Spatial and temporal patterns of overdose in Pennsylvania, 2012-2017: analysis with spatiotemporal scan and Getis-Ord statistics

by

Samuel Robert Wittman

BS, University of Pittsburgh, 2017

Submitted to the Graduate Faculty of the

University of Pittsburgh School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Samuel R. Wittman

It was defended on

November 11, 2020

and approved by

Dr. Jeanine Buchanich, Research Associate Professor, Biostatistics

Dr. Thomas Songer, Assistant Professor, Epidemiology

Dr. Ada Youk, Associate Professor, Biostatistics

Thesis Advisor/Dissertation Director: Dr. Jeanine Buchanich, Research Associate Professor, Biostatistics

Spatial and temporal patterns of overdose in Pennsylvania, 2012-2017: analysis with spatiotemporal scan and Getis-Ord statistics

Samuel R. Wittman, BS

University of Pittsburgh, 2017

The purpose of this thesis was to investigate spatial and temporal patterns of fatal overdoses at the census tract level in the U.S state of Pennsylvania (PA), from 2012 to 2017. During these years, fatal overdose reached historically high rates in PA and became a key public health concern in the state. This thesis identified and visualized fatal overdose clusters and hot/cold spots at a more granular level than previous analyses. This information has important implications for understanding where and how fatal overdoses "spread", and for identifying critical regions for overdose surveillance and intervention.

Individual-level overdose mortality data including decedent age, race/ethnicity, sex, county, and census tract were compiled by the PA Department of Health. SaTScan was used to analyze this data with space-time scan statistics, using a discrete Poisson model. Some SaTScan analyses were stratified by age or sex. The most likely high-rate or low-rate clusters were identified and tested for significance using Monte Carlo methods. ArcGIS was used to analyze the data for hot spots or cold spots using Getis-Ord Gi* statistics. Gi* statistics were generated for each census tract during each quarter of this period, and an animation was generated to show patterns during the entire period.

SaTScan analysis revealed a statistically significant increase in fatal overdose throughout the entire state during 2016-2017. After adjusting for this trend, SaTScan analysis further showed large, statistically significant high-rate clusters in southwestern PA and Philadelphia, as well as

smaller significant high-rate clusters in Pottstown, York, Allentown, and Hazleton. Stratified SaTScan analyses revealed significant clusters in Philadelphia and southwestern PA for all age groups and both sexes; however, there were important demographic-specific differences in these regions and others across the state. In contrast, Getis-Ord analysis was not useful for examining the significance of specific overdose clusters. Rather, this analysis revealed a fine-grained depiction of relative rates and local patterns of overdose in cities and regions throughout the state. Getis-Ord analysis showed that hot spots were sporadic and inconsistent throughout much of the state (including high-risk southwestern PA), while there were a few highly regular hot spots in neighborhoods of northeastern Philadelphia.

Table of Contents

List of Tables

List of Figures

1.0 Introduction

In Pennsylvania (PA) between 2012 and 2017, 17583 deaths were classified as accidental drug poisoning, or overdoses (*University of Pittsburgh - MOIRA*, n.d.). During this period, PA ranked as one of the U.S. states with the highest age-adjusted rate of fatal overdose (Burke & Buchanich, 2018; Centers for Disease Control and Prevention, 2018; Scholl et al., 2018). This period also demonstrated the continuation of a trend of increasing overdose mortality in the state, dating back at least to 1979, and perhaps earlier (Balmert et al., 2016; Buchanich et al., 2016). Researchers have examined recent dynamics of overdose in PA from a variety of perspectives: Balmert et al. (2016) calculated overdose mortality rates in PA from 1979 to 2014 by age group, race, sex, and county, and used rate ratios to investigate patterns during this period; Dwyer et al., (2018) reported increasing overdose deaths from the synthetic opioids acetyl fentanyl in counties in western PA from January 2015 to February 2016; and Burke and Buchanich (2018) employed a systems approach in a review on the overdose epidemic in PA, including death records, hospitalizations, prescription, surveys, economic data, and more. However, none of these articles included spatial analysis at a level more detailed than the county.

In this thesis, I employ methods of spatiotemporal analysis, namely scan statistics and Getis-Ord statistics, in order to identify spatial and temporal patterns of fatal overdoses at the census tract level in PA, from 2012-2017. By identifying specific clusters or "hot spots" in PA when/where risk of fatal overdose was significantly elevated, I present fine-grained insights into the nature of the overdose epidemic in the state and the regions which are central to this epidemic. I also provide a practical comparison of several methods of spatiotemporal analysis.

2.0 Methods

2.1 Data

Individual-level overdose mortality data during 2012-2017 were compiled by the PA Department of Health. For decedents age 15 or older, all deaths with underlying cause of death code in the range X40-X49 (ICD 10) were classified as accidental drug poisonings and included in this analysis. Decedents under age 15 were excluded from analysis due to low numbers. The date and time of deaths were recorded, as well as information on the age, race/ethnicity, sex, county, and census tract of the decedents. Decedent age was categorized into the following groups: younger than 20; 20-34; 35-49; 50-64; 65 or older.

There are 3218 census tracts in PA. Census tract population estimates by age group and sex were obtained from the American Community Survey 2017 5-year estimates (*US Census Bureau*, 2017). Census tract shapefiles and centroids were obtained from the Census Bureau (*US Census Bureau*, 2010).

2.2 Spatiotemporal scan analysis with SaTScan

SaTScan was used to perform space-time scan analysis on these data (Kulldorff, 1997; satscan.org). SaTScan generated cylindrical windows that were defined by a circular spatial window, with time as the height dimension. By changing the size and location of this window,

SaTScan compared all possible spatiotemporal collections or "clusters" of centroids (and by extension, census tracts) in PA during the study period.

SaTScan analysis identified clusters with significantly more (or significantly fewer) fatal overdose cases than expected based on the statewide data. A discrete Poisson model was used for this analysis, where *A* denoted a particular contiguous cluster of census tracts, A^C denoted all census tracts in PA outside of cluster *A*, *p* denoted the overdose mortality rate inside cluster *A*, and *q* denoted the overdose mortality rate in A^C. The null hypothesis was H_0: $p=q$, i.e. that the overdose mortality rate was the same inside and outside cluster *A*. When looking for clusters with high rates, the alternative hypothesis was H_A: $p>q$; when looking for clusters with low rates, the alternative hypothesis was H_A: $p<q$. This model assumed that $N(A) \sim Poi(p*\mu(A))$ and $N(A^C) \sim Poi(q*\mu(A^C))$, where $N(A)$ denoted the number of fatal overdoses in cluster *A*, $N(A^C)$ denoted the number of fatal overdoses in PA outside of cluster *A*, $\mu(A)$ denoted the population of cluster *A*, and $\mu(A^C)$ denoted the population of PA outside of cluster *A*.

SaTScan calculated a likelihood ratio test statistic λ for each cluster. When looking for high rates, this test statistics was given by the formula:

$$\lambda = \frac{(\frac{N(A)}{\mu(A)})^{N(A)}(\frac{N(A^C)}{\mu(A^C)})^{N(A^C)}}{(\frac{N(T)}{\mu(T)})^{N(T)}} I(\frac{N(A)}{\mu(A)} > \frac{N(A^C)}{\mu(A^C)})$$

where $N(T)$ denoted the total number of fatal overdoses in PA, $\mu(T)$ denoted the total population of PA, and $I(\ )$ was the indicator function, which equals one when the inequality $\frac{N(A)}{\mu(A)} > \frac{N(A^C)}{\mu(A^C)}$ is true (i.e. when the rate of fatal overdoses is higher inside cluster *A* than outside cluster *A*) and equals zero when the inequality is false (see Kulldorff 1997 for derivation and more detail). When looking for low rates, the formula is the same, with reversed inequality in the indicator function.

The cluster with the maximum test statistic was designated as the most likely cluster. The p-value of the most likely cluster was calculated with Monte Carlo hypothesis testing. In Monte Carlo testing, replicated data sets were generated under the null hypothesis, and the test statistic for the most likely cluster was recorded for each replication. The p-value was generated by ranking the test statistic from the real data against the replicated test statistics. SaTScan also reported secondary clusters with high likelihoods, and performed Monte Carlo hypothesis testing for these clusters. When generating p-values, the test statistics for the secondary clusters were still compared against the most likely clusters from the replications.

The "No centers in less likely clusters" assumption was used to restrict analysis of secondary clusters. This meant that SaTScan only evaluated secondary clusters which did not contain the center of an earlier-reported cluster with a higher likelihood. Without this restriction, there might be multiple overlapping clusters centered on the same high-risk locations.

In some analyses, adjustments for covariates and temporal trends were used. SaTScan accounted for covariates (age group and sex) by performing covariate adjustment on the expected number of cases within the likelihood calculation. SaTScan also accounted for temporal trends in the data non-parametrically by adjusting the expected number of cases separately for aggregated time intervals, and by randomizing data only within (but not across) aggregated intervals during Monte Carlo simulations. Stratified analyses was also performed by restricting overdose cases and census tract populations by age group or sex, and scanning on these stratified data sets.

Overdose deaths and population estimates were each assigned to the centroid point of the appropriate census tract, creating a discrete set of location points throughout PA corresponding to census tracts. There were 89 overdose cases without a recorded census tract. These cases were excluded from this analysis after inspection to confirm that they were not clustered by date or

county in a way that could potentially affect analysis. Additionally, 2 cases were excluded because they occurred in a census tract where the estimated population was zero, making SaTScan's calculations impossible. This census tract (with code 42101980900) contains industrial areas of southwest Philadelphia along the Schuylkill River, along with the Philadelphia airport, where there is little residential housing.

In these analyses, the spatial dimension was limited so that each cluster could not contain more than 30% of the total population. This limit was chosen based on the reasoning that clusters should occur within regions with substantial social/economic connections; the southeastern PA metropolitan region (Philadelphia, Delaware, Montgomery, Chester, and Bucks counties) is the most populous such region in PA with about 30% of the population. Overdose data were aggregated by month, and therefore one month was the minimum cluster duration.

Analyses were conducted with both purely spatial and temporal clusters, i.e. clusters that spanned the entire time period for a certain set of locations (purely spatial), or clusters that spanned all locations for a certain subset of the time period (purely temporal).

A significance level of 0.01 was used to determine whether a cluster was statistically significant. This level was chosen to be conservative with inference regarding fatal overdose clusters and to avoid false detection, especially given the statewide trends during 2016-2017. No adjustment was made for multiple clusters, so each cluster was evaluated as if it was the only cluster present.

2.3 Hot spot analysis with Getis-Ord statistics

Esri's ArcGIS suite (https://www.esri.com/en-us/arcgis/about-arcgis/overview) was also used to visualize and analyze the data. In particular, the hot spot analysis tools were used to generate Getis-Ord Gi* statistics for each census tract (Getis & Ord, 1992). This statistic was defined for census tract *i* by the formula:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} * x_j - \bar{X} \sum_{j=1}^{n} w_{i,j}}{S \sqrt{\frac{[n \sum_{j=1}^{n} w_{i,j}^2 - (\sum_{j=1}^{n} w_{i,j})^2]}{n-1}}}$$

where n was the total number of census tracts, x_j was the overdose mortality rate in census tract j, $w_{i,j}$ was a spatial weight between tracts i and j, $\bar{X}$ was the mean tract-level overdose mortality rate, and S was the standard deviation of tract-level overdose mortality rate. For this analysis, $w_{i,j} = 1$ when tracts i and j are contiguous, and zero otherwise, following the method used for county-level analysis in Jalal et al. (2018). The Gi* statistic for tract i could be considered as the z-score of the overdose mortality rate in the region defined by tract i and all contiguous tracts.

Analysis was initially performed on data from the entire study duration. In order to investigate how hot and cold spots changed over time, data were also aggregated by calendar quarter, and analyses were performed individually for each quarter (data was not aggregated by month in order to simplify the manual steps of this analysis). Census tracts with test statistics at the 0.1, 0.05, and 0.01 significance level were signified using distinct colors.

3.0 Results

3.1 Data Characteristics

There were 17583 accidental poisoning deaths recorded in PA from 2012 to 2017. The overdose mortality rate over this period was roughly 22.96 deaths per 100,000 people per year, using the 2012 Census Department population estimate for PA. By year, the breakdown was 2007 deaths in 2012, 2058 deaths in 2013, 2270 deaths in 2014, 2265 deaths in 2015, 4118 deaths in 2016, and 4865 deaths in 2017. Overdose deaths increased by 82% from 2015 to 2016, and again by 18% from 2016 to 2017. By sex, 5589 decedents (31.8%) were female, 11992 decedents (68.2%) were male, and 2 decedents had unknown/unmarked sex. By age, 268 decedents (1.5%) were under 20, 6444 decedents (36.6%) were 20-34, 6008 decedents (34.2%) were 35-49, 4402 decedents (25.0%) were 50-64, 460 decedents (2.6%) were over 64, and 1 decedent had unknown/unmarked age. Race was recorded with the levels black, white, other, and unknown; 1940 decedents (11.0%) were black, 14904 (84.8%) decedents were white, 579 decedents (3.3%) were other, and 160 decedents (0.9%) were unknown. Table 1 shows crude mortality rates by age group and race during this period, using population data from the 2017 American Community Survey Age and Sex 5 Year Estimates.

Table 1: Overdose deaths per 100,000 per year for various demographics

Male	**Female**			
31.93	14.26			
Under 20	**20-34**	**35-49**	**50-64**	**Over 64**
1.47	42.82	42.66	27.09	3.52
Black	**White**	**Other**		
22.80	23.93	9.70		

3.2 Spatiotemporal scan

3.2.1 Unstratified high-rate scan

Initially I performed a high-rate scan, adjusted for age and sex, but not adjusted for temporal trends. This scan revealed a purely temporal most likely cluster which contains every census tract in the state from February 2016 through December 2017 (data not shown).

In order to identify spatial clusters against the background of increasing overdose mortality rates statewide, I next performed a high-rate scan, adjusted for age, sex, and temporal trends. The results are shown in Figure 1, with the clusters labeled in order of statistical significance. Data on each cluster, including the estimated population, relative risk, and observed/expected cases is shown in Table 2. The most likely cluster in this analysis is a spatial cluster throughout the entire time window (2012-2017) in western PA, centered on the city of Pittsburgh but including many suburbs, rural areas, and small towns and cities such as Washington, Greensburg, Johnstown,

Butler, and New Castle. This cluster alone accounts for nearly 30% of the overdoses in the state during this time period. The second and third most likely clusters are overlapping spatial clusters throughout the entire time window in Philadelphia, specifically in north and near-northeastern Philadelphia. These two clusters account for 7-10% of the overdoses in the state during the period. The remaining significant clusters are significantly smaller and are mostly limited to the core areas of small towns and cities, including Pottstown, Allentown, York, and Hazelton; note that cluster #6 also contains the rural areas and small towns surrounding Hazleton (see Figure 1 and Table 2 for details).

Table 2: Cluster data from time-adjusted, age-adjusted, sex-adjusted high rate scan

Cluster	Time frame	Population	Observed Cases	Expected Cases	Observed/ Expected	Annual Cases per 100,000
1	2012/1/1-2017/12/31	2747500	5193	3603.23	1.44	31.7
2	2012/1/1-2017/12/31	384706	1220	517.65	2.36	51.9
3	2012/1/1-2017/12/31	287770	906	377.20	2.40	52.8
4	2012/1/1-2017/12/31	13884	58	17.91	3.24	71.3
5	2012/1/1-2017/12/31	25137	82	33.35	2.46	54.1
6	2013/12/1-2016/11/30	74956	99	45.90	2.16	47.5
7	2012/1/1-2017/12/31	21571	72	28.92	2.49	54.8
8	2012/1/1-2017/12/31	38861	104	51.43	2.02	44.5
9	2012/1/1-2017/12/31	8374	38	10.87	3.50	76.9

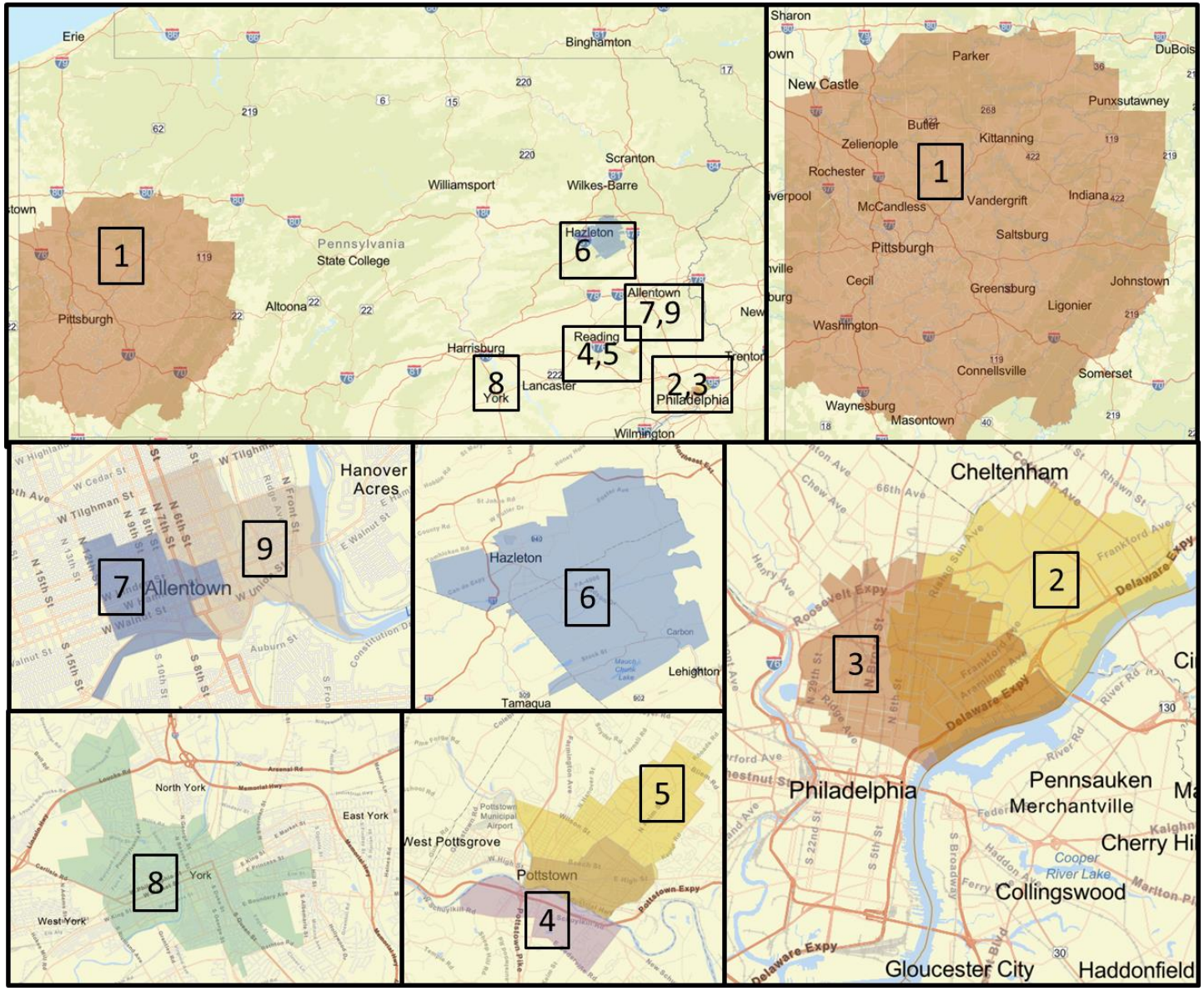

Figure 1: High-rate scan, adjusted for age, sex, and temporal trends

3.2.2 Unstratified low-rate scan

I also performed a low-rate scan, adjusted for age, sex, and temporal trends, in order to identify regions of the state which had significantly lower rates of overdose than the state as a whole. The results of the low-rate scan are shown in Figure 2. Data on each cluster is shown in Table 3. The large low-rate clusters revealed in this scan encompass much of the state during the entire time period, 2012-2017. These low-rate clusters notably do not include the broad region of southwestern PA, parts of Erie County, and an urbanized sliver of southeastern PA including much of Philadelphia, Delaware County, and lower Bucks County. The scan also revealed small low-rate clusters in central neighborhoods of Pittsburgh and Philadelphia.

Table 3: Cluster data from time-adjusted, age-adjusted, sex-adjusted low rate scan

Cluster	Time frame	Population	Observed Cases	Expected Cases	Observed/ Expected	Annual Cases per 100,000
1	2012/1/1-2017/12/31	3961330	3466	5235.32	0.66	14.58
2	2012/1/1-2017/12/31	3968286	3729	5237.05	0.71	15.66
3	2012/1/1-2017/12/31	3350443	3261	4414.30	0.74	16.22
4	2014/9/1-2016/12/31	79260	6	55.74	0.11	3.36
5	2012/1/1-2017/12/31	44137	14	60.96	0.23	5.29
6	2015/2/1-2017/12/31	486261	267	389.42	0.69	18.83

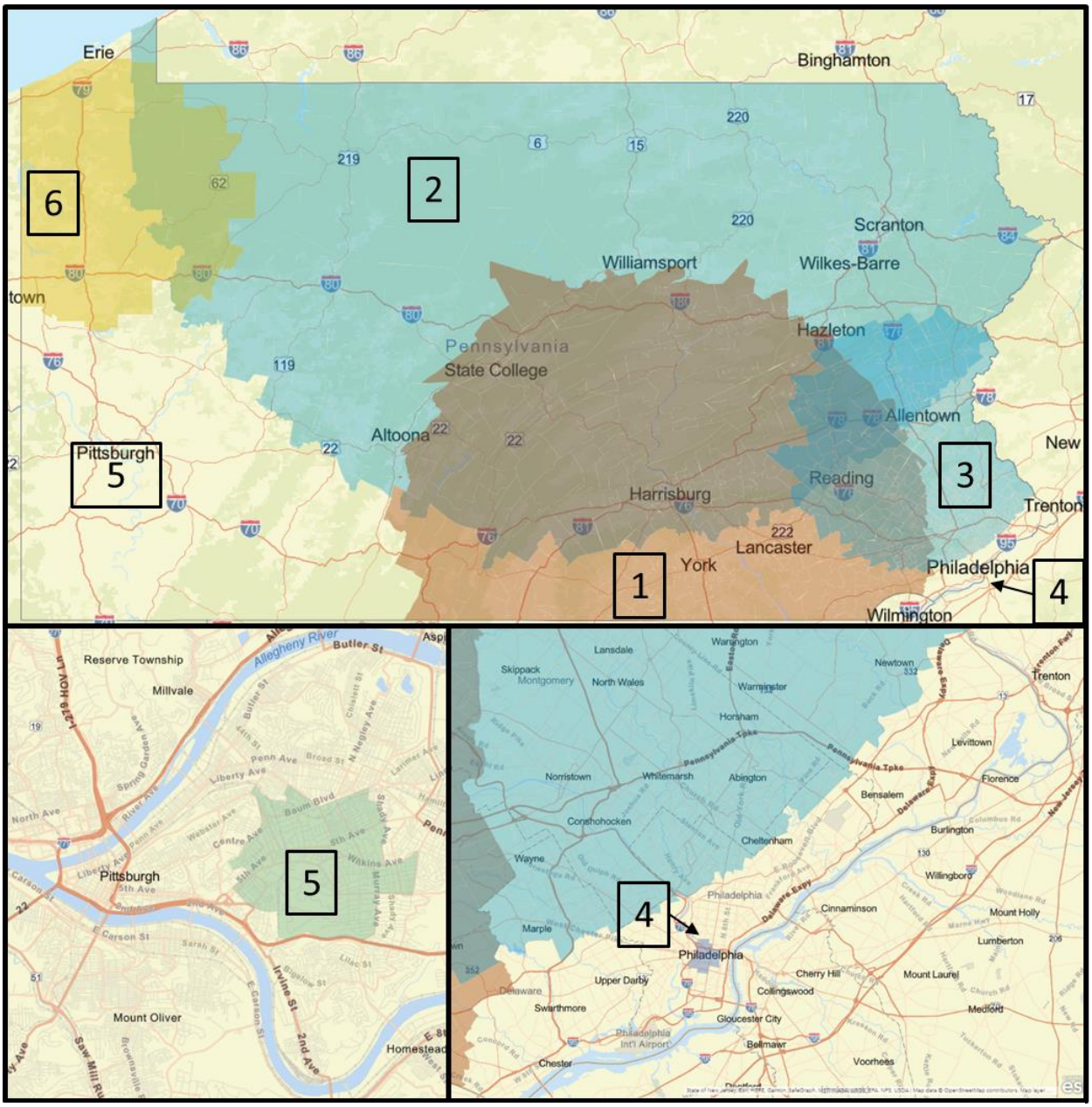

Figure 2: Low-rate scan, adjusted for age, sex, and temporal trends

3.2.3 Male stratified scans

Next I performed a series of stratified scans to explore high and low-rate clusters by age and sex. These scans were also adjusted for temporal trends. The results of the high-rate scan for males are shown in Figure 3, with the clusters labeled in order of statistical significance. Data on each cluster are shown in Table 4.

Table 4: Cluster data from male stratified, time-adjusted, age-adjusted high rate scan

Cluster	Time frame	Population (male)	Observed Cases	Expected Cases	Observed/ Expected	Annual Cases per 100,000
1	2012/1/1-2017/12/31	197048	866	362.77	2.39	71.6
2	2012/1/1-2017/12/31	1339287	3426	2396.14	1.43	42.9
3	2012/1/1-2017/12/31	160924	716	298.06	2.40	72.0

The results of the low-rate scan for males are shown in Figure 4. Note that these clusters are not labelled numerically, and no cluster data are shown. This is also true for the other sex- and age-stratified low-rate scans. It is difficult to differentiate the individual clusters and interpret cluster data for this scan because there are 19 significant clusters, and many of these clusters are large and redundant/overlapping. Additionally, most of these clusters encompass the entire study period, 2012-2017. This pattern is related to the parameters chosen for this analysis (see Discussion). The main takeaway is that much of the state is covered by a statistically significant low-rate cluster for the entirety of 2012-2017. The results from these male sex-stratified scans

resemble the results from the overall unstratified high and low-rate scans: the first- and third-most significant clusters occurred in north and northeastern Philadelphia throughout the entire time period, while cluster #2 occurred in western PA throughout the entire time period. The low-rate clusters again cover much of the state, except for southwestern PA, urban southeastern PA, and parts of Erie County.

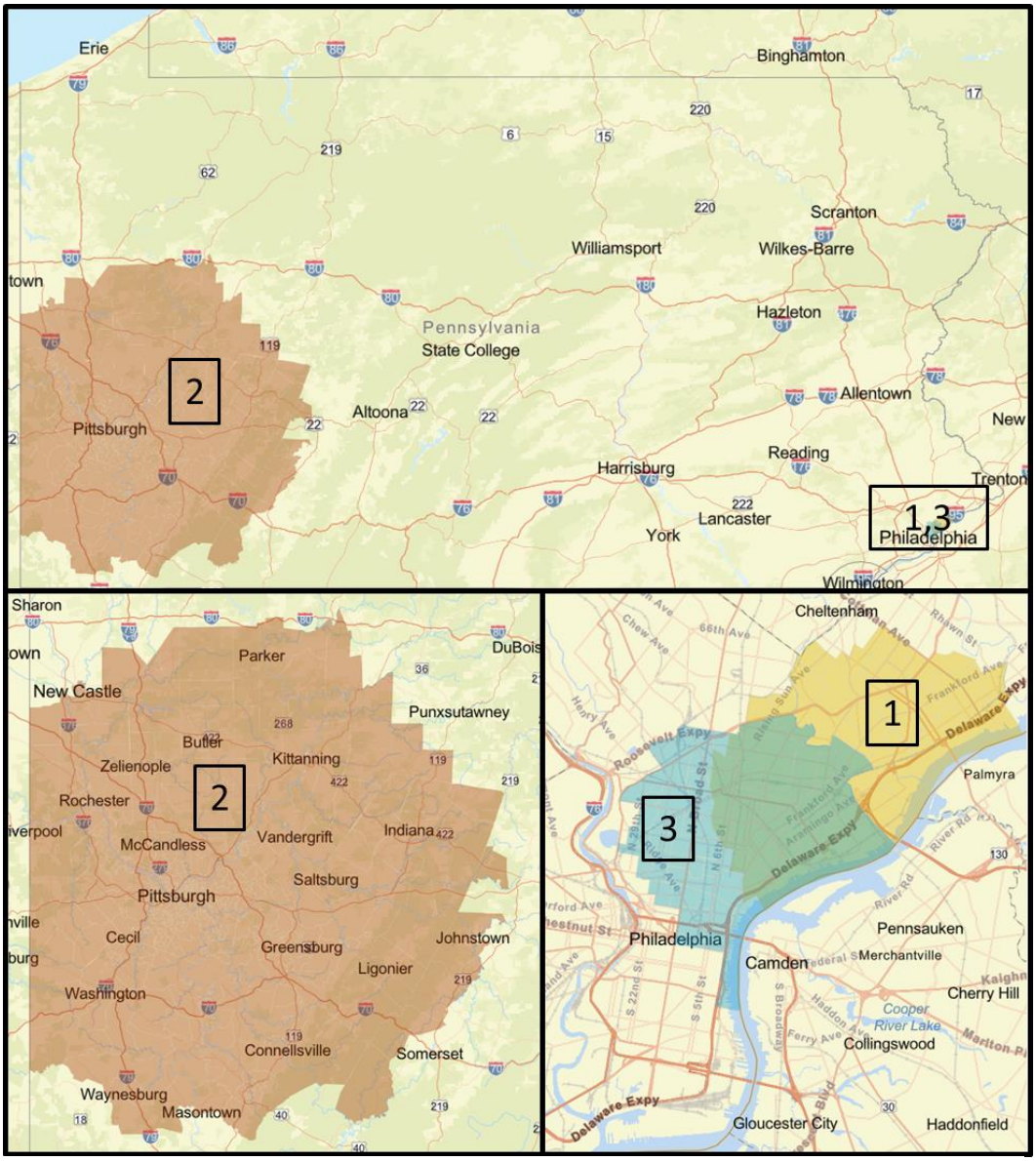

Figure 3: Male stratified high-rate scan, adjusted for age and temporal trends

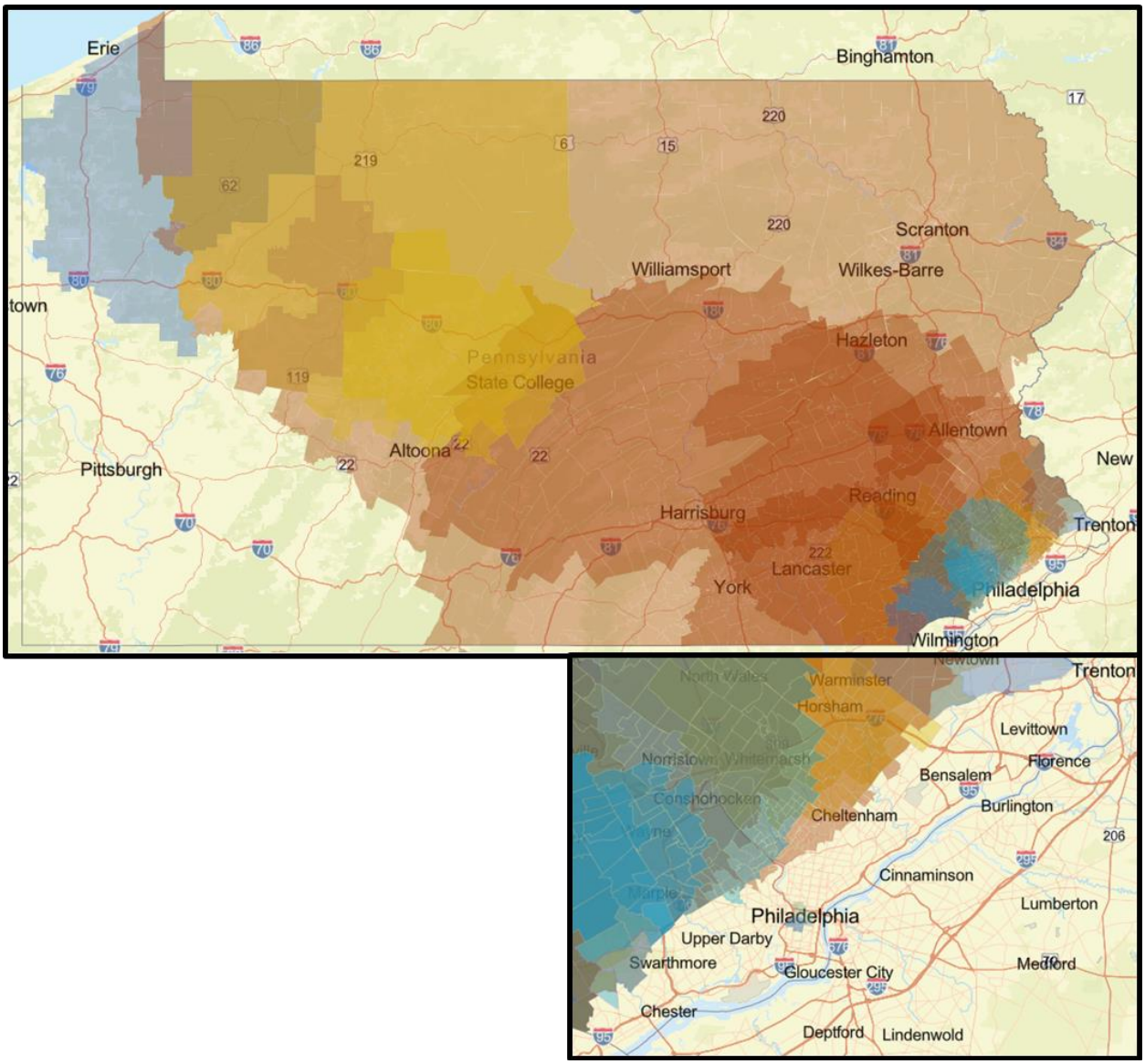

Figure 4: Male-stratified low-rate scan, adjusted for age and temporal trends

3.2.4 Female stratified scans

The results of the high-rate stratified scan for females are shown in Figure 5, and data on each cluster is shown in Table 5. The results of the low-rate scan are shown in Figure 6; again, low-rate clusters are not labeled and data for these clusters are not given.

Table 5: Cluster data from female stratified, time-adjusted, age-adjusted high rate scan

Cluster	Time frame	Population (female)	Observed Cases	Expected Cases	Observed/ Expected	Annual Cases per 100,000
1	2012/1/1-2017/12/31	1408125	1758	1187.94	1.48	20.81
2	2012/1/1-2017/12/31	191328	361	159.16	2.27	31.45
3	2012/1/1-2017/12/31	1500406	1712	1262.09	1.36	19.02
4	2012/1/1-2017/12/31	221414	367	183.69	2.00	27.63
5	2012/1/1-2017/12/31	359802	496	315.29	1.57	22.98

All of these clusters spanned the entire time period. Again, the most significant cluster is in southwestern PA, containing the city of Pittsburgh and surrounding suburban and rural areas. This cluster is overlapped by third-most significant cluster, which contains Pittsburgh but also spans much of rural northwestern PA (which was not covered by a cluster in the unstratified high-rate scan). The second, fourth, and fifth-most significant clusters are in Philadelphia. While the second and fourth-most significant clusters contain areas of near-northeastern and northcentral Philadelphia, the fifth-most significant cluster contains Center City, south Philadelphia and west

Philadelphia (these areas also were not covered by clusters in the unstratified scan). The results from the low-rate scan look similar to the unstratified low-rate scan, except in northwestern PA, which is not covered by any low-rate clusters in the female-only scan.

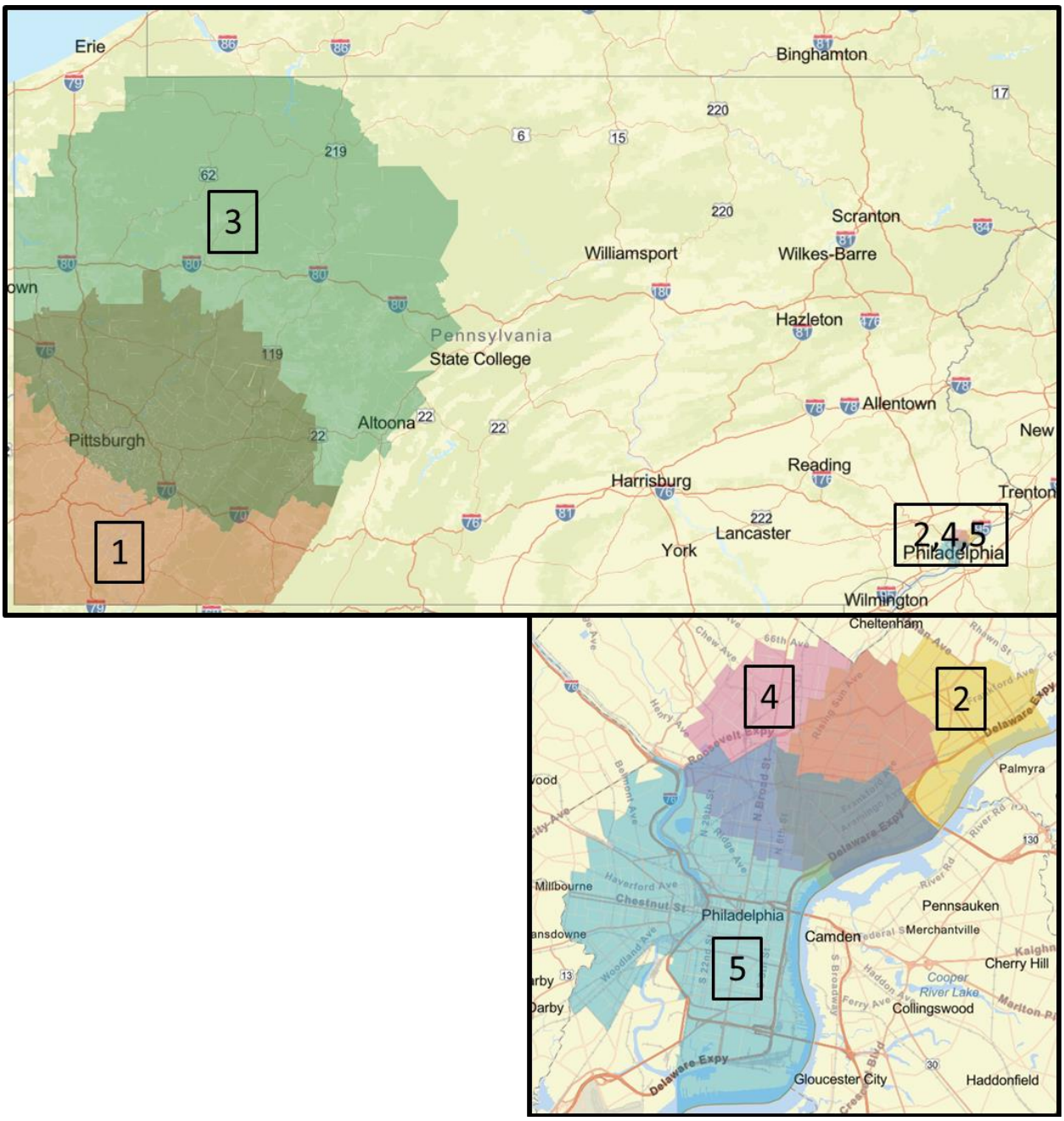

Figure 5: Female stratified high-rate scan, adjusted for age and temporal trends

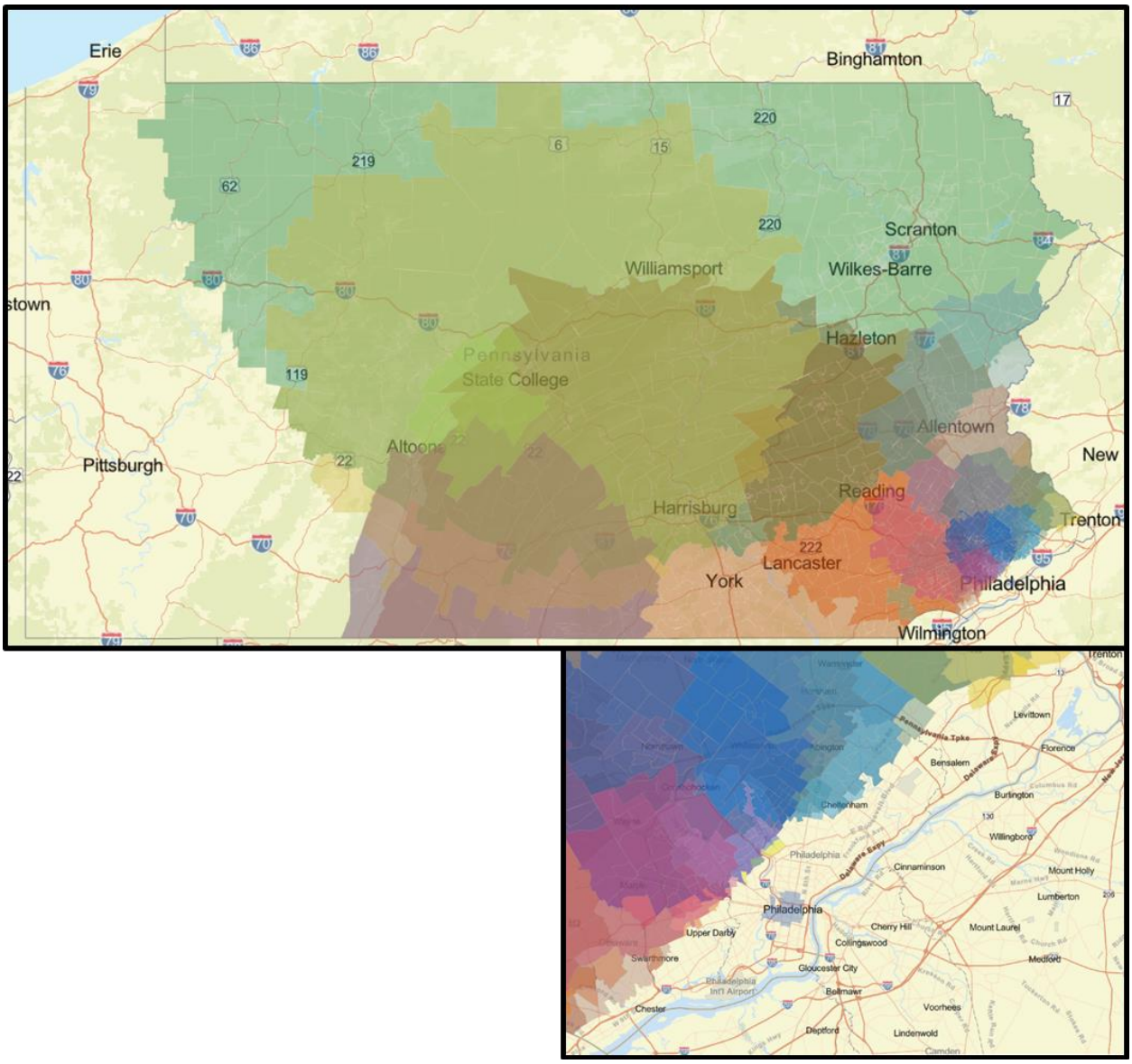

Figure 6: Female stratified low-rate scan, adjusted for age and temporal trends

3.2.5 Under 20 stratified scans

High- and low-rate scans for decedents under age 20 did not reveal any statistically significant clusters (data not shown).

3.2.6 Age 20-34 stratified scans

The results of the high-rate scan for decedents age 20-34 are shown in Figure 7, and data on each cluster are shown in Table 6. The results of the low-rate scan are shown in Figure 8; low-rate clusters are not labeled and data for these clusters are not shown.

Table 6: Cluster data from age 20-34 stratified, time-adjusted, sex-adjusted high rate scan

Cluster	Time frame	Population (20-34)	Observed Cases	Expected Cases	Observed/ Expected	Annual Cases per 100,000
1	2012/1/1-2017/12/31	553421	1933	1330.50	1.45	58.3
2	2012/1/1-2017/12/31	80328	346	190.94	1.81	72.7
3	2012/1/1-2017/12/31	134095	492	322.31	1.53	61.2
4	2012/1/1-2017/12/31	20957	124	49.01	2.53	101.5
5	2012/1/1-2017/12/31	16800	104	39.92	2.61	104.5
6	2012/8/1-2014/12/31	171059	182	107.60	1.69	67.8
7	2015/7/1-2016/8/31	68473	77	33.50	2.30	92.2
8	2012/1/1-2017/12/31	30115	130	72.32	1.80	72.1

Again, there are highly significant clusters in western PA and near-northeastern Philadelphia throughout the entire time period. However, there are also significant clusters in parts of the Philadelphia suburbs, including lower Bucks and Montgomery counties, and in Delaware County. Finally, there is a significant cluster in northeastern PA, spanning from Wilkes-Barre south to Hazleton and Jim Thorpe. Furthermore, some of these clusters did not span the entire time

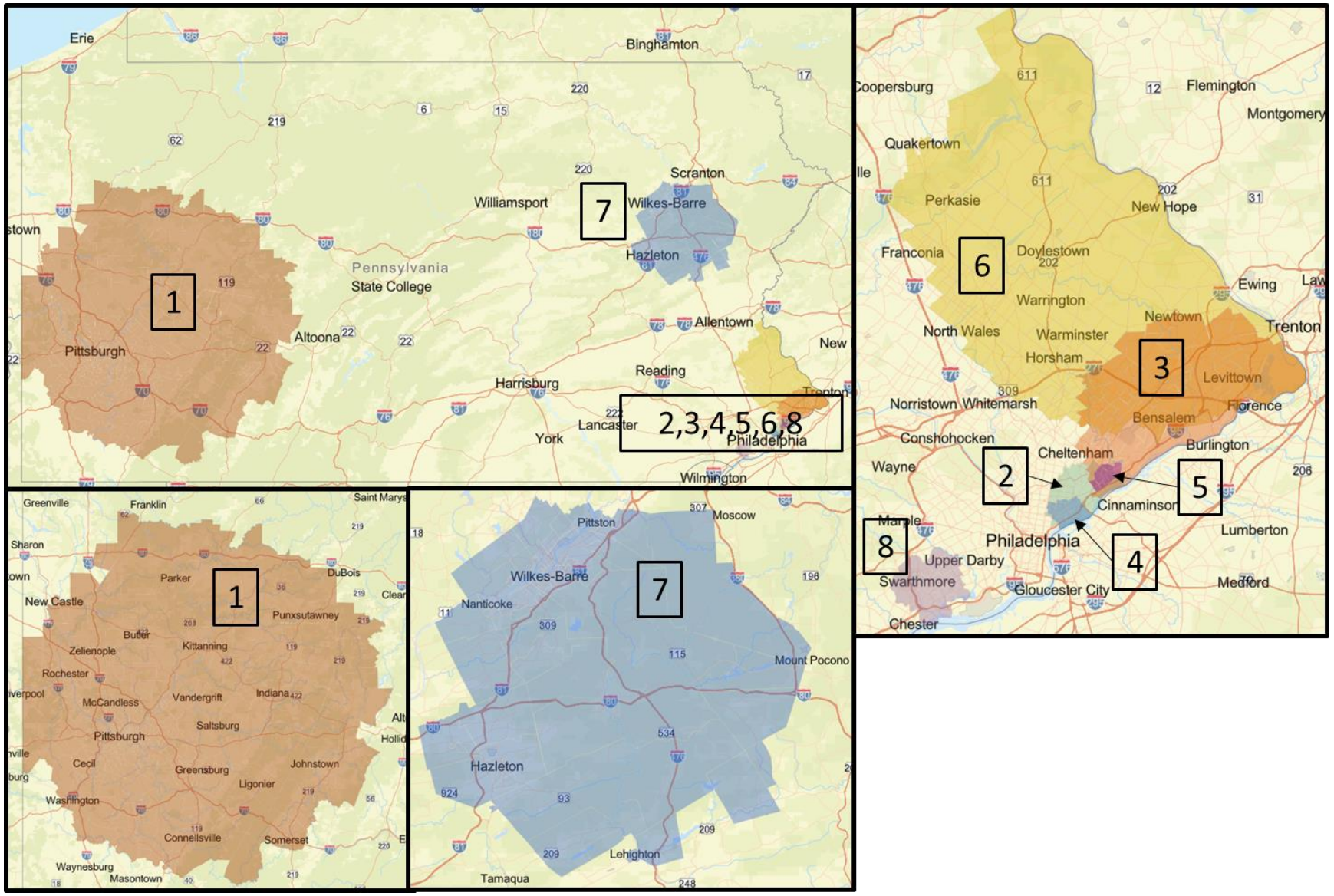

Figure 7: Age 20-34 stratified high-rate scan, adjusted for sex and temporal trends

period (see Table 6), hinting at local dynamics among younger adults that buck the overall statewide trend. The low-rate scan revealed large clusters that cover the middle of the state, while western PA, northeastern PA, and parts of urban southeastern PA remain uncovered. Interestingly, there are small low-rate clusters in parts of Philadelphia and Pittsburgh, indicating the presence of localized factors that may reduce overdose risk in certain communities of young adults.

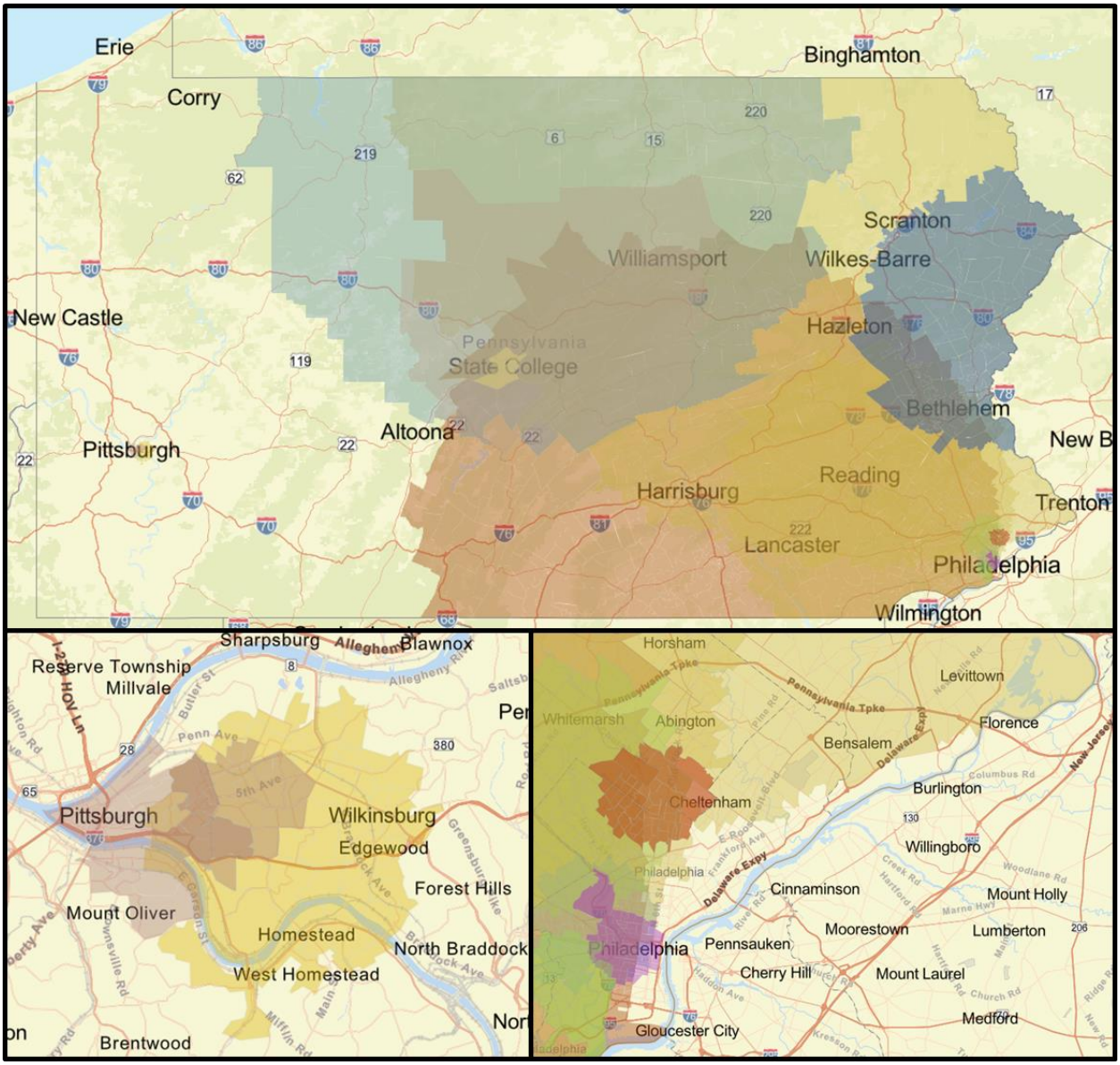

Figure 8: Age 20-34 stratified low-rate scan, adjusted for sex and temporal trends

3.2.7 Age 35-49 stratified scans

The results of the high-rate scan for decedents age 35-49 are shown in Figure 9, and data on each cluster is shown in Table 7. The results of the low-rate scan are shown in Figure 10; low-rate clusters are not labeled and data for these clusters are not shown.

Table 7: Cluster data from age 35-49 stratified, time-adjusted, sex-adjusted high rate scan

Cluster	Time frame	Population (35-49)	Observed Cases	Expected Cases	Observed/ Expected	Annual Cases per 100,000
1	2012/1/1-2017/12/31	163701	768	371.87	2.07	78.7
2	2012/1/1-2017/12/31	527417	1802	1202.01	1.50	57.1
3	2012/1/1-2017/12/31	182555	782	416.54	1.88	71.5
4	2012/1/1-2017/12/31	18720	94	43.28	2.17	82.8
5	2012/1/1-2017/12/31	11556	66	26.42	2.50	95.2
6	2012/1/1-2017/12/31	4995	37	11.27	3.28	125.1

The most significant three high-rate clusters cover most of Philadelphia and western PA, and encompass the entire time period. There are also three small high-rate clusters in Scranton and Allentown, all of which also encompass the entire time period. The low-rate clusters cover the middle of the state, leaving western PA, northeastern PA, and urban southeastern PA uncovered. There are also smaller low rate clusters containing some of the northern suburbs of Pittsburgh.

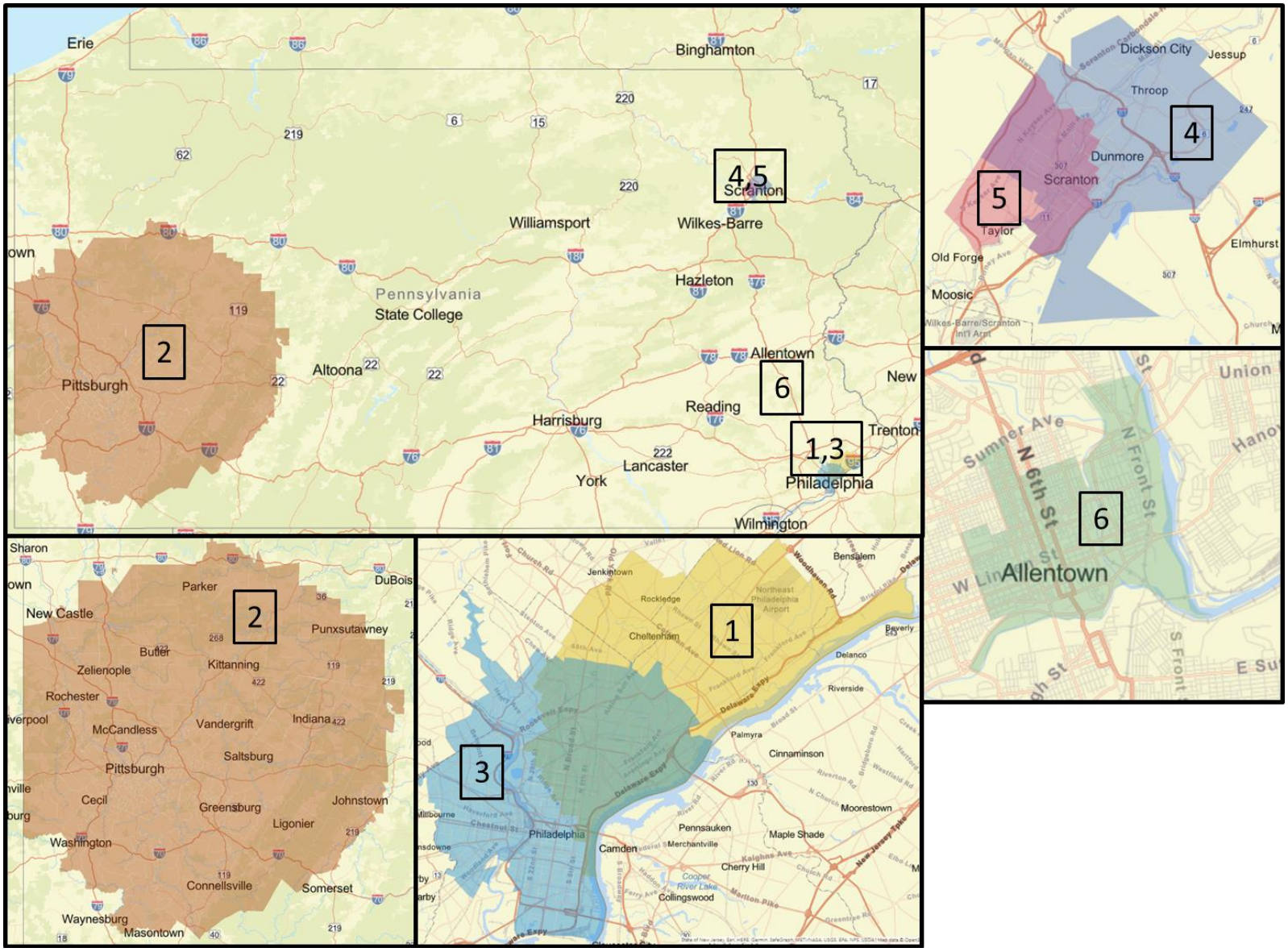

Figure 9: Age 35-49 stratified low-rate scan, adjusted for sex and temporal trends

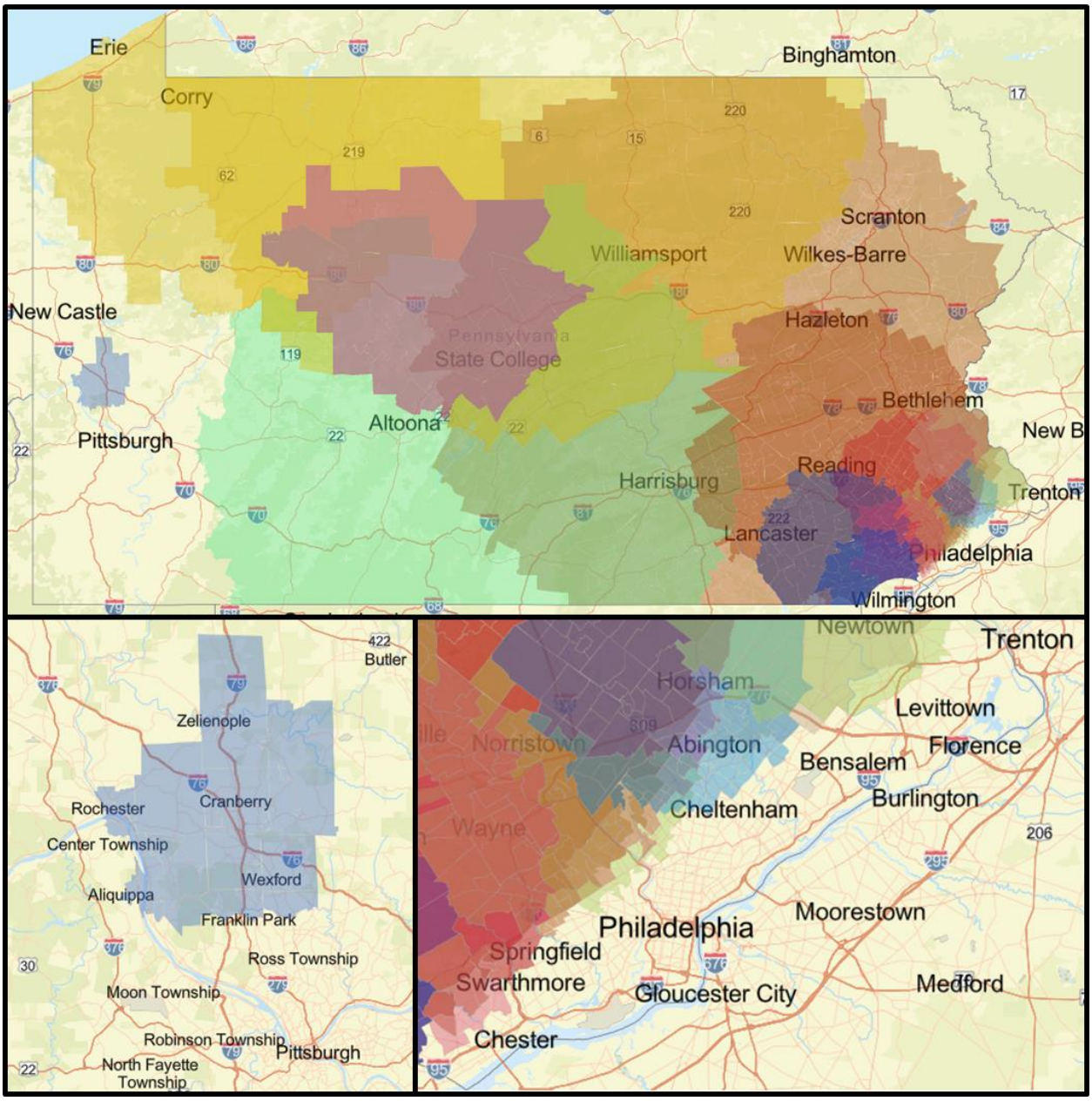

Figure 10: Age 35-49 stratified low-rate scan, adjusted for sex and temporal trends

3.2.8 Age 50-64 stratified scans

The results of the high-rate scan for decedents age 50-64 are shown in Figure 11, and data on each cluster is shown in Table 8. The results of the low-rate scan are shown in Figure 12; low-rate clusters are not labeled and data for these clusters are not shown.

Table 8: Cluster data from age 50-64 stratified, time-adjusted, sex-adjusted high rate scan

Cluster	Time frame	Population (50-64)	Observed Cases	Expected Cases	Observed/ Expected	Annual Cases per 100,000
1	2012/1/1-2017/12/31	208155	915	358.53	2.55	72.2
2	2012/1/1-2017/12/31	371928	1048	621.85	1.69	47.7
3	2012/1/1-2017/12/31	145807	528	242.73	2.18	61.5
4	2012/1/1-2017/12/31	35538	210	59.37	3.54	100.0
5	2014/10/1-2017/9/30	7026	30	7.68	3.91	110.5

The most significant cluster contains much of Philadelphia, and the second-most significant cluster contains much of southwestern PA. There are also smaller clusters in the Pittsburgh area which are entirely contained by the southwestern PA cluster. These clusters all encompassed the entire time period. There is also a small high-rate cluster in Allentown from October 2014-September 2017. The low-rate scan revealed large clusters covering most of the state (notably including northwestern PA), with the exception of southwestern PA and urban southeastern PA. Again, there is a smaller low rate cluster containing the northern suburbs of Pittsburgh.

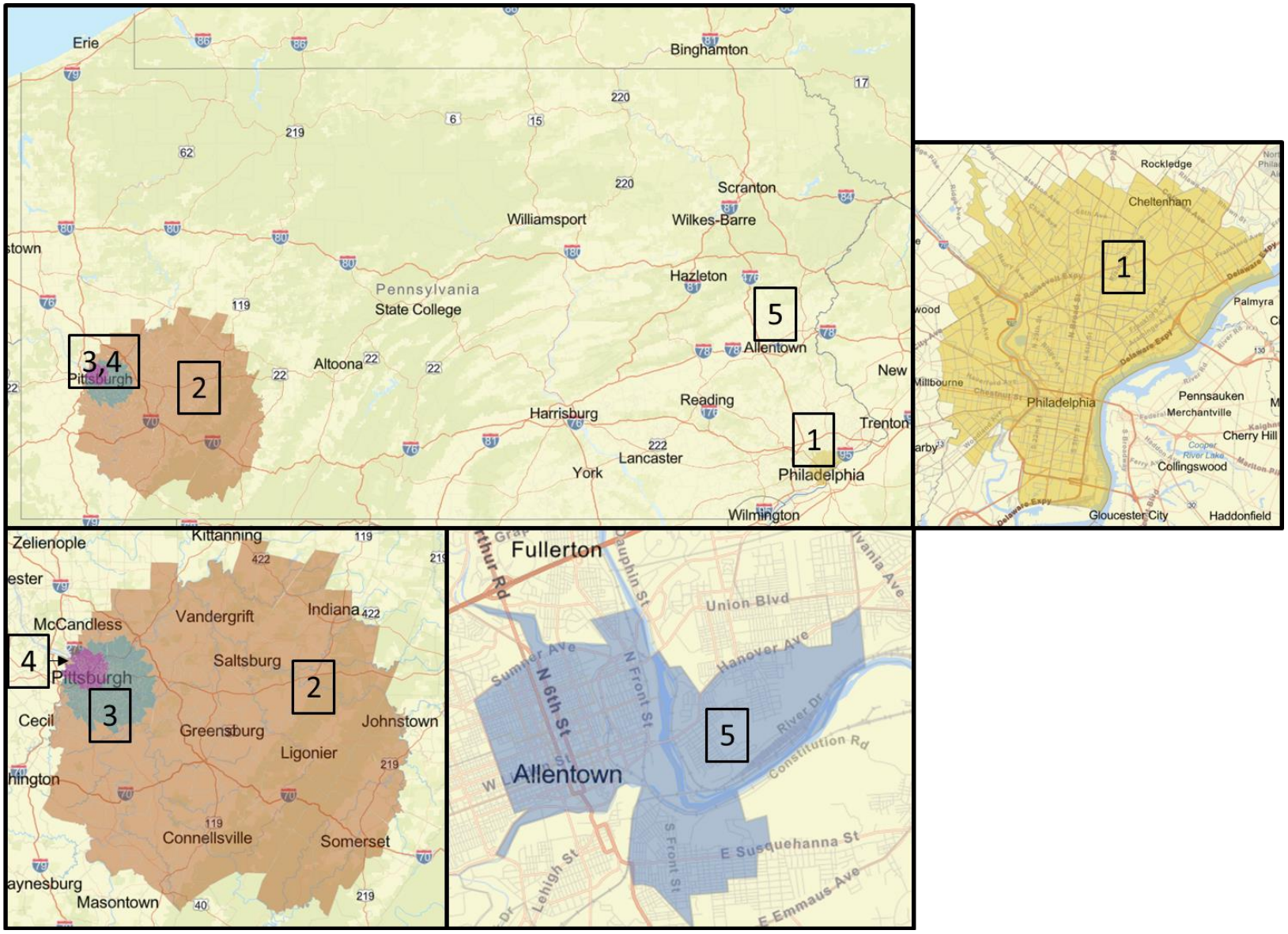

Figure 11: Age 50-64 stratified high-rate scan, adjusted for sex and temporal trends

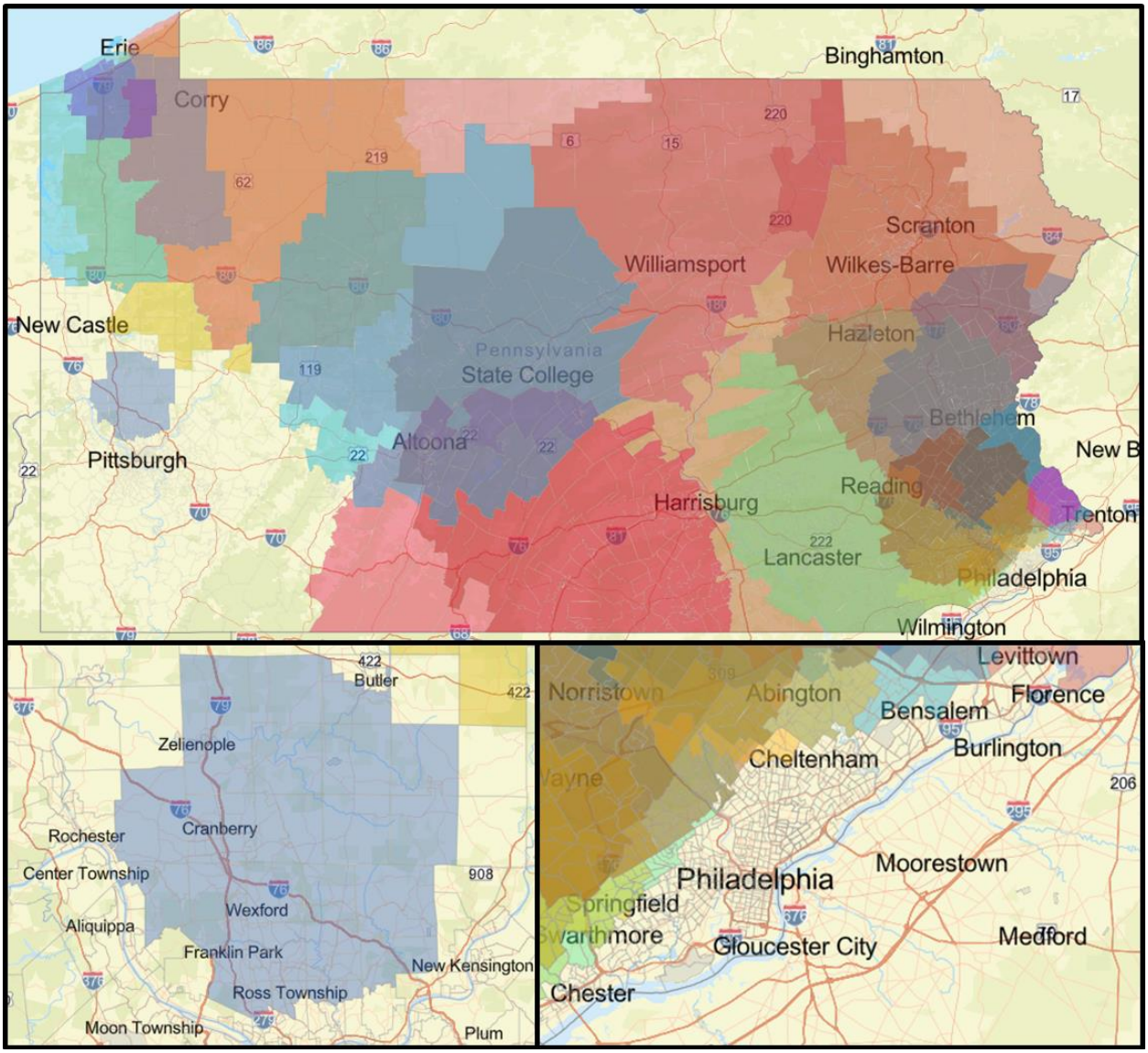

Figure 12: Age 50-64 stratified low-rate scan, adjusted for sex and temporal trends

3.2.9 Age over-64 stratified scans

The results of the high-rate scan for decedents age 65 and older are shown in Figure 13 and data on each cluster is shown in Table 9. The results of the low-rate scan are shown in Figure 14; low-rate clusters are not labeled and data for these clusters are not shown.

Table 9: Cluster data from age over-64 stratified, time-adjusted, sex-adjusted high rate scan

Cluster	Time frame	Population (over-64)	Observed Cases	Expected Cases	Observed/ Expected	Annual Cases per 100,000
1	2012/1/1-2017/12/31	109246	89	21.04	4.23	13.7
2	2015/1/1-2017/12/31	95417	39	10.79	3.62	11.7
3	2015/1/1-2017/9/30	13612	14	1.38	10.15	32.8

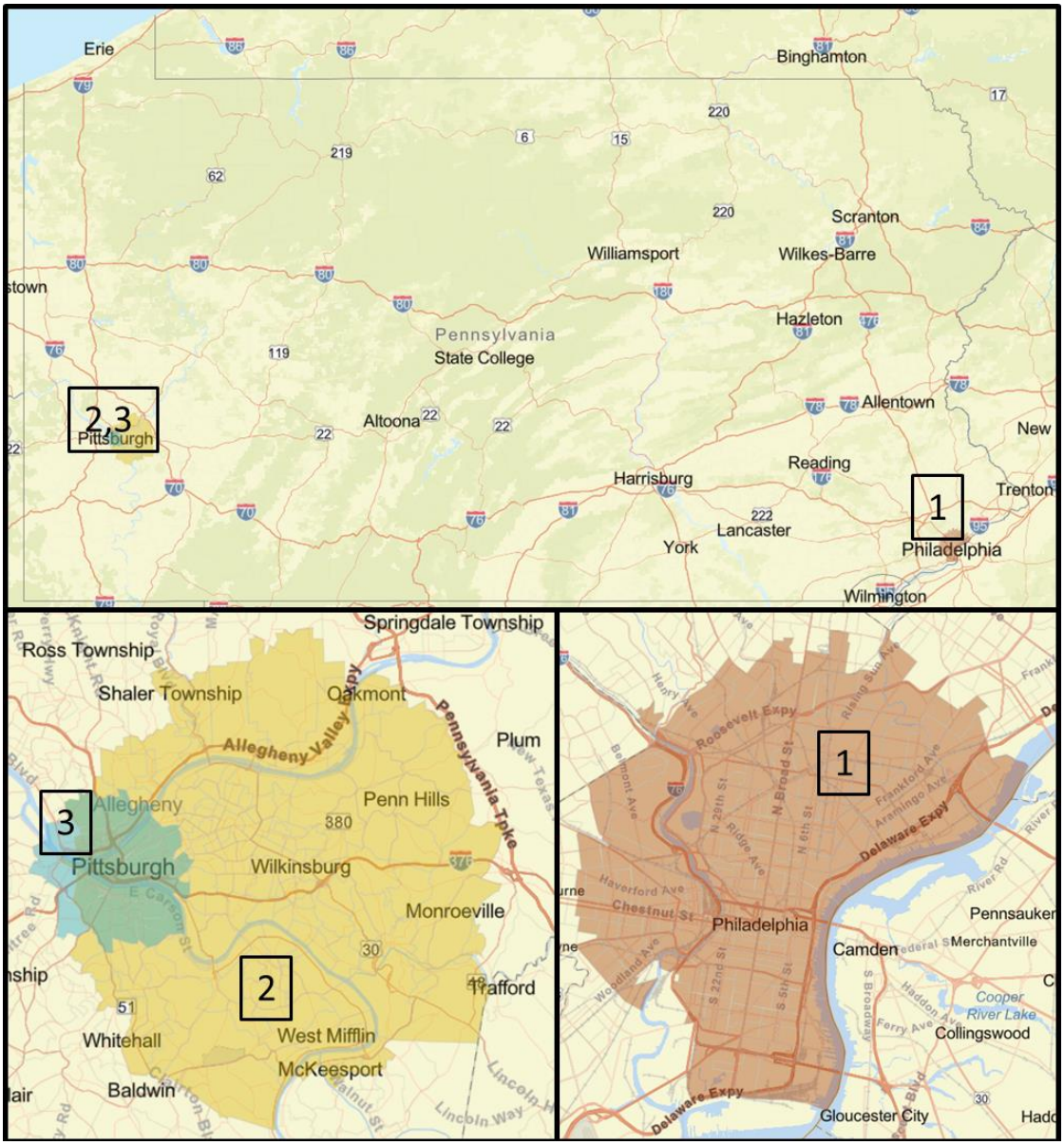

Figure 13: Age over-64 stratified high-rate scan, adjusted for sex and temporal trends

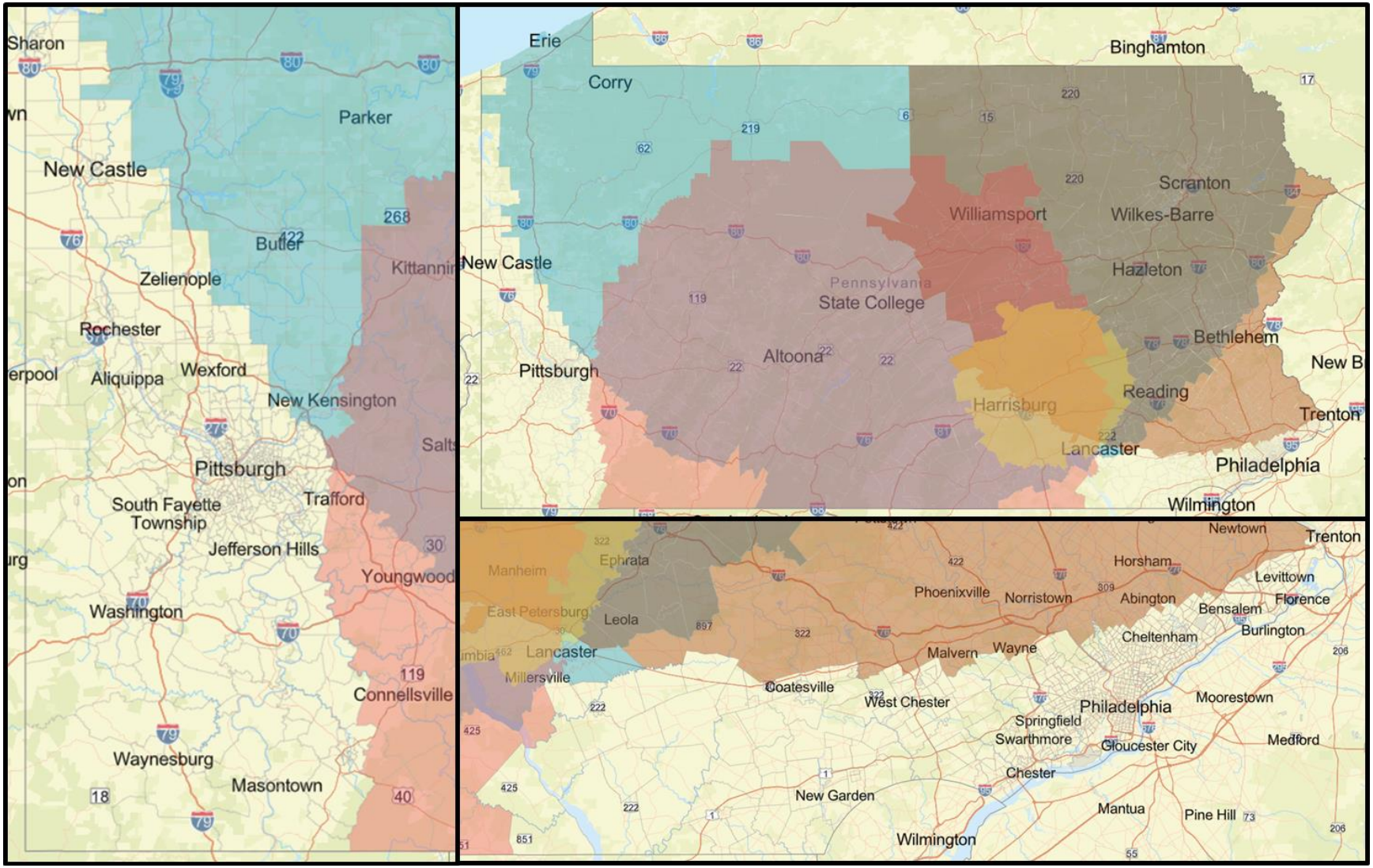

Figure 14: Age over-64 stratified low-rate scan, adjusted for sex and temporal trends

The most significant cluster covers the urban core of Philadelphia during the entire time period, while the second and third-most significant clusters cover urban areas of Pittsburgh during 2015-2017. The low-rate clusters cover most of the state, with the exception of western PA, Philadelphia, and suburban and rural areas of southeastern PA. Interestingly, the latter region is covered by low rate clusters when analyzing all other age groups, as well as the population as a whole.

3.3 Getis-Ord hot spot analysis

For the Getis-Ord hot spot analysis, a map was created for each calendar quarter during the study duration, with census tracts colored according to Gi* statistic. Unshaded areas did not have significant Gi* scores. The three shades of red represent hot spots that are significant at the 0.10, 0.05, and 0.01 levels, with 0.01 being the darkest. The three shades of blue represent cold spots that are significant at the 0.10, 0.05, and 0.01 levels, again with 0.01 being the darkest. There were far more hot spots than cold spots. I combined the results of the quarterly hot spot analysis into an animation. This animation is available as supplemental file Quarterly_hot_spot_whole_state. The supplemental file Quarterly_hot_spot_SWPA is a version of the same animation zoomed in on southwestern PA, and the supplemental file Quarterly_hot_spot_SEPA is a version zoomed in on southeastern PA. This analysis revealed a large number of relatively small hot/cold spots, and therefore the hot/cold spots are not characterized exhaustively here.

The southwestern PA animation shows that there were consistently hot spots throughout southwestern PA, including urban areas around Pittsburgh; less affluent suburbs along the rivers;

outlying towns such as Butler, Indiana, Greensburg, Latrobe, New Stanton, Johnstown, and Uniontown; and rural areas. Interestingly, many different parts of southwestern PA were hot spots at least once, but no single area emerged as a clear, consistent hot spot throughout the entire duration.

Conversely, the southeastern PA animation shows consistent hot spots in neighborhoods of near-northeastern Philadelphia, including Kensington, Port Richmond, Bridesburg, and Frankford. Additionally, the animation shows more sporadic hot spots in part of northcentral, south, and west Philadelphia, and in suburban areas around Chester, Levittown, King of Prussia, Lansdale, Quakertown, Doylestown, and Pottstown. Cold spots are occasionally visible in affluent areas of Chester County, Delaware County, and Montgomery County.

As for the whole state, the animation shows sporadic hot spots in other cities, towns, and rural areas throughout PA. One region of interest is northeastern PA, with hot spots in Scranton, Wilkes-Barre, Hazleton, and the surrounding rural areas. There are also occasional hot spots in the cities of Erie, Harrisburg, and Allentown. However, it is important to note that none of areas experienced hot spots with the same consistency and density as southwestern PA and urban southeastern PA.

4.0 Discussion

Calculating the rates of fatal overdose by sex, age group, and race from 2012-2017 in PA (see Table 1) revealed that men had a higher rate of fatal overdose compared to women (32 vs. 14 per 100,000), and that people aged 20-34 and 35-49 had higher rates of fatal overdose (both ~ 42 per 100,000) compared to adults in other age groups. Black and white residents had similar rates of fatal overdose (23 vs. 24 per 100,000), while residents of other races had a lower rate (10 per 100,000). These findings are consistent with results from Balmert *et al.* (2016) for PA during 2014. That study showed that men had about twice the rate of fatal overdose as women (39 vs. 20 per 100,000); that there were higher rates for those age 25-34 (40 per 100,000), age 35-44 (37 per 100,000), and age 45-54 (34 per 100,000) than for age 15-24 (16 per 100,000) or age 55-64 (20 per 100,000); and that black and white residents had similar rates (27 per 100,000 vs. 31 per 100,000, respectively). As noted in section 3.1, nearly 70% of fatal overdoses during this time had male decedents, and over 70% of fatal overdoses had decedents age 20-49. Nearly 50% of all fatal overdoses had a male decedent age 20-49. Therefore, it is important to keep in mind that the dynamics of fatal overdose in the general population are driven primarily by these groups.

As discussed in section 3.1, the raw data showed that the number of fatal overdoses significantly increased in 2016 and 2017. While fatal overdoses increased by 21.4% in the U.S. as a whole from 2015 to 2016, the 82% increase in PA during this same period far outpaced this (Scholl et al., 2018). The initial high-rate SaTScan analysis (without adjustment for temporal trends) showed that the most likely cluster for PA during 2012-2017 contained the entire state during 2016-2017. The increase in fatal overdoses in the state as a whole was far more statistically significant than the increase in any single geographical cluster. Therefore, the nonparametric

temporal trends adjustment was an important tool for the other SaTScan analyses in order to see geospatial trends while adjusting for the significant temporal trend.

The unstratified high-rate analysis (with adjustments for age, sex, and temporal trends) revealed nine significant clusters (see Figure 1). However, several of these are overlapping clusters during the same time window: clusters #2 and 3 in Philadelphia, clusters #4 and 5 in Pottstown, and clusters #7 and 9 in Allentown. As discussed above, each of these examples can be viewed as a single combined cluster. Thus, this analysis highlighted six key areas: southwestern PA (including urban, suburban, and rural areas); north/northeast Philadelphia; central Pottstown; central York; central Allentown; and the mixed urban/rural area around Hazleton. The clusters in southwestern PA and Philadelphia were much larger in population and were ranked as more significant than the other clusters, which is not surprising based on the influence of cluster population in the likelihood formula (see Methods). Nearly all of these clusters covered the entire time range, 2012-2017 (the cluster in the Hazleton area covered ~2014-2016). These high-rate cluster findings are consistent with county-level results from other studies of overdose in PA during this time period (Balmert et al., 2016; Burke & Buchanich, 2018; Report, 2018). Therefore, it is important to investigate why individuals in these areas experienced a persistently elevated risk of fatal overdose than people in other parts of the state, after considering state-wide trends. Several studies have focused on the role of opioids in these dynamics, including prescription opioids, traditional street drugs like heroin, and new synthetic opioids like fentanyl (Balmert et al., 2016; Burke & Buchanich, 2018; Creppage et al., 2018; Dwyer et al., 2018; Frazier et al., 2017; Jalal et al., 2018; Report, 2018). Drug type was not considered here, but the spatial patterns of fatal overdose for different drugs in PA would be a good topic for future research.

Many of the clusters shown by the sex- and age-stratified high-rate analyses were highly similar to the clusters revealed by the unstratified high-rate analyses. In the male high-rate analysis (see Figure 3), the three significant clusters covered southwestern PA and north/northeast Philadelphia, and were near-identical to clusters from the unstratified analysis. However, the smaller clusters from the unstratified high-rate analysis were not significant in the males-only analysis. The female high-rate analysis (see Figure 5) also revealed three clusters covering southwestern PA and north/northeast Philadelphia; interestingly, this analysis also shows a large cluster spanning from Pittsburgh area to rural northwestern PA, and a cluster covering central/south/west Philadelphia, both areas which were not covered in the unstratified high-rate analysis. An important question is why these areas were significant high-risk clusters for women but not for men.

The age 20-34 high-rate analysis (see Figure 7) showed the familiar clusters in southwestern PA and Philadelphia, a cluster around Hazleton and Wilkes-Barre, and several novel clusters in suburban areas of southeastern PA. It is worth investigating why young adults in suburban southeastern PA experienced significant elevated risk of fatal overdose while other age groups in that area did not. The age 35-49 high-rate analysis (see Figure 9) revealed familiar clusters in southwestern PA, north/northeast Philadelphia, and Allentown, as well as clusters in central/south/west Philadelphia and Scranton. The age 50-64 high-rate analysis (see Figure 11) revealed the familiar cluster in southwestern PA and Allentown, a cluster containing most of Philadelphia, and some smaller clusters around Pittsburgh which are contained by the larger southwestern PA cluster. The over 64 high-rate analysis (see Figure 13) showed clusters containing much of Philadelphia and urban areas around Pittsburgh; notably, a large cluster in southwestern PA does not appear in this analysis. Thus, elderly residents of small-town and rural southwestern

PA do not appear to face the same elevated risk of fatal overdose as residents in other age groups. However, elderly rural PA residents may have a greater chance of having fatal overdoses go undetected due to other comorbidities and the practices of rural coroners and medical examiners. It is also important to investigate how these same discrepancies may impact the data for other demographic groups.

The low-rate analyses are more difficult to interpret at first glance, as they contain many large overlapping clusters. However, as discussed above, these overlapping clusters can be interpreted as a single large cluster, as long as they cover the same time period. Thus, the main finding from the low-rate analyses (see Figures 2, 4, 6, 8, 10, 12, 14) is that most of the state is covered by a low-rate cluster for the entire period 2012-2017, except for southwestern PA and an urbanized strip of southeastern PA. However, in most low-rate analyses there were small low-rate clusters in these uncovered areas, in central urban neighborhoods of Philadelphia and Pittsburgh, and in the northern suburbs of Pittsburgh. The exact boundaries of these clusters vary between the different low-rate analyses, but these differences are not explored in detail here, as the high-rate analyses are more informative.

It is also important to note how the SaTScan analysis parameters influenced the results. For example, I restricted SaTScan to drawing circular cluster windows, even though actual geographical clusters may take other shapes. Thus, clusters may contain low-rate census tracts (especially at the periphery) that were only included due to the circular window shape and proximity to high-rate areas. Several of the high-rate analyses revealed overlapping clusters during the same time window (such as clusters #2 and 3 in Figure 1), which can effectively be considered as a single combined cluster. The “No centers in less likely clusters” option prevented SaTScan from plotting any clusters which contained the center of another already-plotted cluster with higher

likelihood. Thus, high-likelihood clusters could not be contained by larger, less likely clusters, but high-likelihood clusters could overlap or contain less likely clusters, as long as the center of the high-likelihood cluster was not covered. Clusters were also limited to contain no more than 30% of the state population. This did not impact the high-rate scans, as the largest clusters (in southwestern PA) contained 20-25% of the state population. However, in the low-rate scans, much of the state was covered by a mesh of large clusters, as the restriction prevented SaTScan from covering this entire area with one massive cluster. Additionally, I used adjustments for sex and age group, so that SaTScan considered sex and age demographics while calculating the expected number of cases for a prospective cluster. Therefore, significant clusters should represent unique local factors, rather than merely an area with a higher proportion of high-risk residents (such as men age 20-49). In the age-stratified analysis, adjustment for sex was still performed, and vice-versa. Finally, adjustment for multiple clusters was not used during Monte Carlo hypothesis testing. Secondary clusters were assessed as if they were the only cluster present, and therefore calculated p-values were conservative as other high-likelihood clusters were included in the area outside of the assessed cluster. Combined with a significance level of 0.01, this meant that cluster significance was measured quite conservatively. These criteria excluded many smaller clusters which may still be worth investigating.

For the Getis-Ord analysis, it is important to note that a false discovery rate correction was not applied. Since a large number of Gi* scores were calculated, it is guaranteed that some of these scores will falsely indicate a significant result. Thus, the reader should not view the Getis-Ord analysis through the lens of statistical significant. The hot spots should be viewed as areas with higher-than-average rates of fatal overdose, but unclear statistical significance.

As discussed in the Results, the animations revealed hot spots all over the state. Interesting patterns emerged in Philadelphia and southwestern PA, which were the key regions identified by SaTScan analysis. There were persistent hot spots in north/northeast Philadelphia, while in southwestern PA, hot spots appeared at a high density, but not necessarily in the same location from quarter to quarter. Therefore, the dynamics of fatal overdose in the Philadelphia area were driven primarily by a few neighborhoods (Kensington, Port Richmond, Bridesburg, Frankford) with occasional spikes in the nearby urban and suburban areas, while in the southwestern PA, the dynamics of fatal overdose were driven by emerging hot spots all over the region. Exploring this difference is an important question for future work. The hot spot analysis also revealed interesting dynamics in other parts of the state, both around the smaller clusters revealed by SaTScan (Pottstown, Allentown, York, Hazleton) and in other areas. These dynamics are not explored in detail here, but animations are available as a resource to those who are interested (see supplementary files). Note that due to the time and difficulty involved, I did not perform Getis-Ord analysis on any sex- or age-stratified populations, only on the unstratified population. Therefore, stratified Getis-Ord analysis represents another potential avenue for future investigation.

Overall, the novel census tract-level SaTScan and Getis-Ord analyses add more context to previous findings on fatal overdose in PA. The high-rate SaTScan analysis for the whole population revealed that high rates of fatal overdose in Philadelphia and southwestern PA (as shown by county-level analysis) were driven by statistically significant clusters; furthermore, this analysis revealed small statistically significant clusters in Pottstown, York, Allentown, and Hazleton that are less obvious from county-level analysis. Low-rate SaTScan analysis for the whole population revealed that most of the state outside of southwestern PA and the Philadelphia

area was covered by large statistically significant low-rate clusters, despite the fact that these regions had small localized areas with high rates of overdose. SaTScan analyses stratified by age group and sex confirmed the presence of significant high-rate clusters southwestern PA and Philadelphia across all of these demographics; however, there were some interesting demographic-specific clusters, including central/south/west Philadelphia and rural northwestern PA for woman; suburban southeastern PA and Hazleton for those age 20-34; and Scranton for those age 35-49.

The Getis-Ord animation revealed census tracts with high rates of overdose throughout the entire state, including many areas that were not covered by any high-rate clusters in SaTScan analysis. In southwestern PA, hot spots occurred at a high density throughout the entire region, but not consistently in any one location. Conversely, in Philadelphia, a few census tracts in the neighborhoods of Kensington, Port Richmond, Bridesburg, and Frankford were consistently covered throughout the entire period. In contrast to the SaTScan analysis, the Getis-Ord hot spot analysis was not useful in testing the statistical significance of specific overdose clusters. Rather, the unstratified Getis-Ord analysis showed local variations and patterns in fatal overdose rates at the census tract level. Thus, the animation would be a great resource for someone who is interested in relative rates and temporal patterns of overdose in different neighborhoods around a certain city or region in PA during this time period.

Bibliography

Balmert, L. C., Buchanich, J. M., Pringle, J. L., Williams, K. E., Burke, D. S., & Marsh, G. M. (2016). Patterns and trends in accidental poisoning deaths: Pennsylvania's experience 1979-2014. *PLoS ONE*, *11*(3), e0151655. https://doi.org/10.1371/journal.pone.0151655

Buchanich, J. M., Balmert, L. C., Pringle, J. L., Williams, K. E., Burke, D. S., & Marsh, G. M. (2016). Patterns and trends in accidental poisoning death rates in the US, 1979–2014. *Preventive Medicine*, *89*, 317–323. https://doi.org/10.1016/j.ypmed.2016.04.007

Bureau, U. S. C. (2017). *American Community Survey 5-year estimates*. American FactFinder.

Burke, D. S., & Buchanich, J. M. (2018). Drug Overdoses in Pennsylvania: Measuring, Tracking, and Forecasting the Epidemic. *Commonwealth*, *20*(2–3). https://doi.org/10.15367/com.v20i2-3.186

Centers for Disease Control and Prevention. (2018). 2018 Annual Surveillance Drug-Related Risks and Outcomes- United States. *Surveillance Special Report*, 91. https://www.cdc.gov/

Creppage, K. E., Yohannanms, J., Williams, K., Buchanich, J. M., Songer, T. J., Wisniewski, S. R., & Fabio, A. (2018). The rapid escalation of fentanyl in illicit drug evidence in allegheny county, pennsylvania, 2010-2016. *Public Health Reports*, *133*(2), 142–146. https://doi.org/10.1177/0033354917753119

Dwyer, J. B., Janssen, J., Luckasevic, T. M., & Williams, K. E. (2018). Report of Increasing Overdose Deaths that include Acetyl Fentanyl in Multiple Counties of the Southwestern Region of the Commonwealth of Pennsylvania in 2015–2016. *Journal of Forensic Sciences*, *63*(1), 195–200. https://doi.org/10.1111/1556-4029.13517

Frazier, W., Cochran, G., Lo-Ciganic, W. H., Gellad, W. F., Gordon, A. J., Chang, C. C. H., & Donohue, J. M. (2017). Medication-assisted treatment and opioid use before and after overdose in Pennsylvania Medicaid. In *JAMA - Journal of the American Medical Association* (Vol. 318, Issue 8, pp. 750–752). American Medical Association. https://doi.org/10.1001/jama.2017.7818

Getis, A., & Ord, J. K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x

Jalal, H., Buchanich, J. M., Roberts, M. S., Balmert, L. C., Zhang, K., & Burke, D. S. (2018). Changing dynamics of the drug overdose epidemic in the United States from 1979 through 2016. *Science*, *361*(6408), eaau1184. https://doi.org/10.1126/science.aau1184

Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics - Theory and Methods*, *26*(6), 1481–1496. https://doi.org/10.1080/03610929708831995

Report, J. I. (2018). The opioid threat in pennsylvania. *University of Pittsburgh: Pitt Pharmacy*, *September*. https://www.dea.gov/sites/default/files/2018-10/PA Opioid Report Final FINAL.pdf#page=39&zoom=auto,-73,794

Scholl, L., Seth, P., Kariisa, M., Wilson, N., & Baldwin, G. (2018). Drug and Opioid-Involved Overdose Deaths - United States, 2013-2017. *MMWR. Morbidity and Mortality Weekly Report*, *67*(5152), 1419–1427. https://doi.org/10.15585/mmwr.mm675152e1

University of Pittsburgh - MOIRA. (n.d.). Retrieved April 15, 2020, from https://moira.pitt.edu/