

Enhancing Electronic Health Record Data with Geospatial Information

Sherrie Xie, BS¹, Rebecca Greenblatt, BS¹, Michael Z. Levy, PhD¹, Blanca E. Himes, PhD¹
¹Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA, USA

Abstract

Electronic Health Record (EHR)-derived data is a valuable resource for research, and efforts are underway to overcome some of its limitations by using data from external sources to gain a fuller picture of patient characteristics, symptoms, and exposures. Our goal was to assess the utility of augmenting EHR data with geocoded patient addresses to identify geospatial variation of disease that is not explained by EHR-derived demographic factors. Using 2011-2014 encounter data from 27,604 University of Pennsylvania Hospital System asthma patients, we identified factors associated with asthma exacerbations: risk was higher in female, black, middle aged to elderly, and obese patients, as well as those with positive smoking history and with Medicare or Medicaid vs. private insurance. Significant geospatial variability of asthma exacerbations was found using generalized additive models, even after adjusting for demographic factors. Our work shows that geospatial data can be used to cost-effectively enhance EHR data.

Introduction

Electronic Health Records (EHRs) have become an invaluable resource for precision medicine research, including the creation of genetic and tissue biobanks¹⁻⁵, understanding phenotypes and comorbidity relationships in real-life populations⁶⁻⁸, and recruiting subjects for clinical trials and health services studies⁹⁻¹¹. Additionally, EHRs are necessary for the implementation of precision medicine in clinical practice and to enable patient-centered efforts to improve healthcare¹²⁻¹⁵. The success of EHR-derived data research is driven in part because EHR data offers a convenient and low-cost way to obtain detailed cross-sectional or longitudinal information for large numbers of subjects. EHR data does have limitations, however, including its biased nature compared to that of traditional epidemiological studies and the narrow insights it provides into patients' lives. To overcome these limitations, efforts are underway to augment EHR data with complementary personalized data from sources such as social media, portable sensors, and mobile phone apps¹⁶⁻²⁰. Publicly available geospatial data for a wide range of demographic and environmental factors is another resource to cost-effectively gain insights from EHR data in an effort to understand diseases at the community and individual levels^{21,22}.

Asthma is an episodic inflammatory lung disease characterized by airway hyperresponsiveness that affects over 25 million Americans, including 18.7 million adults²³. Exacerbations, which are episodes of worsening asthma symptoms requiring the use of systemic corticosteroids to prevent serious outcomes, are a major cause of morbidity and health care costs^{24, 25}. While asthma cannot be cured, national treatment guidelines exist for preventing exacerbations that are able to control symptoms in most patients²⁶. Despite the availability of these treatment options that are efficacious in clinical trials, asthma prevalence and exacerbation rates in real-life U.S. populations remain very high with persisting and marked disparities by gender, race/ethnicity and socioeconomic status^{27, 28}. These and other factors—poor housing conditions, outdoor and indoor pollutants, educational attainment, health literacy—that are related to asthma exacerbations vary in their geospatial distribution²⁹⁻³¹.

To effectively make progress toward understanding demographic and environmental factors that influence asthma exacerbations in real-life populations, EHR data is valuable as a longitudinal repository of many events affecting diverse populations. In combination with patient address information, EHR data can be augmented to include highly relevant and publicly available asthma-related geospatial data, including air quality, pollen, ultraviolet sensitivity index levels, proximity to major sources of pollution, and housing codes. Here, we analyze the geospatial distribution of asthma patients who received care in the University of Pennsylvania Health System (UPHS) to identify regions in the greater Philadelphia area where patients with frequent exacerbations live and lay a foundation for future studies that identify social and environmental factors associated with these areas and their relationship with genetic and biomarker data.

Methods

Study population. Our study population was obtained from Penn Data Store (PDS), a repository of EHR-derived data that is integrated from across multiple systems within UPHS. PDS supports medical research and patient care

initiatives through the health system and has been used in the past to study atrial fibrillation and acute kidney injury^{10, 11}. We obtained de-identified patient-level data for adult (i.e. aged 18 years or older) encounters that occurred between January 1, 2011 and December 31, 2014 and included an asthma International Classification of Disease, Ninth Revision (ICD-9) diagnosis code (i.e., 493*). Variables extracted included codified demographic information (i.e., gender, age, race/ethnicity, height and weight measures, financial class, smoking history, admission date and type (outpatient visit, emergency department visit, observation, or hospitalization), all ICD-9 codes noted during the encounter, and patient addresses. Medications prescribed during encounters were obtained from codified entries as well as NLP-extracted values from inpatient notes. For reference demographics, we obtained counts of the race/ethnicity and gender of all PDS patients encountered during the study period. The University of Pennsylvania Institutional Review Board approved our study (protocol number 824789).

For the current study, we included patients who were (1) less than 80 years old at first encounter, and (2) were followed for at least 3 years during the study period, as determined by earliest and latest encounter dates among all visits (i.e., visits with or without asthma ICD-9 codes). The latter criterion was implemented to increase the likelihood that study patients resided in the area during the study period such that fewer asthma encounters reflected less exacerbations/severe disease, rather than patients having moved elsewhere. We limited our study to patients who were *non-Hispanic white* and *non-Hispanic black or African American*, as over 91% of patients fell within these categories. Hereafter, we refer to these categories as *white* and *black*. To reduce erroneous weight and height entries, only weight measures within 80 and 700 pounds and height measures within 48 to 84 inches were kept. Patient weight was computed as the mean value of the patient's measures that were within 10 percent of the patient's median weight, and height as the mean of the patient's measures that were within 5 percent of the patient's median height. BMI was computed as the mean weight over mean height squared times 703.0704 and re-leveled into four groups: *not overweight or obese* (<25.0 kg/m²), *overweight* (25.0 to <30.0 kg/m²), *grade 1 obese* (30.0 to <35.0 kg/m²), *grade 2 obese* (35.0 to <40.0 kg/m²), *grades ≥ 3 obese* (≥40.0 kg/m²). Age was calculated as the patient's age at first encounter. Financial class was obtained from a codified billing field used at each patient's most recent hospital encounter, and was re-leveled into three groups: *Private Insurance, Medicare, Medicaid*.

Asthma exacerbation definition. Encounters were classified as asthma exacerbations if they had a primary ICD-9 code for asthma (i.e., 493*) and included prescription of an oral corticosteroid. Cases were defined as those patients who had at least one asthma exacerbation during the study period. Controls were defined as those who had no asthma exacerbations.

Statistical analysis. Statistical analyses were conducted in R³². To determine whether demographic factors (i.e., gender, race, age at first encounter, BMI, smoking status, financial class) were associated with asthma exacerbations, we used logistic regression analyses. Crude odds ratios (ORs) were computed with complete cases for each variable individually, and all factors were included in a multivariate logistic regression model from which adjusted ORs were determined. Differences in gender and race distributions between our study population and all UPHS patients were assessed via Chi-squared tests.

Geocoding. Patient addresses were geocoded to geographic (longitude and latitude) coordinates using the Google Maps Geocoding API run in the R environment³². Addresses were input as a character string in the format: [house number] [street name], [zipcode], and outputs received included geographic coordinates as well as output addresses (street address, city, state, and zip) corresponding to the geocoded coordinates. Apartment and unit numbers were omitted from the input string when present. Geocoding accuracy was determined by comparing input and output zipcodes, whereby a mismatch indicated an incorrect input address, or one not recognized by Google Maps. Because the aim of this study was to determine exacerbation risk at fine spatial scales, addresses were only included if they were correctly geocoded to a street address. Addresses that were only geocoded to the street level (i.e. with no house number), PO boxes, and incorrectly geocoded addresses were excluded from further analysis.

Geospatial analysis. Spatial analysis was performed using a generalized additive model (GAM), which can estimate the log odds of an outcome as a function of location with adjustment for covariates^{33, 34}. 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. The term S(x,y) was smoothed using a loess function, a locally-weighted regression smoother which adapts to local variation in point density. The amount of neighboring points utilized by the loess is determined by a span size optimized by minimizing the Akaike Information Criterion. We used GAM to determine the log odds of asthma exacerbation on a grid of 10,000 points arranged across the study region. The log odds at each point was converted to an odds ratio by

using the odds of exacerbation of the entire study population as a reference group³⁴. Our model included all covariates found to be significant predictors of exacerbation as determined by multivariate logistic regression p-values <0.05. While data was complete in the gender, race, age, and financial class variables, BMI measures were missing for 19% of patients and smoking status was missing in 3% of patients. Missingness of BMI was due to lack of height, rather than weight, measures in patients. To maximize the number of subjects used in spatial analysis, we created “missing” categories for BMI and smoking status. A global test was performed to test the null hypothesis that exacerbation odds did not depend on geographic location by running 999 permutations of the assignment of cases and controls over all patient addresses while keeping the total number of cases and controls the same and preserving the covariate information³⁴. For the observed data and each of the subsequent permutations, a deviance statistic was calculated as the difference of the deviances of the model with the smoothing term and the model without the smoothing term. The p-value of the global test is the rank of the deviance statistic of the observed data relative to the permutations divided by 1000. Given that the global test indicated geographic location to be a significant predictor of exacerbation, a local test for significance was performed to identify areas of significantly increased and 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. Points that ranked in the upper or lower 0.5% of the permutation distributions were defined as “hot spots” (areas of elevated odds) or “cold spots” (areas of decreased odds). Spatial analyses were performed with the R package *MapGAM*, and street maps of the study regions were created using the R package *ggmap*^{35, 36}.

Results

We obtained EHR-derived data for 51,178 patients who utilized UPHS between 2011 and 2014 and whose visits included ICD-9 codes for asthma. Compared to all patients who utilized UPHS during this time, asthma patients were more likely to be *female* and *black* [Table 1]. Of all asthma patients, 27,604 *white* or *black* patients met our inclusion criteria, including having events spanning at least 3 years between 2011 and 2014, and were selected for further study.

Table 1. Demographic characteristics of patients encountered at UPHS during 2011-2014. N (%) are shown. (*p <10⁻⁵)

	Asthma patients (N=51,178)	All patients (N=3,199,282)
Gender*		
Male	15,273 (29.8)	1,154,454 (39.8)
Female	35,905 (70.2)	1,746,632 (60.2)
Race/Ethnicity*		
White	26,991 (52.7)	1,705,967 (58.8)
Black	19,929 (38.9)	842,345 (29.0)

Characteristics of asthma patients with exacerbations. Of the 27,604 asthma patients who were followed ≥ 3 years for the 4-year study period, 2,773 were cases who had at least one exacerbation, and 24,831 were controls who had no exacerbations. The characteristics of cases and controls are presented in Table 2. According to univariate analyses, the gender distribution of cases and controls was not significantly different, but cases were significantly more likely than controls to be *black*, middle-aged or older, overweight or obese, have positive smoking history, and have *Medicare* or *Medicaid* rather than *Private Insurance* (p<0.05) [Table 3]. In multivariate analysis, *female* gender, *black* race/ethnicity, older ages, *grade ≥ 3 obesity*, *passive* or *quit* smoking history, and *Medicare* or *Medicaid* financial class vs. *Private Insurance* were independent predictors of asthma exacerbations [Table 3].

Geospatial distribution of asthma exacerbations. The geospatial distribution of study patients and location of UPHS facilities (i.e., hospitals and clinics) are displayed in Figure 1. Patient density was spatially heterogeneous and generally decreased with distance from UPHS provider sites. To ensure we had an adequate sample size to appropriately assess differences in exacerbation by location, areas with low patient density were excluded from analyses. The final study region was defined as a square area measuring 38 km on each side and centered over the bulk of patients. Of the 27,604 subjects in our study population, 26,402 had addresses that could be appropriately geocoded for geospatial analysis, and 74% of these address locations fell within the study region. GAM analysis with data from the resulting 19,599 patients (1,892 cases and 17,707 controls) found that asthma exacerbations were associated with geographic location. The global test statistic against the null hypothesis that exacerbation odds did not depend on location was highly significant (p <0.001) suggesting that even after adjusting for covariates (i.e. gender, race, age, BMI, smoking status, financial), asthma exacerbations were highly spatially correlated. Local tests identified two areas with significantly increased and two with significantly decreased exacerbation rates (p <0.01) [Figure 2]. Cold spots occurred in North Philadelphia and the bordering region of Camden, NJ as well as the Southwest suburbs of Bryn Mawr, Ardmore, Penn Wynne, and Bala Cynwyd. Hot spots were found in Southwest Philadelphia and in/near Media, PA.

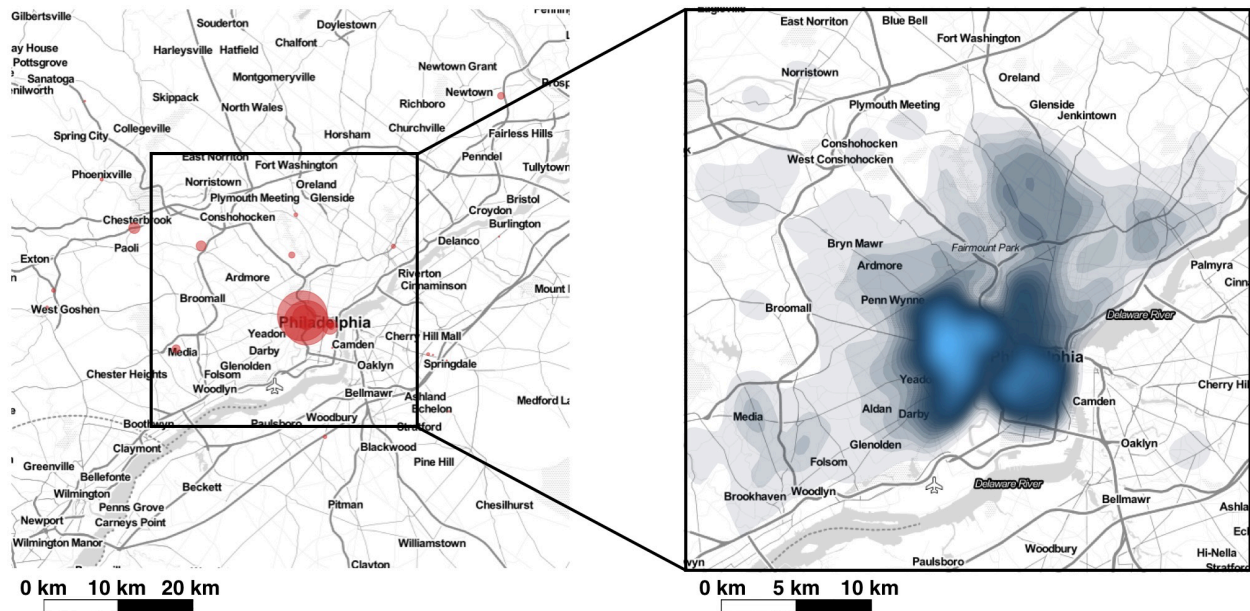


Figure 1. The study region is within the square outlined in black. Left: UPHS encounter sites are represented by red circles that are scaled by encounter volume. Right: density plot of the spatial distribution of patients in the study region; areas with highest patient density are shaded blue.

Discussion

Asthma prevalence and exacerbation disparities by race/ethnicity and socioeconomic status are a well-known problem in the US^{27, 28}. Factors contributing to them include non-modifiable factors such as differences in genetic ancestry³⁷⁻³⁹, and modifiable factors such as poor health, disease risk factors, medication non-adherence, health literacy and limited access to health care due to social, economic and environmental disadvantages^{30, 40, 41}. U.S. racial/ethnic minority groups disproportionately fall within low socioeconomic status categories, resulting in difficulty separating health factors that are heritable from those that are socioeconomic and environmental^{42, 43}. Gender is another documented risk factor for adult asthma, with women disproportionately affected compared to men²⁸. The characteristics of our study's asthma patients relative to all UPHS patients were consistent with these trends: *women* and *black* patients were significantly over-represented in the asthma sample.

The demographic factors we identified as associated with cases (i.e. asthma patients with exacerbations) vs. controls (i.e. asthma patients without exacerbations) were largely consistent with previous reports on exacerbation disparities: *women* were more likely to have exacerbations than *men*, *black* patients were more likely to have exacerbations than *white* patients, and patients with *Medicaid*—a crude indicator of low socio-economic status—were more likely to have exacerbations than those with *Private Insurance*. Our findings were also consistent with the growing body of literature

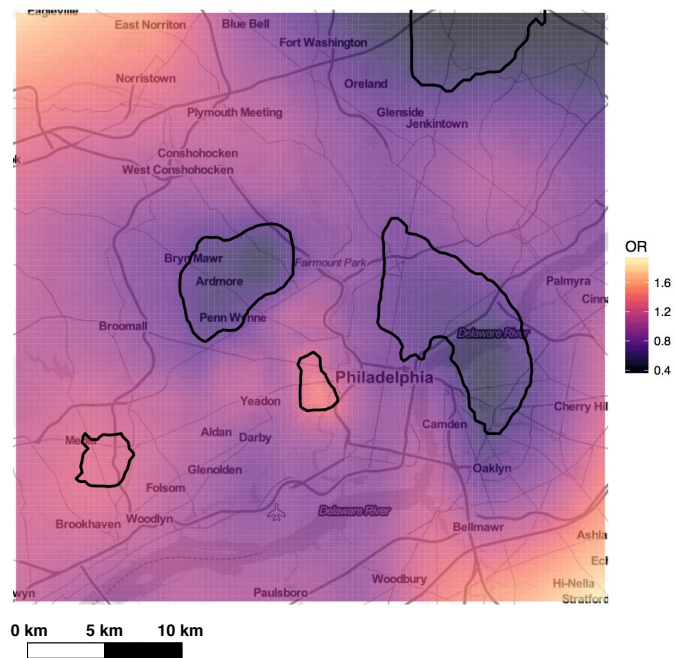


Figure 2. The spatial distribution of the risk of exacerbation, adjusted for gender, race, age, BMI, smoking status, and financial class, computed with an optimal span of 0.15. Significant hot spots and cold spots are indicated by black contour lines.

on the relationship between obesity and asthma⁴⁴, and the known association between smoking and increased risk of asthma⁴⁵. Patients with asthma exacerbations were more likely to be *Grades ≥ 3 obese* vs. all other BMI categories in the adjusted analysis, and to have passive or previous smoking history. Although BMI was missing in ~18% of subjects, the missing rate was comparable between cases and controls, suggesting our results would hold if this data were available. In future work, we will explore imputation approaches for height and use of weight measures alone to reduce potential bias due to missing BMI measures. Interestingly, current smoking status was not significantly associated with exacerbations in adjusted analyses. Whether this association represents bias in reporting or extracting of smoking status for our subjects cannot be determined from our existing data, but future attempts to re-contact UPHS subjects for other projects may provide better estimates of accuracy of codified smoking status in the EHR relative to that provided via interviews and more thorough questionnaires. The strongest factor associated with exacerbations was age: older adult patients (in categories >33 years of age) were 1.88-2.23 times more likely than younger patients (aged 18-33 years) to have exacerbations. The significance of this relationship is likely confounded by other comorbidities, as older adults with asthma are more likely to have other diseases, including chronic obstructive pulmonary disease⁴⁶. In ongoing work, we are performing detailed comorbidity analyses, which might clarify our observed relationship between age and asthma exacerbations.

Beyond measuring demographic associations, we showed that EHR-derived data was enhanced by inclusion of patient-specific geospatial information, as we were able to identify geospatial patterns of asthma exacerbations in the greater Philadelphia area that were independent of known exacerbation predictors. Because asthma rates in Philadelphia have been estimated to be over twice as high as the national average, it is likely that local environmental factors strongly influence disease prevalence and severity⁴⁷. Our results indicate that exacerbation risk varies significantly across the greater Philadelphia area, with predicted odds ratios for the study region ranging from 0.4 to 1.8 and the presence of significant hot spots and cold spots. Whether these differences truly reflect environmental or exposure variability vs. biases in our data remains a question. Independent replication studies using other patient populations in Philadelphia would strengthen our ability to make conclusions based on our current spatial results. Nonetheless, our results raise some interesting hypotheses. For example, an oil refinery is located approximately 2 km south of the hot spot identified in Southwest Philadelphia, where asthma patients were up to 1.6 times more likely to have had an exacerbation compared to the total study region. Thus, it is possible that particulate matter pollution near the refinery is greater than in regions further from it, resulting in increased asthma exacerbations in the nearby area. The two largest cold spots identified corresponded to one of the wealthiest regions in the Philadelphia area (the southwest suburbs of Bryn Mawr, Ardmore, Penn Wynne, and Bala Cynwyd) and one of the poorest (North Philadelphia and neighboring Camden, NJ). In future studies, we will use census data to adjust for socioeconomic status by census block and better understand what factors may be contributing to decreased exacerbations in these areas.

Our data is subject to various limitations, beyond those inherent in EHR-derived data (e.g., missingness and entry-error of measures, phenotyping errors due to use of crude classification schemes such as ICD-9 codes for disease diagnosis and designation of primary ICD-9 codes and oral corticosteroid prescription for exacerbation definition, and ascertainment bias such as including only subjects who more frequently use the healthcare system and are, thus, more likely to be unhealthy). The EHR-derived data we obtained from PDS only records the most current address for each patient, which we used as a proxy for patient's residence over the three- to four-year study period. We limited our analysis to patients with at least three years of UPHS encounters to (1) increase the likelihood that all asthma patients received care within UPHS, and thus, decrease the probability that controls lacked a record of exacerbations due to their seeking care outside of the UPHS system, and (2) to decrease the probability that patients moved out of the Philadelphia area during the study period. Nonetheless, the EHR-derived address data is limited in that we are unable to capture whether patients moved within Philadelphia or actually resided in a location other than the reported address during the study period. In ongoing work, we are incorporating measures related to type of patient housing, neighborhood migration estimates, and density of healthcare providers by location, as well as performing analyses of exacerbation severity to further identify and explicitly address these potential biases. Although capturing more thorough patient information via wearable sensor devices and other health information technologies will be useful to overcome some of the biases our data is subject to¹⁷, the integration of geospatial information that is readily captured via patient geocodes offers a low-cost and convenient alternative to more fully characterize EHR-derived patient information.

Conclusion

We have demonstrated the utility of integrating geospatial information with EHR data to uncover spatial variability of health events. Even after adjustment for demographic factors that are known asthma risk factors, the geospatial

distribution of asthma exacerbations varied significantly across Philadelphia and surrounding regions. Further study of the regions with increased/decreased exacerbations offers promise to understand environmental factors that influence asthma and tailor interventions to effectively decrease exacerbations.

Acknowledgements

We would like to thank Susan Raysor and Yuliya Borovskiy, M.S. from the University of Pennsylvania Penn Data Store for extracting the data used for this project. This work was partially supported by National Institutes of Health (NIH) R00 HL105663 and P30 ES013508.

Table 2. Demographic characteristics of asthma patients with exacerbations (cases) vs. those without exacerbations (controls). N (%) are shown. Table 3 reports statistical differences of these variable distributions between cases and controls.

	Cases (N=2,773)	Controls (N = 24,831)
Gender		
Male	715 (25.8)	6,828 (27.5)
Female	2,058 (74.2)	18,003 (72.5)
Race/Ethnicity		
White	1,359 (49.0)	13,488 (54.3)
Black	1,414 (51.0)	11,343 (45.7)
Age (years)		
18-33	431 (15.5)	6,622 (26.7)
34-47	714 (25.7)	6,143 (24.7)
48-59	859 (31.0)	6,119 (24.6)
60-80	769 (27.7)	5,947 (23.9)
Body Mass Index (BMI)		
Not overweight or obese	620 (22.4)	4,987 (20.1)
Overweight	595 (21.5)	5,520 (22.2)
Grade 1 obese	501 (18.1)	4,237 (17.1)
Grade 2 obese	341 (12.3)	2,722 (11.0)
Grades \geq 3 obese	428 (15.4)	2,862 (11.5)
Missing	450 (16.2)	4,503 (18.1)
Smoking Status		
Never	1,340 (48.3)	13,717 (55.2)
Passive	19 (0.7)	145 (0.6)
Quit	923 (33.3)	6,949 (28.0)
Yes	930 (14.1)	3,275 (13.2)
Missing	101 (4.4)	745 (3.0)
Financial Class		
Private Insurance	1,516 (54.7)	15,817 (63.7)
Medicare	741 (26.7)	5,343 (21.5)
Medicaid	497 (17.9)	4,085 (16.5)
Other	19 (0.7)	216 (0.9)

Table 3. Factors associated with asthma exacerbations. Crude and adjusted odds ratios (ORs) were derived from logistic regression models with exacerbation as the outcome. Shown are ORs and 95% confidence intervals (CIs). *p <0.05, **p<0.001

	Crude ORs (N = 27,604)	Adjusted ORs (N = 21,925)
Gender		
Male	Reference	Reference
Female	1.09 (0.99, 1.19)	1.13 (1.02, 1.26)*
Race/Ethnicity		
White	Reference	Reference
Black	1.24 (1.14, 1.34)**	1.41 (1.27, 1.56)**
Age (years)		
18-33	Reference	Reference
34-47	1.79 (1.58, 2.02)**	1.89 (1.65, 2.18)**
48-60	2.16 (1.91, 2.44)**	2.23 (1.94, 2.56)**
60-80	1.99 (1.76, 2.25)**	1.88 (1.59, 2.22)**
Body Mass Index (BMI)		
Not overweight or obese	Reference	Reference
Overweight	1.17 (1.03, 1.33)*	1.07 (0.94, 1.22)
Grade 1 obese	1.29 (1.13, 1.47)**	1.10 (0.96, 1.26)
Grade 2 obese	1.36 (1.18, 1.58)**	1.12 (0.96, 1.31)
Grades ≥ 3 obese	1.63 (1.42, 1.87)**	1.27 (1.09, 1.47)*
Smoking Status		
Never	Reference	Reference
Passive	1.34 (0.80, 2.11)	1.70 (1.00, 2.71)*
Quit	1.36 (1.24, 1.49)**	1.20 (1.09, 1.33)**
Yes	1.22 (1.08, 1.37)*	1.05 (0.92, 1.20)
Financial Class		
Private Insurance	Reference	Reference
Medicare	1.39 (1.27, 1.52)**	1.16 (1.02, 1.32)*
Medicaid	1.22 (1.09, 1.36)**	1.18 (1.04, 1.34)*
Other	0.88 (0.53, 1.37)	0.91 (0.50, 1.52)

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, Klanderma B, Kohane IS, Weiss ST. Assessing the reproducibility of asthma genome-wide association studies in a general clinical population. *The Journal of allergy and clinical immunology.* 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.
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.
6. Himes BE, Dai Y, Kohane IS, Weiss ST, Ramoni MF. Prediction of chronic obstructive pulmonary disease (COPD) in asthma patients using electronic medical records. *Journal of the American Medical Informatics Association : JAMIA.* 2009;16(3):371-9.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. Warner JL, Rioth MJ, Mandl KD, Mandel JC, Kreda DA, Kohane IS, et al. SMART precision cancer medicine: a FHIR-based app to provide genomic information at the point of care. *J Am Med Inform Assoc.* 2016;23(4):701-10.
13. Smith ME, Aufox S. Biobanking: The Melding of Research with Clinical Care. *Curr Genet Med Rep.* 2013;1(2):122-8.
14. Overby CL, Heale B, Aronson S, Cherry JM, Dwight S, Milosavljevic A, et al. Providing Access to Genomic Variant Knowledge in a Healthcare Setting: A Vision for the ClinGen Electronic Health Records Workgroup. *Clin Pharmacol Ther.* 2016;99(2):157-60.
15. Mandl KD, Kohane IS. Time for a Patient-Driven Health Information Economy? *N Engl J Med.* 2016;374(3):205-8.
16. Eggleston EM, Weitzman ER. Innovative uses of electronic health records and social media for public health surveillance. *Current diabetes reports.* 2014;14(3):468.
17. Himes BE, Weitzman ER. Innovations in health information technologies for chronic pulmonary diseases. *Respir Res.* 2016;17(1):38.
18. 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.
19. 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 Qual Saf.* 2016;25(6):414-23.
20. Dorsey ER, Yvonne Chan YF, McConnell MV, Shaw SY, Trister AD, Friend SH. The Use of Smartphones for Health Research. *Acad Med.* 2016.
21. Rasmussen SG, Ogburn EL, McCormack M, Casey JA, Bandeen-Roche K, Mercer DG, et al. Association Between Unconventional Natural Gas Development in the Marcellus Shale and Asthma Exacerbations. *JAMA internal medicine.* 2016.

22. Gabert R, Thomson B, Gakidou E, Roth G. Identifying High-Risk Neighborhoods Using Electronic Medical Records: A Population-Based Approach for Targeting Diabetes Prevention and Treatment Interventions. *PLoS One*. 2016;11(7):e0159227.
23. Akinbami L, Moorman J, Bailey C, Zahran H, King M, Johnson C, et al. Trends in asthma prevalence, health care use, and mortality in the United States, 2001–2010. Hyattsville, MD: National Center for Health Statistics, 2012.
24. Moorman JE, Rudd RA, Johnson CA, King M, Minor P, Bailey C, et al. National surveillance for asthma--United States, 1980-2004. *MMWR Surveill Summ*. 2007;56(8):1-54.
25. Barnett SB, Nurmagambetov TA. Costs of asthma in the United States: 2002-2007. *J Allergy Clin Immunol*. 2011;127(1):145-52.
26. National Asthma Education Program. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. National Institutes of Health, Bethesda: US Department of Health and Human Services, 2007.
27. Moorman JE, Akinbami LJ, Bailey CM, Zahran HS, King ME, Johnson CA, et al. National Surveillance of Asthma: United States, 2001–2010. . National Center for Health Statistics., 2012.
28. Schiller JS, Lucas JW, Ward BW, Peregoy JA. Summary health statistics for U.S. adults: National Health Interview Survey, 2010. *Vital and health statistics Series 10, Data from the National Health Survey*. 2012(252):1-207.
29. Forno E, Celedon JC. Health disparities in asthma. *Am J Respir Crit Care Med*. 2012;185(10):1033-5.
30. Apter AJ, Wan F, Reisine S, Bender B, Rand C, Bogen DK, et al. The association of health literacy with adherence and outcomes in moderate-severe asthma. *J Allergy Clin Immunol*. 2013;132(2):321-7.
31. Keddem S, Barg FK, Glanz K, Jackson T, Green S, George M. Mapping the urban asthma experience: Using qualitative GIS to understand contextual factors affecting asthma control. *Social science & medicine*. 2015;140:9-17.
32. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2016.
33. 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.
34. 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.
35. L B, SM B, RL B, VM V. ‘MapGAM’: Mapping Smoothed Effects Estimates from Individual-Level Data. R package, version 10; 2016. 2016.
36. D K, H W. ggmap: Spatial Visualization with ggplot2. *The R Journal*. 2013;5(1):144-61.
37. Flores C, Ma S-F, Pino-Yanes M, Wade MS, Pérez-Méndez L, Kittles RA, et al. African ancestry is associated with asthma risk in African Americans. *PloS One*. 2012;7:e26807.
38. Lester LA, Rich SS, Blumenthal MN, Togias A, Murphy S, Malveaux F, et al. Ethnic differences in asthma and associated phenotypes: collaborative study on the genetics of asthma. *The Journal of Allergy and Clinical Immunology*. 2001;108:357-62.
39. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nature genetics*. 2011;43(9):887-92.
40. Desai M, Oppenheimer JJ. Medication adherence in the asthmatic child and adolescent. *Curr Allergy Asthma Rep*. 2011;11(6):454-64.
41. Celedon JC, Roman J, Schraufnagel DE, Thomas A, Samet J. Respiratory health equality in the United States. The American thoracic society perspective. *Annals of the American Thoracic Society*. 2014;11(4):473-9.
42. Fiscella K, Franks P, Gold MR, Clancy CM. Inequality in quality: addressing socioeconomic, racial, and ethnic disparities in health care. *JAMA*. 2000;283(19):2579-84.
43. House J, Williams D. Understanding and reducing socioeconomic and racial/ethnic disparities in health. In: Syme BDSSL, editor. *Promoting health: Intervention strategies from social and behavioral research*. Washington, DC: National Academy Press; 2000. p. 81-125.
44. Kim S-H, Sutherland ER, Gelfand EW. Is there a link between obesity and asthma? *Allergy, Asthma & Immunology Research*. 2014;6:189-95.
45. Perret JL, Dharmage SC, Matheson MC, Johns DP, Gurrin LC, Burgess JA, et al. The interplay between the effects of lifetime asthma, smoking, and atopy on fixed airflow obstruction in middle age. *American Journal of Respiratory and Critical Care Medicine*. 2013;187:42-8.

46. Cosentino J, Zhao H, Hardin M, Hersh CP, Crapo J, Kim V, et al. Analysis of Asthma-Chronic Obstructive Pulmonary Disease Overlap Syndrome Defined on the Basis of Bronchodilator Response and Degree of Emphysema. *Annals of the American Thoracic Society*. 2016;13(9):1483-9.
47. Bryant-Stephens T, West C, Dirl C, Banks T, Briggs V, Rosenthal M. Asthma prevalence in Philadelphia: description of two community-based methodologies siologies to assess asthma prevalence in an inner-city population. *J Asthma*. 2012;49(6):581-5.