

Technical Supplement

Data Sources, and Definitions, and Methods

Data Sources:

- **SEPA HHS:** The Southeastern Pennsylvania Household Health Survey is conducted every 2 years by the Public Health Management Corporation. For more survey details, visit: <http://www.chdbdata.org/>
- **Philadelphia Tobacco Retailer Database:** This database is based on City of Philadelphia tobacco retailer permitting applications. Any tobacco retailers that were active during 2015 were included in the current analysis. Permits granted after Dec 31, 2015 were not included, and the number of “active” establishments fluctuate continuously based on retailer turnover and the permitting processing time.
- **Schools:** This dataset is a comprehensive inventory of public, private, and charter schools in Philadelphia maintained by the Philadelphia City Planning Commission (PCPC).
- **American Community Survey:** The ACS is a mandatory, ongoing statistical survey by the US Census that samples a small percentage of the population every year. Data for population and poverty thresholds (150% and 300%) were obtained from the 2014 ACS 5-year estimates.

Definitions:

- **Federal Poverty Level (FPL):** A measure of income issued annually by the Department of Health and Human Services that is based on family size and income. For example, those making \$36,054 a year for a family of four, or \$18,497 for an individual (under 65 years) were considered to be living at 150% of the federal poverty level in 2015.
- **Mixed-income Communities:** Mixed-income communities were identified first and then low and high-income communities were defined from the non-mixed-income communities using zip codes. Mixed-income communities were defined as the zip codes that had at least >1/4 of its residents in each FPL category (<150% FPL, 150-300% FPL, and >300% FPL). 34.8% of communities were classified as mixed income.
- **Low-income Communities:** Of those not classified as mixed income, low-income communities were defined as the zip codes that had at least 1/3 of its residents with incomes <150% FPL. 28.3% of communities were classified as low income.
- **High-income Communities:** Of those not classified as mixed income, high-income communities were defined as the zip codes that had at least 1/3 of its residents with incomes >300% FPL. 37.0% of communities were classified as high income.
- **Adult smoking:** Adults who have smoked at least 100 cigarettes in their lifetime and currently smoke every day or "some days" as measured in the HHS
- **Shopping frequency:** Adults' responses to how often they would shop at a store if it stopped selling tobacco as measured in the 2014/2015 HHS

- **Proximity to Schools:** Tobacco retailers were counted if they were within a Euclidean distance (500 or 1,000 feet) of the point associated with the school's address. All schools with any grades K-12 that were operating in 2015 were used.
- **Population density:** Population per square mile from the US Census, 2014 American Community Survey 5-year estimates.

Methods:

Data from Philadelphia County in the Southeastern Pennsylvania Household Health Survey (SEPA HHS) from 2010, 2012, and 2014/2015 were combined to provide large enough sample sizes by zip code for zip code level analyses. The average zip code samples size for HHS data on asthma and smoking prevalence was 238 (range 34 – 516). Two zip codes with confidence interval spreads greater than 25 percentage points that had small sample sizes were excluded from zip code level smoking prevalence analyses (19127, 19137). Individual smoking status of HHS respondents was also used in models that did not examine tobacco retailers, which were only available at the zip code level. Income classifications based on 2014 ACS 5-year estimates were created both to capture meaningful income differences as well as to provide an even spread of low, mixed, and high income zip codes for analysis. A map of how zip codes were classified for analyses is shown in Supplemental Figure 1. Classifications based on concentrated poverty (i.e., 20% of residents below 100% FPL) were considered but did not provide a sufficient sample size of non-low-income zip codes. Statistical and spatial analyses were performed in R (v 3.2.3), GeoDa (v 1.6.7), and ArcGIS 10.3.

For the current study, the Philadelphia Tobacco Retailer Database was used for determining active tobacco retailers during the 2015 calendar year. This database is updated on an ongoing basis, such as when new tobacco retail permits are processed, business ownership changes, or permits are renewed among other things. At the time of publication, there were 3,455 active tobacco retailers.

The final zip code level analyses utilized 46 ZIP code areas within Philadelphia. Two non-residential ZIP code areas were excluded (19109, 19112). ZIP code tabulation areas (ZCTAs), census features that provide a ZIP code like unit of aggregation for tabulating summary statistics, were used for spatial analyses. Observed (unadjusted) data were used to create graphs. Low, mixed, and high income zip codes were represented with dummy variables and low income was the reference group. Total tobacco retailers per zip code were examined using an Ordinary Least Squares (OLS) multivariate spatial regression model that adjusted for mean-centered zip code population density and spatial dependence (Supplemental Table 1). Tobacco retailers per capita were examined using rate ratio confidence interval calculation for count data (Supplemental Table 2). A Poisson regression model, adjusted for school size, was used for variables with skewed distributions due to counts (i.e., tobacco retailers within 500 ft. of schools, tobacco retailers within 1000 ft. of schools). SEPA HHS respondents were categorized into low, mixed, or high-income categories based on their home zip code, and ANOVAs were used to examine differences in zip code smoking prevalence by income category (Supplemental Table 3). The association between zip code smoking prevalence and tobacco retailers per zip code, adjusting for mean-centered population density and neighborhood income level, was examined using OLS multivariate spatial regression.

Philadelphia Department of Public Health, Division of Chronic Disease Prevention. Technical supplement to: Tobacco Sales and Neighborhood Income in Philadelphia. CHART 2016;1(2):1-6.

All spatial analyses used queen contiguity weights, defining neighbors as any ZCTAs that share a border or vertex. Assumptions of multicollinearity, heteroscedasticity, normality of residuals, and spatial autocorrelation were examined. Sensitivity analyses examined the impact of outliers, which supported the main analyses. Significance level was set at $p < .05$. Model estimated parameters for all analyses are included in Supplemental Tables 1-3.

Supplemental Table 1. Results from community income status regressed on total tobacco retailer density, retailer proximity to schools, and adult smoking prevalence.

Model	Coefficient	Std.Error	t-Statistic	p-value
Outcome: Total retailers				
(Intercept)	105.05	9.51	11.05	<.001**
Population Density	0.00	0.00	2.75	.009**
Mixed Income	-42.54	12.76	-3.33	.002**
High Income	-69.32	12.71	-5.45	<.001**
Outcome: Odds of Retailers within 500 ft. of a school				
(Intercept)	1.28	0.06	4.22	<.001**
School size	1.00	0.00	-2.68	0.01*
Mixed Income	0.81	0.10	-2.15	0.03*
High Income	0.59	0.12	-4.48	<.001**
Outcome: Odds of Retailers within 1000 ft. of a school				
(Intercept)	5.95	0.03	64.73	<.001**
School size	1.00	0.00	-1.95	0.05
Mixed Income	0.78	0.05	-5.40	<.001**
High Income	0.55	0.06	-10.36	<.001**
Outcome: Adult smoking prevalence (with retailer density)				
(Intercept)	0.13	0.02	7.97	<.001**
Population Density	-0.000003	0.00	-1.98	0.05
Mixed Income	0.06	0.03	2.36	0.02*
High Income	0.07	0.02	3.45	<.001**
Retailer Density	0.0007	0.00	2.90	0.006**

* $p < .05$; ** $p < .01$

Note: The intercept is the estimate for low income communities, adjusting for covariates.

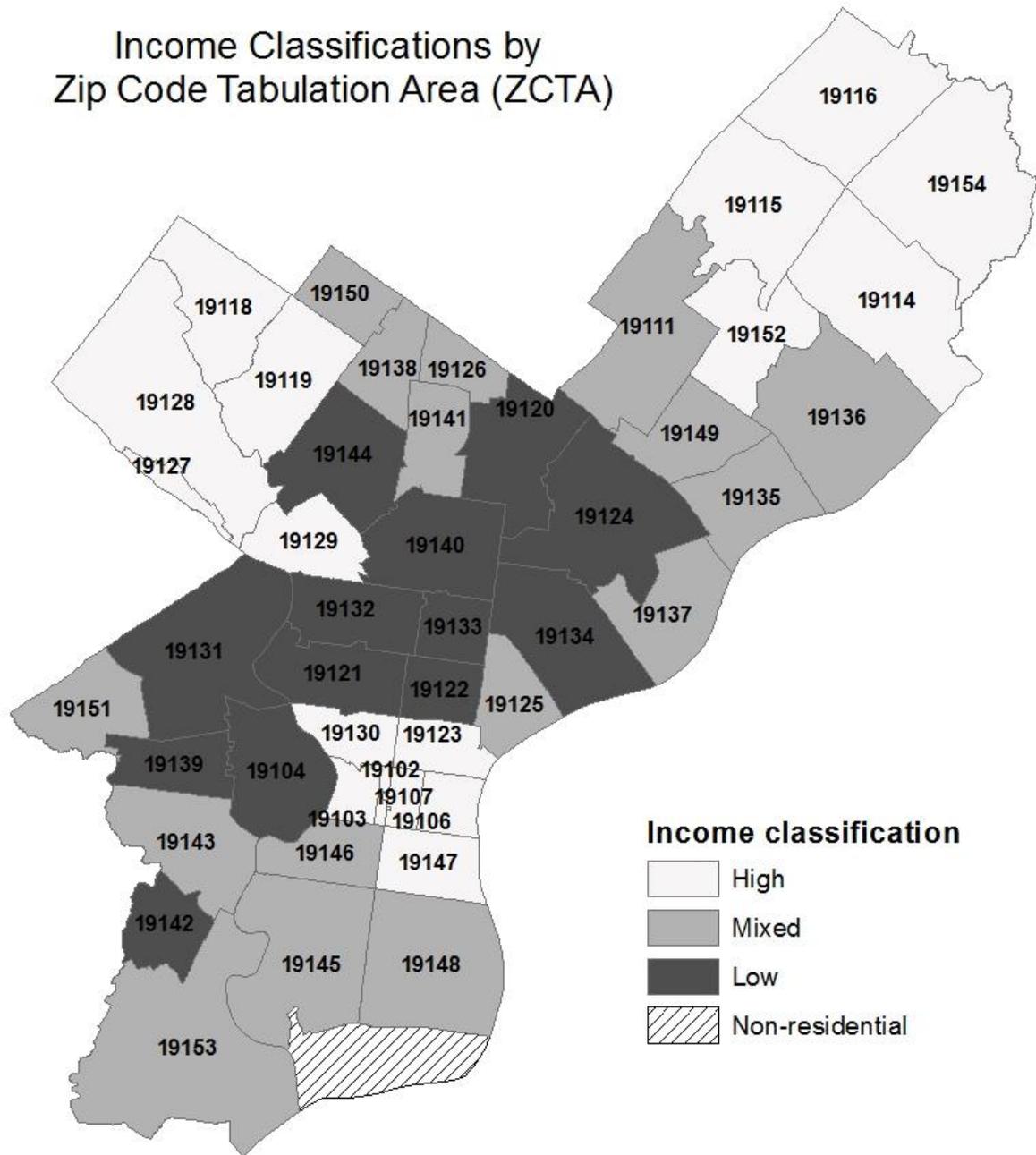
Supplemental Table 2: Results from rate ratio differences in tobacco retailers per capita by income status

	Per Capita	Lower 95% CI	Upper 95% CI	Low vs. High Diff (p-value)	Low vs. Mixed Diff (p-value)	Mixed vs. High Diff (p-value)
Low Income	2.45	2.32	2.57	1.00	0.74	0.26
Mixed Income	1.71	1.60	1.81	(<.001)	(<.001)	(0.002)
High Income	1.45	1.33	1.57			

Supplemental Table 3: Prevalence of adult smoking and adult asthma by community income status.

	Per Capita	Lower 95% CI	Upper 95% CI	Low vs. High Diff (p-value)	Low vs. Mixed Diff (p-value)	Mixed vs. High Diff (p-value)
Adult Smoking						
Low Income	27.09	25.70	28.49	10.34	1.31	9.02
Mixed Income	25.78	24.45	27.11	(<.001)	(.348)	(<.001)
High Income	16.76	15.38	18.13			
Asthma						
Low Income	21.64	20.35	22.93	6.82	3.65	3.17
Mixed Income	17.99	16.82	19.15	(<.001)	(<.001)	(.002)
High Income	14.82	13.52	16.13			

Supplemental Figure 1. Map of low-, mixed-, and high- income classifications by zip code in Philadelphia, PA.



Income categories are mutually exclusive, and mixed income was identified first as zip codes that had at least ¼ of residents <150 of FPL, at least ¼ of residents at 150-300% FPL, and at least ¼ of residents >300% FPL. Non-mixed income zip codes were high (>1/3 of residents >300% FPL) or low (>1/3 of residents <150% FPL) income. Estimates of residents came from the 2010 Census.

