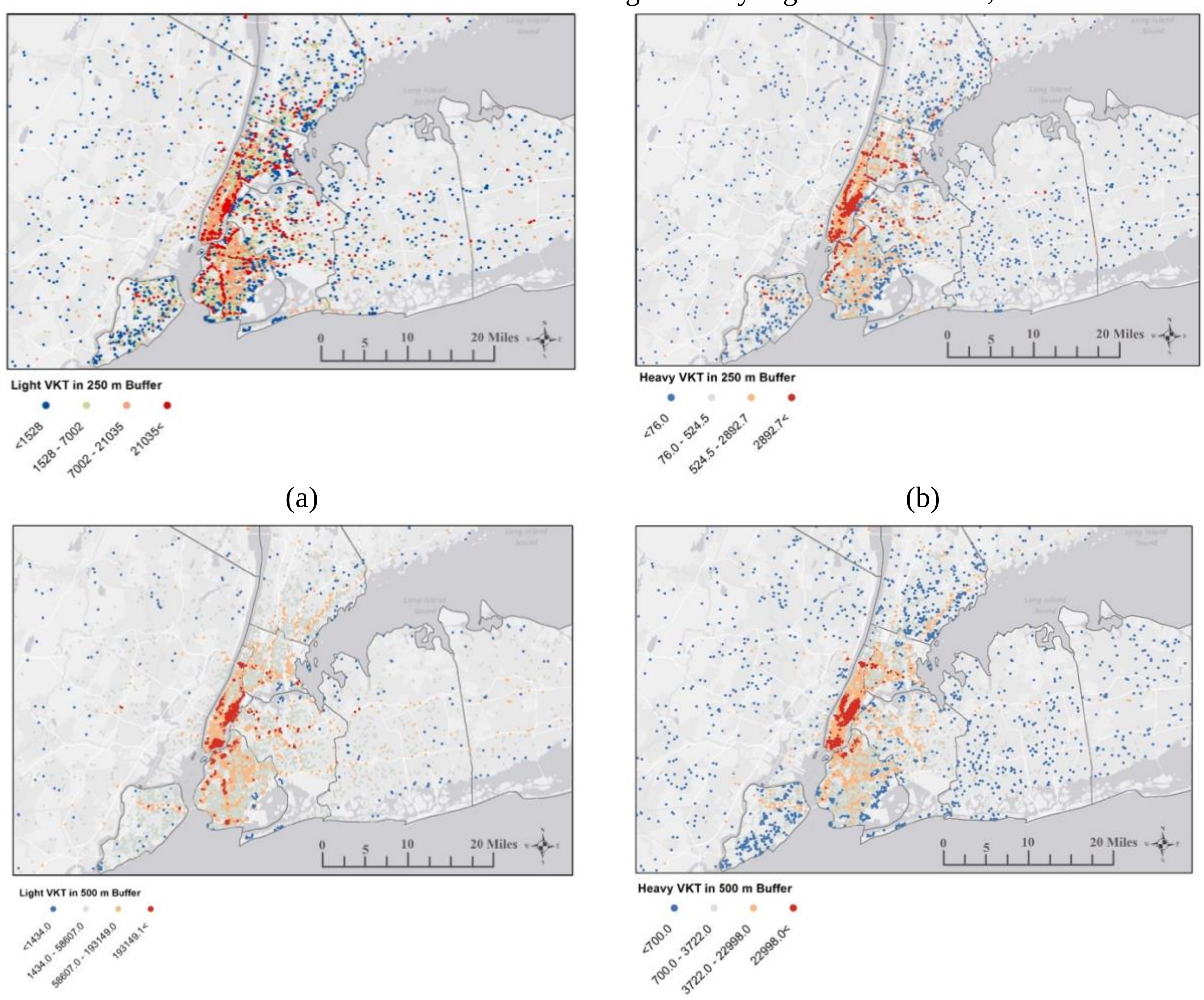
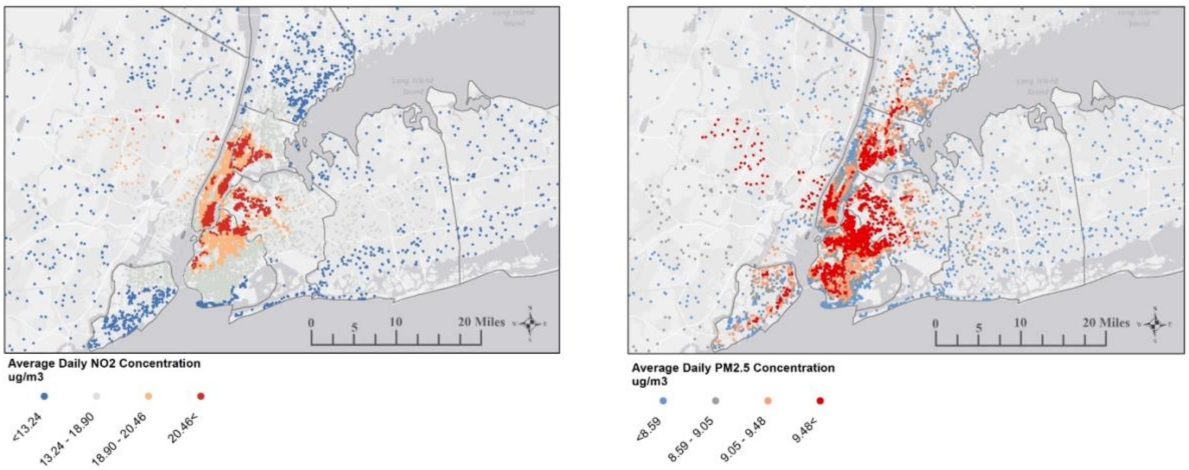


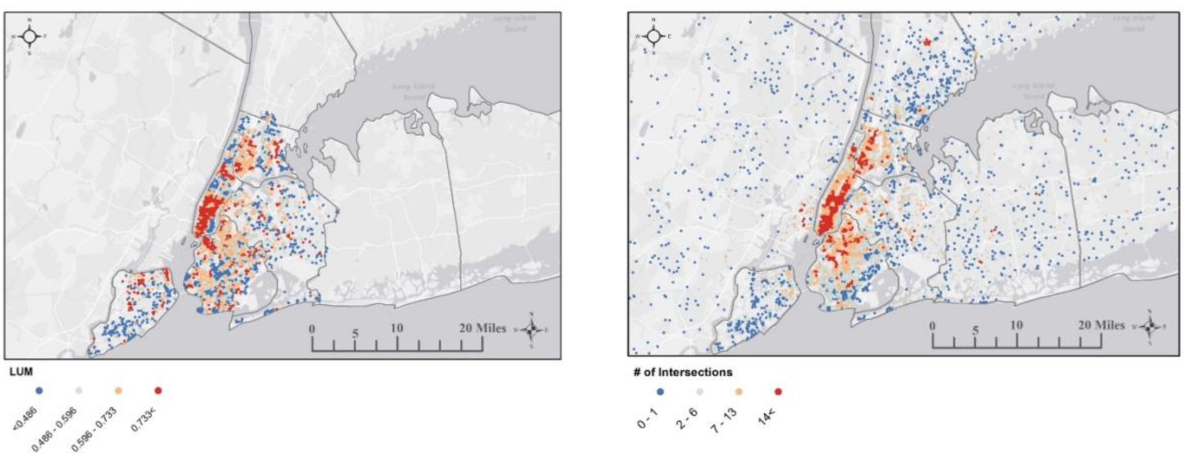
Figure 1 The Vehicle Kilometer Traveled; a) light duty vehicle in 250 m buffer, b) heavy duty vehicle in 250 m buffer, c) light duty vehicle in 500 m buffer, d) heavy duty vehicle in 250 m buffer

More importantly, those who live in the core urban areas with higher LUM index and higher RetFAR, have significantly higher risk of death, 21% and 29%, respectively. Furthermore, we find that there exists a significant association between NO2 concentration and heavy vehicle activity within 500-meter buffers with Pearson correlation of 0.43, and between RetFAR and heavy VKT in 500-meter buffer r=0.47. However, finding no significant ORs for air pollution variables while finding significant correlation between air pollution, land use and transportation covariates, suggest the existence of confounder variables.

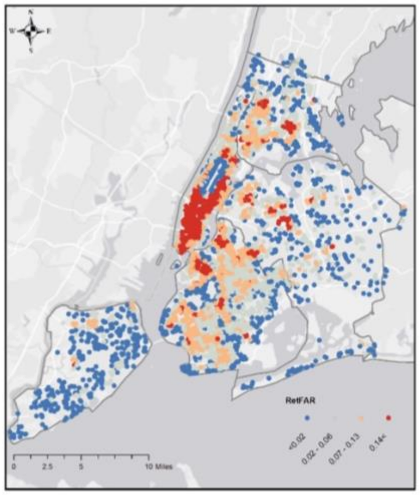


(a) (b)

Figure 2 The average daily air pollution concentration; a) NO2, b) PM2.5



(a) (b)



(C)

Figure 3 The Walkability measures; a) Land Use Mix Index, b) # of Intersections, c) RetFAR

Table 2. Odds Ratio of Death in Heart Failure Patients

	Dead	Alive	Unadjusted ORs (95% confidence interval)
Socioeconomic			
Gender (n = 9908)			
Female	703	3455	1
Male	915	4835	0.93 (0.84, 1.04)
BMI (n = 9274)			
=<18.5	53	147	1.50 (1.08, 2.09)
18.5 – 24.9	545	2276	-
24.9 – 29.9	460	2605	0.74 (0.64, 0.84)
>= 29.9	438	2750	0.67 (0.58, 0.76)
Ethnicity (n=9908)			
Not Hispanic or Latino	1464	7416	1
Hispanic or Latino	154	874	0.89 (0.74, 1.07)
Education(n=5402)			
College or Higher	691	3920	1
Preliminary Education	108	683	1.11 (0.90, 1.39)
Smoking (n=9908)			
Non-smokers	12	49	1
Smokers	702	3934	0.73 (0.39, 1.38)
Median Household Income(n=5402)			
Less than $37,877	185	1184	1
$37,877-$66,172	185	1157	1.02 (0.82, 1.27)
$66,172-$103,797	206	1102	1.19 (0.97, 1.48)
More than $103,797	223	1160	**1.23 (1.00, 1.52)**
Accessibility*(n=7812)*			
Distance to the Nearest Bus Stop (m)			
=< 200	1217	5784	1
200 - 400	106	627	0.80 (0.65, 1.00)
>= 400	12	102	0.56 (0.30, 1.05)

Distance to the Nearest Subway Station (m)			
=< 200	223	1175	1
200 - 400	470	2120	1.17 (0.98, 1.39)
>= 400	642	3182	1.06 (0.93, 1.21)
Distance to the Nearest Park (m)			
=< 200	725	3470	1
200 - 400	417	2089	0.96 (0.84, 1.09)
>= 400	193	918	1.00 (0.83, 1.21)
Distance to the Nearest Bike Facility			
=< 200	925	4404	
200 - 400	189	974	0.92 (0.78, 1.10)
>= 400	221	1099	0.96 (0.77, 1.18)
Transportation*(n=10360)*			
Light Duty Vehicles VKT in 250 m buffer			
=< 1528	457	2133	1
1528 – 7002	410	2180	0.88 (0.76, 1.02)
7002 - 21035	459	2131	1.01 (0.87, 1.16)
>= 21035	495	2095	1.10 (0.96, 1.27)
Heavy Duty Vehicles VKT in 250 m buffer			
=<76	466	2124	1
76-524.5	424	2166	0.89 (0.77, 1.03)
524.5-2892.7	433	2157	0.91 (0.79, 1.06)
>=2892.7	498	2092	1.08 (0.94, 1.25)
Light Duty Vehicles VKT in 500 m buffer			
=< 14343	460	2131	1
14343- 58607	408	2181	0.87 (0.75, 1.00)
58607-193149	436	2154	0.94 (0.81, 1.08)
>=193149	517	2073	**1.16 (1.00, 1.33)**
Heavy Duty Vehicles VKT in 500 m buffer			
=< 700	469	2121	1
700- 3722	416	2174	0.86 (0.75, 1.00)
3722- 22998	424	2166	0.89 (0.77, 1.02)
>=22998	512	2078	**1.11 (0.97, 1.28)**
Air Pollution*(n=10360)*			
Average Daily PM2.5 Concentration (μg/m3) (Monitoring Stations)			
=<7.2	330	1540	
7.2- 7.5	450	2313	0.91 (0.78, 1.06)
7.5- 8.8	526	2454	1.00 (0.86, 1.16)
>=8.8	515	2232	1.07 (0.92, 1.25)
Average Daily PM2.5 Concentration (μg/m3)			
=<8.59	472	2125	1
8.59- 9.05	479	2106	1.02 (0.89, 1.18)
9.05- 9.48	436	2154	0.91 (0.79, 1.05)
>=9.48	434	2154	0.91 (0.79, 1.05)
Average Daily NO2 Concentration (μg/m3)			
=<13.24	465	2148	1
13.24 – 18.90	431	2136	0.93 (0.81, 1.08)
18.90 -20.46	436	2154	0.93 (0.81, 1.07)

>=20.46	434	2154	1.08 (0.94, 1.24)
Land Use			
Ret FAR(n=7839)			
=<0.023	307	1653	1
0.023-0.061	321	1638	1.06 (0.89, 1.25)
0.061-0.128	332	1628	1.10 (0.93, 1.30)
>=0.128	379	1581	**1.29 (1.09, 1.52)**
# of Intersection in 500 m buffer(n=10360)			
=<1	369	1535	1
1-6	510	2581	0.82 (0.71, 1.95)
6-13	389	2093	0.77 (0.66, 0.91)
>=13	553	2330	0.99 (0.85, 1.14)
LUM(n=7839)			
=<0.486	317	1637	1
0.486 – 0.596	320	1640	1.01 (0.85, 1.19)
0.596 – 0.733	330	1640	1.04 (0.88, 1.23)
>=0.733	372	1583	**1.21 (1.03, 1.43)**
Crime *(n=5402)*			
Total Housing Violation per 1,000 Population			
=<0.1	212	1227	1
0.1 -0.29	220	1113	1.14 (0.93, 1.40)
0.29 – 0.61	202	1102	1.06 (0.86, 1.31)
>=0.61	165	1161	0.82 (0.66, 1.02)
Felony per 1,000 Population			
=<51	193	1111	1
51-85	211	1140	1.07(0.86, 1.32)
85-134	196	1173	0.96 (0.78, 1.19)
>=134	199	1179	0.97 (0.78, 1.20)

To control for the potential confounders effects, we then estimate the odds ratio adjusted for age, gender, ethnicity, smoking, and BMI. Table 2 shows that there are only two factors significantly affecting the risk of death in HF patients. Increasing number of intersections by 1 increases the risk of death in HF patients by 1.2%. Surprisingly, increasing the PM2.5 concentration is associated with lowering the risk of death in HR patients. It probably happens because not controlling for income, those who live in the core Manhattan are more likely to be wealthier and have lower risk of death. Comparing the built environment maps also reveal a dissimilarity between the pattern in PM2.5 concentration versus the rest of the built environment measures. Higher Pm2.5 concentration in the Kings county is unique compared to the No2 concentration, traffic activities, LUM, intersection, and RetFAR.

Table 3. Odds Ratio of Death in Heart Failure Patients Adjusted for Age, Gender, BMI, Smoking

	Coefficient	**OR**	**Pr (>|z|)[d]**
Built Environment Models Adjusted for SES			
Accessibility[a]			
Distance to the Nearest Bus Stop	-0.01955	0.994(0.987, 1.002)	0.125
Distance to the Nearest Subway Station	-0.00154	0.999(0.998, 1.001)	0.294
Distance to the Nearest Park	-0.00311	0.999(0.995, 1.003)	0.674
Distance to the Nearest Bike Path	-0.00471	0.998(0.996, 1.001)	0.200
Transportation[b]			
Light Duty Vehicles VKT (250m)	1.348	3.991(0.605, 26.34)	0.151
Heavy Duty Vehicles VKT(250m)	5.223	185.49(0,4.5e08)	0.487

Light Duty Vehicles VKT(500m)	0.183	1.201 (0.991, 1.454)	0.062
Heavy Duty Vehicles VKT(500m)	1.265	3.543(0.780, 16.086)	0.101
Land Use			
Ret Far	0.348	1.417(0.952, 2.109)	0.086
LUM[c]	0.372	1.451(0.931, 2.263)	0.100
Number of Intersections	0.012	1.012(1.003, 1.022)	**0.007**
Air pollution			
NO2- LUR Model (ug/m3)	-0.010	0.990 (0.978, 1.003)	0.127
PM2.5- LUR Model (ug/m3)	-0.108	0.897(0.829, 0.971)	**0.007**
PM2.5- Monitoring (ug/m3)	0.0431	1.044 (0.965, 1.129)	0.283

[a] Odds ratio associated with 100 meter change in accessibly, [b] Odds ratio associated with 1,000,000 VKT change in traffic activity, [d] P-values in **bold** are significant at 95%

The third model calculates the odds of death in HF patients while adjusted for age, gender, smoking, BMI, income, ethnicity, and crime. The number of observations lowered to 5,204 so we could control for all the potential confounding factors. While the measures of accessibility still do not show a significant association with the outcomes, all traffic measures except for one are significantly associated with higher risk of death in HF patients. As the VKT of heavy-duty vehicles increase by 1,000,000 within the 500 meters buffer, the risk of death increases by 1350%. Number of intersections within the 500 meters buffer is the only land use measures that significantly increase the risk of death in HR patients. The NO2 estimated by land use models shows somewhat significant association with HR patients, the modeled PM2.5 is not significantly associated with higher risk of death.

Table 4. Odds Ratio of Death in Heart Failure Patients Adjusted for Age, Gender, BMI, Smoking, Household Income, Education, Ethnicity, Neighborhood Crime

	Coefficient	OR	Pr (>|z|)
Built Environment Models Adjusted for SES			
Accessibility			
Distance to the Nearest Bus Stop	-0.00177	0.995(0.987, 1.002)	0.173
Distance to the Nearest Subway Station	-0.00127	0.999(0.998, 1.001)	0.404
Distance to the Nearest Park	-0.00299	0.999(0.995, 1.004)	0.697
Distance to the Nearest Bike Path	-0.00506	0.998(0.996, 1.001)	0.190
Transportation			
Light Duty Vehicles VKT(250m)	2.267	9.650(1.267, 73.520)	**0.029**
Heavy Duty Vehicles VKT(250m)	13.48	714972(0.025, 2 e13)	0.12
Light Duty Vehicles VKT(500m)	0.305	1.356(1.098, 1.676)	**0.005**
Heavy Duty Vehicles VKT(500m)	2.606	13.545(2.205, 83.192)	**0.005**
Land Use			
Ret Far	0.386	1.472(0.928, 2.333)	0.100
LUM	0.327	1.386(0.852, 2.256)	0.189
Number of Intersections	0.014	1.013(1.003, 1.024)	**0.009**
Air pollution			
NO2- LUR Model (ug/m3)	2.773	1.028(0.999, 1.058)	0.057
PM2.5- LUR Model (ug/m3)	-0.0236	0.977(0.837, 1.140)	0.765
PM2.5- Monitoring (ug/m3)	0.128	1.136(1.122, 1.150)	**0.018**

Discussion

Urban, transportation, social and economic policies shape the built environment that besides the biological factors could affect public health. Built environment might play a crucial role in preventing NCDs. With of 68% of future population living in urban environment by 2050, it is important to frame future cities to be health protective.

We find that exposure to traffic activity, ambient air pollution, and intersection density are significantly associated with higher risk of death in HF patients. While traffic activity and air pollution were believed to have negative heath impacts, number of intersections were previously assumed to have significant association on walking trips (Ewing and Cervero, 2010), and, therefore, be a positive influence on overall health condition. Our findings, however, explained that built environment factors may have controversial direct and indirect impacts on particular health outcomes such as death risk in HF patients. Furthermore, considering the positive, though not statistically significant at 0.05 level, impacts of compact development measures on increasing risk of death in HF patients have significant policy implications. Smart growth strategies including compact land use development are thought to promote active transportation and reduce vehicle dependency and therefore are good for public health. Recently there are controversial evidences that although compact land use development decreases the emission inventories but they might increase population exposure to vehicle emissions (Tayarani et al., 2016). Our findings provide further evidences on the unfavorable side of smart growth strategies in their current simple form. Continuing the smart growth strategies aiming at relieving congestion and curbing transportation emissions is ought to be mingled with new mobility vbuu8 systems such as vehicle electrification and ride sharing.

Furthermore, the community level policies such as enhancing neighborhood walkability and increasing accessibility are synergetic with individual policies such as encouraging more physical activities and keeping health diet. The built environment policies are even believed to be more effective that individual factors in promoting public health (Ana V Diez Roux, 2003). For instance, providing bicycle facilities have been shown to increase bike mode share (Rowangould Gregory M. and Tayarani Mohammad, 2016) and the lack of enough sidewalks affect the number of children walking to their school (Davison and Lawson, 2006). The existing policies, however, have changed the commuting pattern and walk mode share for commuting trips has declined to 2.7% in 2016 from 3.45% in 1989 (Bureau of Transportation Statistics, 2016) and active transportation mode share for children trip to school declined from 27.8% to 12.9% from 1969 to 2001 (McDonald, 2007).

While the literature suggests providing access to active and public transportation might be effective to promote active lifestyle and reduce obesity, and therefore, could potentially improve public health, our findings suggest that they provided no significant help. These facilities might be helpful in reducing fuel use in transportation sector and reduce transportation externalities, they might not be a best place to spend funding money and expect direct impact on public health. It is important to notice that our findings obtained from large study area that include both urban and suburban areas, thus the promotion of public and active transportation and green space on public health should be followed more carefully. The findings once more highlight the need to integrate public health measures into urban policy planning. The more accurate comprehensive policy suggestions will enable policy makers to include cost benefit analysis in the long-range urban planning. In terms of technical issues, the findings from comparing the fully adjusted and semi adjusted models suggest that adjusting the models for potential confounders affect the results and may change both the direction and significance of the impacts. Furthermore, the findings suggest more causes using the built environmental factors act as a proxy for physical activity and correlate the findings to the health outcomes.

Further research is needed to focus on finding health protective urban policy designs that improve public health. Our study could benefit from including more divers case studies, in terms of urban size and environment, so the findings would be more generalizable.

References

Adams, M.A., Sallis, J.F., Conway, T.L., Frank, L.D., Saelens, B.E., Kerr, J., Cain, K.L., King, A.C., 2012. Neighborhood Environment Profiles for Physical Activity Among Older Adults [WWW Document]. https://doi.org/info:doi/10.5993/AJHB.36.6.4

Beelen, R., Hoek, G., van den Brandt, P.A., Goldbohm, R.A., Fischer, P., Schouten, L.J., Armstrong, B., Brunekreef, B., 2008. Long-term exposure to traffic-related air pollution and lung cancer risk. Epidemiology 19, 702–710. https://doi.org/10.1097/EDE.0b013e318181b3ca

Benjamin, E.J., Blaha, M.J., Chiuve, S.E., Cushman, M., Das, S.R., Deo, R., de Ferranti, S.D., Floyd, J., Fornage, M., Gillespie, C., Isasi, C.R., Jiménez, M.C., Jordan, L.C., Judd, S.E., Lackland, D., Lichtman, J.H., Lisabeth, L., Liu, S., Longenecker, C.T., Mackey, R.H., Matsushita, K., Mozaffarian, D., Mussolino, M.E., Nasir, K., Neumar, R.W., Palaniappan, L., Pandey, D.K., Thiagarajan, R.R., Reeves, M.J., Ritchey, M., Rodriguez, C.J., Roth, G.A., Rosamond, W.D., Sasson, C., Towfighi, A., Tsao, C.W., Turner, M.B., Virani, S.S., Voeks, J.H., Willey, J.Z., Wilkins, J.T., Wu, J.H., Alger, H.M., Wong, S.S., Muntner, P., 2017. Heart Disease and Stroke Statistics—2017 Update. Circulation 135, e146–e603. https://doi.org/10.1161/CIR.0000000000000485

Bloom, D.E., Cafiero, E., Jané-Llopis, E., Abrahams-Gessel, S., Bloom, L.R., Fathima, S., Feigl, A.B., Gaziano, T., Hamandi, A., Mowafi, M., O'Farrell, D., Ozaltin, E., Pandya, A., Prettner, K., Rosenberg, L., Seligman, B., Stein, A.Z., Weinstein, C., Weiss, J., 2012. The Global Economic Burden of Noncommunicable Diseases (No. 8712), PGDA Working Papers. Program on the Global Demography of Aging.

Brownson, R.C., Hoehner, C.M., Day, K., Forsyth, A., Sallis, J.F., 2009. Measuring the Built Environment for Physical Activity: State of the Science. American Journal of Preventive Medicine, Measurement of the Food and Physical Activity Environments Enhancing Research Relevant to Policy on Diet, Physical Activity, and Weight 36, S99-S123.e12. https://doi.org/10.1016/j.amepre.2009.01.005

Casagrande, S.S., Whitt-Glover, M.C., Lancaster, K.J., Odoms-Young, A.M., Gary, T.L., 2009. Built Environment and Health Behaviors Among African Americans: A Systematic Review. American Journal of Preventive Medicine 36, 174–181. https://doi.org/10.1016/j.amepre.2008.09.037

Chum, A., O'Campo, P., 2015. Cross-sectional associations between residential environmental exposures and cardiovascular diseases. BMC Public Health 15, 438. https://doi.org/10.1186/s12889-015-1788-0

Davison, K.K., Lawson, C.T., 2006. Do attributes in the physical environment influence children's physical activity? A review of the literature. International Journal of Behavioral Nutrition and Physical Activity 3, 19. https://doi.org/10.1186/1479-5868-3-19

Diez Roux, A.V., 2003. Residential environments and cardiovascular risk. J Urban Health 80, 569–589. https://doi.org/10.1093/jurban/jtg065

Diez Roux, A.V., 2001. Investigating neighborhood and area effects on health. Am J Public Health 91, 1783–1789.

Ewing, R., Cervero, R., 2010. Travel and the Built Environment. Journal of the American Planning Association 76, 265–294. https://doi.org/10.1080/01944361003766766

Ezzati, M., Lopez, A.D., 2003. Estimates of global mortality attributable to smoking in 2000. The Lancet 362, 847–852. https://doi.org/10.1016/S0140-6736(03)14338-3

Finkelstein, M.M., Jerrett, M., Sears, M.R., 2004. Traffic air pollution and mortality rate advancement periods. Am. J. Epidemiol. 160, 173–177. https://doi.org/10.1093/aje/kwh181

Frank, L.D., Andresen, M.A., Schmid, T.L., 2004. Obesity relationships with community design, physical activity, and time spent in cars. Am J Prev Med 27, 87–96. https://doi.org/10.1016/j.amepre.2004.04.011

Hamano, T., Kawakami, N., Li, X., Sundquist, K., 2013. Neighbourhood Environment and Stroke: A Follow-Up Study in Sweden. PLoS One 8. https://doi.org/10.1371/journal.pone.0056680

HEI, 2010. Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects. Health Effects Institute, Boston, MA.

Hoffmann, B., Moebus, S., Möhlenkamp, S., Stang, A., Lehmann, N., Dragano, N., Schmermund, A., Memmesheimer, M., Mann, K., Erbel, R., Jöckel, K.-H., Heinz Nixdorf Recall Study Investigative Group,

2007. Residential exposure to traffic is associated with coronary atherosclerosis. Circulation 116, 489–496. https://doi.org/10.1161/CIRCULATIONAHA.107.693622
Jerrett, M., Finkelstein, M.M., Brook, J.R., Arain, M.A., Kanaroglou, P., Stieb, D.M., Gilbert, N.L., Verma, D., Finkelstein, N., Chapman, K.R., Sears, M.R., 2009. A Cohort Study of Traffic-Related Air Pollution and Mortality in Toronto, Ontario, Canada. Environmental Health Perspectives 117, 772–777. https://doi.org/10.1289/ehp.11533
Kan, H., Heiss, G., Rose, K.M., Whitsel, E.A., Lurmann, F., London, S.J., 2008. Prospective Analysis of Traffic Exposure as a Risk Factor for Incident Coronary Heart Disease: The Atherosclerosis Risk in Communities (ARIC) Study. Environ Health Perspect 116, 1463–1468. https://doi.org/10.1289/ehp.11290
Karner, A.A., Eisinger, D.S., Niemeier, D.A., 2010. Near-Roadway Air Quality: Synthesizing the Findings from Real-World Data. Environ. Sci. Technol. 44, 5334–5344. https://doi.org/10.1021/es100008x
Kim, S.-Y., Bechle, M., Hankey, S., Sheppard, L., Szpiro, A., Marshall, J., 2018. A Parsimonious Approach to National Prediction: Criteria Pollutants in the Contiguous U.S., 1979 - 2015. ISEE Conference Abstracts.
Lavie, C.J., Milani, R.V., Ventura, H.O., 2009. Obesity and Cardiovascular Disease: Risk Factor, Paradox, and Impact of Weight Loss. Journal of the American College of Cardiology 53, 1925–1932. https://doi.org/10.1016/j.jacc.2008.12.068
Malambo, P., Kengne, A.P., De Villiers, A., Lambert, E.V., Puoane, T., 2016. Built Environment, Selected Risk Factors and Major Cardiovascular Disease Outcomes: A Systematic Review. PLoS One 11. https://doi.org/10.1371/journal.pone.0166846
McDonald, N.C., 2007. Active Transportation to School: Trends Among U.S. Schoolchildren, 1969–2001. American Journal of Preventive Medicine 32, 509–516. https://doi.org/10.1016/j.amepre.2007.02.022
Renalds, A., Smith, T.H., Hale, P.J., 2010. A systematic review of built environment and health. Fam Community Health 33, 68–78. https://doi.org/10.1097/FCH.0b013e3181c4e2e5
Rowangould Gregory M., Tayarani Mohammad, 2016. Effect of Bicycle Facilities on Travel Mode Choice Decisions. Journal of Urban Planning and Development 142, 04016019. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000341
Srinivasan, S., O'Fallon, L.R., Dearry, A., 2003. Creating Healthy Communities, Healthy Homes, Healthy People: Initiating a Research Agenda on the Built Environment and Public Health. Am J Public Health 93, 1446–1450. https://doi.org/10.2105/AJPH.93.9.1446
Tayarani, M., Poorfakhraei, A., Nadafianshahamabadi, R., Rowangould, G.M., 2016. Evaluating unintended outcomes of regional smart-growth strategies: Environmental justice and public health concerns. Transportation Research Part D: Transport and Environment 49, 280–290. https://doi.org/10.1016/j.trd.2016.10.011
The Official Website of New York State [WWW Document], 2018. . Welcome to the State of New York. URL https://www.ny.gov/ (accessed 12.31.18).
Tonne, C., Melly, S., Mittleman, M., Coull, B., Goldberg, R., Schwartz, J., 2007. A case-control analysis of exposure to traffic and acute myocardial infarction. Environ. Health Perspect. 115, 53–57.
Witten, K., Blakely, T., Bagheri, N., Badland, H., Ivory, V., Pearce, J., Mavoa, S., Hinckson, E., Schofield, G., 2012. Neighborhood Built Environment and Transport and Leisure Physical Activity: Findings Using Objective Exposure and Outcome Measures in New Zealand. Environ Health Perspect 120, 971–977. https://doi.org/10.1289/ehp.1104584
World Health Organization, 2018. Global status report on noncommunicable diseases. Geneva.

Study 3: Combining rich social determinants data with clinical data*

Introduction

Socioeconomic status (SES) and other social determinants of health (SDH) are predictors of population health and health inequality. Leveraging SDH data to improve care appears increasingly feasible with electronic health records (EHRs) and powerful machine learning capabilities. However, for healthcare organizations, collecting and updating patient income, literacy, social, and environmental factors might be prohibitively labor-intensive (and might not be welcomed by patients). As an alternative, it might be preferable to estimate SES and SDH from publicly available data.

Somewhat surprisingly, studies that have done so have had mixed results. One study showed that "enriching" a model for 30-day readmission rates with community-level SDH variables improved predictive power,[1] while others suggested that SDH made only small[2] or no improvements.[3,4]

We have argued that in relatively homogenous communities, SDH may not vary enough to make a difference in outcomes.[5] New York City, NY, USA, is an ideal setting to explore the impact of SDH because it is extremely diverse, with a population about 44% white, 26% African-American, and 13% Asian. 29% of people of all races identify as Hispanic. Almost one-third of New Yorkers were born outside the US, and the city is also known for extremes of wealth and poverty.

We have also suggested that SDH may produce poor results if estimated imprecisely.[5] This may be a particular problem in dense urban areas such as New York (with more than 8.6 million people residing within 780 km^2). Some previous studies have used the US ZIP (postal) code as the geographic unit, even though ZIP Codes have no standard population and range from thousands to more than 100,000 residents.

Our objective was to construct a large data set that combined clinical data on a cohort of patients with a rich set of localized address-based SDH. This data set will be available for multiple analyses of the impact of SDH on health outcomes.

Methods

Electronic health record data was from NewYork-Presbyterian Hospital-Weill Cornell Medical Center (NYP-WCMC). Weill Cornell's research informatics department maintains NYP data in the Observational Medical Outcomes Partnership (OMOP) Common Data Model for research. Patient addresses were mapped to two US Census Bureau units, the *census tract* (a standard unit with a mean of 4000 people) and the *census block* (which contains less than 1000). We aggregated public data from (1) the US Census decennial census, (2) the American Community Survey, a smaller annual survey by the Census Bureau, (3) the US Environmental Protection Agency's National Air Toxics Assessment, and (4) New York City Open, collections of data sets released by city agencies.

Results

For demonstration purposes, we created a data set to replicate a score that has been demonstrated to predict 30-day readmissions in European clinical data,[6] and then determine the effect of adding SDH. Following methodology from Aubert et al., we selected medical discharges 2015-2017, excluding patients who lived outside NYC, died in hospital, were transferred, or left against medical advice. The data set contains about 4500 discharges, of whom about 650 were readmitted within 30 days.
By mapping patient addresses to geographic regions, we generated an SDH data set for these patients that includes 20 variables mapped to both census track and census block. These represent economic variables (the CDC's social vulnerability index, median income), neighborhood and built environment (crime rate, access to health food, air quality, tree cover, and access to public transportation), and other social and community (percent foreign born). Through public lists, patient addresses can be used to determine whether the patient lives in public

housing or a retirement community. This data can also be joined with SDH elements available in the EHR including race and ethnicity, marital status, and primary preferred language.

Conclusions

By geocoding patient addresses and doing extensive research on publicly available data from multiple sources, it is possible to create research data sets that combine clinical data with rich social determinants data. Currently ongoing analyses will extend previous research on the contribution of SDH to predictive models by assessing its impact in a dense, diverse urban setting using high-granularity geographic mapping.

References

1. Reidhead M. *Including Sociodemographic Factors in Risk-Adjusted Readmission Measures.* Hospital Industry Data Institute; 2016.
2. Ash AS, Mick EO, Ellis RP, Kiefe CI, Allison JJ, Clark MA. Social determinants of health in managed care payment formulas. *JAMA Internal Medicine* 2017;177(10):1424-30.
3. Kasthurirathne SN, Vest JR, Menachemi N, et al. Assessing the capacity of social determinants of health data to augment predictive models identifying patients in need of wraparound social services. *JAMIA* 2018;epub Jan 2018.
4. Bhavsar NA, Gao A, Phelan M, et al. Value of neighborhood socioeconomic status in predicting risk of outcomes in studies that use electronic health record data. *JAMA Network Open 2018;1(5):* e182716.
5. Ancker JS, Kim M-H, Zhang Y, et al. The potential value of social determinants of health in predicting health outcomes. *JAMIA* 2018; 25(8): 1109-10.
6. Aubert CE, Schnipper JL, Williams MV, et al. Simplification of the HOSPITAL score for predicting 30-day readmissions. *BMJ Qual Saf.* 2017;26(10):799-805.

[*]This is an on-going study whose funding was awarded by Weill Cornell Medicine Dean's Office. Data used for study 3 is curated from this CTECH study.

Appendix

This database contains patient encounter information extracted from the EHR at 53 sites of Weill Cornell Medicine and 2 campuses of NewYork-Presbyterian Hospital. It contains over 1,566,917 patients from 2012 to 2018. Number of unique clinical class and drug class are 1814 and 486, respectively. Data are updated every 3 months to provide new encounter information in the database. The database is stored in a Microsoft SQL server securely managed by Weill Cornell Medicine's Information Technology Services. Table A1 lists the encouter frequencies. Table A2 shows the patient breakdown by states. Tables A3, A4, and A5 list the descriptive statatistics for demographics, common conditions, and common drug classes.

Table A1. Encounter statistics

Encounter	**Number**
Emergency Room and Inpatient Visit	165,139
Emergency Room Visit	688,776
Inpatient Visit	236,558
Outpatient Visit	49,969,687

Table A2. Number of patients by state

State	**Count**
NY	599655
NJ	24023
CT	3324
PA	1527
FL	18
CA	10

Table A3. Demographic information in the databases(N=1566917). All variables (apart from mean) are represented as counts (percentages).

Age (mean)		**48.9**
Age group, years	0-19	24,411 (1.6)
	20-39	547,814 (35.0)
	40-59	523,234 (33.4)
	60-79	391,050 (25.0)
	80+	80,408 (5.1)
Gender	Female	913,088 (58.3)
	Male	653,598 (41.7)
	Unknown	230 (0.01)
Race	Asian	121,048(9.1)
	Black or African American	118,425(8.9)
	White	455,330(34.2)

	Multiple Race	210,304(15.8)
	Others	6,955(0.5)
	Unknown	419,843(31.5)

Table A4. Most common diseases

Diagnose	**ICD 10 CM code**	**Count**
Essential (primary) hypertension	I10	73085
Disorders of lipoprotein metabolism and other lipidemias	E78	67651
Abdominal and pelvic pain	R10	44704
Other joint disorder, not elsewhere classified	M25	33844
Type 2 diabetes mellitus	E11	30449
Dorsalgia	M54	29853
Pain in throat and chest	R07	28875
Malaise and fatigue	R53	28284
Other and unspecified soft tissue disorders, not elsewhere classified	M79	25387
Abnormalities of breathing	R06	24766
Long term (current) drug therapy	Z79	23360

Table A5. Most common drug classes (VA class)

Drug class	**Count**
CENTRAL NERVOUS SYSTEM MEDICATIONS	418447
HORMONES/SYNTHETICS/MODIFIERS	340340
ANTIMICROBIALS	321555
GASTROINTESTINAL MEDICATIONS	310652
CARDIOVASCULAR MEDICATIONS	310046
ANALGESICS	275354
DERMATOLOGICAL AGENTS	265955
MUSCULOSKELETAL MEDICATIONS	264233
RESPIRATORY TRACT MEDICATIONS	244489
VITAMINS	240218
ANTIRHEUMATICS	221083

OPIOID ANALGESICS	216895
NONSALICYLATE NSAIs,ANTIRHEUMATIC	201298
GASTRIC MEDICATIONS,OTHER	186168
LAXATIVES	181991
THERAPEUTIC NUTRIENTS/MINERALS/ELECTROLYTES	172122
ELECTROLYTES/MINERALS	169384
ANTILIPEMIC AGENTS	167660
NON-OPIOID ANALGESICS	165715
NASAL AND THROAT AGENTS,TOPICAL	163292