INFORMATICS PROFESSIONALS. LEADING THE WAY.

Methods for Integrating EHRs, Social Determinants of Health, and Built Environment Data for Patient-Centered Research

Detection and Learning Using Multiple Data Sources

S114

Joseph Deferio, MPH

Weill Cornell Medicine

Twitter: @diaferiaj

Acknowledgments

- Evan Sholle, MS — Weill Cornell Medicine
- Kathleen G. Lee, MPH — Weill Cornell Medicine
- John Michael Meddar — CUNY City Tech
- Mohammad Tayarani, PhD — Cornell University
- H. Oliver Gao, PhD, MS — Cornell University
- Jyotishman Pathak, PhD — Weill Cornell Medicine

Disclosure

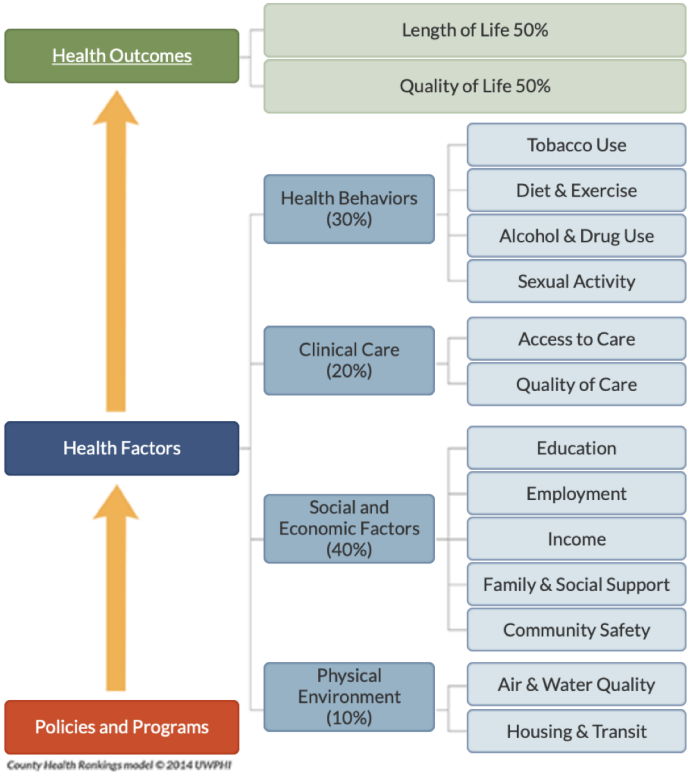
My colleagues and I have no relevant relationships with commercial interests to disclose

Learning Objectives

After participating in this session the learner should be better able to:

- Conceptualize linkage of external datasets with electronic health records
- Improve their understanding of the application and potential utility of social determinants of health (SDOH) in patient-centered research.

Why Social Determinants?



According to this recent framework, **social and economic factors** may contribute up to 40% of health outcomes

County Health Rankings. University of Wisconsin Population Health Institute.
http://www.countyhealthrankings.org/explore-health-rankings/what-and-why-we-rank


NYC OpenData
POLICE DEPARTMENT
CITY OF NEW YORK
Weill Cornell Medicine
WATCH
fitbit
NEW YORK CITY HOUSING AUTHORITY
WEATHER UNDERGROUND

Health and wellness-related data is becoming more ubiquitous electronically

No longer limited to clinical settings

- data on the physical, social (and digital!) environments is coming to the fore

Collectively, this data is often referred to as '**social determinants of health**' or **SDOH**

- Digital / Online behavior (ie. social media)
- Weather, air quality, and environmental factors
- Interpersonal interaction (ie. crime reports)
- Fitness and activity trackers

Collectively, this data is often referred to as '**social determinants of health**' or **SDOH**

- Digital / Online behavior (ie. social media)
- Weather, air quality, and environmental factors
- Interpersonal interaction (ie. crime reports)
- Fitness and activity trackers

Better yet, they should be referred to as '**social influences of health**'

Collectively, this data is often referred to as '**social determinants of health**' or **SDOH**

- Digital / Online behavior (ie. social media)
- Weather, air quality, and environmental factors
- Interpersonal interaction (ie. crime reports)
- Fitness and activity trackers

Better yet, they should be referred to as '**social influences of health**'

We all know that our neighborhoods and communities (physical and/or digital) influence our lives. But…

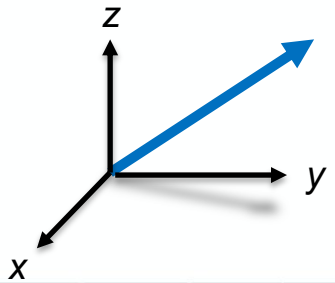
Collectively, this data is often referred to as '**social determinants of health**' or **SDOH**

- Digital / Online behavior (ie. social media)
- Weather, air quality, and environmental factors
- Interpersonal interaction (ie. crime reports)
- Fitness and activity trackers

Better yet, they should be referred to as '**social influences of health**'

We all know that our neighborhoods and communities (physical and/or digital) influence our lives. But

How do they influence us?
What is the magnitude of that influence?



Motivations

Our work was motivated by past work that has investigated the influence of one's environment on health outcomes:

Motivations

Our work was motivated by past work that has investigated the influence of one's environment on health outcomes:

- One study concluded that access to public transportation and dense urban structure could reduce risk of depression[1]

Motivations

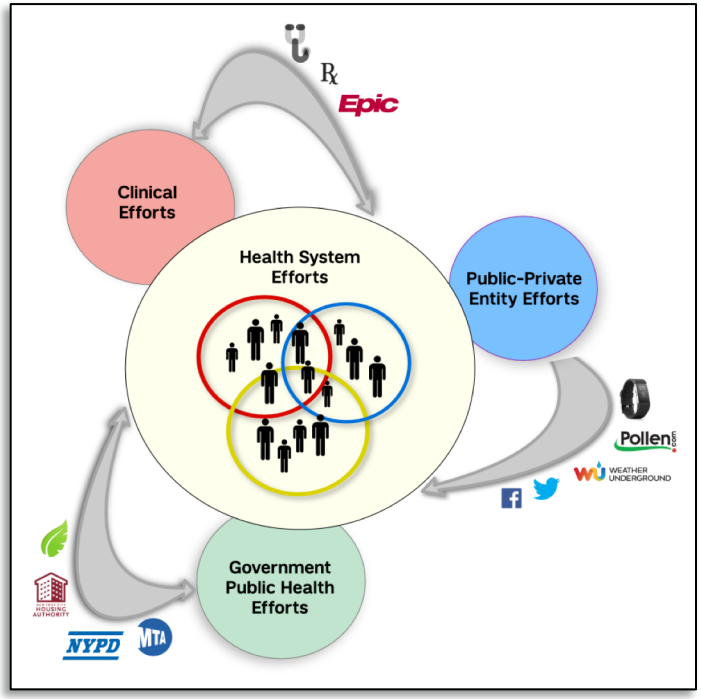
Our work was motivated by past work that has investigated the influence of one's environment on health outcomes:

- One study concluded that access to public transportation and dense urban structure could reduce risk of depression[1]

- Another study concluded that long-term exposure to environments that support physical activity and healthy diets is associated with lower incidence of type-2 diabetes[2]

Now, more than ever, health systems can begin to answer some of these questions!

Now, more than ever, health systems can begin to answer some of these questions!



Health Systems have access to:

- Patient health records
- Publicly available data
  - Public-private entities
  - Government-collected
- Informatics and data science teams

Objectives

Primary:
To address this knowledge gap and achieve a more comprehensive EHR data integration and linkage that incorporates a broad spectrum of information about SDOH and built environment across NYC

Secondary:
To examine risk/protective socio-ecological factors for psychiatric hospitalization in NYC

Journal of the American Medical Informatics Association, 25(4), 2018, 419–422
doi: 10.1093/jamia/ocx117
Advance Access Publication Date: 27 October 2017
Brief Communication

AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

OXFORD

Brief Communication

FACETS: using open data to measure community social determinants of health

Michael N Cantor,[1,2,3] Rajan Chandras,[3] and Claudia Pulgarin[2]

[1]Department of Medicine, [2]Department of Population Health, New York University School of Medicine, [3]Medical Center Information Technology, NYU Langone Health, New York, NY, USA

Corresponding Author: Michael N Cantor, 227 E 30th St, #725, New York, NY 10016, USA. E-mail: michael.cantor@nyumc.org

Received 14 June 2017; Revised 9 August 2017; Editorial Decision 14 September 2017; Accepted 23 September 2017

Census tract-level data from:

- American Community Survey
- Centers for Disease Control and Prevention
- NYC Open Data
- NY State Board of Election

Table 1. Data elements in FACETS

Community-level determinants in FACETS

Measure	Source
Total population	ACS
Urban/rural classification	USDA
Total population	ACS
Racial diversity	ACS
Ethnic diversity (Hispanic/non-Hispanic)	ACS
US citizenship	ACS
Foreign vs native-born	ACS
Educational attainment	ACS
English proficiency	ACS
Poverty rate	ACS
Median household income	ACS
Unemployment rate	ACS
Health insurance status	ACS
Respiratory Hazard Index	EPA
Access to healthy food	USDA
Distance to parks	CDC
Walkability score	BEH
Tobacco retailers/1000 population	NYS
Felony crime/1000 population	NYC
Gini index of inequality	ACS
Social Vulnerability Index	CDC
Housing violations/1000 units	FC
Voter turnout	BOE

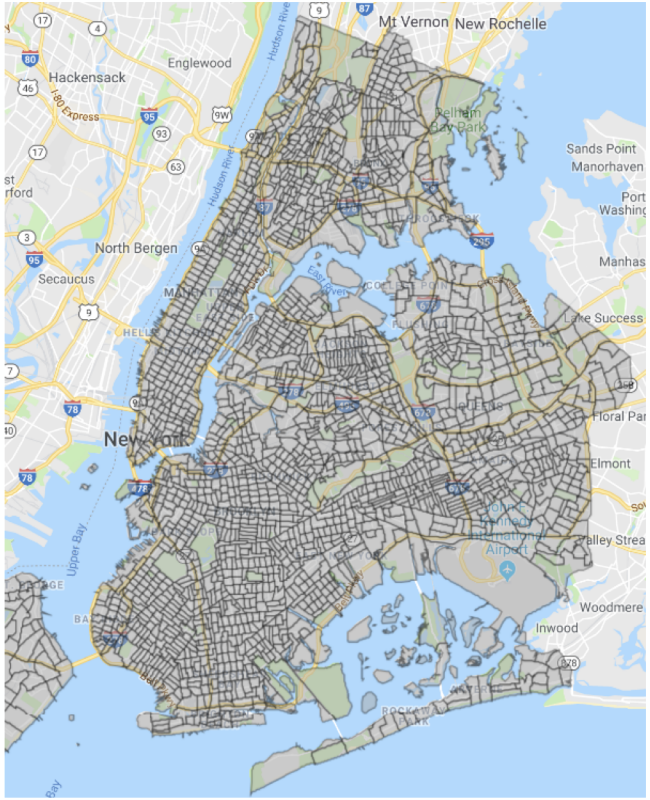
Abbreviations: ACS: American Community Survey; USDA: US Department of Agriculture Food Access Research Atlas; EPA: Environmental Protection Agency National Air Toxics Assessment; CDC: Centers for Disease Control and Prevention; BEH: Columbia Built Environment and Health Research Group; NYS: New York State Open Health Data; NYC: New York City Open Data; FC: Furman Center; BOE: New York State Board of Elections.

2010 NYC Census Tracts

N = 2165

Each tabulation area has 11-digit FIPS code correlating to state, county, and census tract

Ex. 36061011000



New York City Neighborhood Health Atlas

Health outcomes are influenced by social factors, such as education, economic opportunity and environment. Due to these factors, New Yorkers in some neighborhoods live longer, healthier lives than those in other areas. **The New York City Neighborhood Health Atlas** provides data on about 100 measures related to health and social factors for 188 neighborhoods. The data provide a comprehensive and granular view of neighborhood health and its potential determinants, serving as a useful resource for the promotion of health and health equity in our neighborhoods.

Atlas Measures

The Neighborhood Health Atlas' measures relate to health and neighborhood characteristics, including:

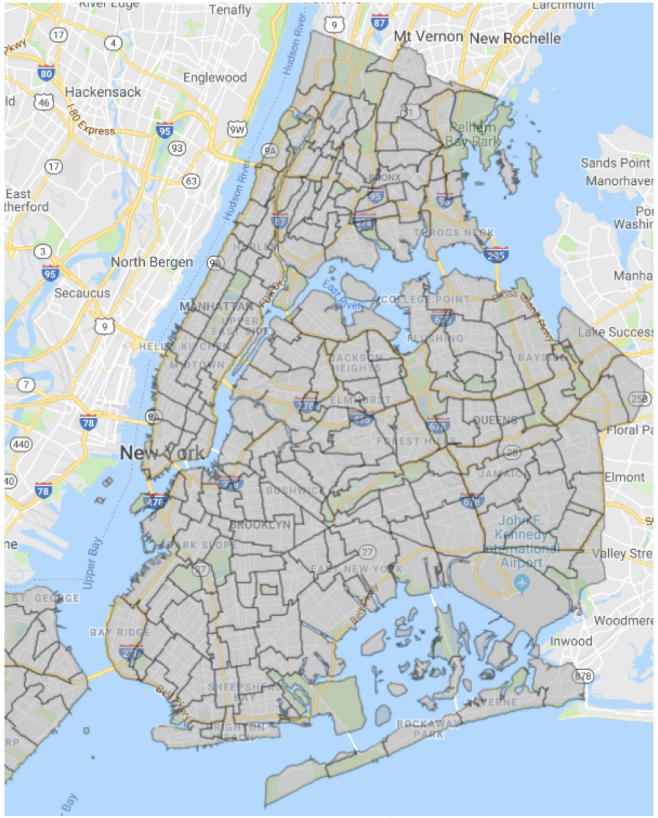
- Demographics (such as race, age, country of origin)
- Social and economic conditions (education, poverty, disabilities)
- Health outcomes (hospitalizations, rates of communicable diseases, premature mortality)
- Health care (health insurance status, number of primary care providers, Medicaid enrollment)
- Housing (density, rent burden)
- Neighborhood conditions (air quality, number of tobacco retailers, crime complaints)

NYC Neighborhood Tabulation Areas (NTAs)

n = 195

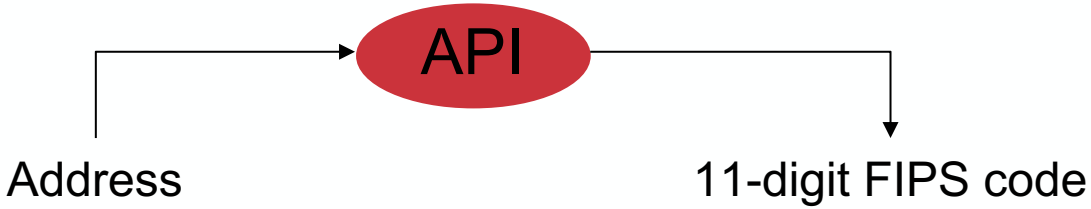
NTAs are based on census tracts ∴ conversion can be done with relative ease

Ex. MN33



The abundance of census tract-level data led us to adopt this unit to facilitate data integration

Therefore, we sought to map patient addresses to their associated census tract FIPS codes

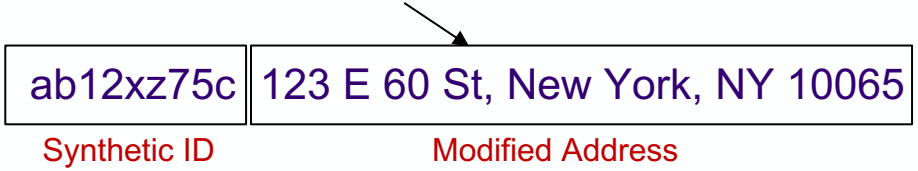


First, we stripped raw address data of apartment-level identifiers and names to allow for maximum preservation of patient privacy.

John Doe | 123 E 60 St Apt 301, New York, NY 10065

First, we stripped raw address data of apartment-level identifiers and names to allow for maximum preservation of patient privacy.

~~John Doe~~ | 123 E 60 St ~~Apt 301~~, New York, NY 10065



First, we stripped raw address data of apartment-level identifiers and names to allow for maximum preservation of patient privacy.

~~John Doe~~ | 123 E 60 St ~~Apt 301~~, New York, NY 10065

ab12xz75c	123 E 60 St, New York, NY 10065
Synthetic ID	Modified Address

Resultant tuple was passed to the **US Census Bureau API**

	Synthetic ID	FIPS
Result →	**ab12xz75c**	**36061011000**

What are the next steps?

In relation to our communities/neighborhoods, we can begin to assess:

How do they influence us?

What is the magnitude of that influence?

What are the next steps?

In relation to our communities/neighborhoods, we can begin to assess:

- How do they influence us?
- What is the magnitude of that influence?

We can now link community-level estimates of social determinants to individual electronic health records

What are the next steps?

In relation to our communities/neighborhoods, we can begin to assess:

- How do they influence us?
- What is the magnitude of that influence?

We can now link community-level estimates of social determinants to individual electronic health records

We can aggregate patient records to census tracts (or other zones) and perform community-level analyses with external data

What are the next steps?

In relation to our communities/neighborhoods, we can begin to assess:

- How do they influence us?
- What is the magnitude of that influence?

We can now link community-level estimates of social determinants to individual electronic health records

We can aggregate patient records to census tracts (or other zones) and perform community-level analyses with external data

We can map patient locations (while maintaining privacy!) and proximity to other features

What are our next steps?

Objective: predict 1st psychiatric hospitalization across NYC in 2014

What are our next steps?

Objective: predict 1st psychiatric hospitalization across NYC in 2014

Who: clinical data from n=1,470,242 patients aged ≥ 18 years in NYC-Clinical Data Research Network (NYC-CDRN)

- n=470,242 (cases) whom have Dx depression OR Rx antidepressant
- n=1,000,000 (controls) whom have no history of schizophrenia, bipolar, depression, other psychoses

What are our next steps?

Objective: predict 1st psychiatric hospitalization across NYC in 2014

Who: clinical data from n=1,470,242 patients aged ≥ 18 years in NYC-Clinical Data Research Network (NYC-CDRN)

- n=470,242 (cases) whom have Dx depression OR Rx antidepressant
- n=1,000,000 (controls) whom have no history of schizophrenia, bipolar, depression, other psychoses

When: diagnosis and encounter data from 2012-2014

What are our next steps?

Objective: predict 1st psychiatric hospitalization across NYC in 2014

Who: clinical data from n=1,470,242 patients aged ≥ 18 years in NYC-Clinical Data Research Network (NYC-CDRN)

- n=470,242 (cases) whom have Dx depression OR Rx antidepressant
- n=1,000,000 (controls) whom have no history of schizophrenia, bipolar, depression, other psychoses

When: diagnosis and encounter data from 2012-2014

Why: to see which clinical features influence predictive power, and if SDH built environment features improve prediction

What are our next steps?

Objective: predict 1st psychiatric hospitalization across NYC in 2014

Who: clinical data from n=1,470,242 patients aged ≥ 18 years in NYC-Clinical Data Research Network (NYC-CDRN)

- n=470,242 (cases) whom have Dx depression OR Rx antidepressant
- n=1,000,000 (controls) whom have no history of schizophrenia, bipolar, depression, other psychoses

When: diagnosis and encounter data from 2012-2014

Why: to see which clinical features influence predictive power, and if SDH built environment features improve prediction

How: supervised learning using logistic regression and 5-fold cross-validation

Inclusion / Exclusion

- Limited to patients with a FIPS code for census tract
  - Not all NYC-CDRN sites reported this data
  - Removed patients with multiple FIPS codes during ‘12-14

Inclusion / Exclusion

- Limited to patients with a FIPS code for census tract
  - Not all NYC-CDRN sites reported this data
  - Removed patients with multiple FIPS codes during ‘12-14

- Removed patients with psych hospitalizations prior to 2014

Inclusion / Exclusion

- Limited to patients with a FIPS code for census tract
  - Not all NYC-CDRN sites reported this data
  - Removed patients with multiple FIPS codes during ‘12-14

- Removed patients with psych hospitalizations prior to 2014

- Included all diagnosis and encounter data 2012-2014 < hospitalization date
  - Patients must have had ≥ 1 encounter between 01/2012-12/2013 & ≥ 1 encounter in 2014

Inclusion / Exclusion

- Limited to patients with a FIPS code for census tract
  - Not all NYC-CDRN sites reported this data
  - Removed patients with multiple FIPS codes during ‘12-14

- Removed patients with psych hospitalizations prior to 2014

- Included all diagnosis and encounter data 2012-2014 < hospitalization date
  - Patients must have had ≥ 1 encounter between 01/2012-12/2013 & ≥ 1 encounter in 2014

- **n_{total} = 337,165**
- **n_{hosp} = 5,185**

Clinical Features

CMS Chronic Conditions Data Warehouse

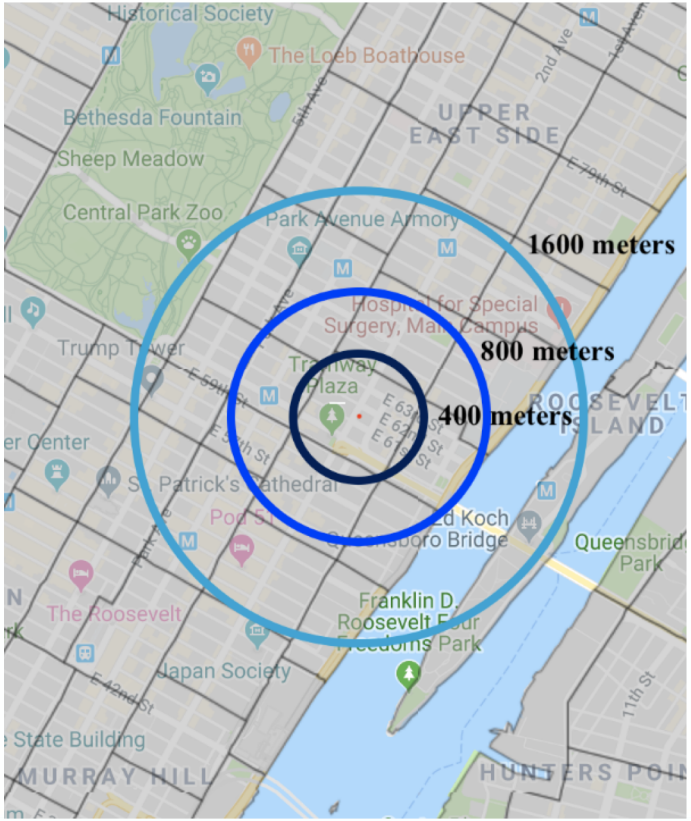
CCW Chronic Conditions	
Acquired Hypothyroidism	Chronic Kidney Disease
Acute Myocardial Infarction	Chronic Obstructive Pulmonary Disease
Alzheimer's Disease	Depression
Alzheimer's Disease, Related Disorders, or Senile Dementia	Diabetes
Anemia	Glaucoma
Asthma	Heart Failure
Atrial Fibrillation	Hip / Pelvic Fracture
Benign Prostatic Hyperplasia	Hyperlipidemia
Cancer, Colorectal	Hypertension
Cancer, Endometrial	Ischemic Heart Disease
Cancer, Breast	Osteoporosis
Cancer, Lung	Rheumatoid Arthritis / Osteoarthritis
Cancer, Prostate	Stroke / Transient Ischemic Attack
Cataract	

66 categories ⇒ representing chronic and other potentially disabling conditions

Additional features:

- Drug/alcohol abuse
- Injury from others/suicide
- Yearly visit counts
  - Inpatient, Outpatient, ER, Ambulatory, Other*

Social Determinants – Built Environment Features

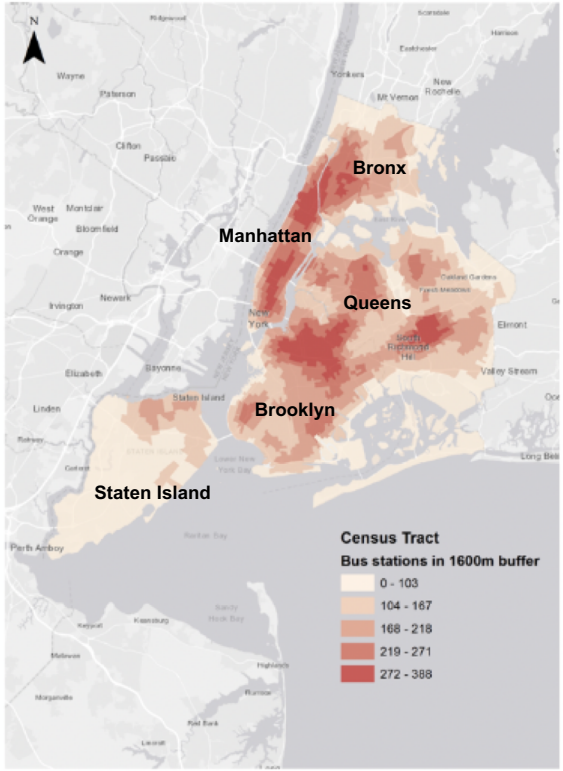


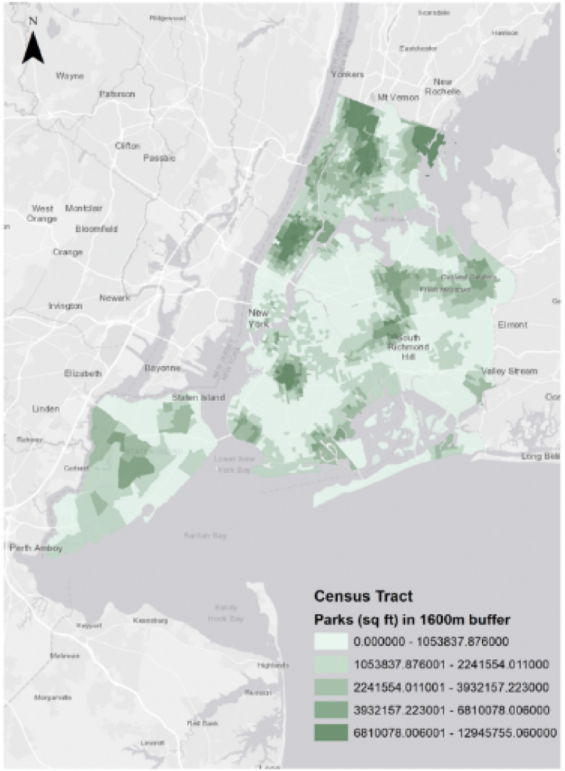
- Using ArcGIS, identified geographic center of each census tract
- Examined SDH features that fell within 400m, 800m, and 1600m from that point

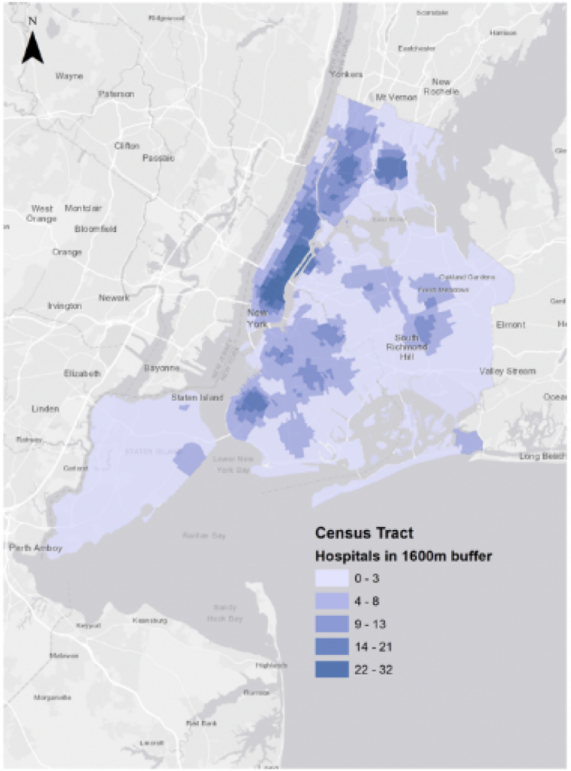
Features Include:

- Public transportation
- Park area
- Hospitals
- Cultural Centers
- Museums
- Schools

Maroko AR, Maantay JA, Sohler NL, Grady KL, Arno PS. The complexities of measuring access to parks and physical activity sites in New York City: a quantitative and qualitative approach. International journal of health geographics. 2009;8:34.


N
Bronx
Manhattan
Queens
Brooklyn
Staten Island
Census Tract
Bus stations in 1600m buffer
0 - 103
104 - 167
168 - 218
219 - 271
272 - 388


N
Census Tract
Parks (sq ft) in 1600m buffer
0.000000 - 1053837.876000
1053837.876001 - 2241554.011000
2241554.011001 - 3932157.223000
3932157.223001 - 6810078.006000
6810078.006001 - 12945755.060000


N
Census Tract
Hospitals in 1600m buffer
0 - 3
4 - 8
9 - 13
14 - 21
22 - 32

Preliminary Results (unpublished)

Clinical Data Only

TN	FN
66,395	1
194	843
FP	TP

AUC: 0.986
Sensitivity: 0.81
Specificity: 1.00

Clinical Data + 400m SDH

TN	FN
66,396	0
199	838
FP	TP

AUC: 0.986
Sensitivity: 0.81
Specificity: 1.00

Clinical Data + 1600m SDH

TN	FN
66,396	0
204	833
FP	TP

AUC: 0.985
Sensitivity: 0.80
Specificity: 1.00

Conclusions

- External data integration can be done with relative ease
- Location-based linkage of non-personal data can be a reasonable proxy for neighborhood factors
- Need to develop composite measures that encompasses social and built-environment factors, such as Neighborhood Socioeconomic Status[1]
    - ie. education, wealth, occupation, and living conditions

'Far Too Early' Conclusions

- Built environment features do not improve prediction models for psychiatric hospitalization

Thank you!

Email me at: jod2033@med.cornell.edu

@diaferiaj