Using Synthetic Populations to Understand Geospatial Patterns in Opioid Related Overdose and Predicted Opioid Misuse

Savannah Bates[1], Vasiliy Leonenko[2], James Rineer[3], and Georgiy Bobashev[4]

[1] *Department of Biomathematics, North Carolina State University, svbates@ncsu.edu*
[2] *High-Performance Computing Department, ITMO University, vnleonenko@corp.ifmo.ru*
[3] *RTI International, jrin@rti.org*
[4] *RTI International, bobashev@rti.org*

1 Introduction and Background

In 2016, opioids were responsible for 58.2% of all unintentional drug overdose deaths [4] and Ohio ranked 2nd worst in the nation for overdose deaths [3]. As the number of deaths continues to rise, it is obvious that the nation is facing an opioid epidemic and Cincinnati is at the heart. "'We continue to be horrified by the tragedy that this epidemic has brought to our community,' Nan Franks, CEO of Addiction Services Council, said. But she added, 'We cannot let ourselves be disheartened. We have to stay committed to eradicating this epidemic'" [3]. Opioid abuse differs from other narcotics in the observed behavioral patterns of the misusers. Tom Synan, co-chair of the interdiction committee of the Hamilton County Heroin Coalition Task Force, stated "While people who bought crack or meth would buy their crack or meth, then drive home and do it, here, because of the physical opiate withdrawal, it is so powerful that often we find the person overdosing near the location where they bought the drug" [5]. In order to halt this epidemic, understanding the different patterns between misuse and overdose is crucial as no clear evidence-based link has been established in the context of demographic and geographic contextual factors. Through this work, we use a synthetic population to predict misuse patterns and establish a relationship between misuse of opioids and overdose in order to suggest intervention strategies aimed at preventing opioid induced overdose events. [1]

2 Objective

The aim of this study is to identify and map areas with extreme ratios of opioid overdoses to model-predicted drug misuse. Opioid overdoses are measured from EMS calls and drug misuse is estimated from the publicly available National Survey on Drug Use and Health (NSDUH). A geographically explicit model that links overdose data from Cincinnati EMS, the RTI-developed synthetic population, and reports of opioid misuse from the NSDUH is used to understand patterns in misuse and overdose in Cincinnati.

[1] The data on overdoses has been perturbed to prevent exact identification of the venue, so all references to specific geographic locations are only suggestive and cannot be used to draw definitive conclusions.

3 Methods

The link between misuse of opioids and overdose events resulting in EMS calls will be established through the use of synthetic populations which allow one to link multiple datasets without violating privacy. Synthetic populations are representations of every household and person in a population. They are produced in a dataset with each individual's coordinates and characteristics. A synthetic population could be viewed as a "scrambled" census. Specifically, at the aggregate level, demographics match the ones in the census; however, at the household and individual level, the data are drawn from complex multivariate distributions to match Public Use Microdata. Previously, RTI created nationwide synthetic populations for the United States [1, 8] and internationally. In sections 3.1–3.6, we outline our implementation of the synthetic population for use with NSDUH and Cincinnati data sets.

3.1 Populating synthetic individuals with the necessary individual and contextual characteristics. See Table 1.

File	Contents
`[prefix]_synth_households.txt`	Contains the location and descriptive attributes for each household. Household records in the `synth_households.txt` file link to individual person records in the `synth_people.txt` table.
`[prefix]_synth_people.txt`	Contains a record for each person, along with his or her age, race, and sex. These synthetic person records link to the `synth_households.txt` file (via the `sp_hh_id` field) and/or to the U.S. Census Public Use Microdata Sample (PUMS) attributes from `pums_p.txt` (via the `serialno` field).
`[prefix]_synth_pums_p.txt`	Contains complete PUMS person records from the original PUMS 5% data. Links to the `[prefix]_synth_people.txt` file the `serialno` field.

Table 1: Table containing synthetic population data used in Section 3.1. Note that the `[prefix]` for Hamilton county Ohio is `2010_ver1_39061`.

3.2 Misuse model. An opioid misuse model was developed from a nationally representative NSDUH dataset and is given by

$$\mathrm{Logit}(P_{misuse}) = b_0 + b_1X_1 + b_2X_2 + \cdots + b_9X_9,$$

where $X_1, \ldots, X_4$; X_5; X_6; and X_7, X_8, X_9 are categorical variables for age, sex, high school education, and race, respectively. See Table 2 for parameter estimates and bin descriptions. Using the model, we assign a probability of opioid misuse based on the NSDUH data, together with other characteristics including treatment facilities from the Substance Abuse and Mental Heath Services Administration (SAMHSA) database.

3.3 Preparing EMS data. The Cincinnati EMS data was downloaded on April 29, 2018. Since EMS records contained in this data set represent all events starting from January 2, 2015,

Parameter	Bin	Estimate
b_0	(Intercept)	-2.11211
b_1	Age 26-34 (Reference: Age 18-25)	-0.20780
b_2	Age 35-49 (Reference: Age 18-25)	-0.59634
b_3	Age 50-64 (Reference: Age 18-25)	-1.03172
b_4	Age 65+ (Reference: Age 18-25)	-2.09052
b_5	Female (Reference: Male)	-0.28753
b_6	Non-graduate (Reference: Graduate)	-0.29243
b_7	White (Reference: Other)	0.25812
b_8	Black (Reference: Other)	-0.05229
b_9	Hispanic (Reference: Other)	-0.02074

Table 2: The statistical misuse model parameters, descriptions, and estimates.

the data needs to be sorted to only contain overdose incidences. Only data tagged as a heroin overdose, overdose/poisoning involving some type of narcotic, or drug induced convulsions is used. Of the remaining records, any data points without latitude and longitude information are ignored. Finally, one data point had latitude and longitude coordinates far outside the bounds of Cincinnati and is removed.

3.4 Obtaining aggregate cell level data from synthetic individuals and EMS calls. First, we prepare the input for the statistical model, matching the data for households (latitude, longitude, size) with individual data for Hamilton county. The resulting table, with every record representing a set of characteristics of one synthetic individual, is saved in `.csv` format. We then apply the statistical model to the input table, calculating for each record the probability an individual is an opioid misuser. The opioid misuse status (0 or 1) for each individual is generated by means of Monte Carlo methods, based on the opioid misuse probability. Next, we convert the coordinates of an individual's location from degrees to meters using Mercator projection. After this we form a grid with cell size 250m $\times$ 250m, defining its bounds with maximum and minimum coordinates of the synthetic individuals. Using a polygon containing the Cincinnati border, which was obtained from open sources, we exclude the grid cells which do not lie within Cincinnati city limits. Then we calculate the overall number of dwellers and opioid misusers for each cell of the grid. Finally, using the EMS calls dataset for Cincinnati, we calculate the overall number of EMS calls for each cell of the grid. This algorithm was implemented as a scripts collection written in Python 3.6 with the libraries `numpy`, `matplotlib`, and `pandas`.

3.5 Categorizing data. Each cell can be defined by three binary variables that indicate the presence or absence of EMS calls, dwellers, and misusers in the cell. From all the possible combinations of those indicators, five have meaningful interpretations and are used to categorize Cincinnati's areas (see Table 3 and Figure 1). Each group has a distinct implication for intervention.

There is a category of cells which cannot be unambiguously interpreted based on our assumptions. These cells, which we call "uninterpreted," are characterized by a sufficient number of both EMS calls and dwellers; however, the predicted number of misusers assessed by the model is negligible. This combination could potentially indicate the underestimation of misusers by the model, still, other explanations may be valid. For instance, migrational drug usage in residential areas (shooting galleries, drug usage in a friends' community) could explain this pattern.

Group	Misusers	Dwellers	EMS Calls
Rural drug use	−	−	+
No misusers/overdoses	−	+	−
Outside home users	+	+	−
Uninterpreted	−	+	+
Trio results	+	+	+
Empty	−	−	−
Impossible	+	−	+
Impossible	+	−	−

Table 3: The five groups of data and the type of data they contain. If +, there is a non-zero value for that cell. If −, there is zero value in that cell. Impossible groups are labeled as such due to the inability to have misusers in a cell without any dwellers.

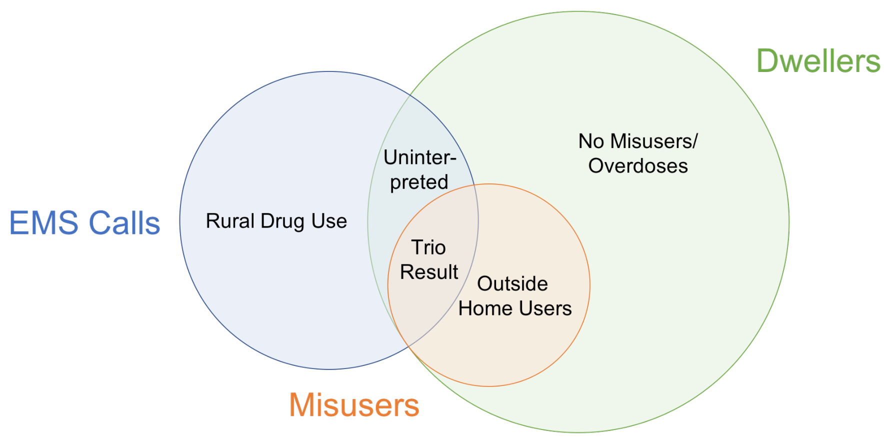

Figure 1: Venn diagram of cell data types.

3.6 Calculating ratios of interest. In order to understand the relationship between the numbers of misusers and the number of EMS calls, we choose to consider the following ratios

$$r_1 = \frac{c+1}{m+1} \quad \text{and} \quad r_2 = \frac{m+1}{d+1}$$

where c is the number of calls in a cell, m is the model predicted number of misusers in a cell, and d is the number of dwellers in a cell based on the synthetic population. These quantities represent modified ratios of calls per misuser from January 2015 and misusers per dweller, respectively. By adding ones to the numerator and denominator we're able to avoid a divide by zero error, and although it provides a small skew in the data, its consistent application across all cells leaves the results in Section 4 and their interpretations unhindered. In this research we assume that the number of EMS calls during the observation time and the number of misusers are of the same order of magnitude and, in absence of external factors, have limited variation on the majority of the territory. We use the ratio r_1 to understand which cells have large differences in the orders of magnitude compared to other cells. The ratio r_2 is used to distinguish between cells that have interesting differences and cells that we believe are uninterpreted by the model.

4 Results

In this section, we focus on results yielded from comparing the Cincinnati EMS data to demographics, the statistical misuse model, and different geographic features of Cincinnati. We do this to obtain visualizations of drug use patterns and understand exceptional values of r_1.

4.1 Visualizing data categorization and the ratios of interest. Using cell level data and the categorization set up in Section 3.5, we plot cells according to their classification in Figure 2.

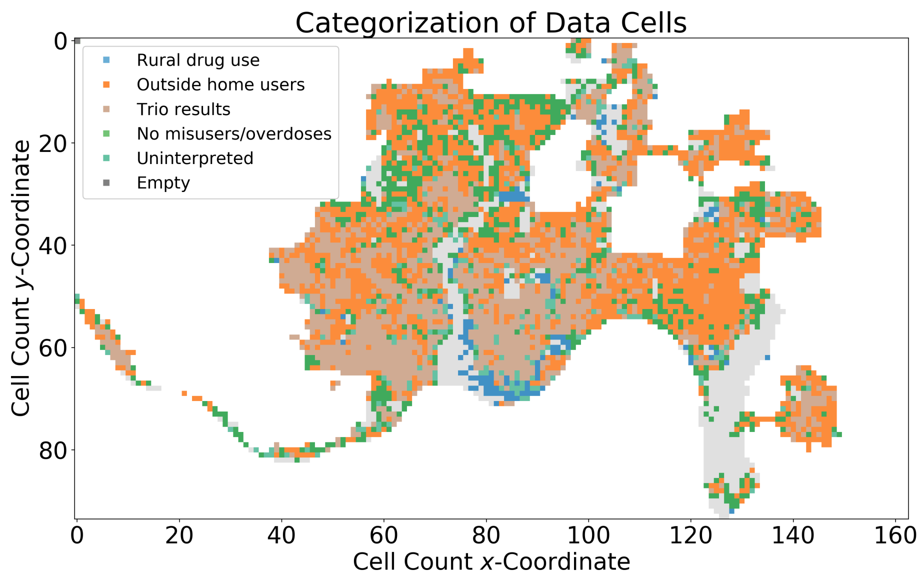

Figure 2: Categorical depiction of cell level data using the zero threshold. White cells are those outside the Cincinnati city limits.

While the strict non-zero threshold value is useful for visualizing the categorized data, in order to better understand the nature of the relationship between our misuse model and EMS calls, we want to consider a more flexible statistic. In a sense, we allow ourselves to blur the borders of the Venn diagram in Figure 1 and consider categories on a gradient. Thus, the non-zero thresholds for c, m and d are replaced by ratio thresholds for r_1 and r_2 (see Table 4 for examples of categorization according to both thresholds). Particularly, uninterpreted points, which represent cells with a large population and many EMS calls, but few to zero misusers according to our statistical model, are found to have $r_1 > 5$ and $r_2 < 0.1$. These cutoff values were chosen based on a careful examination of the data to ensure data points were classified correctly. Figure 3 provides a heat map of r_1 (from Section 3.6) which represents a modified ratio of calls per misuser from January 2015 to April 29, 2018 with uninterpreted cells being shown in gray. From this figure, we see that there are cells with exceptionally high (red cells) or exceptionally low (blue cells) ratios of EMS calls to misusers compared to the average r_1.

Misusers	Dwellers	EMS Calls	Non-zero Threshold	Ratio Thresholds
0	73	13	True	True
0	2	70	True	False
2	64	56	False	True
20	173	1	False	False

Table 4: A sample of representative cells and their counts for classifying data as a uninterpreted.

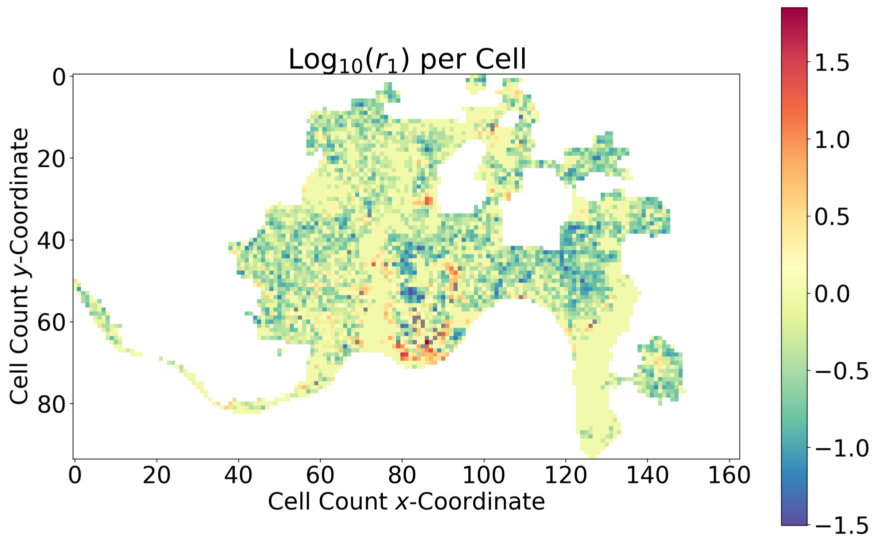

Figure 3: Heat map of r_1 – the modified ratio of calls per misuser from January 2015 to April 29, 2018. The r_1 values for cells are plotted on a log scale. Cells in red have a higher number of calls per misuser and conversely cells in blue have lower number of calls per misuser. Cells which we determined were uninterpreted by the misuse model based on r_1 and r_2 are shaded in grey.

4.2 Analyzing contents of exceptional cells. In Figure 3, we see that there is a number of cells with values of r_1 notably higher or lower than the average. To determine if there are observed geographic similarities in the contents of cells with comparable r_1 values, we first examined cells with exceptionally low ratio values based on a scatter plot of r_1 and recorded their contents in Table 5. We noticed that these cells are almost entirely residential. This gives support for the observation that unlike other narcotics, opioid misusers often use drugs near locations of drug purchase rather than at their homes.

Next, we examined cells with exceptionally high ratio values and recorded their contents in Table 6. Note that one of the exceptional cells is classified as uninterpreted by the model and is denoted with an asterisk. We included this cell despite its classification because it belongs to a large cluster of red cells. In examining these cells, we noticed that cells with high r_1 often contain secluded commercial areas which might be preferential for drug use. These cells also often contain public structures or commercial areas, including several chains of a certain fast food restaurant. These public places could be locations where drug trade takes place. Interestingly, fast food establishments such as this have a reputation for being locations of drug trade and abuse. For

r_1	Type	Contents
0.0313	Residential	Victor St, Stratford Ave, Chichasaw St
0.0385	Residential	Ohio Ave
0.0417	Residential	Senator Pl
0.0417	Residential	Hardisty Ave and Delta Ave
0.0435	Residential	Torrence Ln, a possible new construction or damaged home
0.0435	Residential	Strand Ln, an elementary school

Table 5: Cells with the lowest r_1 values and their contents.

example, in January 2017, a mother and father overdosed on heroin at a fast food restaurant just outside Cincinnati [7]. In February 2018, a man abducted an 84 year old woman and forced her to drive him to several locations including a fast food restaurant where the man bought something from two other men and returned to the car and injected himself with a drug [6]. While we only discovered a pattern with this certain fast food restaurant specifically, we doubt this behavior is unique to this establishment. However, it is clear that these public places could be locations where drug trade takes place and could be key potential locations for drug intervention. In Figure 4 we plot r_1 on a map of Cincinnati and show the locations of this specific fast food restaurant in the Cincinnati area. We chose to include two of these restaurants that are outside Cincinnati but are included due to their proximity to exceptional cells.

r_1	Type	Contents
71.0000	Non-residential	A public library, a parking garage, an empty building, public transportation and parking
35.0000	Non-residential	A homeless shelter, parking, shipping containers, a seemingly abandoned building
26.0000	Non-residential	An electric company, a warehouse, shipping containers, covered parking for large trucks
24.0000	Non-residential	A visitor center, a library, a parking garage, hotels, restaurants
20.0000	Non-residential	An employment agency, a gas station, a veterans center, a certain fast food restaurant
20.0000	Non-residential	Train tracks, a manufacturing company, a halfway house, a certain fast food restaurant (nearby)
20.0000*	Non-residential	A corporate office, parking garage, a credit union, a certain fast food restaurant

Table 6: Cells with the highest r_1 values and their contents. Cells marked with an asterisk are uninterpreted by the model.

4.3 Visualizing r_1 on the zoning map of Cincinnati. One can see a pattern from Tables 6 and 5 that the cells in each table have similar contents. Cells with high r_1 values seem to be in more commercial or industrial areas while cells with low r_1 values are almost entirely residential. In order to examine if this pattern holds for all cells, we compare r_1 values with a simplified zoning

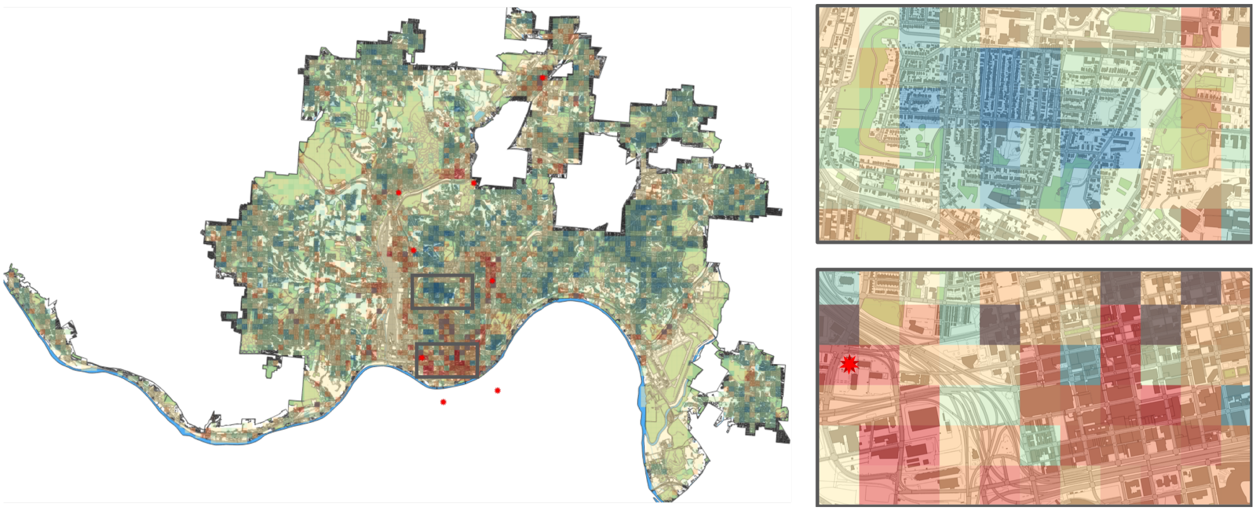

Figure 4: Heat map of $\log_{10}(r_1)$ plotted on the map of Cincinnati with uninterpreted cells shaded in gray. Regions containing the highest and lowest ratios have been highlighted. The locations of a certain fast food restaurant are shown by red stars.

map of Cincinnati. Using the original zoning codes for Cincinnati [2, 9], we classify regions as commercial, industrial areas and parks, or residential but maintain the original zoning borders (see Table 7). This result can be seen in Figure 5. The relationship between r_1 and the simplified zoning map appears strongest in more populous areas like downtown Cincinnati.

Simplified Zone	Original Zone
Commercial	Commercial Community, Commercial Neighborhood, Commercial General, Office General, Office Limited, Downtown Development District, Riverfront Commercial
Industrial & Parks	Manufacturing, Riverfront Manufacturing, Planned Development District, Parks and Recreation, Urban Mix District
Residential	Residential, Institutional-Residential, Riverfront Residential, Transect Zones

Table 7: Simplified Cincinnati zones and the originals.

5 Discussion

The main results show a dramatic variation in values of a ratio between overdose events registered as EMS calls and the model predictions of the number of misusers (r_1). This is a crucial observation. Misuse of opioids and opioid overdose deaths are hypothesized to be related; however, extremely high ratios signal that factors other than opioid misuse might be in play in these areas. Our study suggests that interventions should pay special attention to these areas where the number of overdoses is higher than could be expected from the misuse model.

Figure 5: On the left, a heat map of $\log_{10}(r_1)$ with uninterpreted cells shaded in gray is plotted with the Cincinnati zoning borders. On the right, a simplified zoning map of Cincinnati is shown. Red areas represent commercial zones, yellow represent industrial areas and parks, and blue represents residential areas.

Identification of fast food establishments in the high risk areas poses an intervention dilemma. On one hand, keeping their bathrooms locked so that drug trade or drug use is not as easily accomplished could lead to less overdoses in their vicinity but that might not solve the overdose death problem. On the other hand, providing employees with training on how to administer naloxone will be a harm-reduction intervention that saves lives but could attract drug users. Nevertheless community awareness and some basic training of staff could provide inexpensive and easy to implement practices that could have a significant impact on reducing the number of opioid related deaths in Cincinnati. Other cities suffering from a growing number of opioid overdoses could implement these policies in similar fast food chains.

Beyond the scope of fast food restaurants, Figure 5 shows a definite relationship between r_1 and the zoning code of the area. We conclude that, especially in urban areas, intervention strategies to prevent overdose events should be focused in commercial zones and their surrounding areas. Alternatively, education or outreach programs targeted at reducing the number of opioid misusers would be better placed in urban residential areas where the number of misusers is high.

The study has a number of limitations. One is that we have not accessed the levels of uncertainty and did not formally test the boundaries between different cells of interest. Although we have used inferential methods in similar studies before, this work is left for future research. Another limitation is the use of the NSDUH to estimate the probability of opioid misuse. The majority of opioid misusers in the survey misuse prescription opioids, only a fraction report using heroin and fentanyl. At the same time a large proportion of deaths could be attributed to fentanyl and "bad heroin". As more local data is available, higher accuracy could be achieved in the model.

For this study we linked a simple predictive model to the synthetic population via a handful of predictors. Cincinnati, according to our statistical model, has a rather homogeneous distribution of potential misusers. That is, the average probability for an individual to be a misuser does not vary a lot among different cells. As a result, using a ratio r_3 that measures EMS calls per dweller, i.e. $r_3 = \frac{c+1}{d+1}$, would yield similar results as the results obtained using r_1. However, this homogeneous mixing might not be witnessed in other US cities; hence, we are confident that more meaningful results will be achieved through analyzing r_1 instead of r_3, in general.

The described interpretation of a ratio between EMS calls and the assessed number of misusers also shows the necessity to account for possible dynamics in movements of opioid misusers acquiring and using the drug when planning intervention strategies. More work is needed to understand socially what intervention strategies would be best to stop the opioid epidemic as it is crucial to explore the connection between misuse and overdose, along with the factors that might influence it. The presented work is the first step in that possible direction of studies.

Acknowledgements

Savannah Bates received support from the Research Training Group in Mathematical Biology, funded by a National Science Foundation grant RTG/DMS - 1246991. Vasiliy Leonenko was supported by the Fullbright Visiting Scholar Program.

References

[1] Cajka, James C., Philip C. Cooley, and William D. Wheaton. "Attribute assignment to a synthetic population in support of agent-based disease modeling." *Methods report (RTI Press)* 19.1009 (2010): 1.

[2] "Chapter 1703 Form-Based Code." *Cincinnati Zoning Codes*, City of Cincinnati, 15 Feb. 2013, www.cincinnati-oh.gov/planning/assets/File/CFBC_1703_FBC_FinalDraft_021513_web(1).pdf.

[3] DeMio, Terry. "Ohio's Overdose Deaths Soar Again, as State Ranks 2nd Worst in the Nation." *Cincinnati.com*, Cincinnati Enquirer, 21 Dec. 2017, www.cincinnati.com/story/news/2017/12/21/ohios-overdose-deaths-soar-again-state-ranks-2nd-worst-nation/973220001/.

[4] "Fentanyl and Related Drugs Like Carfentanil as Well as Cocaine Drove Increase in Overdose Deaths." *2016 Ohio Drug Overdose Data: General Findings*, Ohio Department of Health, 2016, www.odh.ohio.gov/health/vipp/data/rxdata.aspx.

[5] "How Can Ohio Cities Step Up Their Response To Overdoses? Make The Data Public." *WVXU*, WOSU 89.7 NPR News, 10 Apr. 2018, wvxu.org/post/how-can-ohio-cities-step-their-response-overdoses-make-data-public#stream/0.

[6] Knight, Cameron, and Sarah Brookbank. "Deters: Man Who Kidnapped 84-Year-Old 'Deserves to Stay in Prison for a Long Time'." *Cincinnati.com*, Cincinnati Enquirer, 26 Feb. 2018, www.cincinnati.com/story/news/crime/crime-and-courts/2018/02/19/police-man-made-woman-buy-him-wendys-drive-him-drug-deal-knifepoint/352954002/.

[7] Rotuno-Johnson, Michelle. "Ohio Parents Overdose on Heroin at [redacted] [2] While Kids Are with Them." *WCMH*, WCMH, 8 Jan. 2017, www.nbc4i.com/news/ohio-parents-overdose-on-heroin-at-[redacted] [2]-while-kids-are-with-them/1064825372.

[8] Wheaton, William D., et al. "Synthesized population databases: A US geospatial database for agent-based models." *Methods report (RTI Press)* 2009.10 (2009): 905.

[9] "Zoning Map." *CAGISOnline*, City of Cincinnati, cagisonline.hamilton-co.org/cagisonline/index.html?zoning=cincinnati.

[2] The name of the restaurant has been redacted for confidentiality reasons.