

Approaches to Link Geospatially Varying Social, Economic, and Environmental Factors with Electronic Health Record Data to Better Understand Asthma Exacerbations

Sherrie Xie, BS¹, Blanca E. Himes, PhD¹

¹Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA, USA

Abstract

Electronic health record (EHR)-derived data has become an invaluable resource for biomedical research, but is seldom used for the study of the health impacts of social and environmental factors due in part to the unavailability of relevant variables. We describe how EHR-derived data can be enhanced via linking of external sources of social, economic and environmental data when patient-related geospatial information is available, and we illustrate an approach to better understand the geospatial patterns of asthma exacerbation rates in Philadelphia. Specifically, we relate the spatial distribution of asthma exacerbations observed in EHR-derived data to that of known and potential risk factors (i.e., economic deprivation, crime, vehicular traffic, tree cover). Areas of highest risk based on integrated social and environmental data were consistent with an area with increased asthma exacerbations, demonstrating that data external to the EHR can enhance our understanding of negative health-related outcomes.

Introduction

Electronic health records (EHRs) offer convenient and low-cost access to longitudinal data for large numbers of patients, which have made them an invaluable resource for biomedical research. EHR-derived data has been used to drive a number of research pursuits, including the creation of genetic and tissue biobanks,¹⁻⁵ recruitment of subjects for health studies,⁶⁻⁸ and the establishment of phenotype and comorbidity relationships in real-life populations.⁹⁻¹¹ Despite advances made in these areas, EHR-derived data is less often used in studies that consider social and environmental exposures due to the relatively narrow scope of EHR-derived variables and potential biases in patient populations utilizing health systems compared to all those residing at specific locations. Currently, efforts are underway to augment EHR-derived data by capturing personal data from social media, apps, and wearable devices.¹²⁻¹⁶

Here, we use a complementary approach to enhancing EHR-derived data in which residential addresses and other location data derived from EHRs are used to link clinical data to rich and diverse sources of social, economic and environmental variables to aid in the identification of factors driving health risks beyond what is possible with EHR data alone. The utility of this approach is illustrated by visualizing, analyzing, and integrating several sources of geospatial data to better understand asthma exacerbation rates in Philadelphia observed in EHR-derived data.

Linking Geospatially Varying Data to EHR-Derived Data

There are four steps involved in linking EHR-derived data to geospatially varying data: (1) identify and obtain geospatial variables from the EHR, (2) select external sources of geospatially varying data, (3) integrate EHR-derived data to external data sources via a geospatial or spatiotemporal link, and (4) assess associations between health outcomes derived from the EHR to geospatial trends. The first step is generally straightforward in terms of computational feasibility, as geospatial data variables that are collected by the health system can be found in codified EHR entries (e.g., residential addresses obtained for communication and billing purposes). Additionally, information about a person's place of work or education may be located in clinical notes where it can be mined, or in the case of school-aged children, place of education could be inferred from residential addresses and school district zone maps. Of utmost importance, prior to using geospatial information for patients, appropriate authorization by an Institutional Review Board (IRB) must be granted. Patient addresses are Protected Health Information under U.S. law, and cannot be used for research unless proper authorization is obtained. Depending on the research question, coarse spatial resolution at the level of census tracts or zip codes may suffice, while in other cases, geocodes are preferable. Ideally, any necessary geocoding is performed by a designated group at an institution with expertise and authority to determine whether and how to use tools for this purpose (e.g., *can an external resource like Google Maps API be used?*).

Next, researchers must determine which external sources of geospatial data are useful and available to complement specific EHR-based studies. A key consideration in the selection of external data sources is data quality. Incorporating data from sources of questionable quality can greatly compromise the validity of a study's findings, and researchers should read the documentation of data sources carefully, screening for potential bias and random error. There is a diverse range of high quality and resolution U.S. geospatial data on social, economic, and environmental factors that

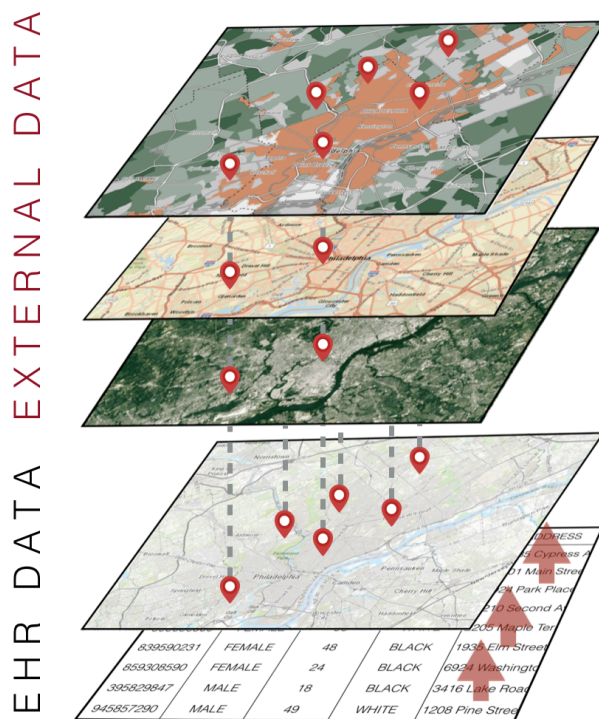


Figure 1. Geospatial variables derived from the EHR can be used to link health data to external sources of geospatial data on the social and physical environment.

exist in the public domain. Socioeconomic status (SES) variables are often obtained for people in epidemiologic studies by linking their geographic residence to data that the U.S. Census Bureau releases, including the American Community Survey (ACS) and the American Housing Survey (AHS). ACS data has been used to detect socioeconomic gradients for a variety of health outcomes, including cancer incidence¹⁷ and childhood obesity.¹⁸ AHS data, which contains information on U.S. housing quality and amenities, has been utilized to study the relationship between housing quality and asthma incidence,¹⁹ as well as the mitigating effect of air conditioning on the adverse health effect of particle pollution.²⁰ The U.S. Environmental Protection Agency (EPA) releases several data products on the U.S. physical environment, including EPA Air Data, which have been used in health studies to estimate exposure to air pollution.^{19, 21} State and local governments also release data that are useful for the study of health outcomes. For example, the City of Philadelphia shares municipal-level information on an open data website²² that includes an urban tree canopy dataset that was used to measure associations between tree cover and gun assault.^{23, 24} Crime data released by the Philadelphia Police Department has been used to study the association between crime density and adolescent firearm homicide.^{25, 26} Data on traffic maintained by state departments of transportation have been used to estimate exposure to traffic-related air pollution.²⁷

The third step, data integration, occurs after geospatial variables have been obtained from EHRs and external sources. Methods for linking EHR-derived data to geospatial data are determined, in part, by the geospatial data type. Broadly, these data types include: point data, line data, raster data, area-based measures, and measurements taken at discrete locations. We assume that the geospatial variables obtained from EHRs are geocoded residential addresses, but similar approaches exist for linking broader-scale data (i.e., residential census tracts or zip codes). Point data, such as locations of crime incidents, traffic accidents or alcohol outlets, can be linked to residential address by summing the number of events within a circular buffer of the residence²⁸ or by estimating the local density of events through kernel density estimation.^{25, 26} Line data, which represent linear features such as rivers or roads, are linked to residential location via creation of indicator variables that are assigned according to residence proximity to a linear feature (e.g., whether a patient residence is located within 100 m of a highway). Road network data may contain additional values, such as traffic count, which allow for calculation of more refined variables, like traffic density.²⁷ Raster data, composed of values stored in grid cells, includes aerial images, land cover, and land use. A simple way of integrating raster data to residential location is to quantify a raster feature within a circular buffer of a residential location, such as the percentage of tree cover within a 250-m buffer of a subject's residential address.²⁹ Area-based measures, such as SES variables documented by the ACS, can be readily linked to residential address by determining which boundaries the residence is located within (e.g., the census tract or block group in which the patient resides).^{17, 18} Measurements taken at discrete locations, including data from stationary air pollution monitors operated by the EPA, can be used to spatially interpolate air pollution exposure at a residential location via inverse distance weighting.²¹

Finally, once data integration is complete, the association of EHR-derived health outcomes to geospatial trends can be measured. External geospatial data variables are commonly treated as additional covariates in regression models. Alternatively, spatial autocorrelation of these external variables can be modeled explicitly.^{30, 31}

Known Relationships Between Asthma Exacerbations and Social and Environmental Variables

Asthma is a common chronic disease that affects one in every twelve children and one in every thirteen adults nationally.^{32, 33} Exacerbations, which are episodes of worsening asthma symptoms requiring the use of systemic corticosteroids to prevent serious outcomes, significantly impact patients' quality of life and are a source of substantial

healthcare costs.³⁴⁻³⁷ Many social and environmental factors are known to predispose asthma patients to exacerbations. Traffic-related air pollution, believed to induce airway inflammation and hyper-responsiveness, is a well-documented trigger of asthma exacerbations.³⁸ Asthma morbidity has been associated with several factors representative of low SES, including low educational attainment,³⁹ working class occupation,⁴⁰ unemployment,⁴¹ and low household income.^{41, 42} Crime has also been documented as a risk factor for asthma morbidity, with studies in children linking violence and violent crime to asthma symptom severity,^{43, 44} and studies in adults associating residence in regions of high crime with decreased adherence to asthma medication.⁴⁵ The physiologic mechanism underlying increased morbidity as a result of exposure to crime is thought to be stress, which upregulates mediators of allergic response and alters airway structure and functioning.^{44, 46}

Urban trees have many documented health benefits, including reduction of air pollution and heat; promotion of physical activity and mental health; and reduction in rates of violence and crime.⁴⁷⁻⁴⁹ However, their effect on asthma is less clear. The ability of urban trees to mitigate air pollution have led some to hypothesize that they could benefit respiratory health,^{50, 51} but the link between tree pollen and the exacerbation of asthma have raised concerns that urban trees could, instead, exert a negative effect.^{29, 52} Studies examining the association between tree cover and asthma have yielded mixed results: one study reported a positive association between residential tree cover and asthma prevalence,²⁹ while others have reported no or even negative (i.e. protective) associations.^{53, 54} Further studies are needed to better understand the association between urban trees and asthma exacerbations.

Geospatial Distribution of Asthma Exacerbations in Philadelphia According to EHR-Derived Data

In previous work, we used EHR-derived data to characterize the geospatial distribution of asthma exacerbations in the Philadelphia area.⁵⁵ We found significant exacerbation hot spots and cold spots that could not be explained by demographic characteristics and other EHR-derived covariates, which suggest that spatially-varying factors in the region contribute to asthma exacerbation rates. Philadelphia is a city with a heterogeneous landscape characterized by diverse neighborhoods and environmental features, many of which could affect the risk of asthma exacerbations. Here, we compare the geospatial distribution of asthma exacerbations to the distributions of three known exacerbation risk factors (i.e., economic deprivation, crime, vehicular traffic) and tree cover, which has an undetermined association with asthma exacerbations.

Methods

Study population. Characteristics of this study population have been described previously.⁵⁵ Briefly, subjects were adults with asthma encounters at the University of Pennsylvania Health System (UPHS) between 2010 and 2014. Inclusion criteria were: at least one primary ICD-9 code for asthma, at least one prescription for a short-acting beta₂-agonist inhaler, and residence within Philadelphia. An encounter was classified as an asthma exacerbation if it had a primary ICD-9 code for asthma (i.e. 493*) and a prescription for an oral corticosteroid, a definition that attempted to capture with EHR-derived data what is considered an exacerbation according to U.S. national asthma management guidelines.⁵⁶ Cases were defined as patients with at least one asthma exacerbation during the study period, and controls were defined as those with no asthma exacerbations. The University of Pennsylvania Institutional Review Board approved our study (protocol number 824789).

Socioeconomic conditions. We extracted twelve SES variables for all U.S. census block groups in Philadelphia from the 2010-2014 ACS 5-year estimates. These variables have been validated in prior studies to detect associations between SES and a range of health outcomes, including cancer incidence, low birth weight, and mortality.^{17, 57, 58} To adjust for multicollinearity among SES variables, we performed dimension reduction via factor analysis to derive a composite SES score. We examined factor loadings to determine whether the composite score represented a familiar construct (e.g. poverty, affluence) to assist with interpretation of results. To visualize the geospatial distribution of SES in Philadelphia, we used the factor loadings to calculate the SES score for each block group and mapped our results as a raster image. Block groups were chosen as the geographic level of data extraction and analysis because they are the finest level of geography that ACS reports, containing an average population of 1,000 residents. Raster and spatial data manipulations were performed using the *raster* and *rgdal* R packages.⁵⁹⁻⁶¹

Exposure to crime. A publicly available database of all crime incidents in Philadelphia is maintained by the Philadelphia Police Department.⁶² The database includes the date, type of offense, and geocoded location for each incident. Geocoordinates in the database were used to plot the point locations of all crime incidents that occurred in 2010-2014. These point data were then summed using a 500-m circular moving window to estimate the density of crime incidents across Philadelphia. Population density was estimated from 2010 U.S. Census population counts at

the block group level with a 500-m circular moving window. Crime rate was then calculated across the city as the crime density divided by the population density.

Exposure to vehicular traffic. We obtained annual average daily traffic (AADT) measurements for all major road segments in Philadelphia from the Pennsylvania Department of Transportation.⁶³ AADT is a measure of the total number of vehicles that travel on both sides of a road segment on a typical day, for all days in a week, for a one-year period. The AADT for a segment of road can be multiplied by the length of that segment to obtain the daily vehicle miles traveled (DVMT). We estimated the geospatial distribution of traffic exposure by calculating the DVMT using a 250-m circular moving window across Philadelphia.

Exposure to tree cover. A detailed and accurate (0.5-m resolution) assessment of tree canopy in Philadelphia was performed for the year 2008 by the University of Vermont Spatial Analysis Laboratory using an automated object-based analysis approach that combined orthophotography, high resolution light detection and ranging (LiDAR) data provided by the U.S. Department of Agriculture Forest Service, and ancillary GIS data (building footprints, road polygons, and hydrography) provided by the City of Philadelphia.²⁴ Tree cover exposure was estimated as the percent of land area composed of tree canopy in a circular 250-m moving window across Philadelphia.

Evaluating asthma exacerbation hotspots using EHR-derived data. We identified hotspots of asthma exacerbations using methods described previously.^{55,64,65} Briefly, we used a generalized additive model (GAM), which can estimate the log odds of an outcome as a function of location with adjustment for covariates. The GAM takes the form $\text{logit}[p(x,y)] = \gamma * z + S(x,y)$, where $\text{logit}[p(x,y)]$ is the log odds of exacerbation at latitude and longitude coordinates (x, y) , z is a vector of covariates with corresponding regression parameters γ , and $S(x,y)$ is a bivariate smooth of latitude and longitude that represents the spatial variation unexplained by the covariates. Using a GAM, we assessed the effect of geographic residence extracted from EHRs on risk of asthma exacerbation, with simultaneous smoothing of geographic location and adjustment for patient-level covariates. The variables selected for adjustment were significant independent predictors of exacerbation as previously determined by multivariable logistic regression: gender, race, age, BMI, smoking status, and insurance billing class.⁵⁵ The GAM was used to determine the log odds of asthma exacerbation across a point grid of Philadelphia, which were then converted to an odds ratio by using the odds of exacerbation of the entire study population as a reference group. A global test was performed against the null hypothesis that exacerbation odds did not depend on geographic location by permuting the assignments of cases and controls over residential geocoordinates. Given a significant global test, a local test was performed to identify areas of increased or decreased odds of exacerbation. The model fitted to the permutations was used to produce a distribution of log odds at each point on the grid, and points that ranked in the upper or lower 0.5% of the

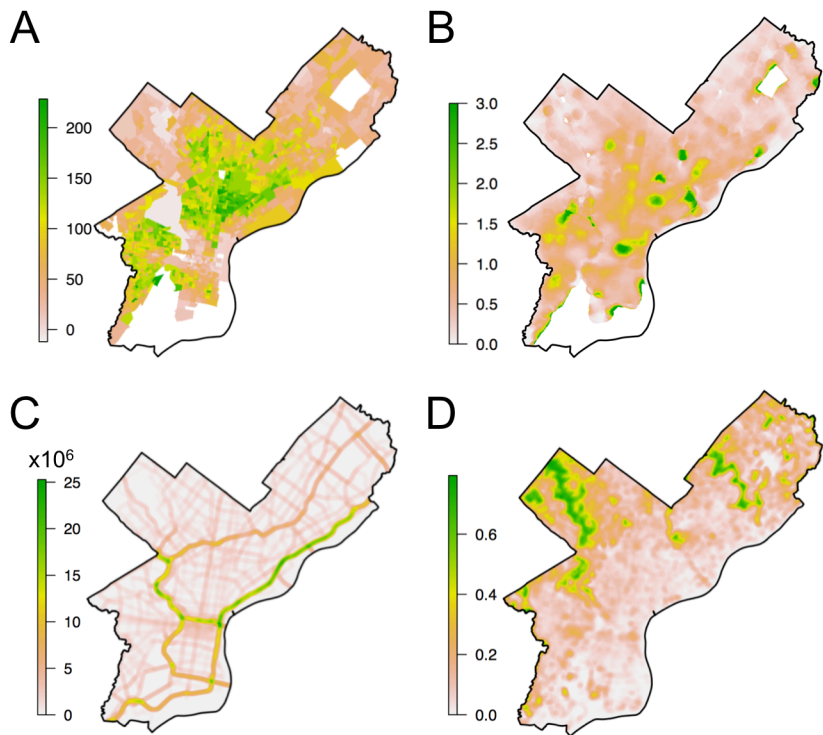


Figure 2. The geospatial distribution of known and potential risk factors for asthma exacerbations in Philadelphia. (A) Composite deprivation score for U.S. Census block groups. (B) Crime rate normalized by population density calculated with a 500-m radius circular moving window. (C) Traffic exposure estimated as the daily vehicle miles traveled (DVMT) in a 250-m radius circular moving window. (D) Percentage tree cover calculated with a 250-m radius circular moving window.

permutation distributions were defined as “hotspots” or “cold spots,” respectively. Spatial GAM analyses were performed using the *MapGAM* R package.⁶⁶

Identification of high-risk areas using social and environmental data. We integrated variables with a literature-supported association with asthma exacerbations (low SES, crime, and vehicular traffic) to create a variable that indicated areas of residence where the risk of asthma exacerbations is expected to be highest on their basis. Our *ad hoc* risk variable was created by intersecting the greatest socioeconomic deprivation and highest crime rate quintiles with the highest risk tertile for vehicular traffic. Vehicular traffic was divided into tertiles rather than quintiles because the road density in Philadelphia is high, and thus, a third of the city by land area likely experiences significant exposure to vehicular traffic.

Results

Geospatial distribution of SES, crime, vehicular traffic, and tree cover. The twelve block group-level SES measures that were extracted from the 2010-2014 ACS 5-year estimates and utilized in factor analysis are given in Table 1. Overall, there was less than 1% missingness in SES measures by block group. Measures associated with low SES (e.g., % low-income, % poverty, % unemployed) loaded positively on the composite factor, whereas measures associated with high SES (i.e., % high-income and % high education) loaded negatively. Thus, the composite score can be interpreted as an overall measure of economic deprivation, with higher scores representing greater disadvantage. Deprivation scores for Philadelphia block groups ranged from -12 to 228 (median score = 79), and were not calculated for nine block groups with missing data. Block groups with the highest deprivation score were found to be clustered in the northern and western parts of the city (Figure 2A). A total of 946,326 crime incidents occurred in Philadelphia in 2010-2014. After normalizing for local population density, pockets of high crime (> 2 incidents per person) were found scattered throughout north and west Philadelphia (Figure 2B). Exposure to vehicular traffic was estimated to be highest near major roads and freeways, which bisect the city longitudinally and laterally, and run along the southern and western borders (Figure 2C). Tree cover was found to be relatively low throughout the city (Figure 2D). Over three-quarters of the city had less than 20% tree cover within a 250-m radius, and over a quarter of the city had less than 5% tree cover. Areas of high tree cover (>50% within a 250-m radius) were few, totaling less than 4% of total land area, and were concentrated in the northwest and northeast.

Asthma exacerbation hotspots. A total of 4,238 subjects met our inclusion criteria, of which 1,509 were cases (i.e. had at least one asthma exacerbation). The global test statistic against the null hypothesis that the odds of exacerbations do not vary by location was highly significant ($p < 0.001$), suggesting that even after adjusting for EHR-derived covariates (i.e., gender, race, age, BMI, smoking status, insurance billing class), asthma exacerbations are spatially correlated. Local tests identified a hotspot in West Philadelphia and a cold spot in North Philadelphia (Figure 3).

Identification of high-risk areas. Block groups in areas with the highest quintile of economic deprivation are concentrated in North and West Philadelphia, and the top quintile for crime ranges throughout the city, with the largest clusters occurring in North and West Philadelphia (Figure 4A). The top tertile of vehicular traffic exposure extends throughout the city, following the distributions of major road segments (Figure 4A). The intersection of the highest risk quantiles for these known exacerbation risk factors identified regions in North and West Philadelphia as posing the greatest risk for asthma patients (Figure 4B). The high-risk areas identified in West Philadelphia

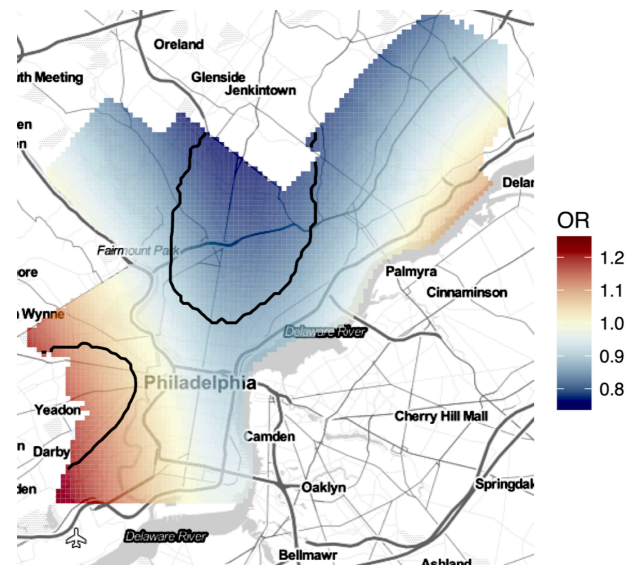


Figure 3. The spatial distribution of the odds of exacerbation, adjusted for gender, race, age, BMI, smoking status, and financial class. Significant hotspots and cold spots ($p < 0.01$) are indicated by black contour lines.

coincide with the exacerbation hotspot identified using GAM analysis, while the high-risk areas in North Philadelphia partially coincide with the exacerbation cold spot.

Table 1. Factor loadings for composite block group-based SES indices constructed using data from ACS 2010-2014 5-year estimates

Variable Name	ACS Measure	Factor Loading
% Low-income	Percentage of households earning < 50% the national median household income for 2010 (i.e. < \$25,000)	0.904
% High-income	Percentage of households earning $\geq$ 400% the national median household income for 2010 (i.e. $\geq$ \$200,000)	-0.548
% Poverty	Percentage of households below the federal poverty threshold	0.889
% SNAP	Percentage of households utilizing the Supplemental Nutrition Assistance Program (SNAP)	0.818
% Unemployed	Percentage of people aged $\geq$ 16 years in the work force who are unemployed	0.510
% Uninsured	Percentage of people (all ages) without health insurance coverage	0.276
% Low Education	Percentage of people aged $\geq$ 25 years with less than 12 th -grade education	0.662
% High Education	Percentage of people aged $\geq$ 25 years with a master's, professional school, or doctorate degree	-0.651
% Crowded	Percentage of households containing more than one person per room	0.136
% Incomplete Plumbing	Percentage of households with incomplete plumbing facilities	0.381
% Incomplete Kitchen	Percentage of households with incomplete kitchen facilities	0.398
% Incomplete Phone	Percentage of households lacking phone or cellphone service	0.222

Discussion

We visualized the distributions of known and potential risk factors of asthma exacerbations in Philadelphia and compared them to geospatial patterns of asthma exacerbations in the same region. We identified potential high-risk regions for asthma exacerbations in West Philadelphia and North Philadelphia where economic deprivation, crime, and traffic density were highest. The high-risk areas identified in West Philadelphia overlapped with the exacerbation hotspot observed in EHR-derived data, suggesting that economic deprivation, crime, and traffic density contribute to the high rates of asthma exacerbation in this region. This hypothesis can be tested in future studies via regression analyses to find effect estimates for these social and environmental variables while controlling for potential confounders. In addition, sample stratification and the incorporation of interaction terms to regression models can assess for effect modification by variables like age and smoking status. In apparent contrast to this result, a high-risk area identified in North Philadelphia was associated with the cold spot where rates of asthma exacerbation were low. This discrepancy is likely due to bias in provider coverage inherent in our data: although UPHS, from which our study population was derived, is the largest health provider in Philadelphia, North Philadelphia is primarily served by a

separate provider network (i.e., Temple Health). Thus, the low rates of asthma exacerbation we observed in North Philadelphia are likely due to the selection bias that results from utilizing a health system-based cohort in an area with competing providers. Although the spatial odds calculated by the GAM model controlled for spatial variation in patient density (the odds of being a case are divided by the odds of being a control at each grid point), residual confounding may still be present in areas of low patient density. In future analyses, we will consider the density of UPHS providers as a variable to select our study area and/or adjust results.

Raster images of economic deprivation, crime, traffic density, and tree cover in Philadelphia demonstrated that use of these social and environmental variables derived from external data sources can complement EHR-derived health data and provide an overview of potential risk areas. Assigning values for these variables to each subject on the basis of their residential address is more computationally efficient than creating raster layers. For example, assigning tree cover to each subject entailed calculating percentage tree cover for 4,238 residential address locations. In contrast, creating a raster image of tree cover involved performing this calculation for each raster cell - over 1×10^8 computations. We selected a 250-m radius circular moving windows to calculate vehicular traffic and tree cover exposure and a 500-m circular moving window to calculate crime rate, because similar buffer sizes have been used to calculate these quantities in other studies.^{28, 29, 67} For brevity, we showed a single buffer size for each exposure, though studies typically include at least one other buffer size, which allows for detection of a dose-response relationship with proximity of the exposure.^{28, 29, 67} In future analyses, a range of buffer sizes will be used to determine the sensitivity of results to buffer size. Furthermore, while the present study selected exposure variables with documented effects on asthma, future studies will incorporate hypothesis-free social and environmental data sources. This broader, data-driven approach that leverages multiple external data sources in conjunction with EHR-derived data may permit discovery of novel relationships not available via any other means, although additional considerations during analyses will be necessary to reduce false positive associations that may appear when considering large search spaces.

Use of residential addresses to link EHR-derived data to geospatial data provides some insights into social and environmental conditions, but residence makes up only a portion of an individual's geographic exposures. Extracting a person's places of work or education from EHR notes or other sources would allow for a fuller characterization of

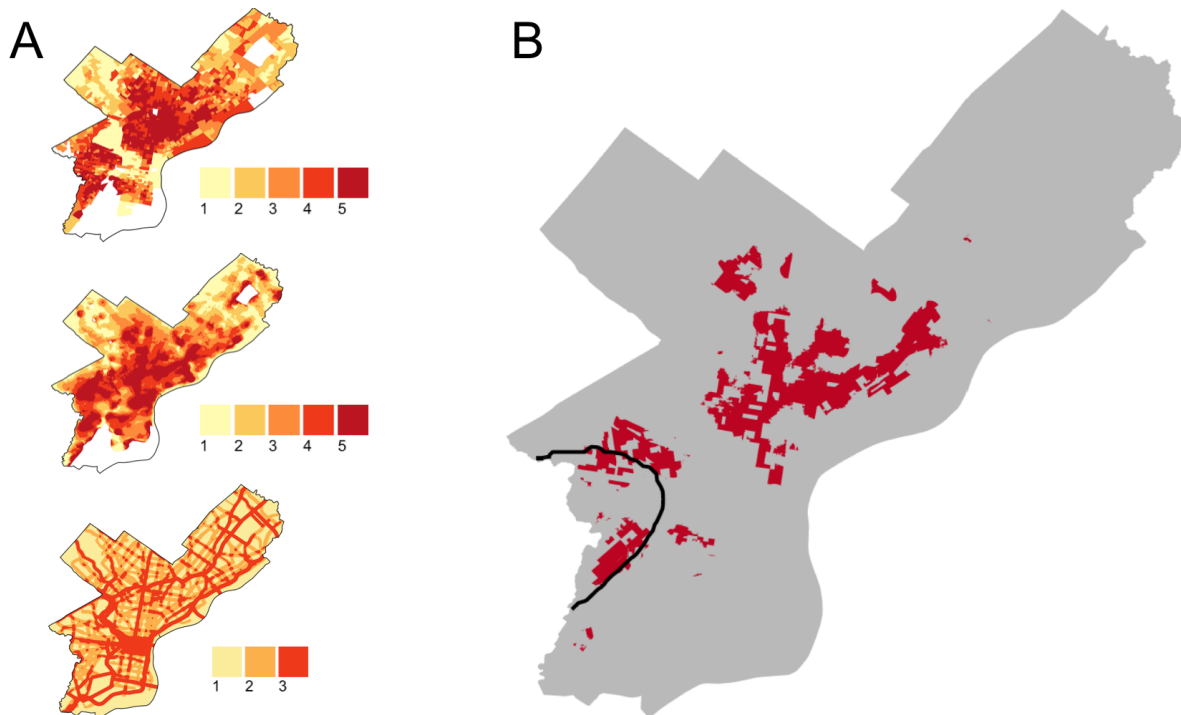


Figure 4. High-risk areas for asthma exacerbations were found by intersecting highest-risk quantiles from known exacerbation risk factors. (A) The geospatial distribution of economic deprivation (top), crime rate (middle) and traffic exposure (bottom) split into quintiles (deprivation and crime) and tertiles (traffic). (B) High-risk areas are colored red. The exacerbation hotspot derived from GAM analysis are indicated by the black contour line.

their geographic range. Full activity paths could also be captured using GPS-enabled devices,⁶⁸ although doing so requires considerable resources and is currently not scalable for a study utilizing a large EHR-derived cohort.

Conclusion

High resolution geospatial data on social, economic, and environmental factors, much of which exist in the public domain, can be linked to EHR-derived data to aid in the identification of factors driving health outcomes beyond what is possible with the use of EHR data alone. We used public data on economic deprivation, crime, traffic density, and tree cover to visualize the geospatial distributions of these known and potential asthma exacerbation risk factors and compared them to geospatial patterns of asthma exacerbations inferred from EHR-derived data. A hotspot of asthma exacerbation coincided with areas containing highest densities of economic deprivation, crime, and traffic density.

Acknowledgements

We would like to thank Yuliya Borovskiy, M.S. from the University of Pennsylvania Penn Data Store for extracting the EHR data used for this project. This work was partially supported by National Institutes of Health (NIH) R01-HL133433 and P30 ES013508.

References

1. Ritchie MD, Denny JC, Crawford DC, Ramirez AH, Weiner JB, Pulley JM, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *Am J Hum Genet.* 2010;86(4):560-72.
2. Himes BE, Klanderman B, Kohane IS, Weiss ST. Assessing the reproducibility of asthma genome-wide association studies in a general clinical population. *J Allergy Clin Immunol.* 2011;127(4):1067-9.
3. Munoz M, Pong-Wong R, Canela-Xandri O, Rawlik K, Haley CS, Tenesa A. Evaluating the contribution of genetics and familial shared environment to common disease using the UK Biobank. *Nat Genet.* 2016;48(9):980-3.
4. Gainer VS, Cagan A, Castro VM, Duey S, Ghosh B, Goodson AP, et al. The Biobank Portal for Partners Personalized Medicine: A Query Tool for Working with Consented Biobank Samples, Genotypes, and Phenotypes Using i2b2. *J Pers Med.* 2016;6(1).
5. Carey DJ, Fetterolf SN, Davis FD, Faucett WA, Kirchner HL, Mirshahi U, et al. The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. *Genet Med.* 2016;18(9):906-13.
6. Shaibi G, Singh D, De Filippis E, Hernandez V, Rosenfeld B, Otu E, et al. The Sangre Por Salud Biobank: Facilitating Genetic Research in an Underrepresented Latino Community. *Public Health Genomics.* 2016;19(4):229-38.
7. Patel PJ, Borovskiy Y, Killian A, Verdino RJ, Epstein AE, Callans DJ, et al. Optimal QT interval correction formula in sinus tachycardia for identifying cardiovascular and mortality risk: Findings from the Penn Atrial Fibrillation Free study. *Heart Rhythm.* 2016;13(2):527-35.
8. Wilson FP, Yang W, Machado CA, Mariani LH, Borovskiy Y, Berns JS, et al. Dialysis versus nondialysis in patients with AKI: a propensity-matched cohort study. *Clin J Am Soc Nephrol.* 2014;9(4):673-81.
9. Himes BE, Dai Y, Kohane IS, Weiss ST, Ramoni MF. Prediction of chronic obstructive pulmonary disease (COPD) in asthma patients using electronic medical records. *J Am Med Inform Assoc.* 2009;16(3):371-9.
10. Siebert S, Lyall DM, Mackay DF, Porter D, McInnes IB, Sattar N, et al. Characteristics of rheumatoid arthritis and its association with major comorbid conditions: cross-sectional study of 502 649 UK Biobank participants. *RMD Open.* 2016;2(1):e000267.
11. Pike MM, Decker PA, Larson NB, St Sauver JL, Takahashi PY, Roger VL, et al. Improvement in Cardiovascular Risk Prediction with Electronic Health Records. *J Cardiovasc Transl Res.* 2016;9(3):214-22.
12. Eggleston EM, Weitzman ER. Innovative uses of electronic health records and social media for public health surveillance. *Current diabetes reports.* 2014;14(3):468.
13. Himes BE, Weitzman ER. Innovations in health information technologies for chronic pulmonary diseases. *Respiratory research.* 2016;17:38.
14. Padrez KA, Ungar L, Schwartz HA, Smith RJ, Hill S, Antanavicius T, et al. Linking social media and medical record data: a study of adults presenting to an academic, urban emergency department. *BMJ quality & safety.* 2016;25(6):414-23.
15. Dorsey ER, Yvonne Chan YF, McConnell MV, Shaw SY, Trister AD, Friend SH. The Use of Smartphones for Health Research. *Acad Med.* 2017;92(2):157-60.

16. Rumsfeld JS, Brooks SC, Aufderheide TP, Leary M, Bradley SM, Nkonde-Price C, et al. Use of Mobile Devices, Social Media, and Crowdsourcing as Digital Strategies to Improve Emergency Cardiovascular Care: A Scientific Statement From the American Heart Association. *Circulation*. 2016;134(8):e87-e108.
17. Yu M, Tatalovich Z, Gibson JT, Cronin KA. Using a composite index of socioeconomic status to investigate health disparities while protecting the confidentiality of cancer registry data. *Cancer Causes Control*. 2014;25(1):81-92.
18. Eagle TF, Sheetz A, Gurm R, Woodward AC, Kline-Rogers E, Leibowitz R, et al. Understanding childhood obesity in America: Linkages between household income, community resources, and children's behaviors. *Am Heart J*. 2012;163(5):836-43.
19. Bell ML, Ebisu K, Peng RD, Dominici F. Adverse health effects of particulate air pollution: modification by air conditioning. *Epidemiology*. 2009;20(5):682-6.
20. Hughes HK, Matsui EC, Tschudy MM, Pollack CE, Keet CA. Pediatric Asthma Health Disparities: Race, Hardship, Housing, and Asthma in a National Survey. *Acad Pediatr*. 2017;17(2):127-34.
21. Volk HE, Lurmann F, Penfold B, Hertz-Picciotto I, McConnell R. Traffic-related air pollution, particulate matter, and autism. *JAMA Psychiatry*. 2013;70(1):71-7.
22. City of Philadelphia. Open Data Philly 2018 [Available from: <https://www.opendataphilly.org/>].
23. Kondo MC, South EC, Branas CC, Richmond TS, Wiebe DJ. The Association Between Urban Tree Cover and Gun Assault: A Case-Control and Case-Crossover Study. *Am J Epidemiol*. 2017;186(3):289-96.
24. USDA Forest Service; University of Vermont Spatial Analysis Laboratory. High Resolution Landcover, Philadelphia PA, 2008. 2011.
25. Culyba AJ, Jacoby SF, Richmond TS, Fein JA, Hohl BC, Branas CC. Modifiable Neighborhood Features Associated With Adolescent Homicide. *JAMA Pediatr*. 2016;170(5):473-80.
26. Hohl BC, Wiley S, Wiebe DJ, Culyba AJ, Drake R, Branas CC. Association of Drug and Alcohol Use With Adolescent Firearm Homicide at Individual, Family, and Neighborhood Levels. *JAMA Intern Med*. 2017;177(3):317-24.
27. Gunier RB, Hertz A, Von Behren J, Reynolds P. Traffic density in California: socioeconomic and ethnic differences among potentially exposed children. *J Expo Anal Environ Epidemiol*. 2003;13(3):240-6.
28. Duncan DT, Hatzenbuehler ML. Lesbian, gay, bisexual, and transgender hate crimes and suicidality among a population-based sample of sexual-minority adolescents in Boston. *Am J Public Health*. 2014;104(2):272-8.
29. Lovasi GS, O'Neil-Dunne JP, Lu JW, Sheehan D, Perzanowski MS, Macfaden SW, et al. Urban tree canopy and asthma, wheeze, rhinitis, and allergic sensitization to tree pollen in a New York City birth cohort. *Environ Health Perspect*. 2013;121(4):494-500.
30. Pope CA, 3rd, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA*. 2002;287(9):1132-41.
31. Havard S, Deguen S, Zmirou-Navier D, Schillinger C, Bard D. Traffic-related air pollution and socioeconomic status: a spatial autocorrelation study to assess environmental equity on a small-area scale. *Epidemiology*. 2009;20(2):223-30.
32. Center for Disease Control and Prevention. Most Recent Asthma Data: National Current Asthma Prevalence (2015) 2017 [updated June 2017. Available from: https://www.cdc.gov/asthma/most_recent_data.htm].
33. Akinbami LJ, Moorman JE, Bailey C, Zahran HS, King M, Johnson CA, et al. Trends in asthma prevalence, health care use, and mortality in the United States, 2001-2010. *NCHS data brief*. 2012(94):1-8.
34. Bai TR, Vonk JM, Postma DS, Boezen HM. Severe exacerbations predict excess lung function decline in asthma. *Eur Respir J*. 2007;30(3):452-6.
35. O'Byrne PM, Pedersen S, Lamm CJ, Tan WC, Busse WW, Group SI. Severe exacerbations and decline in lung function in asthma. *Am J Respir Crit Care Med*. 2009;179(1):19-24.
36. Lloyd A, Price D, Brown R. The impact of asthma exacerbations on health-related quality of life in moderate to severe asthma patients in the UK. *Prim Care Respir J*. 2007;16(1):22-7.
37. Barnett SB, Nurmagambetov TA. Costs of asthma in the United States: 2002-2007. *J Allergy Clin Immunol*. 2011;127(1):145-52.
38. Guarnieri M, Balmes JR. Outdoor air pollution and asthma. *Lancet*. 2014;383(9928):1581-92.
39. Bacon SL, Bouchard A, Loucks EB, Lavoie KL. Individual-level socioeconomic status is associated with worse asthma morbidity in patients with asthma. *Respir Res*. 2009;10:125.
40. Hedlund U, Eriksson K, Ronmark E. Socio-economic status is related to incidence of asthma and respiratory symptoms in adults. *Eur Respir J*. 2006;28(2):303-10.
41. Beck AF, Moncrief T, Huang B, Simmons JM, Sauers H, Chen C, et al. Inequalities in neighborhood child asthma admission rates and underlying community characteristics in one US county. *J Pediatr*. 2013;163(2):574-80.
42. Keet CA, McCormack MC, Pollack CE, Peng RD, McGowan E, Matsui EC. Neighborhood poverty, urban residence, race/ethnicity, and asthma: Rethinking the inner-city asthma epidemic. *J Allergy Clin Immunol*. 2015;135(3):655-62.

43. Wright RJ, Mitchell H, Visness CM, Cohen S, Stout J, Evans R, et al. Community violence and asthma morbidity: the Inner-City Asthma Study. *Am J Public Health*. 2004;94(4):625-32.
44. Wright RJ, Steinbach SF. Violence: an unrecognized environmental exposure that may contribute to greater asthma morbidity in high risk inner-city populations. *Environ Health Perspect*. 2001;109(10):1085-9.
45. Williams LK, Joseph CL, Peterson EL, Moon C, Xi H, Krajenta R, et al. Race-ethnicity, crime, and other factors associated with adherence to inhaled corticosteroids. *J Allergy Clin Immunol*. 2007;119(1):168-75.
46. Vig RS, Forsythe P, Vliagoftis H. The role of stress in asthma: insight from studies on the effect of acute and chronic stressors in models of airway inflammation. *Ann N Y Acad Sci*. 2006;1088:65-77.
47. Nowak DJ, Crane DE, Stevens JC. Air pollution removal by urban trees and shrubs in the United States. *Urban Forestry & Urban Greening*. 2006;4(3-4):115-23.
48. Hartig T, Mitchell R, de Vries S, Frumkin H. Nature and health. *Annu Rev Public Health*. 2014;35:207-28.
49. Kondo MC, South EC, Branas CC, Richmond TS, Wiebe DJ. The association between urban tree cover and gun assault: a case-control and casecrossover study. *Am J Epidemiol*. 2017.
50. McPherson EG, Simpson JR, Xiao QF, Wu CX. Million trees Los Angeles canopy cover and benefit assessment. *Landscape Urban Plan*. 2011;99(1):40-50.
51. Rao M, George LA, Rosenstiel TN, Shandas V, Dinno A. Assessing the relationship among urban trees, nitrogen dioxide, and respiratory health. *Environ Pollut*. 2014;194:96-104.
52. Dales RE, Cakmak S, Judek S, Dann T, Coates F, Brook JR, et al. Influence of outdoor aeroallergens on hospitalization for asthma in Canada. *J Allergy Clin Immunol*. 2004;113(2):303-6.
53. Pilat MA, McFarland A, Snelgrove A, Collins K, Waliczek TM, Zajicek J. The Effect of Tree Cover and Vegetation on Incidence of Childhood Asthma in Metropolitan Statistical Areas of Texas. *Horttechnology*. 2012;22(5):631-7.
54. Lovasi GS, Quinn JW, Neckerman KM, Perzanowski MS, Rundle A. Children living in areas with more street trees have lower prevalence of asthma. *J Epidemiol Community Health*. 2008;62(7):647-9.
55. Xie S, Greenblatt R, Levy MZ, Himes B. Enhancing electronic health record data with geospatial information. *AMIA Joint Summits March 28, 2017; San Francisco, CA2017*.
56. National Asthma E, Prevention P. Expert Panel Report 3 (EPR-3): Guidelines for the Diagnosis and Management of Asthma-Summary Report 2007. *J Allergy Clin Immunol*. 2007;120(5 Suppl):S94-138.
57. Yost K, Perkins C, Cohen R, Morris C, Wright W. Socioeconomic status and breast cancer incidence in California for different race/ethnic groups. *Cancer Causes Control*. 2001;12(8):703-11.
58. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. *Am J Public Health*. 2005;95(2):312-23.
59. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2017.
60. Hijmans RJ. *raster: Geographic Data Analysis and Modeling*. 2016.
61. Bivand R, Keitt T, Rowlingson B. *rgdal: Bindings for the Geospatial Data Abstraction Library*. 2017.
62. Crime Incidents (2006 - Present) [Internet]. City of Philadelphia. 2018. Available from: <https://www.opendataphilly.org/dataset/crime-incidents>.
63. RMSTRAFFIC (Traffic Volumes) [Internet]. 2018 [cited 4 March 2018]. Available from: <https://data-pennshare.opendata.arcgis.com/datasets/rmstraffic-traffic-volumes>.
64. Vieira V, Webster T, Weinberg J, Aschengrau A, Ozonoff D. Spatial analysis of lung, colorectal, and breast cancer on Cape Cod: an application of generalized additive models to case-control data. *Environ Health*. 2005;4:11.
65. Webster T, Vieira V, Weinberg J, Aschengrau A. Method for mapping population-based case-control studies: an application using generalized additive models. *Int J Health Geogr*. 2006;5:26.
66. Bai L, Bartell S, Bliss R, Vieira V. MapGAM: Mapping Smoothed Effect Estimates from Individual-Level Data. 2016.
67. Lin S, Munsie JP, Hwang SA, Fitzgerald E, Cayo MR. Childhood asthma hospitalization and residential exposure to state route traffic. *Environ Res*. 2002;88(2):73-81.
68. Byrnes HF, Miller BA, Morrison CN, Wiebe DJ, Woychik M, Wiehe SE. Association of environmental indicators with teen alcohol use and problem behavior: Teens' observations vs. objectively-measured indicators. *Health Place*. 2017;43:151-7.