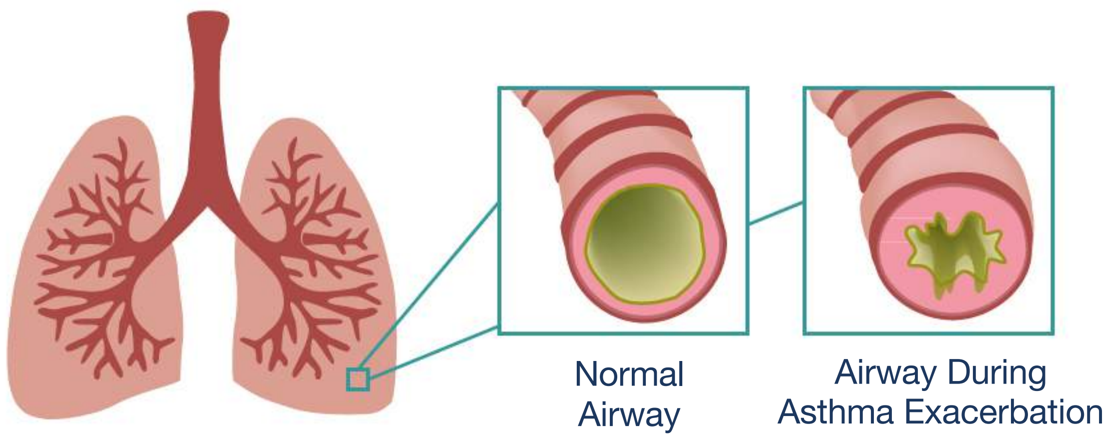
Enhancing Electronic Health Record (EHR)-Derived Data with Data from Secondary Sources to Address Multifactorial Problems in Real Life Populations

Blanca E. Himes, Ph.D.
Assistant Professor of Informatics

Asthma and its Treatment

- Asthma is a chronic inflammatory lung disease characterized by episodes of airway obstruction
- β_2-agonists and glucocorticoids are commonly used drugs
- Treatment according to clinical guidelines decreases asthma symptoms, exacerbations and improves lung function in most patients



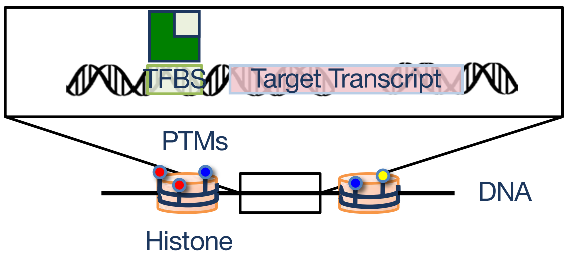
Influence of Genetics on Asthma is Substantial

- Heritability estimated to be 50-90%
- Several reproducible asthma-associated loci have been identified
  - *ORMDL3/GSDMB*
  - *TSLP*
  - *HLA-DQA1*
  - *IL33*
  - *IL1RL1*
  - *IL13*
  - *SMAD3*
  - *RORA*

Ober C, *Ann Am Thorac Soc* (2016)

Asthma Management is a Multifactorial Problem

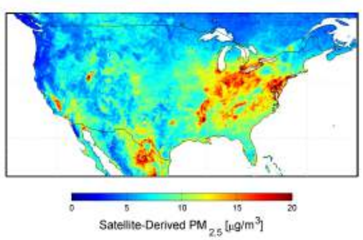
Genomics, Transcriptomics, Epigenomics



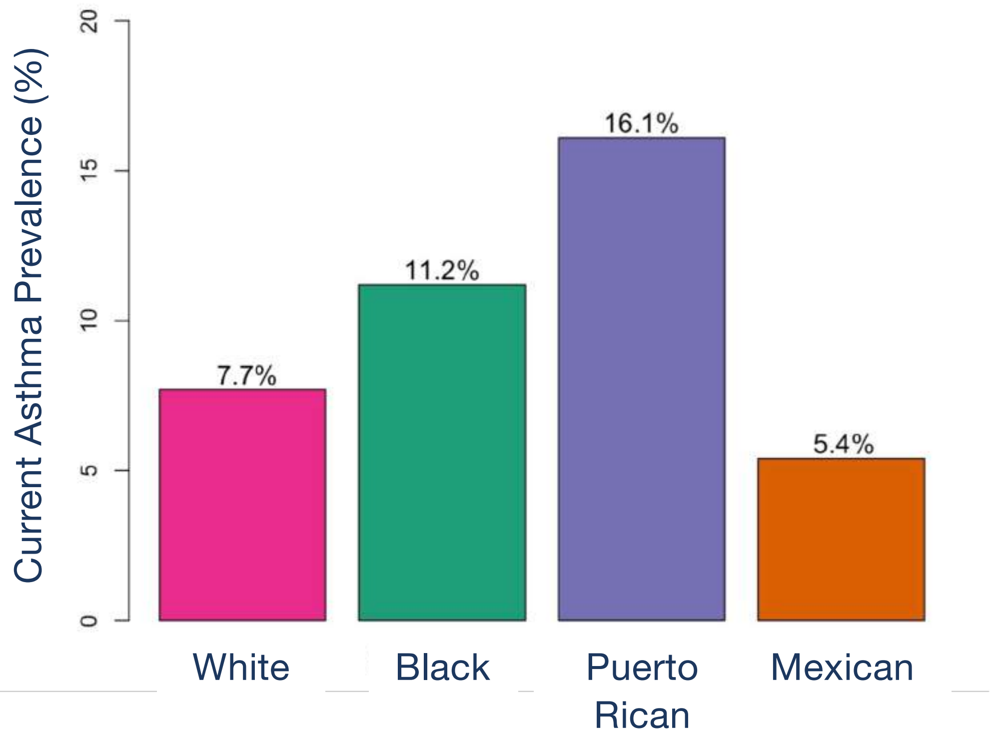
Medications

Social Factors

Environmental Exposures



U.S. Disparities in Asthma Prevalence by Race/Ethnicity

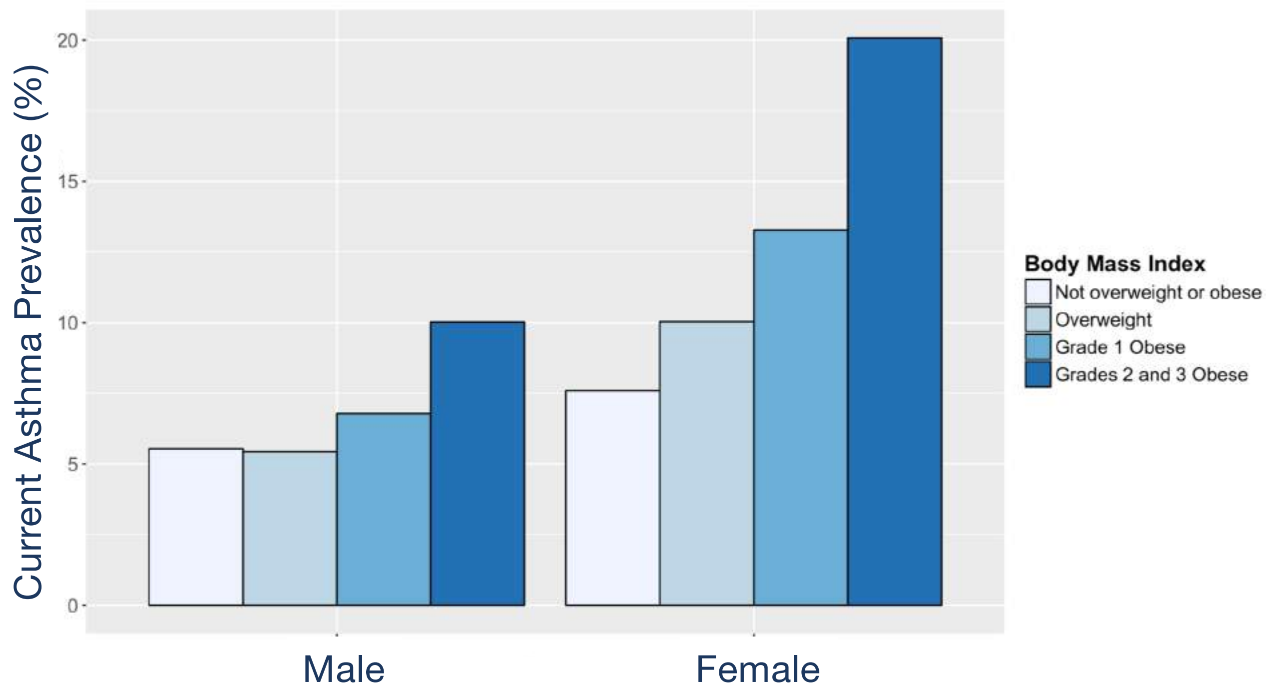


Data Source: National Health Interview Survey (NHIS) 2008-2010

Sex Disparities in Asthma Prevalence and Risk Factors

1,003,894 subjects 21+ years of age from Behavioral Risk Factor Surveillance Study (BRFSS) years 2007-2012

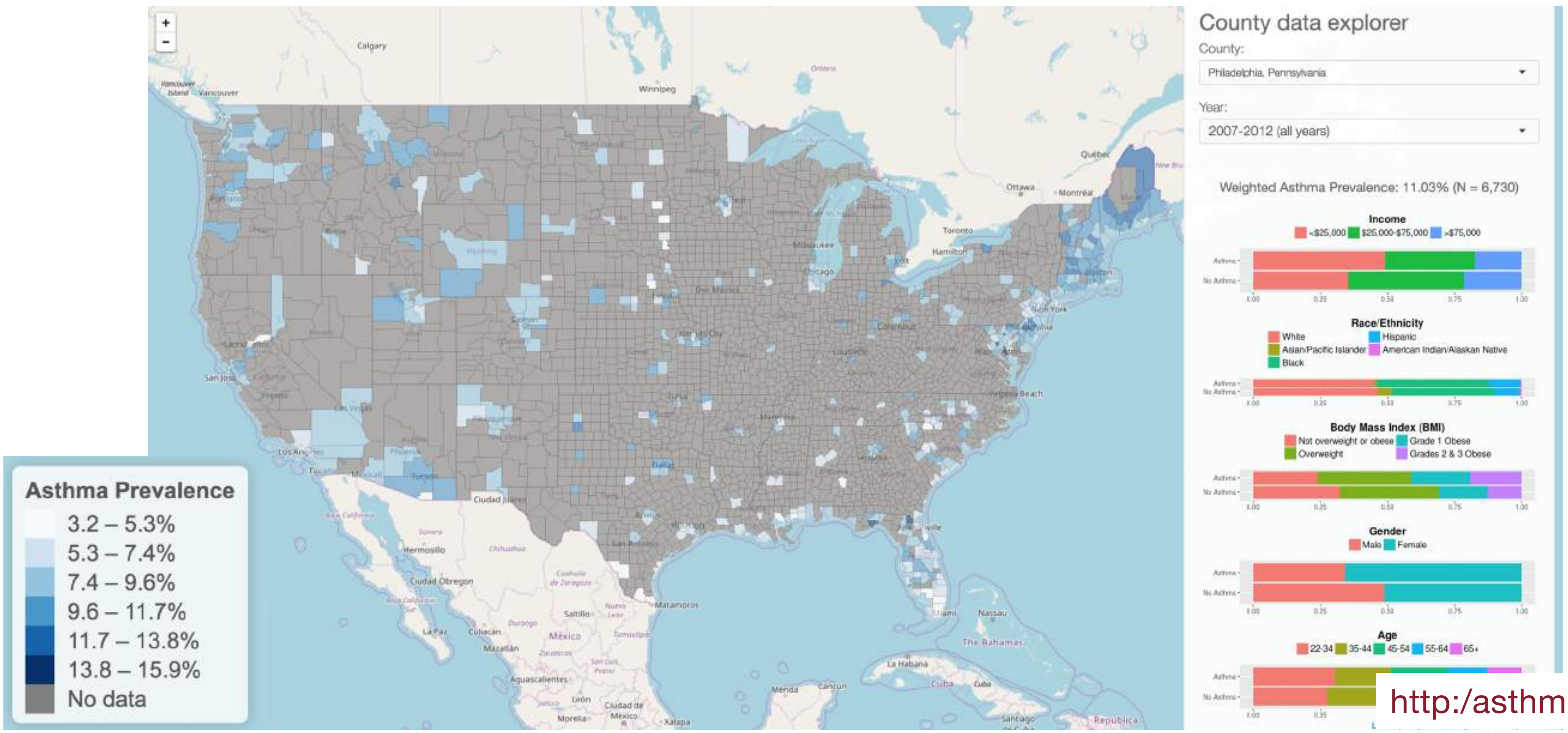
Sex	Respondents with Current Asthma
Male	26,490 (36.8%w)
Female	66,958 (63.2%w)



Greenblatt R *et al*, Asthma Res Pract (2017)

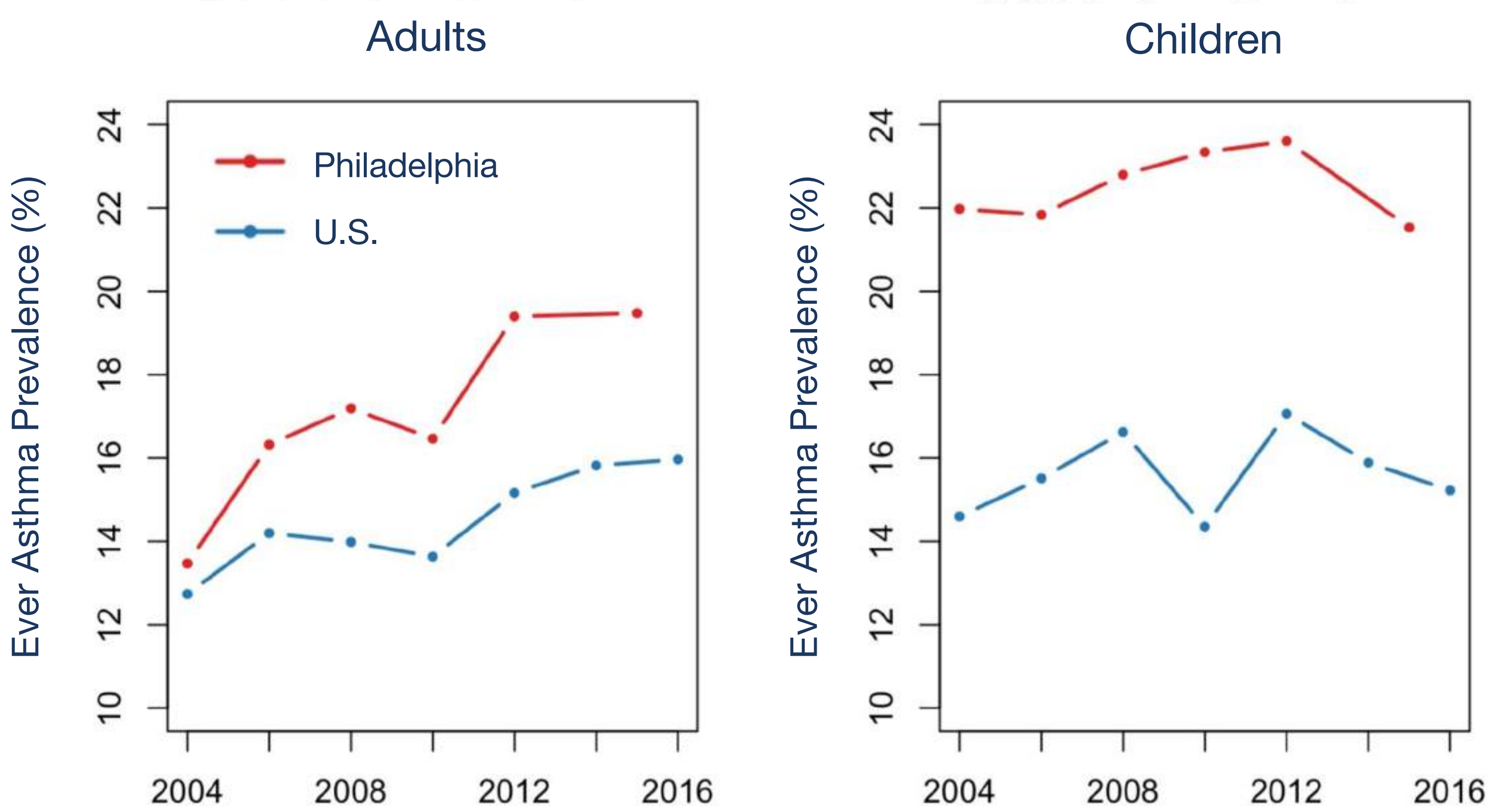
U.S. Adult Asthma Prevalence Varies Geographically

1,003,894 subjects 21+ years of age from BRFSS years 2007-2012



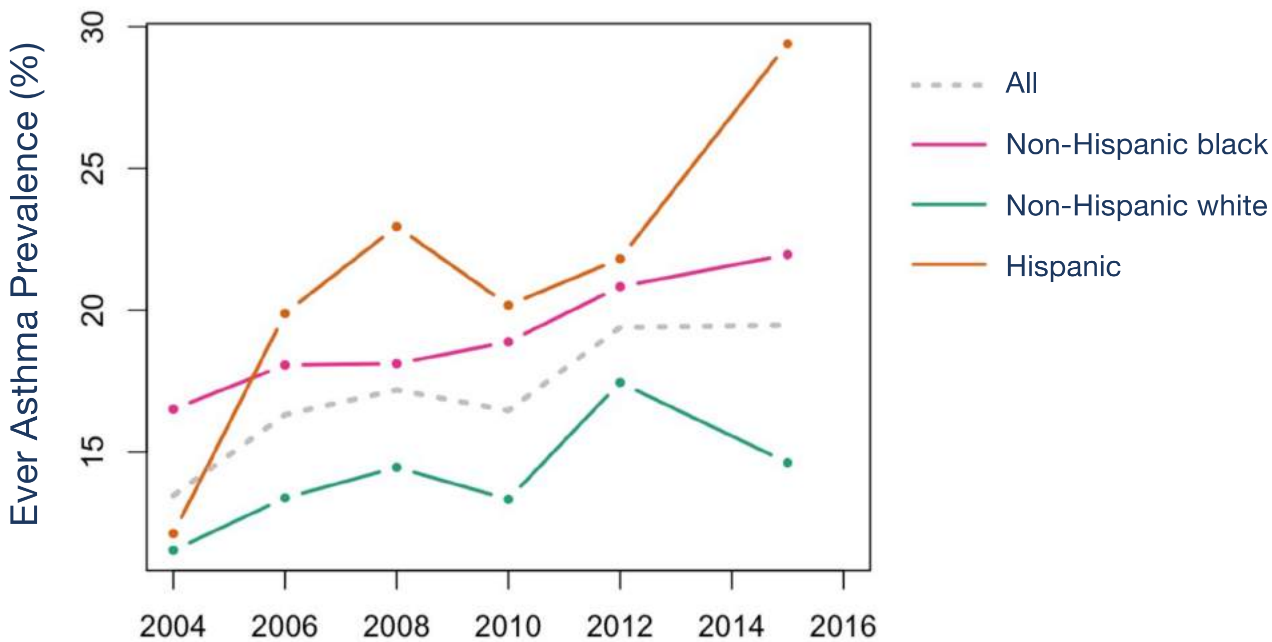
http:/asthmamaps.org

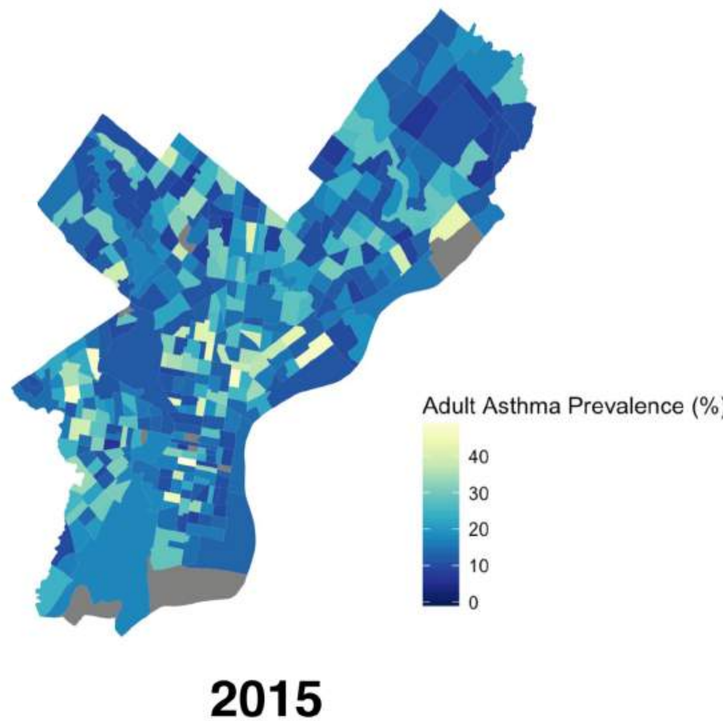
Asthma Prevalence in Philadelphia is Higher than U.S.



Data Sources: National Health and Nutrition Examination Survey (NHANES) 2003-2016, Southeastern Pennsylvania Household Health Survey 2004-2015

Demographic Risk Factors in Philadelphia Mirror National Ones



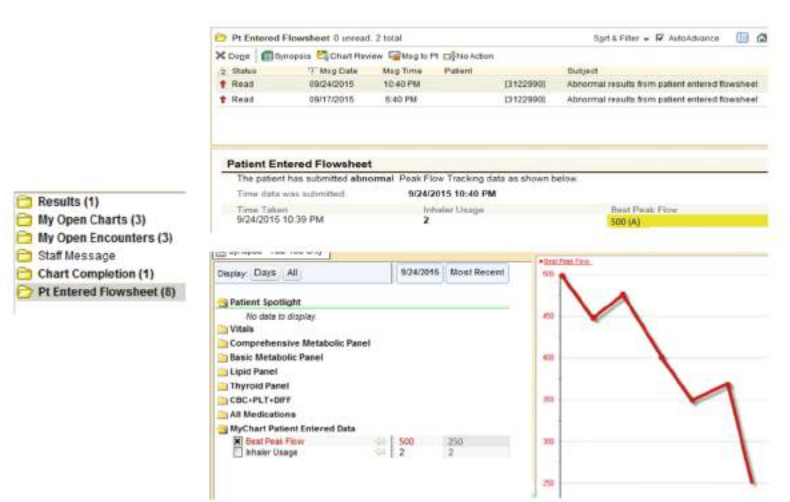


Southeastern Pennsylvania Household Health Survey 2004-2015

EHRs are Valuable for Research

Provide convenient and low-cost access to longitudinal information of many patients that represent *real-life* populations

- Facilitate contact of people for research studies
- Enable improvement of clinical workflows at the point-of-care
- Derived data can be used for primary research
- Essential for creation of large biobanks for omics studies



EHR-Derived Data for the Study of Asthma in Philadelphia

University of Pennsylvania Health System (UPHS)

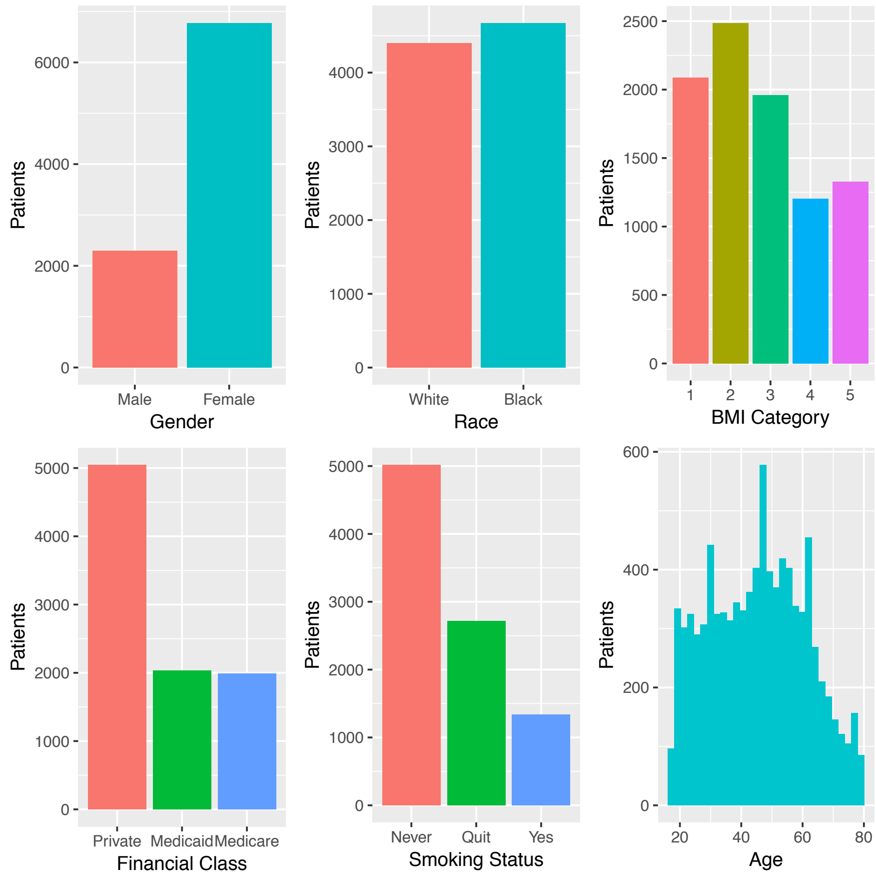
- Serves the greater Philadelphia area
- Maintains a clinical data warehouse with demographic and clinical data

UPHS Asthma Patient Characteristics

- Obtained 2011-2014 encounter and demographic data for asthma patients (ICD-9 493*)
- Inclusion criteria:
  - 18-80 years old
  - Followed for at least 3 years
  - Prescription for albuterol
  - At least 1 primary asthma ICD-9 code

→ 9,068 complete cases

UPHS Asthma Patient Characteristics



	All UPHS patients (N=3,199,282)
Gender	
Male	1,154,154 (39.8%)
Female	1,746,632 (60.2%)
Race/Ethnicity	
White	1,705,967 (58.8%)
Black	842,345 (29.0%)

Characteristics of Asthma Patients with Exacerbations

- Defined as prescription for oral steroid & primary asthma ICD-9 code (493*)

Number of Exacerbations	0	1-2	3-4	5+
Number of Subjects (%)	6,042 (66.63)	2,639 (29.10)	273 (3.01)	114 (1.26)

- Measure associations with other variables

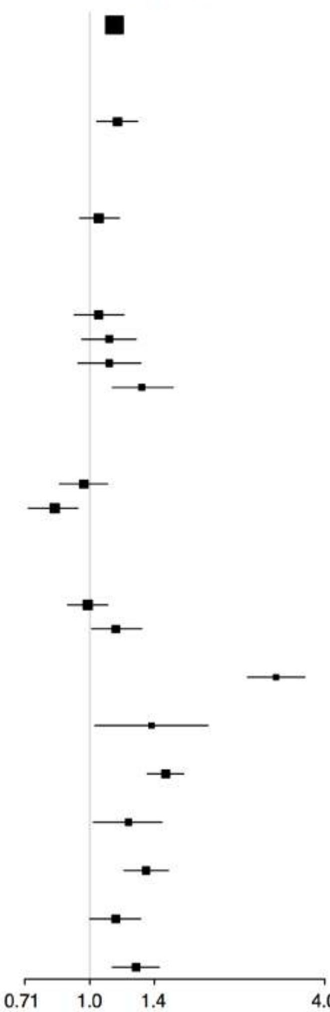
Race				
White	3,034 (50.22)	1,252 (47.44)	99 (36.26)	12 (10.53)
Black	3,008 (49.78)	1,387 (52.56)	174 (63.74)	102 (89.47)
Sex				
Male	1,576 (26.08)	623 (23.61)	68 (24.91)	28 (24.56)
Female	4,466 (73.92)	2,016 (76.39)	205 (75.09)	86 (75.44)

Greenblatt RE *et al* Asthma Res Prac (2019)

Factors Associated with Exacerbations among Adults with Asthma According to EHR Data

	Crude OR	Adjusted OR	P-Value
Age (10 years)	1.21 (1.18, 1.25)	1.14 (1.10, 1.18)	2.6e−13
Race			
White	Reference	Reference	
Black or African American	1.28 (1.17, 1.4)	1.16 (1.04, 1.29)	6.9e−03
Sex			
Male	Reference	Reference	
Female	1.13 (1.02, 1.25)	1.05 (0.95, 1.17)	3.6e−01
BMI (kg/m²)			
Not Overweight Or Obese	Reference	Reference	
Overweight (25 to < 30)	1.09 (0.96, 1.24)	1.05 (0.92, 1.20)	4.7e−01
Class 1 Obese (30 to < 35)	1.28 (1.13, 1.46)	1.11 (0.96, 1.28)	1.6e−01
Class 2 Obese (35 to < 40)	1.35 (1.16, 1.56)	1.11 (0.94, 1.31)	2.2e−01
Class 3 Obese (= or > 40)	1.67 (1.45, 1.93)	1.32 (1.13, 1.56)	7.4e−04
Health Insurance Type			
Private Insurance	Reference	Reference	
Medicaid	1.21 (1.09, 1.35)	0.97 (0.85, 1.10)	6.4e−01
Medicare	1.53 (1.37, 1.7)	0.83 (0.72, 0.94)	6.2e−03
Smoking Status			
Never	Reference	Reference	
Quit Smoking	1.36 (1.24, 1.5)	0.99 (0.89, 1.10)	8.5e−01
Current Smoker	1.42 (1.26, 1.61)	1.15 (1.01, 1.32)	4.1e−02
Chronic Bronchitis	4.04 (3.52, 4.64)	2.70 (2.32, 3.15)	4.0e−37
Emphysema	3.21 (2.43, 4.25)	1.39 (1.03, 1.88)	3.2e−02
Sinusitis	1.47 (1.34, 1.6)	1.50 (1.36, 1.65)	2.0e−16
Pulmonary Circulation Disorder	2.21 (1.88, 2.61)	1.23 (1.02, 1.47)	2.6e−02
Fluid & Electrolyte Disorder	1.97 (1.78, 2.19)	1.35 (1.20, 1.52)	6.5e−07
Obstructive Sleep Apnea	1.81 (1.62, 2.03)	1.15 (1.00, 1.31)	4.2e−02
Diabetes	2.05 (1.84, 2.29)	1.28 (1.13, 1.45)	1.0e−04

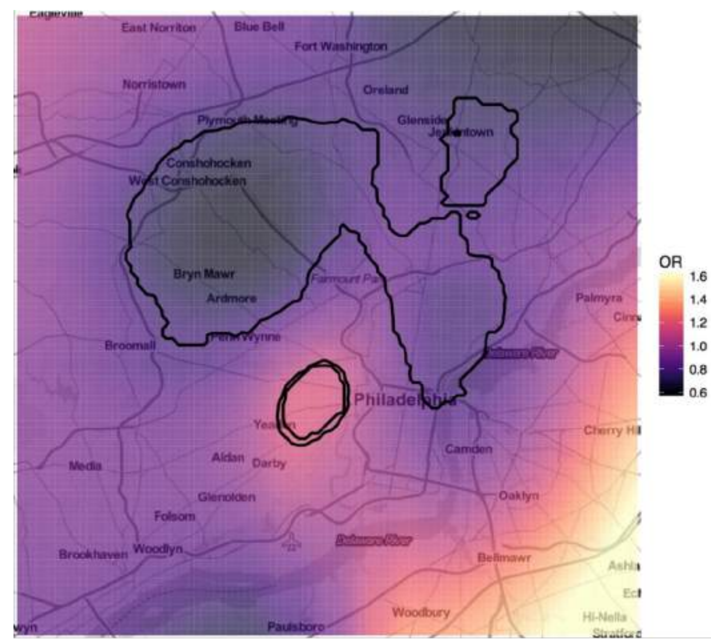
Plot of Adjusted OR



Greenblatt RE *et al* Asthma Res Prac (2019)

Spatial Distribution of Asthma Exacerbations

- Used residential address to map asthma exacerbation rates
- Generalized additive models measured differences while adjusting for race, age, BMI, smoking status, insurance class
- Global spatial heterogeneity present
- Significant hot spots and cold spots indicated by contour lines ($p<0.01$)

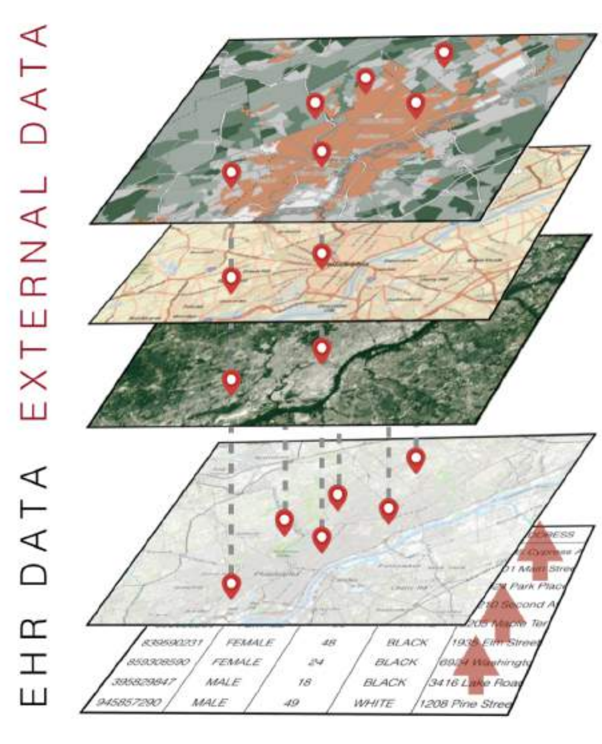


Xie S *et al*, AMIA Joint Summ Proc (2017)

Important Health-Related Information is Missing from EHRs

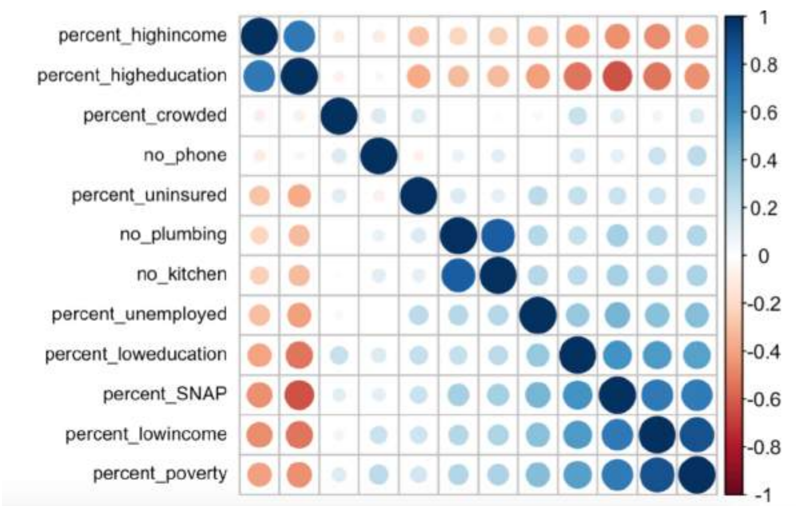
- Socioeconomic status
- Education and health literacy
- Environmental exposures

This information can be captured by external sources and linked to EHR data via patient-specific geocodes

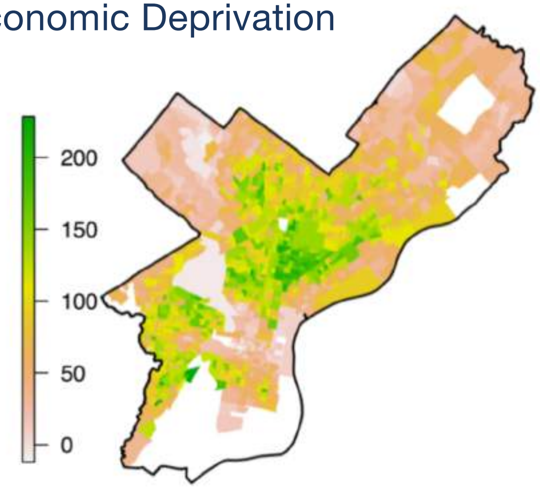


Geospatially Varying Socioeconomic Conditions

- Data source: American Community Survey
- Twelve SES variables for all U.S. census block groups in Philadelphia from 2010-2014
- Composite score of "economic deprivation" obtained via factor analysis



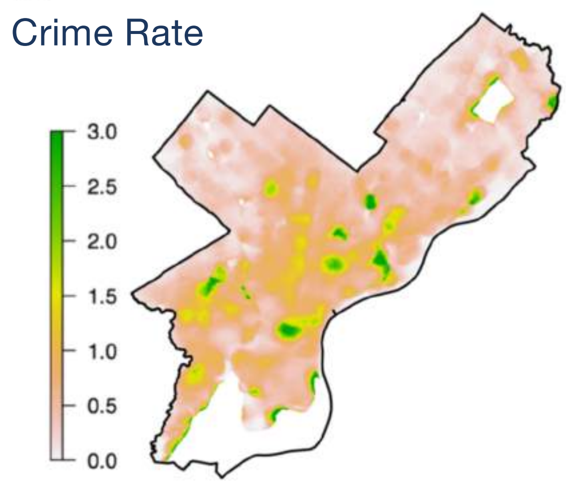
Economic Deprivation



Xie S and Himes BE AMIA Symp Proc (2018)

Geospatially Varying Exposure to Crime

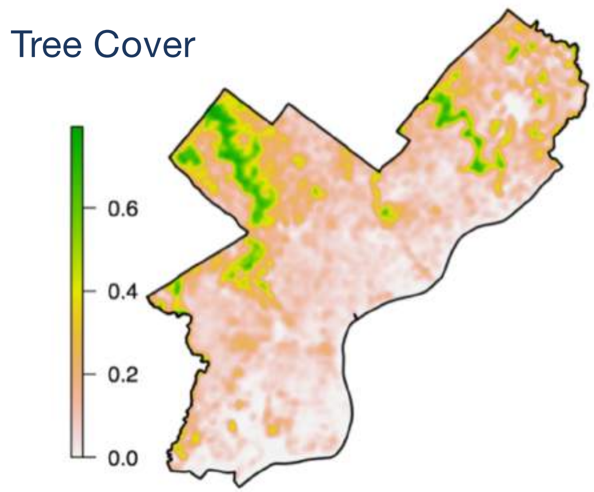
- Data source: publicly available database from Philadelphia Police Department of all crime incidents in Philadelphia
- Crime rate calculated as crime density divided by population density



Xie S and Himes BE AMIA Symp Proc (2018)

Geospatially Varying Exposure to Tree Cover

- Data source: assessment of tree canopy by University of Vermont Spatial Analysis Laboratory using automated object-based analysis approach that combined high resolution light detection and ranging (LiDAR) data and ancillary GIS data (building footprints, road polygons)
- Tree cover estimated as percent of land area composed of tree canopy in a circular 250-m moving window

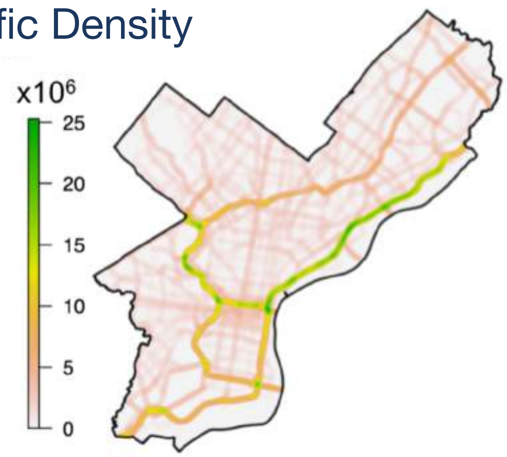


Xie S and Himes BE AMIA Symp Proc (2018)

Geospatially Varying Exposure to Vehicular Traffic

- Data source: annual average daily traffic (AADT) measurements for all major road segments in Philadelphia from the Pennsylvania Department of Transportation
- Traffic density calculated based on daily vehicle miles traveled (DVMT) using a 250-m circular moving window

Traffic Density

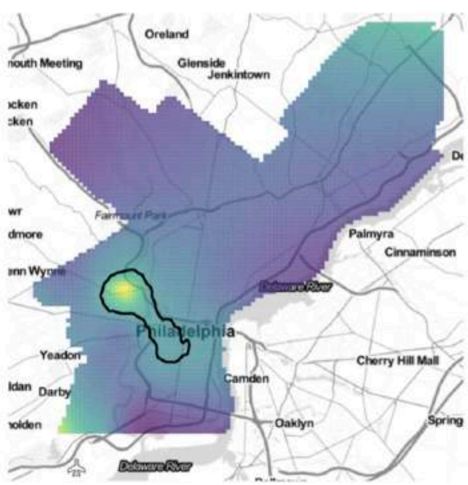


Xie S and Himes BE AMIA Symp Proc (2018)

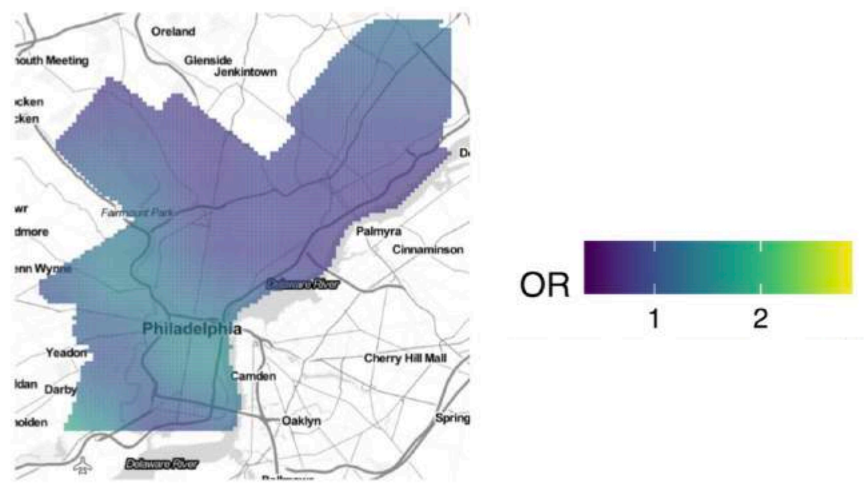
Spatial Heterogeneity of Asthma Exacerbations Decreased with Adjustment of Externally Sourced Variables

- EHR-derived data (2014-2016) restricted to 1,568 patients who had ≥1 outpatient encounter with a primary ICD-9/10 code for asthma and a prescription for albuterol.
- Generalized additive models of asthma exacerbations (570 cases and 998 controls) performed using:

EHR-derived variables (age, sex, race, BMI, smoking status, insurance class)



EHR-extracted variables + neighborhood deprivation, crime, and vehicular traffic



Global test for spatial heterogeneity

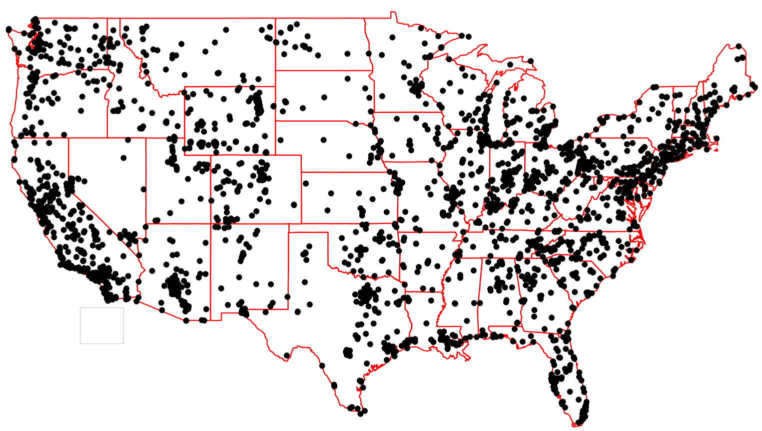
p=0.005

p=0.064

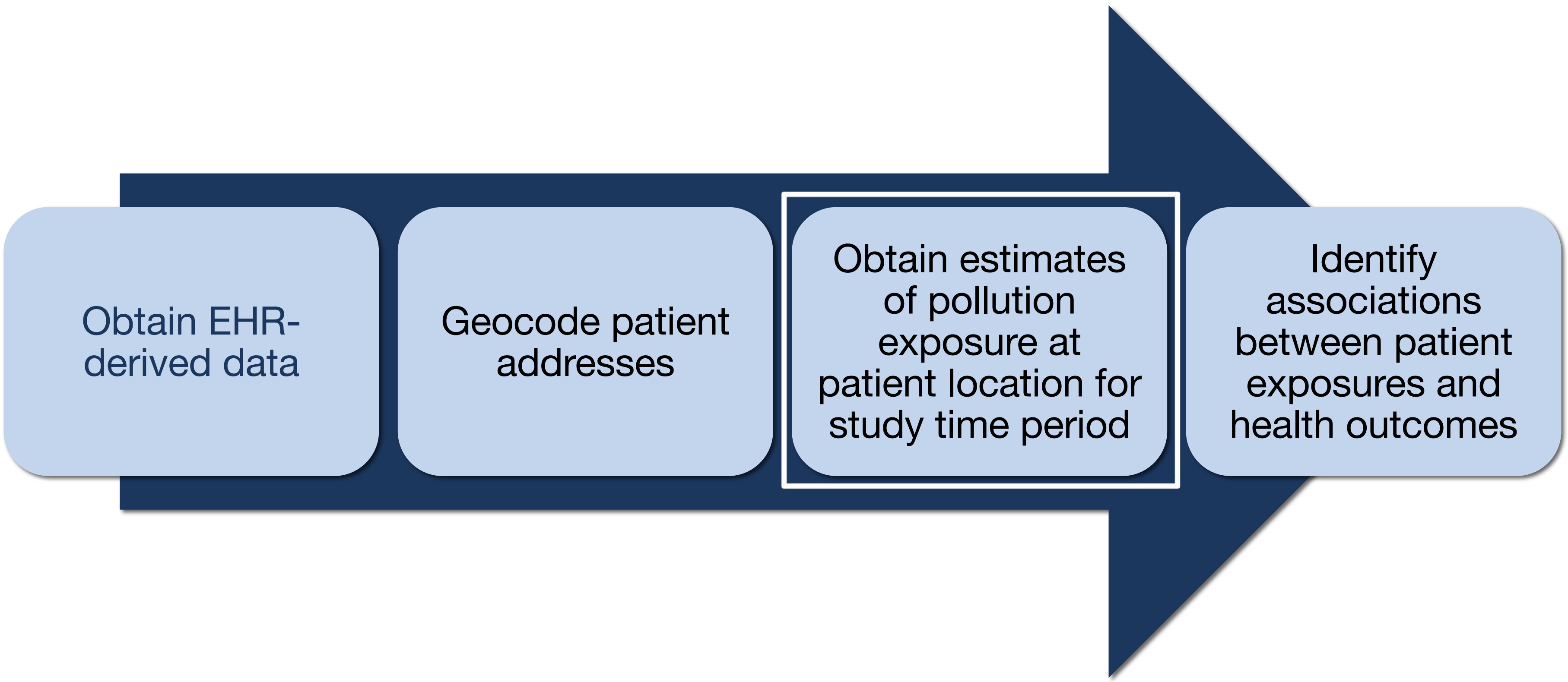
Air Pollution

- Associated with many adverse health outcomes, including asthma
- U.S. Environmental Protection Agency Data monitors air pollution using > 2,000 regulated monitors across the U.S. and provides data to the public

    $PM_{2.5}$, CO, Ozone, NO_2, SO_2

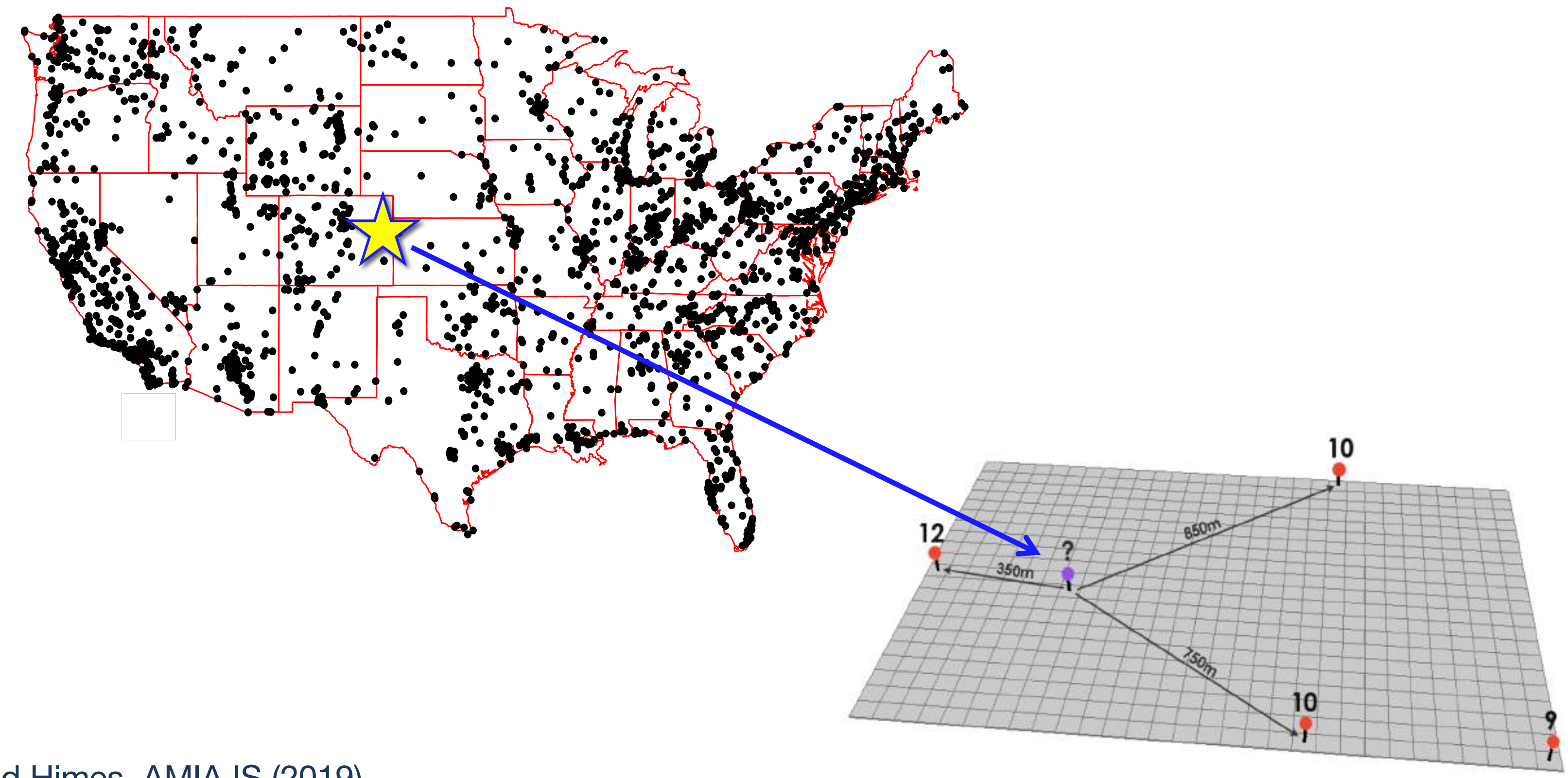


Workflow of Linked EHR and Pollution Studies



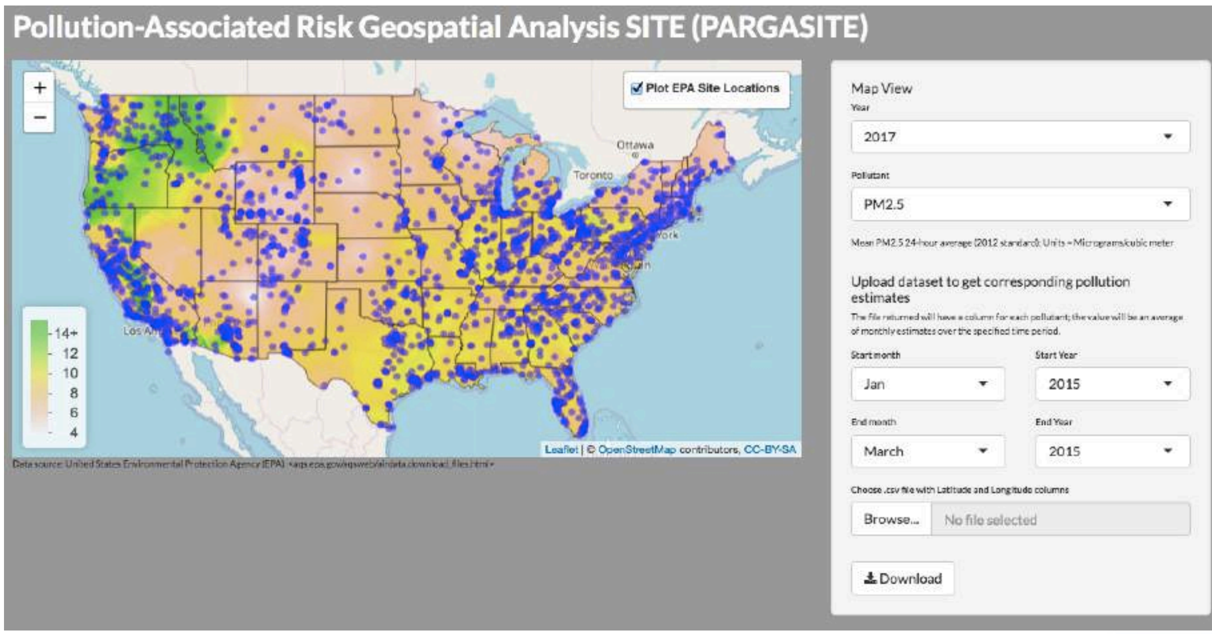
Greenblatt and Himes, AMIA IS (2019)

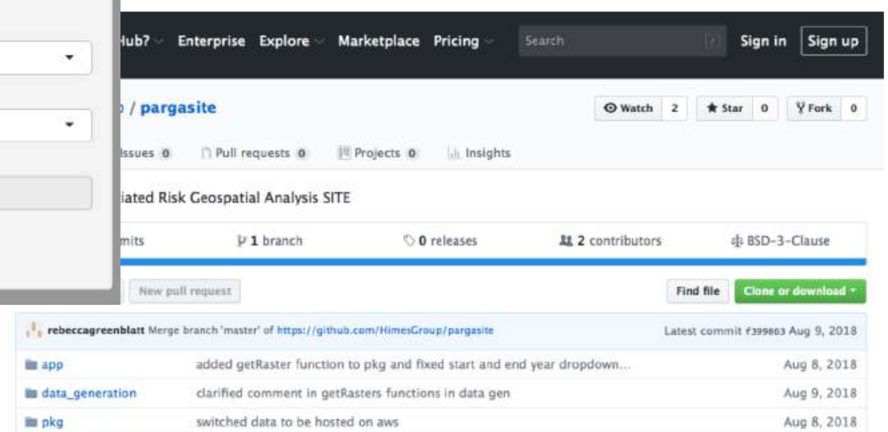
Exposure at a Given Geocoordinate is Interpolated from Sites with Measures



Greenblatt and Himes, AMIA IS (2019)

Pollution-Associated Risk Geospatial Analysis SITE (PARGASITE)



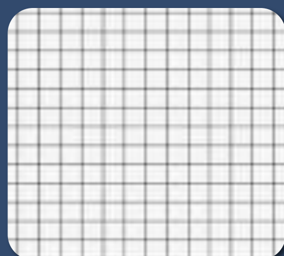


http://pargasite.org

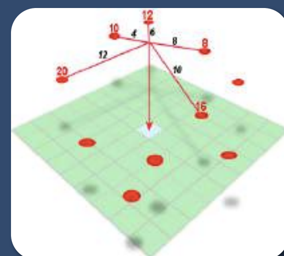
http://GitHub.com/HimesGroup/Pargasite

Greenblatt and Himes, AMIA IS (2019)

PARGASITE has Pre-Computed Estimates of Air Pollution Measures Across the U.S.



10km x 10km grid across contiguous U.S.



Inverse distance weighted estimates from 5 nearest EPA monitoring sites

$PM_{2.5}$

Ozone

NO_2

SO_2

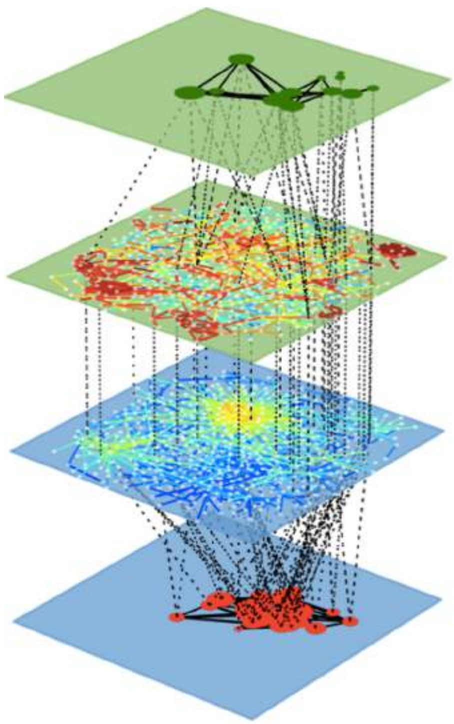
CO

Monthly and yearly estimates

Jan 2005 - Dec 2017

Greenblatt and Himes, AMIA IS (2019)

Plans to Expand PARGASITE

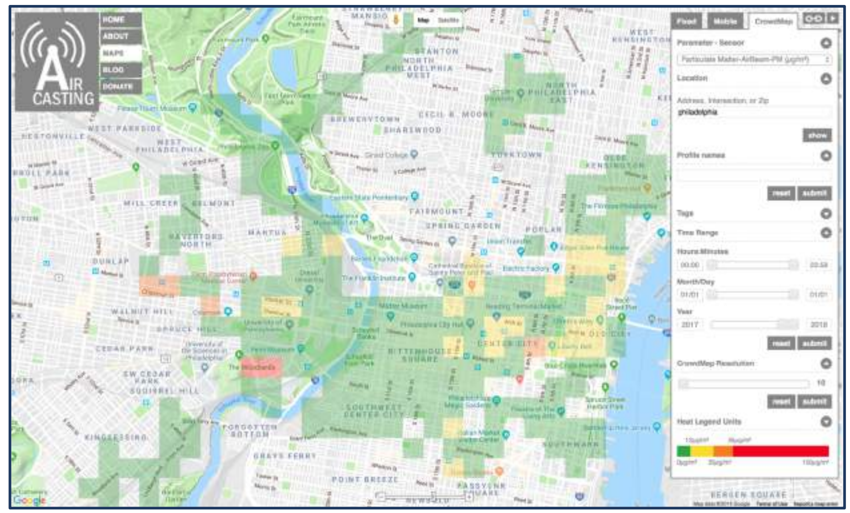


- Environmental data: vegetation, elevation, wind patterns, temperature
- Comparative studies of pollution measures: satellite data sources, portable pollution sensors
- Alternative geospatial interpolation methods: kriging, spline interpolation

Personal Monitoring of Fine Particulate Matter

AirBeam PM2.5 Sensors

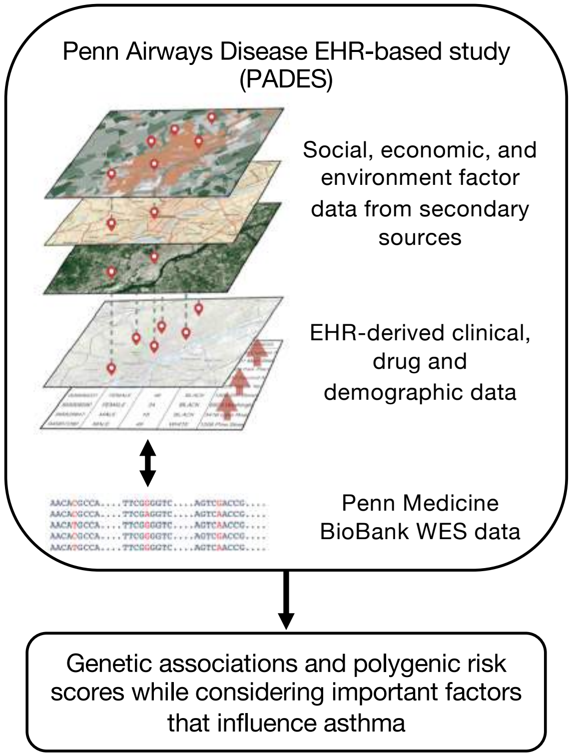
Greater Spatial Resolution than EPA Monitors



http://aircasting.org

Future Directions

- Integrate Penn Medicine Biobank whole exome sequencing data into EHR-based studies
- Determine contribution of genetics vs. social, economic and environmental factors to asthma



Acknowledgements

Himes Lab
Mengyuan Kan, PhD
Jaehyun Joo, PhD
Sherrie Xie
Rebecca Greenblatt
Avantika Diwadkar, MS
Colin Christie
Su Ly
Lena Leszinsky

Funding
NHLBI R01 HL133433, R01 HL141992
F31 HL142153
Center of Excellence in Environmental Toxicology (CEET) Pilot Award NIEHS P30-ES013508
UPenn McCabe Fund
UPenn Faculty Opportunity Fund

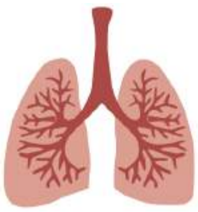
Collaborators

University of Pennsylvania
Andrea Apter, MD
Frances Barg, PhD
Casey Greene, PhD
Rebecca Hubbard, PhD

Harvard School of Public Health
Quan Lu, PhD
Xiaofeng Jiang, PhD
Maoyun Sun, PhD

Rutgers University
Reynold A Panettieri Jr, MD
Cynthia Koziol-White, PhD
William Jester

University of California, San Francisco
Esteban Burchard, MD
Angel Mak, PhD
Sam Oh, PhD



http://himeslab.org