

Assessment of asthma hospitalization, cancer incidence, and cancer mortality in areas near KCBX North and South

MAY 29, 2015

ToxStrategies

Innovative solutions
Sound science

Assessment of asthma hospitalization, cancer incidence, and cancer mortality in areas near KCBX North and South

MAY 29, 2015

PREPARED FOR:

Koch Companies

PREPARED BY:

ToxStrategies, Inc.
9390 Research Blvd
Suite 100
Austin, Texas

Table of Contents

1. Introduction	4
2. Methods	4
2.1 Identification of study area	4
2.2 Data Sources.....	4
2.2.1 Selection of cities for city-to-city comparison.....	4
2.2.2 Health outcome data	5
2.2.2.1 Health outcomes by city.....	5
2.2.2.2 Health outcomes by Chicago zip code	6
2.2.2.3 Health outcomes by Chicago community area.....	6
2.2.3 Demographic data	7
2.2.3.2 Demographics by community area	7
2.3 Statistical Analyses	8
3. Results	9
3.1 City-to-City Comparison.....	9
3.2 Statistical Analysis of Health Outcomes	9
4. Conclusion.....	16
5. References	18

Supplemental Materials

1. Introduction

Community members have expressed concerns about living near the petroleum coke piles at the KCBX North and South Terminals and, in particular, have stated that cancer and asthma rates are higher in neighborhoods near petroleum coke piles than elsewhere in the City of Chicago. In response to these concerns, a study was designed to evaluate cancer and asthma endpoints in these areas. As a preliminary step, cancer and asthma rates in the City of Chicago were compared to those in similar U.S. cities in an effort to establish a baseline for the city as a whole, as well as to guide subsequent analyses. Then, as zip code 60617 and community areas East Side and South Deering are the closest to the petroleum coke piles, these areas were considered exposed to those facilities and all other zip codes and community areas were collectively considered to be unexposed in an ecological assessment of asthma hospitalizations, cancer incidence, and cancer mortality. Regression models provided the baseline relations between demographic variables and the health outcomes; then the exposed areas were evaluated for elevated levels of these outcomes.

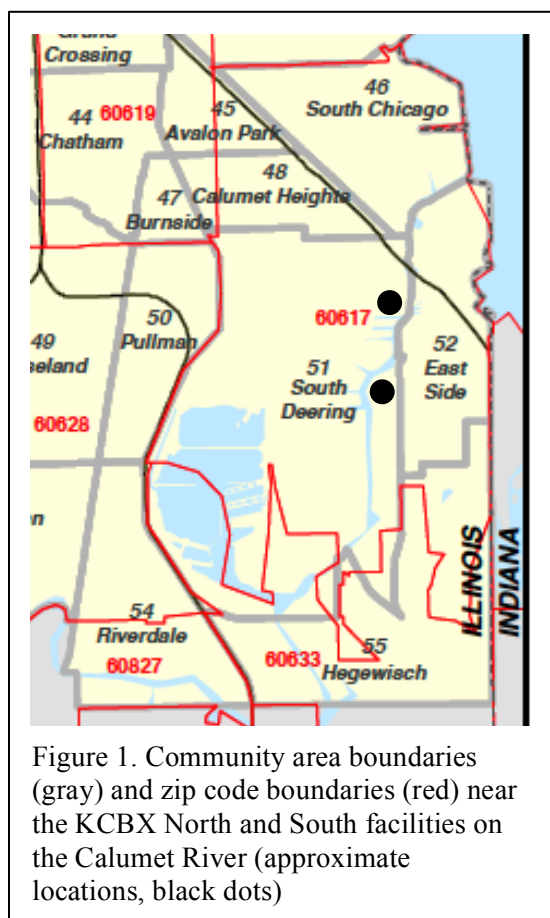


Figure 1. Community area boundaries (gray) and zip code boundaries (red) near the KCBX North and South facilities on the Calumet River (approximate locations, black dots)

2. Methods

2.1 Identification of study area

Petroleum coke is stored at the KCBX North and South bulk materials facilities in Southeast Chicago. These facilities are located in zip code 60617, in and near the South Deering and East Side community areas, respectively (Figure 1). These areas, zip code 60617 and the South Deering and East Side community areas (evaluated together), were identified as the areas of interest in this study.

2.2 Data Sources

2.2.1 Selection of cities for city-to-city comparison

For the preliminary city-to-city comparison, the Centers for Disease Control and Prevention's (CDC) Community Health Status Indicators tool (CHSI, 2014) was used to identify peer counties with similar demographic indicators (*e.g.*, population size, household income, overall poverty) (CDC, 2015a). The City of Chicago lies within the boundaries of Cook County, IL and partially within DuPage County; it is the county seat of Cook County. Using the CDC's CHSI tool, 42 peer counties were identified as being comparable to Cook County based on demographics.

Then, the county seat (which is by definition the government center of the county, which is also typically a city) of each CHSI peer county was identified (National Association of Counties, 2015), which in turn yielded a list of 42 cities considered to be generally comparable to Chicago with respect to demographics.

This list of cities was then further refined based on similarities with respect to industrial history. First, a series of general Internet searches were conducted using terms such as “20th century industrial cities,” “US manufacturing belt,” and “Rust Belt.” Applying the results of this search to the list of cities identified using the CHSI tool narrowed the list of similar cities down to the following five: St. Louis, MO, Detroit, MI, Cleveland, OH, Pittsburgh, PA, and Philadelphia, PA. Additionally, several cities identified as being similar to Chicago in the general internet search on industrial history were added to the list of cities to be compared, even though they were not identified as demographically similar via the peer county comparison tool, to account for the possibility that industry, more than demographics, may have been related to the health endpoints. These additional cities included Akron, OH, Indianapolis, IN, Baltimore, MD, and Boston, MA, yielding a total of 10 cities, including Chicago, available for comparison.

2.2.2 Health outcome data

Health outcome data were limited to publicly available information; no health data were collected for the purposes of these analyses. General Internet searches were conducted to identify available datasets. Datasets were then screened and selected based on the most currently available data for each health outcome of interest and the most granular geographic area available as appropriate to the analysis (i.e., metropolitan statistical area for the city-to-city comparison, zip code for the ecological assessments of asthma hospitalizations and cancer incidence, and community area for the ecological assessment of cancer mortality). Importantly, because the zip code demographics were only available as summarized data across several years, the health outcome data used in the modeling had to be summarized across the same years. In one case the years summarized could not be matched exactly (i.e., for the Chicago community area cancer mortality analysis, outcome data were available for 2006-2010 and demographic data were available for 2007-2011).

2.2.2.1 Health outcomes by city

The CDC’s *Behavioral Risk Factor Surveillance System: Selected Metropolitan/Micropolitan Area Risk Trends* (BRFSS-SMART) was selected as the data source for asthma for the city-to-city comparisons, as it reported data by metropolitan statistical areas (MSAs) corresponding to the selected cities (CDC, 2015b). Adults (>18 years) were surveyed in 2010, and asthma prevalence, as the percentage of respondents who self-reported ever having been diagnosed with asthma, was used in this assessment. None of the available yearly datasets contained prevalence rates for all of the MSAs of interest in this assessment, due to the survey’s response requirements. The 2010 dataset was chosen, due to availability of data for 9 of the 10 MSAs. Prevalence data is not available in the BRFSS-SMART database for Boston-Cambridge-Newton, MA-NH, due to a small sample size, thus it has been omitted from this asthma prevalence comparison.

The CDC’s *Wide-ranging Online Data for Epidemiologic Research* (WONDER) tool was selected as the data source for cancer incidence for the city-to-city comparisons, as it provides access to official federal statistics on newly diagnosed cases of cancer by year as reported in

United States Cancer Statistics (USCS) (CDC, 2015c). The USCS in turn reflects cancer incidence data from two programs - the CDC's National Program of Cancer Registries (NPCR) and the National Cancer Institute's (NCI) Surveillance, Epidemiology and End Results (SEER). Similar to the asthma data, the MSA cancer incidence rates represent the statistical areas defined by the Office of Management and Budget (OMB), allowing for a consistent comparison of MSAs. This assessment specifically focused on the most currently available data for three types of cancer outcomes: combined cancer incidence, incidence of oral and pharynx cancer sites, and incidence of lung and bronchus cancer sites. Data in WONDER are provided as age-adjusted average incidence (per 100,000) rates for the years 2007 – 2011 for all 10 MSAs of interest.

The USCS also includes federal statistics on cancer mortality. This assessment specifically focused on two types of cancer mortality: cancer mortality (all sites) and mortality due to lung and bronchus cancers. Data were provided in WONDER as age adjusted average mortality (per 100,000) rates for the years 2007 – 2011 for all 10 MSAs of interest (CDC, 2015c). These years were selected for consistence with cancer incidence.

2.2.2.2 Health outcomes by Chicago zip code

Asthma hospitalization rates used for the regression modeling were obtained from the Chicago Data Portal (City of Chicago Data Portal, 2015). These data were generated by the Chicago Department of Public Health (CDPH) using the Illinois Department of Public Health (IDPH) annual hospital discharge data. The CDPH found some zip codes to have very small populations and very few cases and, as such, the CDPH aggregated those zip codes into larger sets (60601, 60602, 60603, 60604, 60605, and 60611; 60606, 60607, and 60661; 60610 and 60654; 60622 and 60642; 60707 and 60635; and 60827 and 60633), to avoid instability in zip code assignment or biased estimates due to low population. The Chicago Data Portal provides annual age-adjusted rates (per 10,000) for persons aged 5-64 years through 2011 by zip code and zip code aggregate. For each zip code or zip code aggregate, annual hospitalization rates were then averaged for the years 2007-2011 to obtain the average hospitalization rate for 2007-2011.

Cancer incidence data used for the regression modeling were obtained from the Illinois Department of Public Health (IDPH) Illinois State Cancer Registry (ISCR) (IDPH, 2015). This dataset contains records of individual cancer cases in the State of Illinois. While many types of cancer are reported in the registry, total cancer, lung/bronchus cancer, and oral/pharynx cancer counts were selected for this analysis. The dataset contained data that were grouped into 5-year intervals in the registry dataset. Data for the most recent time period available, 2008-2012, were obtained. For all zip codes in the city of Chicago boundaries, cancer counts were converted into crude cancer incidence (per 100,000 persons) by dividing by zip code population. Cancer incidence rates were calculated for total cancers, lung/bronchus cancers, and oral/pharynx cancers separately. Finally, the same six zip code aggregates created by the CDPH for asthma rates were created for cancer incidence rates, to avoid instability in zip code assignment or biased estimates due to their low populations.

2.2.2.3 Health outcomes by Chicago community area

Cancer mortality rates used for the regression modeling were obtained from the City of Chicago Data Portal (City of Chicago Data Portal, 2015b). Mortality cause was determined based on selected underlying causes of death; specifically, the Chicago Department of Public Health (CDPH) calculated mortality rates using the Illinois Department of Public Health (IDPH)

geocoded annual death certificate dataset. For this analysis, causes of death reported for total cancers (all sites) and lung cancer were selected. In contrast to our analysis of cancer incidence by zip code, which included subsets of lung/bronchus combined and oral cavity/pharynx combined, the only relevant cancer subset available for cancer mortality was lung cancer alone. Data were available annually by community area; data for the most recent years available, 2006-2010, were selected and averaged across the 5-year period.

2.2.3 Demographic data

To match demographics to health outcomes for the various ecological assessments, demographic data were obtained at both the zip code and community area levels, as described below. All data were obtained via publicly available sources such as the U.S. Census Bureau and the Chicago Department of Public Health. Demographic data were available in 1-, 3-, and 5-year summaries; 5-year summaries were chosen to match the 5-year data provided for cancer incidence and cancer mortality. For asthma hospitalizations, because we had selected the 5-year period as providing more stability in results than smaller periods would, demographic data was chosen to match that same 5-year period.

2.2.3.1 Demographics by zip code

Population demographic data were obtained by zip code from the U.S. Census Bureau *American Community Survey (ACS)* 5-Year Estimates (U.S. Census Bureau, 2015a and 2015b). Five-year periods of demographic data were selected to match the periods associated with the most recent health outcome data: 2007-2011 for asthma hospitalizations and 2008-2012 for cancer incidence. All available demographic variables or sets of variables available by zip code tabulation area were obtained for use in the statistical analyses. These included: percentage employed, method of commuting to work (*e.g.*, driving, public transportation, walking, etc.), occupation (*e.g.*, management, service, production, etc.); industry (*e.g.*, agriculture, manufacturing, finance, etc.), class of worker (*e.g.*, private, government, self-employed, etc.), household income (*e.g.*, \$35,000-\$49,999, \$50,000-\$74,999, etc.), type of income (*e.g.*, percentage of households reporting earnings income, social security income, retirement/pension income, etc.), percentage in poverty, gender, age (*e.g.* 25-34, 35-44, etc.), and race/ethnicity (*e.g.* Hispanic-Mexican, Non-Hispanic-Asian, etc.).

As noted previously, the Chicago Department of Public Health had aggregated some zip codes due to the reported zip codes having very small areas or populations. In order to evaluate these aggregates, weighted averages of the demographic data were generated for the six aggregated zip codes (60601, 60602, 60603, 60604, 60605, and 60611; 60606, 60607, and 60661; 60610 and 60654; 60622 and 60642; 60707 and 60635; and 60827 and 60633).

2.2.3.2 Demographics by community area

Demographic data were obtained for community areas from two sources, both of which are maintained by the City of Chicago: 1) the Chicago Data Portal, which maintains open-access data on community areas obtained from the Chicago Department of Public Health (CDPH) (Chicago Data Portal, 2015c), and 2) the Social IMPACT Research Center which tracks change in poverty and poverty-related indicators (Social IMPACT Research Center, 2015). Data in these sources were available for the years 2007 to 2011. Health outcome data were based on the years 2006-2010, and thus there was not an exact match between the datasets. However, the overall impact on the analyses was expected to be negligible as rates of mortality and demographic

characteristics were not expected to change significantly from the period 2006-2010 to the period 2007-2011 (particularly considering that the data are summarized across five years also reduced variability). The Chicago Data Portal provided demographic information related to poverty, housing, dependents, education, income, and unemployment, generated from census tract-level counts and estimates from the U.S. Census Bureau's 2000 and 2010 Decennial Census, and the 2007-2011 American Community Survey (City of Chicago 2014). The Social IMPACT Research Center provided demographic information related to race, poverty, income, cash/SNAP assistance, housing costs, and education from the U.S. Census Bureau's 2000 Decennial Census and the 2007-2011 5-year American Community Survey (Social IMPACT Research Center, 2013). The demographic information in each of these databases were combined to form a more complete set of publicly available demographic data by community area. The combined dataset contained the following variables: race/ethnicity, poverty/hardship, housing or cash assistance, education level, income, employment status, and number of dependents. All variables were utilized in the statistical analyses.

2.3 Statistical Analyses

The preliminary city-to-city assessment involved comparing means and 95% confidence intervals around those means (all values provided by the CDC: CDC, 2015b; and CDC, 2015c) for each MSA. The data were then copied into Microsoft Excel and used for graphical comparisons across MSAs.

For the statistical analysis of health outcomes by zip codes and community areas, the areas were first tested by themselves for any direct association with the outcomes using univariate regression. Then optimal regression models of relations between each of the outcomes and demographic data were built. Each regression model quantified how much variation the significant demographic variables explained in the health outcome rate among zip codes or community areas in the city of Chicago. First, ordinary least squares (OLS) regression was used to determine which of the single demographic variables (e.g. percentage of households reporting earnings income, etc.) or sets of variables (e.g. percentage of households in income bands \$35,000-\$49,999; \$50,000-\$74,999; etc.) were significantly related to the health outcome. Conventional significance tests (at the $p < 0.05$ level) were used to determine the subset of variables to be tested in the multivariate model. Each subset of significant demographic variables was ranked according to the Akaike Information Criterion (AIC) (Akaike, 1974) statistic, which guided the order of entry of variables into each model respectively. Model building began with the most important variable or set of variables; then subsequent covariates were added in importance order and retained if they accounted for additional statistically significant variance in the health outcome rate and set aside if they did not. Periodically the models were evaluated using a backwards elimination, to ensure that redundant variables were not retained, in the interest of developing the most parsimonious model and preventing over-adjustment (Schisterman, Cole, & Platt, 2009).

To test whether the areas near the KCBX North and South Terminals experienced higher rates than predicted based on their demographics, an exposed/unexposed (1/0) indicator was added to the final multivariate model to determine whether it made a significant contribution toward explaining the health outcome. Each regression model was examined for validity with respect to regression assumptions such as normality of residuals and variance constancy.

3. Results

3.1 City-to-City Comparison

The Chicago MSA had similar asthma prevalence, cancer incidence, and cancer mortality rates to MSAs of similar demographics and/or industrial history (See Supplemental Materials). As Chicago did not appear systematically different, this provided a baseline level of understanding of these health outcomes in Chicago.

3.2 Statistical Analysis of Health Outcomes

The regression modeling resulted in a unique set of significant demographic covariates associated with each health outcome. Descriptive statistics for the health outcomes and demographic data retained as significant in at least one of the six regression models are summarized in Tables 1 and 2. Descriptive statistics are not shown for the demographic variables determined to be not significant (note: 174 demographic data points were evaluated in the asthma and cancer incidence models, and 26 demographic data points were evaluated in the mortality models).

Table 1. Descriptive statistics of asthma hospitalization and cancer incidence data and demographics retained as significant in at least one regression model, for all zip codes and for exposed zip code 60617. All data are from 2008-2012 unless otherwise noted.

	Mean	(SD)	Range	60617
<i>Health outcomes¹</i>				
Asthma hospitalization rate	24.9	(19.4)	4.9 - 75.1	32.2
Cancer rate, all sites	2324.7	(626.3)	1301.1 - 3469.0	2550.6
Cancer rate, lung & bronchial	317.2	(125.8)	79.6 - 557.8	381.7
Cancer rate, oral & pharynx	61.8	(18.2)	27.5 - 99.6	53.0
<i>Demographics (percent in zip code)</i>				
Hispanic	22.7	(22.1)	0.9 - 83.1	37.3
Hispanic - Mexican	16.4	(18.9)	0.4 - 77.2	33.2
Hispanic - Cuban	0.3	(0.3)	0.0 - 1.0	0.1
Hispanic – Puerto Rican ²	3.1	(4.2)	0.1 - 17.5	2.1
Non-Hispanic White ²	34.9	(28.7)	0.6 - 87.1	7.5
Non-Hispanic Black	35.3	(36.0)	0.8 - 97.5	53.9
Non-Hispanic Black ²	35.3	(35.9)	0.6 - 97.7	53.5
Income below poverty line ²	21.1	(10.4)	2.7 - 45.5	25.1
Reported earnings income	78.3	(8.0)	58.4 - 90.2	71.2
Age 25-34 years	18.3	(7.7)	8.9 - 39.2	12.3
Age 62+ years	13.7	(4.0)	6.5 - 23.1	16.8
<i>Employment Industry (percent in zip code)</i>				
Manufacturing	9.0	(4.8)	3.1 - 22.0	10.5
Wholesale trade	2.3	(0.9)	0.6 - 4.8	2.5
Education, health care, social assistance	24.1	(6.7)	11.6 - 42.4	24.8
Arts, entertainment, accommodation, food service	10.1	(3.2)	5.1 - 17.7	8.9
Services, except public administration	5.3	(1.3)	3.1 - 9.2	6.0
<i>Employment Occupation (percent in zip code)</i>				
Service	21.4	(6.1)	6.1 - 32.3	24.7
Sales and office	24.2	(3.2)	17.3 - 32.3	24.2

¹ Asthma hospitalization rates (2007-2011) are given per 10,000 and cancer incidence rates are per 100,000.

² Data are from 2007-2011 to match the most recent data for asthma hospitalization.

Table 2. Descriptive statistics of cancer mortality data and demographics retained as significant in at least one regression model, for all community areas in Chicago and for East Side and South Deering community areas.

	Mean	(SD)	Range	East Side	South Deering
<i>Health outcomes (rate per 100,000)</i>					
Cancer mortality rate (all sites)	193.2	(44.1)	126.0 - 287.2	182.7	207.7
Cancer mortality rate (lung)	50.3	(15.6)	15.5 - 85.3	42.2	58.5
<i>Demographics (percent in zip code)</i>					
Hispanic	25.1	(27.7)	0.0 - 89.3	79.8	35.1
Non-Hispanic Black	40.5	(40.2)	0.2 - 99.5	3.9	60.8
Households receiving cash assistance	4.3	(3.3)	0.5 - 18.3	2.2	2.7
Persons in poverty	22.3	(11.9)	2.9 - 60.3	23.3	28.2
Persons in extreme poverty	10.5	(7.0)	0.8 - 36.0	11.6	14.9
Unemployed	14.3	(7.2)	4.6 - 35.9	12.7	12.8
<i>Education (percent in zip code)</i>					
Some college	19.9	(6.3)	7.9 - 31.7	16.6	27.8
High school diploma	25.8	(8.8)	4.4 - 39.7	33.6	31.9
No schooling completed	1.6	(1.2)	0.0 - 5.0	2.3	2.0

Note. Cancer mortality rates are from 2006-2010; demographics are from 2007-2011.

The univariate regression with exposed zip code 60617 alone indicated that the zip code was not significantly related to increased rates of asthma hospitalization ($b = 7.46$, $p = 0.71$). When taking the demographic differences among the zip codes into account, statistical modeling of asthma hospitalization rates indicated that it was significantly associated with the percentage of the population identifying themselves as Puerto Rican, non-Hispanic White, or non-Hispanic Black, and the percent of families or individuals whose income fell below the poverty level in the last 12 months (Table 3). The study area zip code 60617 was not significantly related to asthma hospitalization rate ($b = 0.02$, $t(42) = 0.38$, $p = 0.70$).

In the analysis of cancer incidence, the univariate regression with exposed zip code 60617 alone indicated that the zip code was not significantly related to increased rates of total cancers (all sites) ($b = 96.57$, $p = 0.88$), oral and pharynx cancers ($b = -4.35$, $p = 0.83$), or lung and bronchus cancers ($b = 47.5$, $p = 0.70$). When taking demographic differences among the zip codes into account, statistical modeling indicated that the incidence of total cancers (all sites) was related to age, the industries in which the population was employed, and the percentage of households reporting earnings income (Table 4). The study area zip code 60617 was not significantly related to the incidence rate of cancers (all sites) ($b = -0.04$, $t(32) = -1.61$, $p = 0.12$). Similar findings were observed for lung and bronchus cancer incidence ($b = 0.00$, $t(35) = 0.07$, $p = 0.95$) and oral cancer incidence ($b = -0.08$, $t(36) = -1.17$, $p = 0.25$); for both of these outcomes, study area zip code 60617 was not significantly related to the outcome (Tables 5 and 6, respectively). Incidence rates of lung and bronchus cancers were found to be related to Hispanic ethnicity, age,

households reporting earnings income, and occupation (Table 5). Incidence rates of oral and pharynx cancers were found to be related to Hispanic – Mexican and Hispanic – Cuban ethnicities, Non-Hispanic Black race, and households reporting earnings income (Table 6).

During the evaluation of cancer mortality data, it was noted that 5 of the 77 community areas were described by the City of Chicago Data Portal as having unreliable lung cancer rates due to the low number of cancer deaths in those areas. To ensure that these 5 community areas were not biasing results, statistical modeling was conducted both with and without them. Significant results were unchanged; therefore the 5 community areas were included in the final models reported here. In the analysis of cancer mortality, the univariate regression with exposed community areas South Deering and East Side alone indicated that the community areas were not significantly related increased rates of total cancer mortality (all sites) ($b = 2.07$, $p = 0.95$) or lung cancer mortality ($b = 0.07$, $p = 1.0$). When taking demographic differences among the zip codes into account, statistical modeling indicated that cancer mortality from all sites was related to the percent Black population, households receiving cash assistance, and percent population achieving less than a college education (Table 7), while mortality from lung cancer was found to be related to percent Hispanic population, poverty, and having had no further education after high school (Table 8). The community areas of interest (i.e., East Side and South Deering) were not significantly related to cancer mortality, either for all cancer sites ($b = 0.01$, $t(70) = 0.26$, $p = 0.80$) or for lung cancer ($b = 0.03$, $t(70) = 0.48$, $p = 0.63$), (Tables 7 and 8, respectively).

All multivariate regression models were assessed for violations of assumptions for the appropriateness of linear regression modeling: normality of residuals, variance constancy, and variance homogeneity. All models passed all tests: Shapiro-Wilk, inspection of model residuals plotted against model fitted values, and the Breusch-Pagan tests, respectively.

Table 3. Multivariate OLS regression model of associations between Chicago zip codes' age-adjusted asthma hospitalization rate (per 10,000) and significant demographics, with exposed zip code 60617.

	<i>B</i>	(SE)	<i>b</i>	t(42)	<i>p</i> -value
Intercept	-23.47	(5.38)		-4.36	< 0.0001
<i>Demographics (percent in zip code)</i>					
Hispanic – Puerto Rican	1.00	(0.24)	0.21	4.23	0.0001
Non-Hispanic White	0.22	(0.06)	0.32	3.60	0.0008
Non-Hispanic Black	0.43	(0.05)	0.80	9.54	<0.0001
Income below poverty line	1.06	(0.15)	0.57	7.14	<0.0001
<i>Exposed zip code</i>					
60617	2.28	(6.42)	0.02	0.38	0.7034
<i>Model fit</i>					
<i>R</i> ²	0.92				

Table 4. Multivariate OLS regression model of associations between Chicago zip codes' cancer incidence rate (all sites, per 100,000) and significant demographics, with exposed zip code 60617.

	<i>B</i>	(SE)	<i>b</i>	t(32)	<i>p</i> -value
Intercept	3704.78	442.47		8.37	<0.0001
<i>Demographics (percent in zip code)</i>					
Aged 25-34 years	-21.76	4.35	-0.27	-5.00	<0.0001
Aged 62+ years	93.27	6.80	0.60	13.72	<0.0001
Reporting earnings income	-11.63	3.47	-0.15	-3.35	0.0018
<i>Industry (percent in zip code)</i>					
Manufacturing	-32.75	7.86	-0.25	-4.17	0.0002
Wholesale trade	-67.59	29.37	-0.10	-2.30	0.0270
Education, health care, social assistance	-14.48	5.64	-0.16	-2.57	0.0144
Arts, entertainment, accommodation, food service	-21.28	6.90	-0.11	-3.08	0.0038
Services, except public administration	-62.52	16.00	-0.13	-3.91	0.0004
<i>Exposed zip code</i>					
60617	-188.58	117.22	-0.04	-1.61	0.1159
<i>Model fit</i>					
<i>R</i> ²	0.97				

Table 5. Multivariate OLS regression model of associations between Chicago zip codes' lung and bronchus cancer incidence rate (per 100,000) and significant demographics, with exposed zip code 60617.

	<i>B</i>	(SE)	<i>b</i>	t(35)	<i>p</i> -value
Intercept	96.42	169.04		0.57	0.5715
<i>Demographics (percent in zip code)</i>					
Hispanic	-1.60	0.39	-0.28	-4.08	0.0002
Aged 62+ years	13.24	1.73	0.42	7.64	<0.0001
Reporting earnings income	-3.01	1.52	-0.19	-1.98	0.0541
<i>Occupations (percent in zip code)</i>					
Service	6.79	1.60	0.33	4.23	0.0001
Sales and office	6.87	2.11	0.17	3.25	0.0023
<i>Exposed zip code</i>					
60617	2.62	40.30	0.00	0.07	0.9484
<i>Model fit</i>					
R^2	0.92				

Table 6. Multivariate OLS regression model of associations between Chicago zip codes' oral cancer incidence rate (per 100,000) and significant demographics, with exposed zip code 60617.

	<i>B</i>	(SE)	<i>b</i>	t(36)	<i>p</i> -value
Intercept	250.20	24.86		10.06	<0.0001
<i>Demographics (percent in zip code)</i>					
Hispanic – Mexican	-0.59	0.08	-0.61	-7.62	<0.0001
Hispanic – Cuban	-15.23	6.37	-0.22	-2.39	0.0214
Non-Hispanic Black	-0.35	0.07	-0.70	-5.10	<0.0001
Reporting earnings income	-2.06	0.30	-0.90	-6.90	<0.0001
<i>Exposed zip code</i>					
60617	-10.32	8.85	-0.08	-1.17	0.2505
<i>Model fit</i>					
R^2	0.81				

Table 7. Multivariate OLS regression model of associations between Chicago community areas (CAs)' mortality rates from cancer (all sites, per 100,000) and significant demographics, with exposed area East Side and South Deering communities.

	<i>B</i>	(SE)	<i>b</i>	t(70)	<i>p</i> -value
Intercept	116.35	10.88		10.70	<0.0001
<i>Demographics (percent in CA)</i>					
Black	0.53	0.11	0.49	4.73	<0.0001
Receiving cash assistance	2.20	0.92	0.16	2.38	0.0199
<i>Education (percent in CA)</i>					
Some college	1.51	0.72	0.21	2.08	0.0410
High school diploma	1.14	0.34	0.23	3.40	0.0011
No school	-8.43	2.15	-0.24	-3.91	0.0002
<i>Exposed community areas</i>					
East Side and South Deering	3.52	13.66	0.01	0.26	0.7974
<i>Model fit</i>					
R^2	0.84				

Note. East Side and South Deering were also tested individually; neither was significantly related to the health outcome.

Table 8. Multivariate OLS regression model of associations between Chicago community areas (CAs)' mortality rates from lung cancer (per 100,000) and significant demographics, with exposed area East Side and South Deering communities

	<i>B</i>	(SE)	<i>b</i>	t(70)	<i>p</i> -value
Intercept	27.88	3.18		8.76	<0.0001
<i>Demographics (percent in CA)</i>					
Hispanic	-0.23	0.05	-0.41	-4.95	<0.0001
In poverty	-0.68	0.28	-0.52	-2.39	0.0197
In extreme poverty	1.42	0.44	0.63	3.23	0.0019
Unemployed	0.83	0.26	0.38	3.21	0.0020
<i>Education (percent in CA)</i>					
High school diploma	0.64	0.15	0.36	4.39	<0.0001
<i>Exposed community areas</i>					
East Side, South Deering	3.03	6.31	0.03	0.48	0.6320
<i>Model fit</i>					
R^2	0.73				

Note. East Side and South Deering were also tested individually; neither was significantly related to the health outcome.

4. Conclusion

As Chicago as a whole did not appear to differ significantly from other cities with similar demographics or history of industry, the regression analysis focused on whether areas nearest the pet coke piles at the KCBX North and South Terminals were associated with increased rates of cancer and asthma health outcomes. The series of regression analyses demonstrated that when significant demographic variables were accounted for, the zip code and community areas classified as “exposed” to the KCBX facilities were not significantly related to asthma hospitalization, cancer incidence (all sites, oral and pharynx, and lung and bronchus), or cancer mortality (all sites, lung). The modeling did, however, demonstrate that the demographic characteristics of the zip code and community areas of interest were strongly significantly related to their health outcome rates.

Some factors known to be related to cancer, such as population percentage tobacco use and genetic make-up, were not available by zip code or community area and thus could not be included in an ecological analysis. Even so, the regression models, validated for appropriateness and fit, yielded a squared correlation coefficient upwards of 0.70, indicating that roughly 70+% of the variance in health outcomes was explained by the retained demographic variables. Poverty was one of the demographics that played a significant role, as did education and percentage of households reporting earnings income, even with age controlled for. The importance of demographic factors in both ecological and individual studies has also been demonstrated elsewhere. Several studies have reported that living in areas associated with greater pollution, such as urban environments, is not related to cancer incidence or mortality, once individual factors such as age, race, and poverty were taken into account (reviewed by Meilleur et al., 2013). Likewise the role of education in cancer mortality rates has been identified as an important variable in other ecological studies of the U.S. population (Albano et al., 2007).

Asthma hospitalization rates may be associated with variables that were unavailable for modeling in the current analyses, such as medical insurance coverage. Additionally, asthma hospitalization is not the equivalent of incidence or prevalence; a visit to the hospital may be related to other factors such as the availability of primary care or insurance coverage. However, when controlling for poverty in this analysis, the absence of a significant relation between zip code 60617 and asthma hospitalizations does provide evidence that acute events are not occurring at greater rates than expected, given the demographic make-up of the zip code. A similar null finding was reported recently in a U.S. population-based (N = 23,065) epidemiological study of differences in asthma in inner city vs. non-inner city environments, in which inner city residence was not associated with increased asthma-related outcomes (Keet et al., 2014). These authors reported that asthma exacerbations and related emergency department visits were strongly associated with Black race, Puerto Rican ethnicity, and lower household income in univariate analyses, mirroring the findings presented in this analysis.

As with all ecological studies, the present analysis illuminates relations among summarized data: the association between zip code or community area demographic characteristics and those areas’ health outcome rates. Individual exposure to petroleum coke or individual outcomes were not measured or evaluated in this study. However, ecological studies using geographically aggregated data such as this one can quickly test for relations with a wide range of potential

covariates, identify or confirm patterns in population health trends, and narrow the scope for future studies on specific hypotheses.

5. References

Albano, J.D., Ward, E., Jemal, A., Anderson, R., Cokkinides, V.E., Murray, T., Henley J., Liff, J., Thun, M.J. (2007). Cancer mortality in the United States by education level and race. *J Natl Cancer Inst* 99:1384-94.

Centers for Disease Control and Prevention (CDC). 2015a. Community Health Status Indicators (CHSI). Available from: <http://wwwn.cdc.gov/CommunityHealth/homepage.aspx>. Accessed February 20, 2015.

Centers for Disease Control and Prevention (CDC). 2015b. SMART: BRFSS City and County Data. Available from: <http://apps.nccd.cdc.gov/brfss-smart/index.asp>. Accessed January 29, 2015.

Centers for Disease Control and Prevention (CDC), 2015c. CDC WONDER. Available from: <http://wonder.cdc.gov/cancer.HTML>. Accessed January 29, 2015 (asthma and cancer incidence data) and February 24, 2015 (cancer mortality data).

City of Chicago Data Portal. 2015a. Public Health Statistics - Asthma hospitalizations in Chicago, by year, 2000 – 2011. Available from: <https://data.cityofchicago.org/Health-Human-Services/Public-Health-Statistics-Asthma-hospitalizations-i/vazh-t57q>. Accessed May 16, 2014.

City of Chicago Data Portal. 2015b. Public Health Statistics – Selected underlying causes of death in Chicago, 2006-2010. Available from: <https://data.cityofchicago.org/Health-Human-Services/Public-Health-Statistics-Selected-underlying-cause/j6cj-r444>. Accessed November 14, 2014.

City of Chicago Data Portal. 2015c. Census Data – Selected socioeconomic indicators in Chicago, 2007-2011. Available from: <https://data.cityofchicago.org/Health-Human-Services/Public-Health-Statistics-Selected-underlying-cause/j6cj-r444>. Accessed May 16, 2014.

Illinois Department of Public Health (IDPH). 2015. Illinois State Cancer Registry: Public Data Set. Available from: <http://www.idph.state.il.us/cancer/statistics.htm>. Accessed May 8, 2015.

Keet, C.A., McCormack, M.C., Pollack, C.E., Peng, R.D., McGowan, E., & Matsui, E.C. (2014). Neighborhood poverty, urban residence, race/ethnicity, and asthma: rethinking the inner-city asthma epidemic. *Journal of Allergy and Clinical Immunology*, 135:655-662.

Meilleur, A., Subramanian, S.V., Plascak, J.J., Fisher, J.L., Paskett, E.D. & Lamont, E.B. (2013). Rural residence and cancer outcomes in the United States: issues and challenges. *Cancer Epidemiology, Biomarkers & Prevention*, 22:1657-1667.

National Association of Counties (NACO). 2015. County Seats. Available from: <http://www.naco.org/Counties/Pages/CountySeats.aspx>. Accessed February 20, 2015.

Schisterman, E.F., Cole, S.R., & Platt, R.W. (2009). Overadjustment bias and unnecessary adjustment in epidemiological studies. *Epidemiology* 20:488-495.

Social IMPACT Research Center. 2013. Chicago Neighborhood Indicators 2000-2011. Data available from:

http://socialimpactresearchcenter.issuelab.org/resource/chicago_neighborhood_indicators_2011. Accessed November 6, 2014.

US Census Bureau. 2015a. American FactFinder: American Community Survey 5-Year Estimates; DP05: ACS Demographic and Housing Estimates. Available for zip codes in Illinois from: http://factfinder.census.gov/faces/nav/jsf/pages/download_center.xhtml Accessed May 21, 2014 and May 11, 2015.

US Census Bureau. 2015b. American FactFinder: American Community Survey 5-Year Estimates; DP03: Selected Economic Characteristics. Available for zip codes in Illinois from: http://factfinder.census.gov/faces/nav/jsf/pages/download_center.xhtml Accessed May 21, 2014 and May 11, 2015.

Supplemental Materials

City-to-City comparison: Asthma prevalence, cancer incidence, and cancer mortality. Data for asthma prevalence, cancer incidence (all sites, oral and pharynx, and lung and bronchus), and cancer mortality (all sites and lung and bronchus) were retrieved from publicly available data sources for the 10 cities selected and were graphed, as follows (Figures 1 – 7) (note: asthma prevalence data was unavailable for Boston- Cambridge-Newton MA-NH):

Figure 1. Lifetime Asthma Prevalence Rates Among Adult Responders For Select Metropolitan Statistical Areas, 2010. Asthma prevalence rates for adults having ever been diagnosed, as a percentage of total responses to the 2010 BRFSS. Error bars represent the 95% confidence interval for each rate.

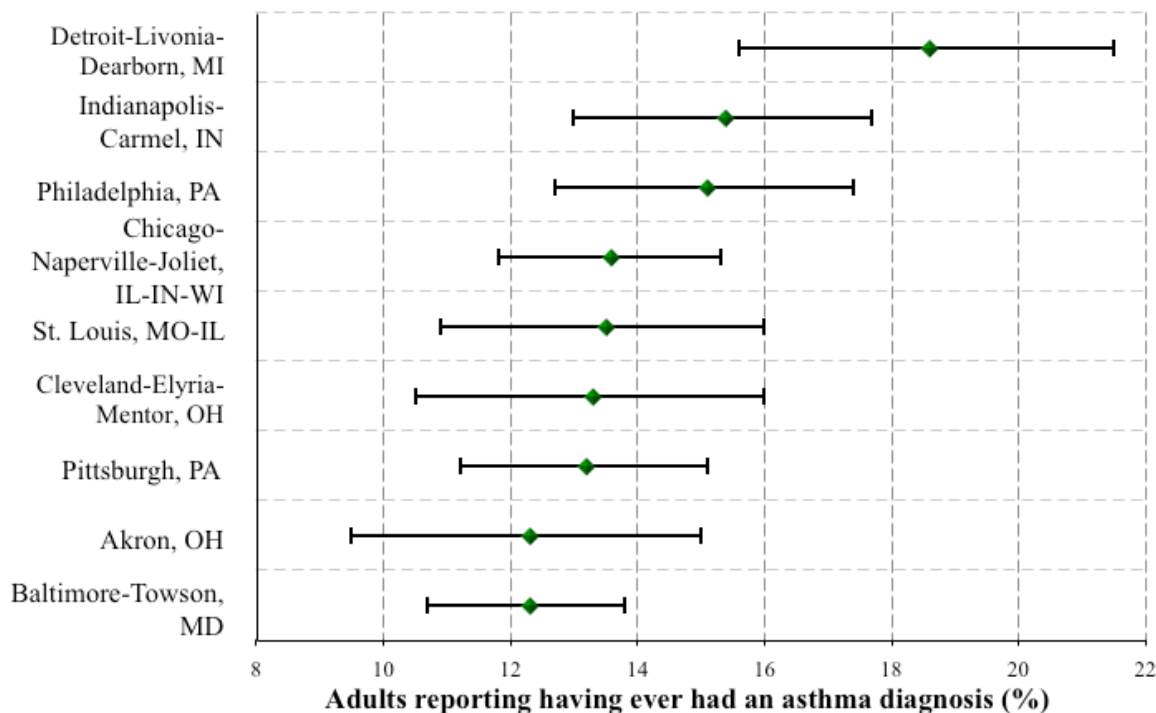


Figure 2. Current Asthma Prevalence Rate Among Adult Responders For Select Metropolitan Statistical Areas, 2010. Asthma prevalence rates for adults who reported a current asthma diagnosis, as a percentage of total responses to the 2010 BRFSS by MSA. Error bars represent the 95% confidence interval for each rate.

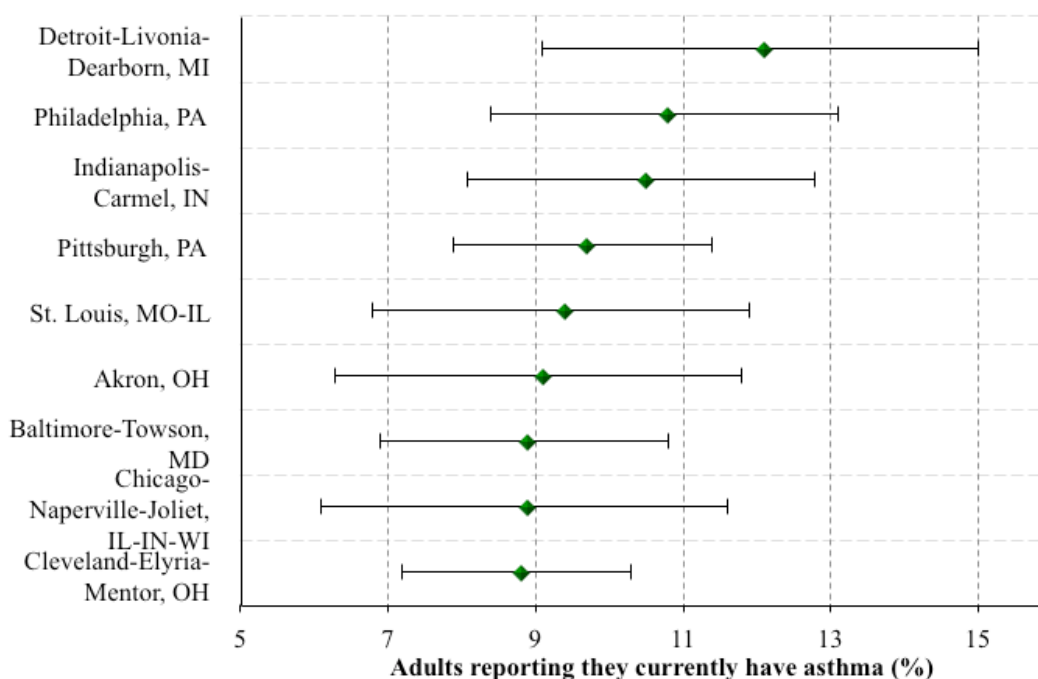


Figure 3. Age-Adjusted Average Cancer Incidence (All Sites) in Select Metropolitan Statistical Areas, 2007-2011. Age-adjusted average annual cancer incidence (all sites combined) per 100,000 people, for the years 2007-2011, by MSA. Error bars represent the 95% confidence interval of each age-adjusted average cancer incidence.

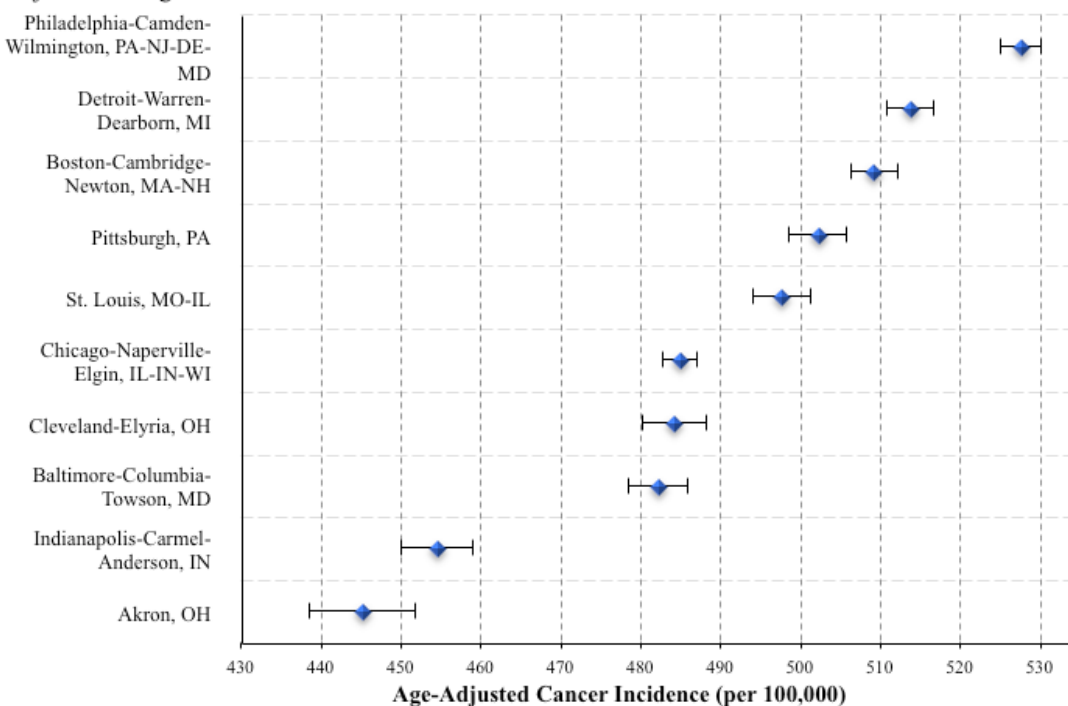


Figure 4. Age-Adjusted Average Cancer Incidence (oral and pharynx sites) in Select Metropolitan Statistical Areas, 2007-2011. Age-adjusted, average annual cancer incidence (oral and pharynx sites) per 100,000 people, for the years 2007-2011, by MSA. Error bars represent the 95% confidence interval of each age-adjusted average cancer incidence.

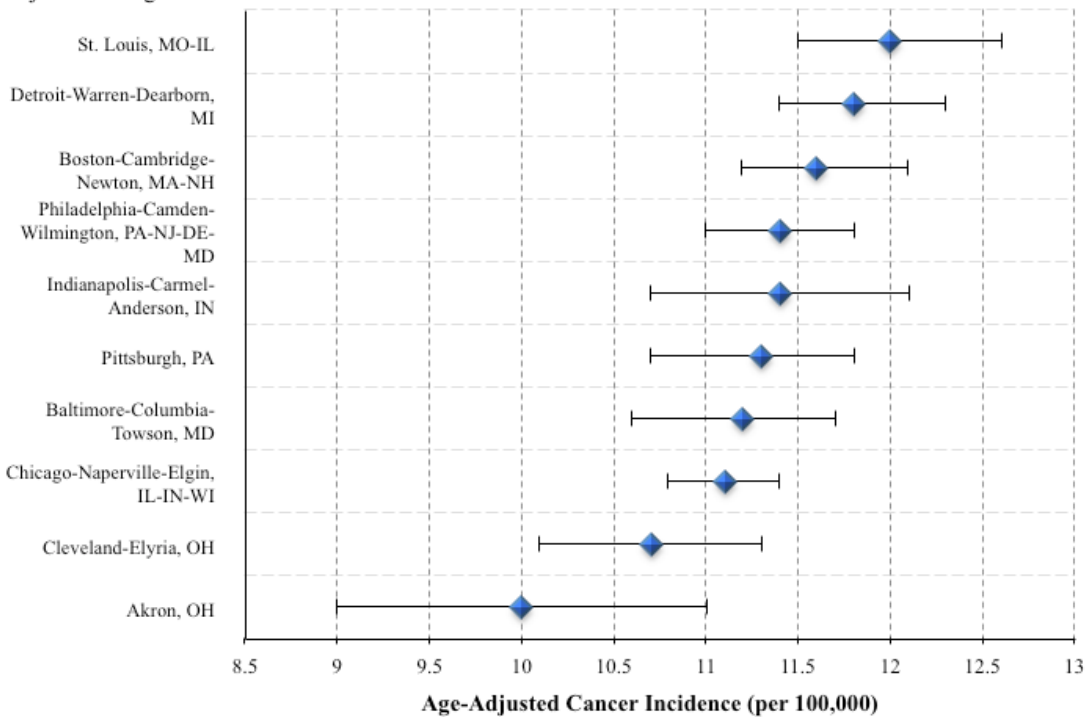


Figure 5. Age-Adjusted Average Cancer Incidence (lung and bronchus sites) in Select Metropolitan Statistical Areas, 2007-2011. Age-adjusted, annual average cancer incidence (lung and bronchus sites) per 100,000 people, for the years 2007-2011, by MSA. Error bars represent the 95% confidence interval of each age-adjusted average cancer incidence.

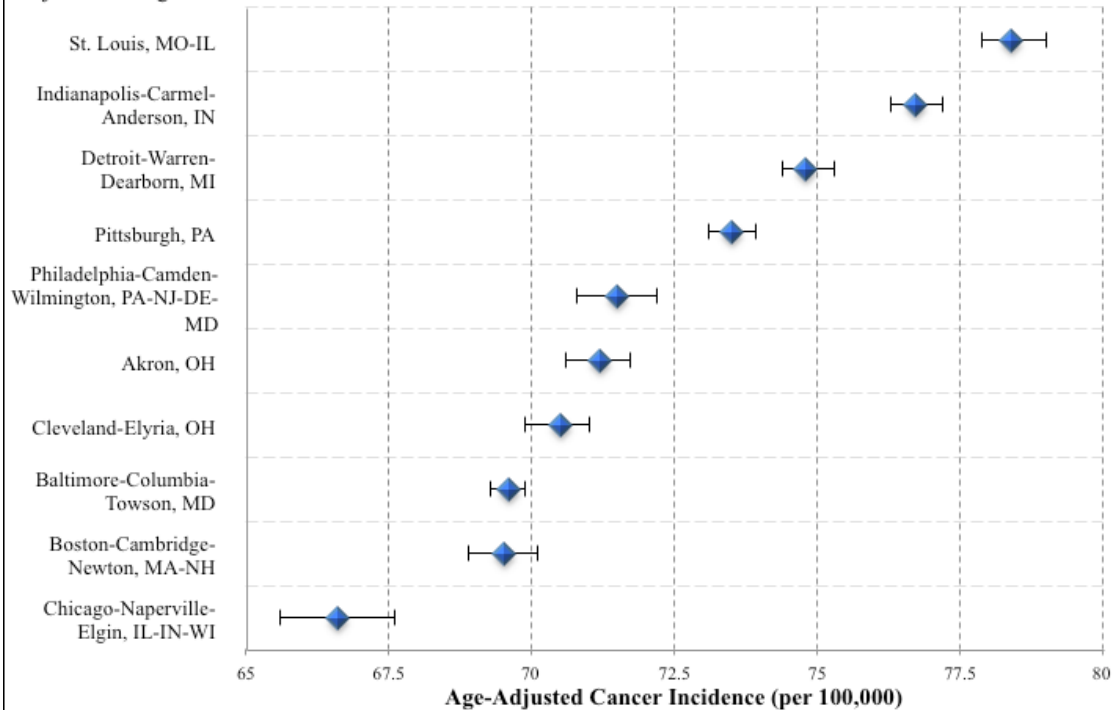


Figure 6. Cancer Mortality (all sites) in Select Metropolitan Statistical Areas, 2007 – 2011. Age-adjusted, average annual cancer mortality for all cancer sites (per 100,000) for the years 2007-2011, by MSA. Error bars represent the 95% confidence interval of each age-adjusted average cancer incidence.

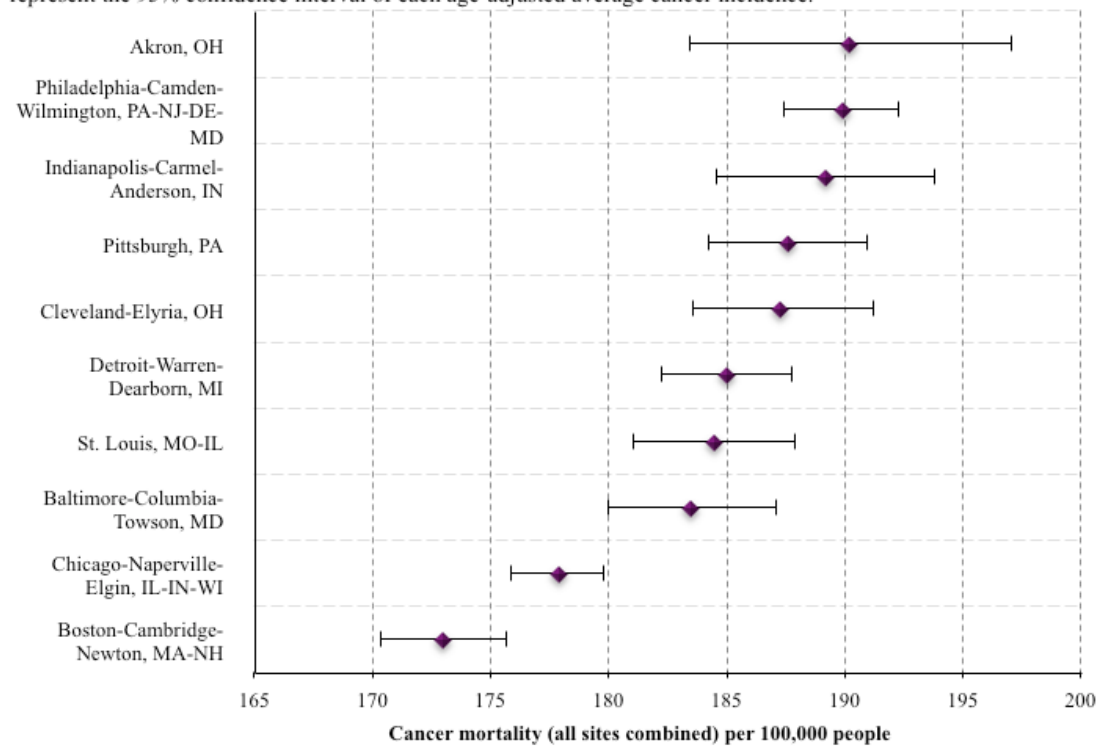


Figure 7. Cancer Mortality (lung and bronchus sites) in Select Metropolitan Statistical Areas, 2007 – 2011. Age-adjusted, average annual cancer mortality for lung and bronchus sites (per 100,000) for the years 2007-2011, by MSA. Error bars represent the 95% confidence interval of each age-adjusted average cancer incidence.

