

Analysis of Spatial Trends in Smoking Status Among Patients with Obstructive Airway Diseases Highlight Potential for Targeted Smoking Cessation Interventions

Sherrie Xie, BS, Rebecca A. Hubbard, PhD, and Blanca E. Himes, PhD

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

Abstract

Because chronic obstructive airway diseases like asthma, chronic obstructive pulmonary disease (COPD) and asthma-COPD overlap (ACO) increase individual susceptibility to the harmful effects of cigarettes, smoking cessation programs could strengthen their public health impact by targeting smokers with these conditions. We performed spatial analyses on data derived from the Electronic Health Records (EHRs) of 25,119 asthma, 3,323 COPD, and 3,620 ACO patients and a community-based health survey of 18,740 residents to identify regions in the Greater Philadelphia Area with a high density of current smokers among patients with obstructive airway diseases and the general population. We identified areas in North and West Philadelphia with high prevalence of current smokers across all patient groups and community members that should be prioritized in smoking cessation initiatives. Neighborhood deprivation, which was linked to patient data using residential geocodes, was associated with greater smoking prevalence in these regions.

Introduction

Cigarette smoking is a leading cause of morbidity and premature death in the U.S. and has been a target of many federal and local policy initiatives.¹⁻⁴ Asthma and chronic obstructive pulmonary disease (COPD) are common obstructive airway diseases that increase individual susceptibility to the harmful effects of cigarettes. Among individuals with asthma, smoking increases symptom severity, accelerates lung function deterioration, and decreases therapeutic response to corticosteroids.⁵⁻⁸ Most COPD-related deaths are caused by smoking, and the mortality risk for smokers with COPD is twice that of smokers with other conditions.^{1,9} People who have both asthma and COPD, referred to as asthma-COPD overlap (ACO), have more comorbidities and exacerbations than those with asthma or COPD alone.¹⁰⁻¹² While ACO had been under-studied partly because asthma and COPD trials commonly excluded patients with both conditions, the recognition that people with ACO have worse outcomes has resulted in it being a topic of recent interest.^{13,14}

While no study has explicitly considered the effects of smoking cessation on people with ACO, the beneficial effects of smoking cessation on those with asthma or COPD are well documented. Smokers with asthma who quit smoking experience decreased symptoms and improved lung function,^{7,15} while smokers with COPD who quit smoking experience slower lung function decline and improved survival^{16,17} and smoking cessation is considered the most effective treatment for COPD.¹⁷ Given the high prevalence of obstructive airway diseases in the U.S.^{18,19} and the substantial benefits that smokers with these conditions stand to gain by quitting, smoking cessation initiatives can increase their public impact by reaching these vulnerable individuals.

Here, we leveraged data derived from Electronic Health Records (EHRs) and a community-based health survey to understand how smoking cessation campaigns could geographically target patients with asthma, COPD, and ACO. Our study aimed to (1) determine how smoking status among patients with asthma, COPD, or ACO compares to that of a representative community sample, (2) identify factors associated with smoking status among patients vs. community members, and (3) identify geographical areas with high densities of current smokers that should be prioritized for smoking cessation interventions.

Methods

Study population. Our study population consisted of (1) adults with asthma, COPD, or ACO encountered at the University of Pennsylvania Health System (UPHS) in 2012-2016, and (2) residents of Southeastern Pennsylvania who were part of the Southeastern Pennsylvania Household Health Survey (SEPA-HHS, henceforth referred to as *HHS*). UPHS operates several large hospitals, specialty medical centers, and satellite clinics in the Greater Philadelphia Area, including five counties in Southeastern Pennsylvania (Philadelphia, Bucks, Chester, Delaware, and Montgomery counties, henceforth referred to as *SEPA*). We obtained de-identified, patient-level data from Penn Data Store, the

UPHS clinical data warehouse, which we have used previously to study spatial trends in asthma exacerbations.^{20, 21} Variables extracted included geocoded patient residential addresses, codified demographic information (i.e., gender, birth year, race/ethnicity, height and weight measures), as well as codified smoking history (i.e., whether a patient is a current smoker, has quit smoking, or never smoked). Medications prescribed during encounters were obtained from codified entries as well as NLP-extracted values from clinical notes. Patient insurance class was obtained from an encounter-level codified billing field that was re-categorized into four groups: *Private Insurance*, *Medicare*, *Medicaid*, and *Other/no insurance*. Patients were assigned the most frequently occurring insurance class among all encounters. Patient age was defined as age at first encounter.

For the current study, patients were included if they had: (1) at least one ICD-9 or ICD-10 code for asthma (i.e. ICD-9 493* or ICD-10 J45*) or COPD (i.e. ICD-9 491*, 492*, or 496*; ICD-10 J41*, J43*, or J44*), (2) non-missing smoking history, (3) non-missing residential geocode and (4) residence within SEPA. Patients were classified as having asthma if they had at least one ICD code for asthma, prescription for a short-acting β_2 -agonist, and did not meet the criteria for COPD. Patients were classified as having COPD if they had at least one ICD code for COPD, prescription for a short-acting β_2 -agonist or short- or long-acting muscarinic antagonist, and did not meet the criteria for asthma. Patients were classified as having ACO if they met the diagnosis and medication criteria for both asthma and COPD.

The *HHS*, a community-based survey of SEPA residents conducted by the Public Health Management Corporation (PHMC) in 2012 and 2015, was a random-digit-dialing mobile and home telephone survey that aimed to capture the health status and behaviors of the local population and included survey items on individual-level demographics (e.g., age, gender, race, BMI, health insurance status) and smoking history. PHMC recorded residential street addresses or nearby cross-streets of survey respondents, which were used to match respondents to residential census tracts.

Neighborhood disadvantage metrics. While little information on socioeconomic status (SES) is collected in EHRs, residential geocodes can be used to link patient data to high-resolution information on neighborhood environments.²⁰ We sourced data on neighborhood SES from the Neighborhood Atlas, an online tool that provides a standardized score for neighborhood disadvantage, termed the area deprivation index (ADI), for all block groups in the U.S.^{22, 23} ADI is a factor-based score that captures multiple dimensions of SES (poverty, education, employment, housing quality) by combining several measures from the 2013 American Community Survey Five Year Estimates.²² It has been validated for detecting socioeconomic gradients in a number of health outcomes, including cancer mortality²⁴ and hospital readmission risk.^{25, 26} Block group ADI is reported as a national percentile ranking (taking integer values 1-100) indicating each block group's economic deprivation relative to all others in the U.S., with higher values representing greater disadvantage. The geographic distribution of ADI in the SEPA region is shown in Figure 1. We determined neighborhood disadvantage for each patient by using residential geocodes to link their EHR-derived data to their block group ADI, which was then transformed to an ordinal variable with four levels (1-15, 16-50, 51-84, and 85-100).

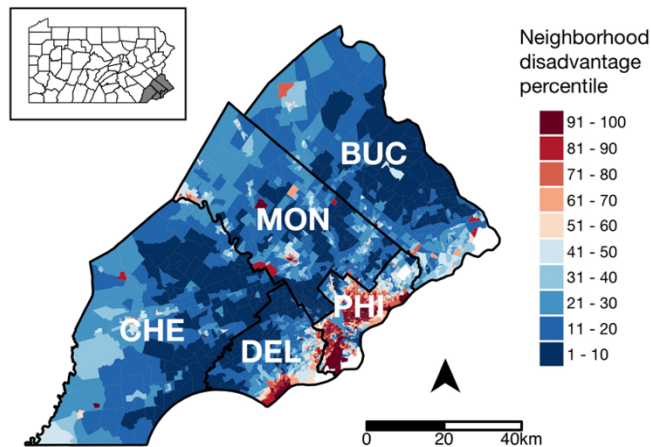


Figure 1. SEPA block groups shaded by neighborhood disadvantage as a national percentile ranking. Red colors denote more and blue colors denote less disadvantage than the national median. County boundaries are represented by black lines, and county names are abbreviated as follows: PHI = Philadelphia, BUC = Bucks, CHE = Chester, DEL = Delaware, MON = Montgomery. The inset map in the upper left-hand corner shows Pennsylvania counties with the SEPA region shaded.

Block group ADI is reported as a national percentile ranking (taking integer values 1-100) indicating each block group's economic deprivation relative to all others in the U.S., with higher values representing greater disadvantage. The geographic distribution of ADI in the SEPA region is shown in Figure 1. We determined neighborhood disadvantage for each patient by using residential geocodes to link their EHR-derived data to their block group ADI, which was then transformed to an ordinal variable with four levels (1-15, 16-50, 51-84, and 85-100).

Statistical analysis. Three smoking-related rates were computed for each patient group (asthma, COPD, ACO) and HHS: current smoker prevalence, ever smoker prevalence, and smoking cessation rate. *Current smoker prevalence* was defined as the proportion of subjects who currently smoke, and *ever smoker prevalence* was defined as the proportion of subjects with a positive smoking history. *Smoking cessation rate* was defined as the proportion of subjects who quit smoking among those with a positive smoking history. Note that while the total population served as the denominator

for current and ever smoker prevalence, only those with a positive smoking history formed the denominator for smoking cessation rate.

Rates for community members were calculated using survey weights provided by HHS to correct for nonresponse bias, allowing estimates to more accurately reflect the SEPA population. Patient rates were standardized by gender, race, sex, and county residence using weighted HHS as the standard population via the direct method.²⁷ Briefly, patients within each disease group were cross-tabulated by gender, race/ethnicity (“Non-Hispanic black,” “Non-Hispanic white,” “other”), age (18-34, 35-54, 56-74, >75), and county residence, for a maximum of 120 cells, and cells were collapsed in some patient groups to avoid cells with count 0. Subsequently, standardized rates for each patient group were calculated as the weighted average of cell-specific rates, where weights represented the relative frequency of each demographic combination in the weighted HHS sample.

We produced choropleth maps for rates aggregated at the census tract level to visualize the geographic distribution of smoking behaviors across the three patient groups and HHS. To correct for instability of rates estimated from small samples (median census tract $n = 7$, IQR = 3-15), we applied an empirical Bayes estimation method for small area prevalence of non-rare conditions, where crude rates were corrected towards local, county-wide means to a degree inversely proportional to the sample size.²⁸ Regional differences in patient smoking behavior compared to that of local community members were visualized by subtracting HHS rates from those of each patient group for all census tracts.

We performed two-tailed Wald tests to test the null hypothesis that the difference between each smoking-related rate for each patient group vs. HHS equaled 0. To determine whether demographic factors (i.e., gender, race, age, BMI category, insurance billing class, county residence, and neighborhood disadvantage) were associated with smoking cessation for each group, we used multivariable logistic regression analyses. A separate model was constructed for each group (asthma, COPD, ACO, and HHS). Statistical analyses were conducted with a type I error rate of 0.05 and performed in R version 3.5.1.²⁹

Geospatial analysis. We identified significant smoking cessation cold spots (i.e., regions where the relative density of current smokers was high) for each patient group and HHS using methods described previously.³⁰ Briefly, we used generalized additive models (GAM) to estimate the log odds of smoking cessation among subjects with a lifetime smoking history as a function of their location of residence before and after adjustment for individual-level covariates. Because residential geocodes were not provided by HHS, the residential location of each HHS respondent was approximated with the mean center of population for their census tracts of residence.³¹ The GAM included a bivariate smooth term $S(x,y)$ smoothed using a loess function, a locally-weighted regression smoother which adapts to local variation in patient/respondent density. The amount of neighboring points utilized by the loess was determined by a span size optimized by minimizing the Akaike Information Criterion.

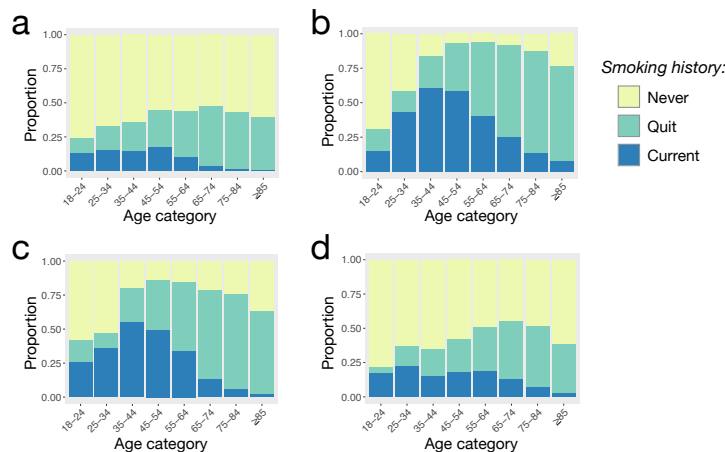


Figure 2. Smoking history by age category among (a) asthma, (b) COPD, and (c) ACO patients and (d) HHS respondents. The proportion of “ever” smokers is the combined height of the proportion of “current” smokers and those who have “quit” smoking.

We used GAM to determine the log odds of smoking cessation on a grid of 10,000 points arranged across SEPA, which was converted to an odds ratio by using the global odds of smoking cessation within each patient group as reference. Two GAM models were fit for each patient group and for HHS respondents: one with and one without individual-level covariate adjustment, for a total of eight models. Variables selected for adjustment were the significant independent predictors of smoking cessation as determined by multivariable logistic regression: gender, race, age at first encounter, BMI category, insurance billing class and neighborhood disadvantage. County residence was not included in adjusted models to prevent over-fitting. Global tests were performed to test the null hypothesis that smoking cessation rates did not vary by geographic location by running 999

Table 1. Characteristics of 2012-2016 UPHS patients with asthma, chronic obstructive pulmonary disease (COPD) and asthma-COPD overlap (ACO) and respondents surveyed by the 2012 & 2015 Southeastern Pennsylvania Household Health Survey (HHS). Shown are median (IQR) for age and counts (%) for other variables.

	Asthma	COPD	ACO	HHS
Total	25,119	8,323	3,620	18,740
Age (IQR)	40 (28, 54)	67 (59, 76)	61 (53, 71)	53 (43, 64)
Female	18,041 (71.8)	4,181 (50.2)	2,380 (65.7)	11,638 (62.1)
<i>Race</i>				
Non-Hispanic White	11,654 (46.4)	5,073 (61.0)	1,446 (39.9)	13,095 (69.9)
Non-Hispanic Black	10,178 (40.5)	2,645 (31.8)	1,924 (53.1)	3,900 (20.8)
Hispanic	1,201 (4.8)	117 (1.4)	101 (2.8)	936 (5.0)
Other	2,086 (8.3)	488 (5.9)	149 (4.1)	809 (4.3)
<i>County residence</i>				
Philadelphia	13,721 (54.6)	4,427 (53.2)	2,425 (67.0)	7,260 (38.7)
Bucks	1,646 (6.6)	685 (8.2)	179 (4.9)	2,743 (14.6)
Chester	3,700 (14.7)	1,209 (14.5)	362 (10.0)	2,758 (14.7)
Delaware	3,367 (13.4)	1,061 (12.7)	363 (10.0)	2,821 (15.1)
Montgomery	2,685 (10.7)	941 (11.3)	291 (8.0)	3,158 (16.9)
<i>Smoking status</i>				
Never	15,773 (62.8)	809 (8.1)	707 (17.9)	10,432 (55.7)
Quit	6,007 (23.9)	5,007 (50.2)	1,885 (47.7)	5,252 (28.0)
Current	15,773 (13.3)	809 (25.1)	707 (26.0)	3,056 (16.3)

permutations of the assignment of cases (those who quit smoking) and controls (those who did not) over all residential locations. Given a significant global test, a local test was performed to identify areas of increased or decreased odds of smoking cessation. The model fit 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 and lower 0.5% of the permutation distributions were defined as “hot spots” and “cold spots”, respectively. GAM analyses were performed in R using the *MapGAM* package.³²

Results

Characteristics of patient groups and community members. A total of 71,029 patients encountered by UPHS between 2012 and 2016 met the criteria for asthma and/or COPD. Of these, 16,530 (23%) were excluded for missing residential geocodes, 12,655 (18%) were excluded for residence outside SEPA, and 4,782 (7%) were excluded for missing smoking history, leaving 37,062 patients for geospatial analysis and the calculation of smoking-related rates. Of

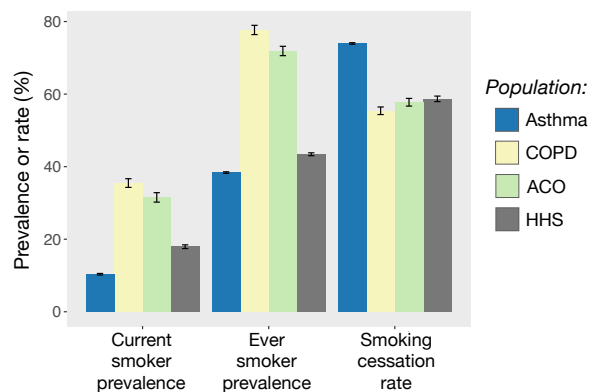


Figure 3. Standardized current smoker prevalence, ever smoker prevalence, and smoking cessation rate colored by patient group or HHS. Error bars represent 95% confidence intervals.

included patients, 25,119 (68%) were classified as having asthma, 8,323 (22%) were classified as having COPD, and 3,620 (10%) were classified as having ACO. A total 20,066 HHS respondents were surveyed in 2012 and 2015. Of these, 102 (0.5%) were excluded for missing smoking history, 555 (3%) were excluded for missing residential census tract information, and 669 (3%) were excluded for missing demographic information (age or race/ethnicity), leaving 18,740 respondents for estimation of local population rates.

Demographic characteristics of UPHS patients and HHS respondents are provided in Table 1. The distribution of characteristics vary considerably between patients and HHS respondents, accentuating the importance of adjusting group-specific rates by demographic factors so that they can be compared appropriately. The median

ages of COPD and ACO patients were older than that of HHS respondents, while the median age of asthma patients was younger. Asthma and ACO patients were disproportionately female (asthma 71.8%, ACO 65.7% versus COPD 50.2%) and Non-Hispanic black (asthma 40.5%, ACO 53.1% versus COPD 31.8%), consistent with findings from previous studies of asthma and COPD in UPHS and other patient populations.^{10, 12, 21} All groups were enriched for subjects that lived in Philadelphia relative to other counties, though this enrichment was more pronounced in patient groups than HHS.

Differences in smoker characteristics between patient groups and community members. Smoking history varies markedly by age among asthma, COPD and ACO patients, and HHS respondents (Figure 2). Although COPD and ACO patients have higher proportions of current and ever smokers for all age groups, the distributions of ever smokers show a roughly unimodal distribution across all groups, with highest rates in middle to older age (45-74). The distribution of current smokers is also roughly unimodal across all groups, with peaks that occur at relatively younger ages (25-54) compared to ever smoker distributions.

Standardized rates of smoking and smoking cessation vary by patient group (Figure 3). Current smoker prevalence was higher among COPD (35.5 %) and ACO patients (31.5%) compared to HHS respondents (17.9%), ($p < 0.001$ for both comparisons). However, current smoker prevalence was lower among asthma patients (10.3%, $p < 0.001$). Similarly, ever smoker prevalence was higher in COPD and ACO patients compared to HHS respondents (COPD = 77.7% and ACO = 71.9% vs. HHS = 43.4%, $p < 0.001$ for both comparisons), while ever smoker prevalence was

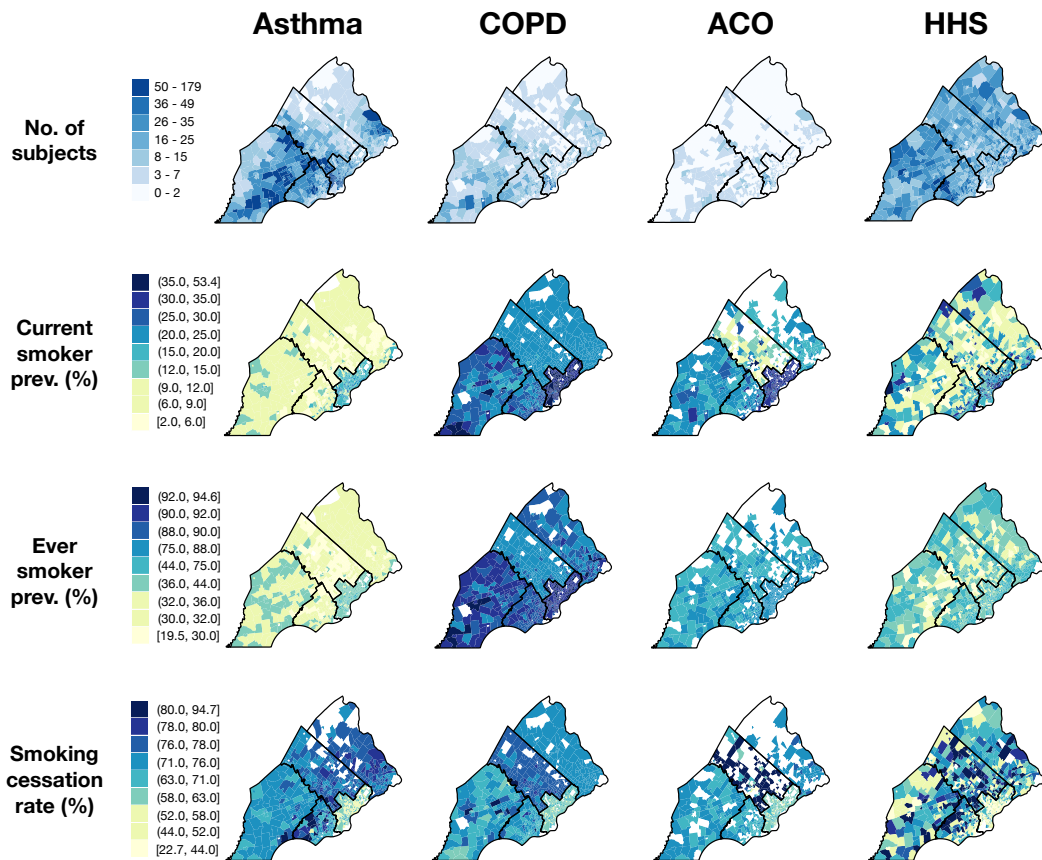


Figure 4. Choropleth maps of SEPA census tracts shaded according to subject counts (top row) or smoking-related rates (bottom three rows). Rates were corrected for small sample variability via an empirical Bayes method in which crude estimates were corrected toward county means to a degree inversely proportional to tract-level sample size. Tracts with no data to calculate rates are colored white, and county boundaries are represented by black lines.

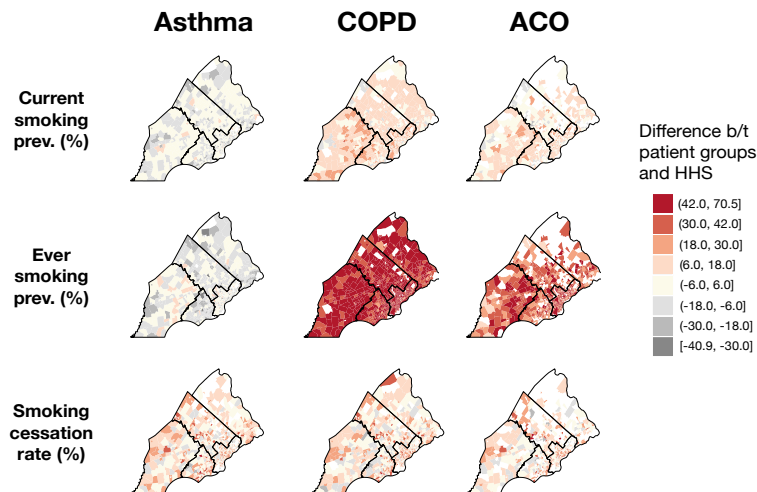


Figure 5. Choropleth maps of SEPA census tracts shaded according to rate differences between patient groups and HHS, with red colors denoting tracts in which rates estimated from patient data are *greater* than rates estimated from HHS and grey colors denoting tracts in which rates estimated from patient data are lower. Tracts with no data are colored white, and county boundaries are represented by black lines.

smoking rates appeared to be higher in Chester and Delaware compared to Bucks and Montgomery, especially among COPD patients, though this trend was less striking among HHS respondents. As a result, positive rate differences were evident in Chester and Delaware counties, where smoking rates tended to be higher in patient groups compared to HHS (Figure 5), though we did not test for statistical significance. The higher smoking rates among COPD and ACO groups compared to HHS extended across all SEPA census tracts, and this *difference* in rates does not exhibit a discernible spatial trend (Figure 5).

Neighborhood disadvantage measures. Block group-level neighborhood disadvantage measures are shown in Figure 1 for the five SEPA counties studied. Philadelphia county encompasses many highly disadvantaged neighborhoods, with over one-fifth of its block groups ranking among the nation's most disadvantaged (91-100th percentile). The surrounding suburbs (Bucks, Chester, Montgomery, Delaware counties) are markedly more affluent, with large regions that rank among the nation's bottom 10th percentile in neighborhood disadvantage.

Factors associated with smoking cessation. Multivariable analyses found that older age and overweight/obese BMI were consistent independent predictors of smoking cessation across patient groups and HHS (Table 2). On the other hand, Medicaid insurance (a crude indicator of low individual-level SES) and neighborhood disadvantage (an indicator of low neighborhood-level SES) were consistent independent predictors of subjects not quitting smoking.

Spatial trends in smoking cessation rates. Smoking cessation rates were heterogeneous in SEPA across all groups (Figure 6). The global test statistic for spatial heterogeneity was significant for all 8 GAMs ($p < 0.01$), indicating there was significant spatial heterogeneity in smoking cessation rates, and these trends were not fully explained by the covariates considered. The cold spots identified by the crude models (top row of Figure 6) showed considerable overlap, particularly in North and West Philadelphia, suggesting that smoking cessation rates were low (i.e., relative proportion of current smokers was high) for all patient groups and community members in these regions. While spatial trends were still present in the adjusted models (bottom row of Figure 6), the cold spots in North and West Philadelphia were absent, suggesting that the decreased smoking cessation rates in these regions were explained by spatial variation of the covariates considered.

Discussion

A strong evidence base supports the effectiveness of community-based interventions for smoking cessation.^{33,34} Policy makers interested in maximizing the impact of smoking cessation programs determine how to target groups that are most susceptible to the harmful effects of cigarettes and identify regions with the greatest number of current smokers.

lower among asthma patients (38.1%, $p < 0.001$). Rates of smoking cessation were similar between ACO patients and HHS respondents (57.8% vs. 58.7%, $p = 0.47$). Smoking cessation rates were also similar between COPD patients and HHS respondents, though the difference reached marginal significance (55.4% vs. 58.7%, $p = 0.03$). In contrast, smoking cessation rates among asthma patients were significantly higher than that of HHS respondents (74.0% vs. 58.7%, $p < 0.001$).

Geographic distribution of chronic airway disease patients by smoking status. Current and ever smoker prevalence estimates for Philadelphia census tracts were generally higher than those of the surrounding counties, and this trend held across patient groups and HHS (Figure 4). In addition, smoking cessation rates were generally lower in Philadelphia census tracts than those of the surrounding counties across all groups. Among the suburban counties,

We leveraged data derived from EHRs and a community-based survey to understand the smoking characteristics of patients with asthma, COPD and ACO relative to those of local populations, and we determined where smokers in these groups live. We found smoking rates among ACO and COPD patients to be nearly twice as high as those estimated from community members, and smoking rates among asthma patients to be somewhat lower than local community estimates. In addition, while smoking cessation rates were not substantially different between COPD and ACO patients vs. community members, cessation rates were significantly higher among patients with asthma. The high prevalence of smoking among patients with COPD and ACO highlights the need for smoking cessation programs that target these patients. Although smoking rates were lower among patients with asthma, the number of smokers with asthma were substantial, accounting for over 90% of smokers in our patient population. Because asthma is highly prevalent in Philadelphia,³⁵ smokers with asthma should not escape the notice of local policy makers.

Table 2. Factors associated with smoking cessation among UPHS patients with a positive smoking history, stratified by patient group and HHS. Adjusted odds ratios (ORs) were derived from multivariable logistic regression models with smoking cessation as the outcome. Shown are ORs and 95% confidence intervals, with significant results in bold.

	Asthma (N = 9,066)	COPD (N = 7,967)	ACO (N = 2,825)	HHS (N = 8,185)
Sex				
Male	Reference	Reference	Reference	Reference
Female	1.52 (1.36, 1.69)	1.09 (0.98, 1.21)	0.85 (0.70, 1.02)	1.00 (0.90, 1.11)
Age				
18-24	Reference	Reference	Reference	Reference
25-34	1.35 (1.15, 1.59)	0.40 (0.04, 3.93)	0.68 (0.13, 4.00)	2.60 (1.69, 4.10)
35-44	1.41 (1.19, 1.67)	0.43 (0.05, 3.71)	0.99 (0.22, 5.23)	4.65 (3.08, 7.23)
45-54	1.70 (1.44, 2.00)	0.68 (0.08, 5.78)	1.58 (0.37, 8.16)	4.44 (2.96, 6.84)
55-64	3.27 (2.72, 3.95)	1.43 (0.17, 12.13)	3.04 (0.70, 15.62)	6.63 (4.42, 10.22)
65-74	12.7 (9.4, 17.5)	2.73 (0.32, 23.22)	9.17 (2.08, 47.86)	17.1 (11.1, 26.9)
75-84	46.2 (23.7, 103.9)	5.91 (0.69, 50.48)	23.9 (5.2, 128.9)	37.5 (23.7, 60.7)
≥ 85	109.4 (23.7, 1,915.6)	10.64 (1.22, 92.66)	97.6 (14.3, 988.0)	65.7 (35.3, 128.4)
Race				
Non-His White	Reference	Reference	Reference	Reference
Non-His Black	0.82 (0.71, 0.94)	0.98 (0.85, 1.13)	1.19 (0.93, 1.53)	0.68 (0.58, 0.79)
Hispanic	0.96 (1.21, 1.53)	1.18 (0.74, 1.89)	0.92 (0.54, 1.58)	1.03 (0.80, 1.32)
Other	1.15 (0.93, 1.43)	1.05 (0.83, 1.34)	1.23 (0.73, 2.10)	1.02 (0.78, 1.35)
BMI category				
Not overweight or obese	Reference	Reference	Reference	Reference
Overweight	1.28 (1.12, 1.47)	1.29 (1.13, 1.47)	1.30 (1.02, 1.66)	1.46 (1.30, 1.65)
Obese class 1	1.19 (1.03, 1.37)	1.46 (1.25, 1.70)	1.58 (1.21, 2.05)	1.72 (1.49, 1.98)
Obese class 2	1.60 (1.35, 1.89)	1.58 (1.29, 1.94)	2.23 (1.65, 3.02)	1.89 (1.53, 2.32)
Obese class 3	1.65 (1.40, 1.94)	1.98 (1.57, 2.50)	2.33 (1.73, 3.15)	2.45 (1.93, 3.11)
Insurance class				
Private	Reference	Reference	Reference	Reference
Medicare	0.51 (0.43, 0.60)	0.92 (0.79, 1.07)	0.74 (0.57, 0.96)	0.44 (0.38, 0.52)
Medicaid	0.38 (0.33, 0.42)	0.63 (0.52, 0.75)	0.47 (0.36, 0.61)	0.55 (0.47, 0.65)
Missing/no insurance	0.14 (0.02, 0.61)	<i>No data</i>	<i>No data</i>	0.48 (0.39, 0.58)
County residence				
Philadelphia	Reference	Reference	Reference	Reference
Bucks	1.06 (0.83, 1.36)	1.38 (1.10, 1.74)	1.43 (0.88, 2.35)	0.93 (0.78, 1.10)
Chester	0.94 (0.79, 1.12)	1.06 (0.89, 1.28)	1.28 (0.89, 1.85)	1.02 (0.85, 1.23)
Delaware	1.13 (0.96, 1.34)	1.24 (1.04, 1.48)	1.34 (0.96, 1.88)	0.88 (0.75, 1.03)
Montgomery	1.02 (0.83, 1.26)	1.50 (1.22, 1.85)	2.06 (1.35, 3.20)	1.12 (0.94, 1.33)
Neighborhood deprivation				
1-15 (least disadvantaged)	Reference	Reference	Reference	Reference
16-50	0.60 (0.52, 0.71)	0.84 (0.71, 0.98)	0.69 (0.49, 0.95)	0.57 (0.49, 0.66)
51-84	0.54 (0.44, 0.65)	0.75 (0.61, 0.91)	0.68 (0.47, 0.97)	0.40 (0.33, 0.49)
85-99 (most disadvantaged)	0.45 (0.37, 0.55)	0.63 (0.51, 0.79)	0.65 (0.44, 0.94)	0.33 (0.26, 0.41)

Factors associated with smoking cessation were similar across patient groups and HHS. Consistent with reports from the U.S. and other developed nations that the most deprived groups have the highest smoking rates and the lowest rates of quitting,³⁶⁻³⁸ those with a positive smoking history across all groups were less likely to quit if they had low individual-level SES (coarsely identified by *Medicaid* insurance status) or lived in disadvantaged neighborhoods. These results reaffirm the need to target smoking cessation interventions toward disadvantaged communities. Obesity also had a robust association with subjects quitting smoking, though this was likely an instance of reverse causation, as smoking suppresses appetite and individuals who quit smoking tend to gain weight.^{39, 40}

Our unadjusted geospatial models identified consistent cold spots for smoking cessation in North and West Philadelphia (Figure 6 top row), regions with high neighborhood disadvantage relative to SEPA and nationally (Figure 1). The loss of these cold spots in adjusted models confirmed the association of smokers' neighborhood socioeconomic environment with quitting smoking status. Additionally, our results are promising from a policy perspective because the consistency of these cold spots indicates that initiatives that target areas with low smoking cessation rates in the general population may also target similar high-needs areas for populations with specific diseases. However, overlapping hotspots in Bucks and Montgomery county remained in the adjusted models for asthma and COPD patients, and a partially overlapping hotspot remained in the adjusted model for HHS. These hotspots suggest that the higher rates of smoking cessation in these regions cannot be fully explained by their relative affluence or the spatial distribution of the other covariates considered. Investigating the smoking-related policies in these communities could uncover the drivers of high smoking cessation rates in these regions and provide insights on effective steps that can be pursued in regions with low cessation rates.

Our findings are subject to limitations, including those inherent to analyses of EHR-derived data. Phenotyping errors can result from the use of ICD codes for the classification of disease status, though the combination of diagnosis codes with additional EHR components, such as medication information (e.g. prescriptions of short-acting β_2 -agonist or short- or long-acting muscarinic antagonist used in the present analysis), can improve phenotyping performance.⁴¹ Future analyses will incorporate other EHR data, such as information from clinical notes, to further improve phenotyping accuracy. Another limitation common to EHR-derived data is data missingness. The variable with the highest degree of missingness in our EHR-derived data was patient residential geocodes, which were missing for 23% of patients. While the distributions of basic demographic characteristics (age, race, gender) were similar between patients with and without residential geocode information (data not shown), the lack of geocode information for patients in the latter group prevented us from ascertaining information about their neighborhood SES. Thus, we could not determine whether neighborhood socioeconomic environments differed between patients with missing and available geocode information. Given the strong associations between neighborhood disadvantage and smoking

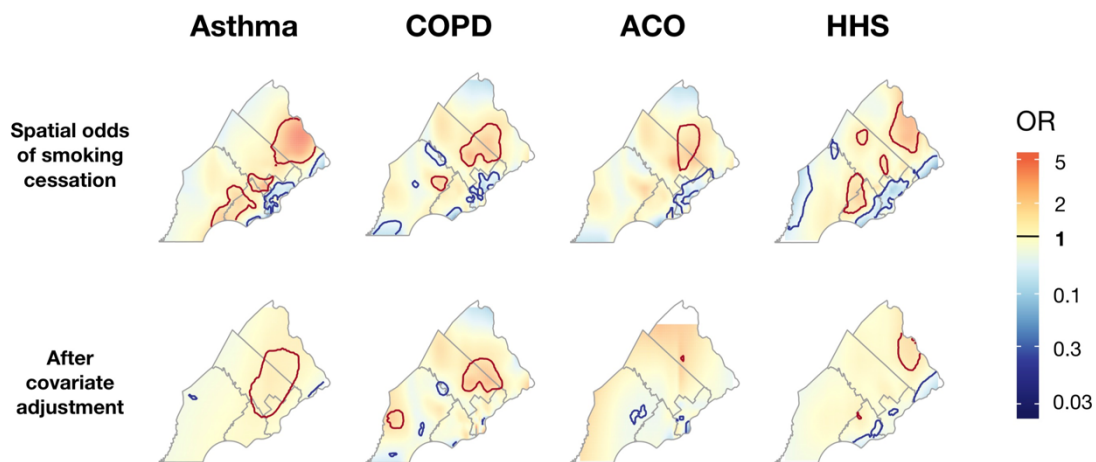


Figure 6. The spatial odds ratio for smoking cessation calculated for asthma, COPD, and ACO patients and HHS respondents who had a positive smoking history. The top row represents odds ratio surfaces estimated from unadjusted models, while plots in the bottom row include adjustment for individual-level covariates (i.e., gender, race/ethnicity, age, BMI, insurance billing class, and neighborhood disadvantage). Regions shaded blue represent areas where the odds of quitting smoking are below the global mean and areas shaded red represent areas where the odds of quitting smoking are greater than the global mean. County boundaries are represented by grey lines. Significant cold spots (with decreased rates of smoking cessation, $p < 0.005$) are indicated by blue contour lines, and significant hot spots (with increased rates of smoking cessation, $p < 0.005$) are indicated by red contour lines.

behaviors noted in the present analysis, it is unclear how the exclusion of patients without geocodes may have biased our results. A smaller, though not insignificant, proportion of patients (~7%) were excluded for missing smoking history. A disproportionate percentage of patients who were excluded for missing smoking history were Non-Hispanic black (41.3% vs. 30.1%) and lived in more disadvantaged neighborhoods (median ADI = 69 vs. 36), indicating a systematic bias in the ascertainment of smoking history among UPHS providers. Though the relatively small number of patients excluded for missing smoking history is unlikely to change our primary results, this bias in the ascertainment of smoking history warrants further investigation. The GAM models used in our analyses mapped the spatial distribution of smoking cessation rates without explicitly modeling spatial autocorrelation, and future analyses can incorporate geographically weighted regression or spatial autoregressive models to account for spatial dependence directly. In addition, future analyses can incorporate patients with non-respiratory health conditions, such as coronary heart disease and hypertension, to understand smoking characteristics across other susceptible patient populations.

A commonly cited limitation of EHR-derived data is a lack of variables, such as socioeconomic factors, that are relevant to the study of many health outcomes.¹² We leveraged residential geocodes available in EHR-derived data to link patient data to high-resolution data on neighborhood disadvantage. The robust associations that we found between patients' neighborhood disadvantage and their smoking status mirrored similar trends found using data from a population-based survey, thus demonstrating the utility of augmenting EHR-derived data with external geospatial variables through the use of residential geocodes. Factors that would be of interest to studies of other health-related behaviors that can be linked to EHR-derived data via similar methods include data on green spaces, walkability, crime, and locations of tobacco or alcohol outlets.

Conclusion

The efficacy of smoking cessation programs can be greatly enhanced by targeting regions with high prevalence of current smokers, particularly among those with obstructive airway diseases, which increase individual susceptibility to the harmful effects of cigarettes. We leveraged EHR-derived data on asthma, COPD, and ACO patients and data from a community-based health survey to identify several coinciding regions in North and West Philadelphia that had high densities of current smokers across patient groups and community members. These areas should be prioritized in smoking cessation initiatives.

Acknowledgements

We would like to thank Sunil Thomas, M.B.A., T.M. 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) F31 HL142153, R01 HL133433, R01 HL141992, and P30ES013508.

References

1. US Department of Health and Human Services. The health consequences of smoking - 50 years of progress: a report of the Surgeon General. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control; 2014.
2. Food and Drug Administration H. Deeming tobacco products to be subjected to the Federal Food, Drug, and Cosmetic Act, as amended by the Family Smoking Prevention and Tobacco Control Act; restrictions on the sale and distribution of tobacco products and required warning statements for tobacco products. Final rule. 81. 2016;90(Federal register):28973.
3. Family Smoking Prevention and Tobacco Control Act, PL 111-31, 123 Stat. 1776, (2009).
4. Philadelphia Department of Public Health. Smoke Free Philly 2019 [Available from: <http://smokefreephilly.org/>].
5. Siroux V, Pin I, Oryszczyn MP, Le Moual N, Kauffmann F. Relationships of active smoking to asthma and asthma severity in the EGEA study. Epidemiological study on the Genetics and Environment of Asthma. *Eur Respir J.* 2000;15(3):470-7.
6. James AL, Palmer LJ, Kicic E, Maxwell PS, Lagan SE, Ryan GF, et al. Decline in lung function in the Busselton Health Study: the effects of asthma and cigarette smoking. *Am J Respir Crit Care Med.* 2005;171(2):109-14.
7. Thomson NC, Chaudhuri R, Livingston E. Asthma and cigarette smoking. *Eur Respir J.* 2004;24(5):822-33.
8. Thomson NC, Spears M. The influence of smoking on the treatment response in patients with asthma. *Curr Opin Allergy Clin Immunol.* 2005;5(1):57-63.
9. Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years' observations on male British doctors. *BMJ.* 2004;328(7455):1519.
10. Hardin M, Silverman EK, Barr RG, Hansel NN, Schroeder JD, Make BJ, et al. The clinical features of the overlap between COPD and asthma. *Respir Res.* 2011;12:127.
11. Bonten TN, Kasteleyn MJ, de Mutsert R, Hiemstra PS, Rosendaal FR, Chavannes NH, et al. Defining asthma-COPD overlap syndrome: a population-based study. *Eur Respir J.* 2017;49(5).
12. Greenblatt RE, Zhao EJ, Henrickson SE, Apter AJ, Hubbard RA, Himes BE. Factors associated with exacerbations among adults with asthma according to electronic health record data. *Asthma Res Pract.* 2019;5:1.

13. Sin DD, Miravitlles M, Mannino DM, Soriano JB, Price D, Celli BR, et al. What is asthma-COPD overlap syndrome? Towards a consensus definition from a round table discussion. *Eur Respir J*. 2016;48(3):664-73.
14. Woodruff PG, van den Berge M, Boucher RC, Brightling C, Burchard EG, Christenson SA, et al. American Thoracic Society/National Heart, Lung, and Blood Institute Asthma-Chronic Obstructive Pulmonary Disease Overlap Workshop Report. *Am J Respir Crit Care Med*. 2017;196(3):375-81.
15. Chatkin JM, Dullius CR. The management of asthmatic smokers. *Asthma Res Pract*. 2016;2:10.
16. Godtfredsen NS, Lam TH, Hansel TT, Leon ME, Gray N, Dresler C, et al. COPD-related morbidity and mortality after smoking cessation: status of the evidence. *Eur Respir J*. 2008;32(4):844-53.
17. Laniado-Laborin R. Smoking and chronic obstructive pulmonary disease (COPD). Parallel epidemics of the 21 century. *Int J Environ Res Public Health*. 2009;6(1):209-24.
18. 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. In: *Statistics NCfH*, editor. Hyattsville, MD2012.
19. Tilert T, Dillon C, Paulose-Ram R, Hnizdo E, Doney B. Estimating the U.S. prevalence of chronic obstructive pulmonary disease using pre- and post-bronchodilator spirometry: the National Health and Nutrition Examination Survey (NHANES) 2007-2010. *Respir Res*. 2013;14:103.
20. Xie S, Himes BE. Approaches to Link Geospatially Varying Social, Economic, and Environmental Factors with Electronic Health Record Data to Better Understand Asthma Exacerbations. *AMIA Annu Symp Proc*. 2018;2018:1561-70.
21. Xie S, Greenblatt R, Levy MZ, Himes BE. Enhancing Electronic Health Record Data with Geospatial Information. *AMIA Jt Summits Transl Sci Proc*. 2017;2017:123-32.
22. Kind AJH, Buckingham WR. Making Neighborhood-Disadvantage Metrics Accessible - The Neighborhood Atlas. *N Engl J Med*. 2018;378(26):2456-8.
23. University of Wisconsin School of Medicine and Public Health. Area Deprivation Index [Internet]. 2018 [cited 1 August 2018]. Available from: <https://www.neighborhoodatlas.medicine.wisc.edu/>.
24. Singh GK, Williams SD, Siahpush M, Mulhollen A. Socioeconomic, Rural-Urban, and Racial Inequalities in US Cancer Mortality: Part I-All Cancers and Lung Cancer and Part II-Colorectal, Prostate, Breast, and Cervical Cancers. *J Cancer Epidemiol*. 2011;2011:107497.
25. Hu J, Kind AJH, Nerenz D. Area Deprivation Index Predicts Readmission Risk at an Urban Teaching Hospital. *Am J Med Qual*. 2018;33(5):493-501.
26. Kind AJ, Jencks S, Brock J, Yu M, Bartels C, Ehlenbach W, et al. Neighborhood socioeconomic disadvantage and 30-day rehospitalization: a retrospective cohort study. *Ann Intern Med*. 2014;161(11):765-74.
27. Wilcosky TC, Chambless LE. A comparison of direct adjustment and regression adjustment of epidemiologic measures. *J Chronic Dis*. 1985;38(10):849-56.
28. Martuzzi M, Elliott P. Empirical Bayes estimation of small area prevalence of non-rare conditions. *Stat Med*. 1996;15(17-18):1867-73.
29. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. p. Retrieved from <https://www.R-project.org>.
30. 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.
31. U.S. Census Bureau. Centers of Population by Census Tract: 2010 [available at <https://www.census.gov/geo/reference/centersofpop.html>] Accessed 5 March 2019 [Internet]. 2011.
32. Bai L, Bartrell S, Bliss R, Vieira V. MapGAM: Mapping smoothed effect estimates from individual-level data. R package at <https://rdrr.io/cran/MapGAM/man/MapGAM-package>. 2016.
33. NIH State-of-the-Science Panel. National Institutes of Health State-of-the-Science conference statement: tobacco use: prevention, cessation, and control. *Ann Intern Med*. 2006;145(11):839-44.
34. Hopkins DP, Briss PA, Ricard CJ, Husten CG, Carande-Kulis VG, Fielding JE, et al. Reviews of evidence regarding interventions to reduce tobacco use and exposure to environmental tobacco smoke. *Am J Prev Med*. 2001;20(2 Suppl):16-66.
35. Bryant-Stephens T, West C, Dirl C, Banks T, Briggs V, Rosenthal M. Asthma prevalence in Philadelphia: description of two community-based methodologies to assess asthma prevalence in an inner-city population. *J Asthma*. 2012;49(6):581-5.
36. Kanjilal S, Gregg EW, Cheng YJ, Zhang P, Nelson DE, Mensah G, et al. Socioeconomic status and trends in disparities in 4 major risk factors for cardiovascular disease among US adults, 1971-2002. *Arch Intern Med*. 2006;166(21):2348-55.
37. Corsi DJ, Boyle MH, Lear SA, Chow CK, Teo KK, Subramanian SV. Trends in smoking in Canada from 1950 to 2011: progression of the tobacco epidemic according to socioeconomic status and geography. *Cancer Causes Control*. 2014;25(1):45-57.
38. Giskes K, Kunst AE, Benach J, Borrell C, Costa G, Dahl E, et al. Trends in smoking behaviour between 1985 and 2000 in nine European countries by education. *J Epidemiol Community Health*. 2005;59(5):395-401.
39. Audrain-McGovern J, Benowitz NL. Cigarette smoking, nicotine, and body weight. *Clin Pharmacol Ther*. 2011;90(1):164-8.
40. Williamson DF, Madans J, Anda RF, Kleinman JC, Giovino GA, Byers T. Smoking cessation and severity of weight gain in a national cohort. *N Engl J Med*. 1991;324(11):739-45.
41. Wei WQ, Teixeira PL, Mo H, Cronin RM, Warner JL, Denny JC. Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. *J Am Med Inform Assoc*. 2016;23(e1):e20-7.