Leveraging Statistical Methods for an Analysis of Demographic Factors of Opioid Overdose Deaths

Amna Alalawi[1], Daniel Fooks[2], Les Sztandera[3], and Sean Zakrzewski[4]

[1]DMgmt in Strategic Leadership Program, Thomas Jefferson University, Philadelphia, PA, USA
Email: Amna.Alalawi@Jefferson.edu

[2]School of Business, Thomas Jefferson University, Philadelphia, PA, USA
Email: Daniel.Fooks@Jefferson.edu

[3]Kanbar College of Design, Engineering, and Commerce, Thomas Jefferson University, Philadelphia, PA, USA
Email: Les.Sztandera@Jefferson.edu

[4]School of Business, Thomas Jefferson University, Philadelphia, PA, USA
Email: Sean.Zakrzewski@Jefferson.edu

Abstract - Deaths from drug overdose including opioid overdose have been increasing at an alarming rate, and authorities still find tackling this problem an acute challenge. This paper applies Artificial Intelligence and statistical techniques to big data to identify the demographic and socio-economic factors that have led to the increasing number of drug overdose deaths in Allegheny County, Pennsylvania, United States. Using Artificial Intelligence software, we analyzed a dataset of over 3,500 patients alongside demographic and socio-economic variables to gain detailed insights into the issue, insights that we can generalize to craft solutions to this problem in both domestic and global communities. Our findings revealed patterns ranging from possible psychological and behavioral factors and drug use on weekends, as well as a direct market supply effect on the number of deaths. These findings imply the need for authorities to offer educational workshops to individuals and their families about the dangers of the current drug epidemic and to design an effective policy for the oversight of drug market supply that includes taking firm action against violators.

Keywords – data analytics; big data; opioids; drug overdose

I. INTRODUCTION

Deaths caused by the opioid crisis have reached epidemic levels in the US, with more people dying from opioid overdose than by either motor vehicle incidents, gun violence, or HIV [2]. In 2016, more than 42,000 Americans lost their lives as the result of a drug overdose, including 613 Allegheny County residents, a rise in deaths of over 44 percent. This has largely been attributed to the presence of newer, stronger drugs such as fentanyl in US communities [7]. Local health officials have found it difficult to keep pace with the new drugs being introduced into common usage, especially the presence of fentanyl in most of the heroin sold on the streets. Fentanyl is extremely potent and can cause a fatal overdose on the first try for some users [7]. Today, drug abuse is indeed a major problem in the US. A study by the Center for Disease Control (CDC) clearly evidences the existence of this crisis, stating that in 2017, more than 47,000 people died of a drug overdose in the country [6]. It was found that more than 2,000,000 Americans live with drug addiction, such as opioid addiction [9].

This study uses Artificial Intelligence (AI) and machine learning techniques to explore a dataset containing information on 3,551 fatal incidents of opioid overdose in Allegheny County, Pennsylvania, with the aim of finding ways to minimize the consequences of the opioid crisis in communities in the US and globally.

Opioid overdose has become an epidemic in the US as well as Allegheny County [13]. In 2015, the country experienced more than 400 deaths, and from then, the trend continued to increase. Data showed that individuals affected were within the age range of 25 to 54 years old. It was found that there are plenty of opportunities that have not been taken full advantage of in terms of intervention for opioid users [13]. The authors recommended that screening for opioid and other drugs among adults involved in child welfare should be improved. Further, enhancing the ability of the direct care staff to identify opioid use and risk of overdose, as well as access to expert consultants should be increased to improve the effectiveness of the care mechanism.

In Section 2 of this paper, we discuss and highlight different statistical tools and methods used to carry out the investigation to study several demographic factors of those affected in Allegheny County. While Section 3 focuses on providing a thorough analysis of the data collected through various sources, Section 4 highlights key findings of the study as well as trends in relation to the subject matter. Section 5 provides conclusion and recommendations

 ISBN: 978-1-61208-741-2

pertaining to drug overdose in the community that can be used for prevention intervention.

II. METHODOLOGY

To gain greater insight into the social and economic factors that have increased the risk of fatal opioid overdose in Allegheny County, PA, we used multiple approaches utilizing AI and statistical software and programming languages, including IBM SPSS Analytics, IBM Watson Analytics, Microsoft Power BI, and Python, to explore the dataset, which contained information on 3,551 fatal incidents of opioid overdose in Allegheny County.

The dataset, which covers opioid overdose deaths from the year 2008 to 2017, includes fatal accidental overdoses in the country and contains information on the date of death, the time of death, the manner of death, the age, gender, and race of the decedent, the seven most prevalent drugs found in overdose victims, the zip code of the overdose incident, and the zip code of the decedent's residence [1]. To look further into this issue, we searched for other variables that may have a relationship with opioid overdose rates. These other variables can be divided into two general categories: climate and economic.

The climate variables we examined were: monthly average temperature and temperature departure from mean levels for the 2008-2017 base period. This data was retrieved using the National Oceanic and Atmospheric Administration "Climate at a Glance" tool.

The economic indicators we examined were: county unemployment rate and income inequality in the county (measured as a ratio of the mean income of the highest quintile of earners divided by the mean income of the lowest quintile of earners in the county). These two datasets were collected from the FRED Economic Data service of the St. Louis Federal Reserve Bank. The final variable examined in this category was the uninsured rate, data on which was sourced using the U.S. Census Bureau's Small Area Health Insurance Estimates program [12].

III. ANALYSIS

It has proven valuable to first observe the overdose dataset by itself to get a full understanding of the situation and to provide a baseline against which to compare the climate and economic variables examined.

The timespan of this dataset is from 01/03/08 to 12/31/17, a total of 9 years, 11 months and 28 days. The most surprising and saddening takeaway from the data is the age range of those affected by fatal drug overdose, which is an astonishing 1 to 91. The n for this dataset is 3,460.

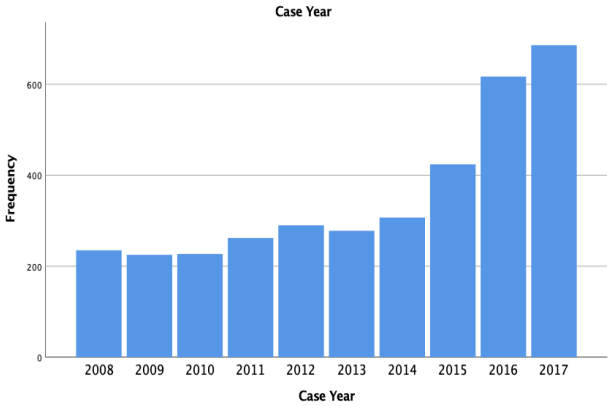

Figure 1. Overdose deaths in Allegheny Country, PA (2008-2017)

Figure 1 indicates overdose deaths over the past several years have been increasing at an exponential rate. This cannot be explained simply by an increase in population as the population of Allegheny County has fluctuated over the past decade and has, in fact, not grown at all. The graph uses data from 2010 (the earliest year in which reliable data was collected) to 2017, to illustrate the rising trend in overdose deaths.

Across all ten years covered by the dataset, it becomes clear that the crisis appears to have two distinct peaks in terms of age: one among those in their 20s and 30s and another among those in their 50s. The most common age for someone in the county to die of an overdose is 51, indicating that there is a slight skew toward the older of the two age peaks. Moreover, the standard deviation is 12.5, showing that most deaths occur within an approximately 24-year timeframe in mid-life.

It is interesting to observe how age distribution across drug overdose deaths changes over time. The double peaks are not initially pronounced, but develop over time until 2017, when they seem to disappear. It is also interesting to see how age range widens over the years as well, and how age distribution starts to skew toward younger people. Nonetheless, the opioid epidemic is more prevalent among men than women given that women accounted for a comparably low 31 percent of fatal drug overdoses in this timeframe.

It is also clear that the overdose deaths in this county occur overwhelmingly among Caucasians. However, it is worth looking at the demographics of the county overall to identify any major disparities. In Table 1, we compare overdose deaths with county population data that closely aligns with U.S. Census data. While there may be some disparities in the total percentages due to non-matching categories, for the purposes of a "sanity check" on the proportions of overdose deaths in the dataset, the table serves its purpose.

 ISBN: 978-1-61208-741-2

TABLE I. PERCENTAGE OVERDOSE DEATHS BY RACE WITH COUNTY POPULATION DATA

Race	Percent of County Population	Percent of Overdose Deaths	Difference
Asian	4.0	.2	3.8%
Black or African American	13.4	14	-.06%
Hispanic	2.1	.2	1.9%
White	78.6	85.5	-6.9%

It appears that Asian and Hispanic ethnic groups are comparatively less affected by overdose deaths to a small degree and African Americans are more affected, again by a small amount. Meanwhile, Caucasian people are overrepresented as victims of overdose deaths by a larger difference than any other, although still not by a significant degree.

It is interesting to note that April and August are the most common months for overdose deaths to occur. The reasons for this are not obvious and warrant further investigation. However, other than August, there is a decline in deaths during the summer months, which may suggest a possible seasonal element in overdose deaths.

Figure 2 shows the number of deaths by day of the week, the results collected here are perhaps unsurprising. Friday, Saturday, and Sunday see the highest number of overdose deaths. People are less likely to be working on these days and will thus have more leisure time. These days are also when people are most social, which, depending on the person, can involve alcohol or recreational drug use.

The spike in overdose deaths that occurs at around 5 pm is also of interest. This is when many people get out of work and there may be a connection here. The fact that most overdose deaths occur in the middle of the day is interesting as well, particularly the spike observed at 1 pm.

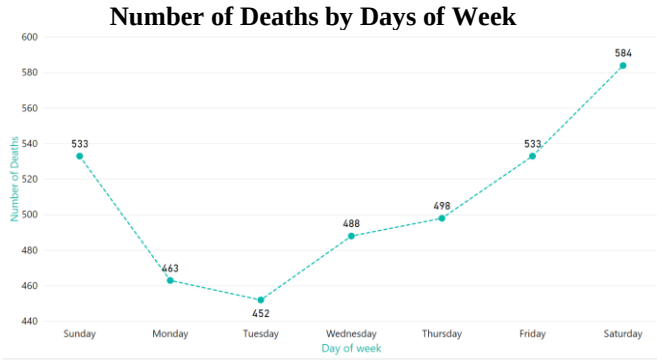

Figure 2. Number of deaths due to drug overdose by day of week

		Frequency
Valid	Total	3522
	15210	184
	15212	165
	15136	112
	15216	108
	15132	93

Figure 3. Top five zip codes for the locations of overdose deaths in the county during the period examined

Figure 3 shows the top five zip codes for the locations of overdose deaths in the county during the period examined. They account for a substantial proportion of all overdose deaths, but overall distribution remains quite wide and dispersed.

		Frequency
Valid	Total	3493
	15210	171
	15212	154
	15136	114
	15216	97
	15132	94

Figure 4. Top five zip codes for the residences of people who suffered a fatal drug overdose in Allegheny County during the period examined

Figure 4 shows the top five zip codes for the residences of people who suffered a fatal drug overdose in Allegheny County during the period examined. They are the same as the zip codes identified as the top five locations of fatal overdoses. This overlap indicates that people tend to overdose in the zip code in which they live, which seems reasonable. Nonetheless, there are some differences in frequency and a far larger range of zip codes cover the residences of the decedents. So, some people do travel and then overdose, including from as far away as West Virginia and even Minnesota.

IV. FINDINGS AND TRENDS

This section highlights key findings of the study, as well as trends in relation to the subject matter as per the demographic variables tested.

A. Temperature and Overdose Deaths

The graph in Figure 5 charts temperatures and uninsured rates and reveals a higher death rate for uninsured people in the warmer months. In contrast, in the graph above, a

 ISBN: 978-1-61208-741-2

higher death rate is observed to occur in the colder months. This is both fascinating and difficult to explain. It is possible that uninsured people are less motived to go to the hospital in the warmer months, believing that whatever ailment they may have will pass on its own since they don't have the cold weather working against them. Moreover, illnesses, such as flu and pneumonia, tend to arise in the colder months, and this again suggests people are less concerned about illness in the warmer months. Therefore, there is no significant dependence on temperature in relation to drug overdose deaths.

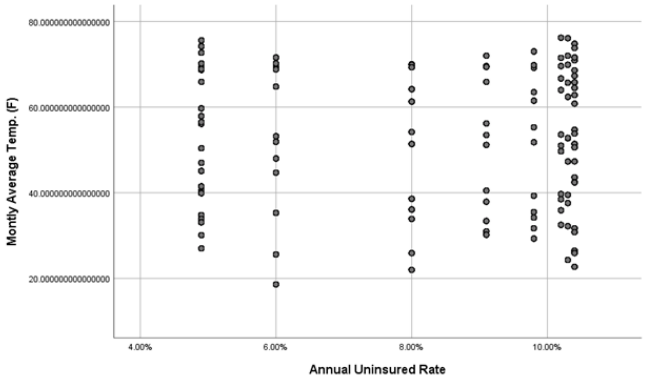

Figure 5. A chart showing how the monthly average temperature and annual uninsured rate influenced the death rate

Figure 6 analyzes the distribution between male and female in overdose deaths in Allegheny County. The study revalidates that overdose deaths are more amongst men, which make up around 69 percent of the total deaths. Women account for 31 percent of the total deaths.

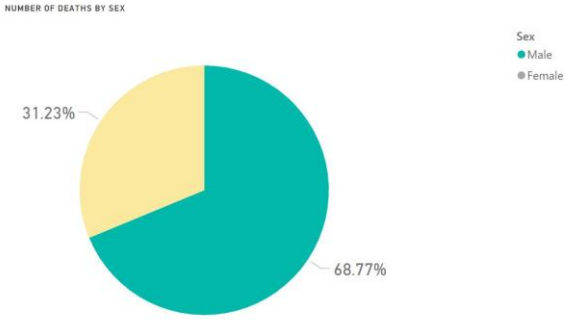

Figure 6. A chart showing the number of deaths by sex in Allegheny County, PA

Figure 7 identifies the number of deaths by race. It is seen that Whites represent the highest segment of the population that is affected among drug users and number of fatalities due to the opioid crisis.

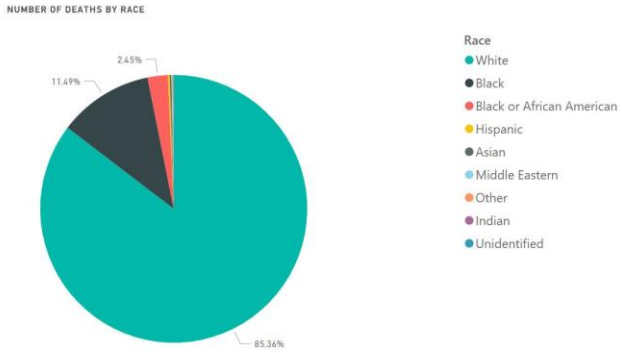

Figure 7. A pie chart showing number of deaths by race in Allegheny County, PA

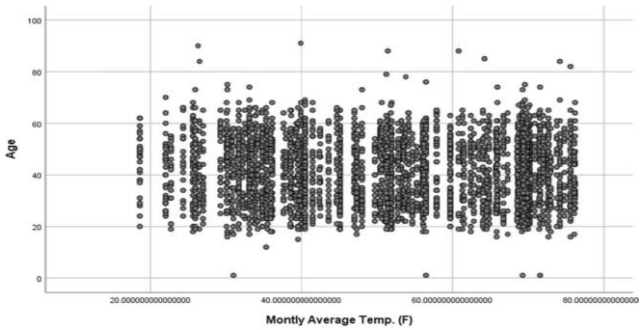

Figure 8. Shows how the rate of overdose deaths is determined by age and the monthly average temperatures

The graph in Figure 8 plots average monthly temperatures against the age of the decedents. The graph shows that there are strong clusters and then gaps around certain temperatures as compared with age, which shows a distinct range. It is interesting to observe certain gaps around temperatures at which no overdose deaths appear to have occurred versus clusters where many overdose deaths occurred. Again, this graph further supports our finding that there is no significant dependence on temperature in relation to drug overdose deaths.

B. Month to Month Fluctuations

Looking at every year separately on IBM Watson, a cloud system for data analytics, another pattern emerges whereby months with abnormally high death counts are followed by months with abnormally low death counts, and vice versa. This may be related to the availability of drugs in the market, reflecting the high level of addiction to drugs containing fentanyl and heroin specifically. The authorities have successfully taken down various platforms that were previously used to sell and buy drugs, but not a lot of information has been gathered in terms of the effects of such an action [4].

 ISBN: 978-1-61208-741-2

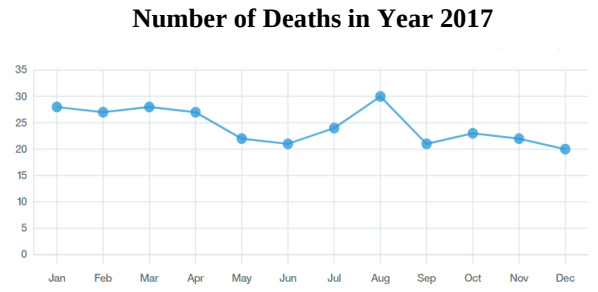

Figure 9. A visualization of the death rates of each month of the year 2017

In 2017, the aforementioned pattern was most visible in the months of August and September. August saw a high number of deaths and was followed by September, a month with lower deaths, which is the pattern we recognized through IBM Watson which we concluded that this could be related to market supply. When one particular month has a high death number, the following month is much lower, suggesting that the supply is lower in the market. The year ended with relatively low death rates in comparison with previous months.

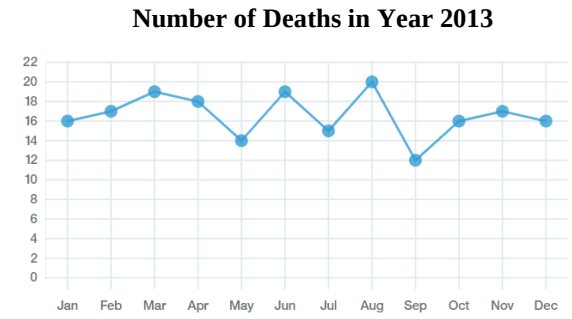

Figure 10. A visualization of the death rates of each month of the year 2013

As per the visualization in Figure 10, the year 2013 is a very good example of the potential oscillating trend, in which months with very high fatal drug overdose rates are followed by months with significantly lower overdose death rates.

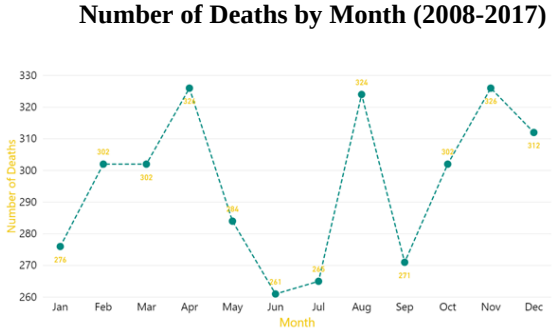

Figure 11. A visualization of the number of deaths for each month of all the years of observation

Looking at all the years examined together in Figure 11, we can confirm the pattern that shows that it may be due to the supply of drugs in the market which causes fluctuation in deaths by months. Looking at August for example, we see a peak in deaths, of which then September shows much lower deaths. As previously mentioned, and can be seen from Figures 9 and 10, this was also apparent when we studied each year separately on IBM Watson. There was a high indication that after a month of high deaths, the following month sees much lower deaths. The visualization in Figure 11 was generated using Microsoft Power BI.

V. CONCLUSION AND RECOMMENDATIONS

This data analytics study provides an expanded exploration of the problem of opioid drug overdoses in Allegheny County. Applying several statistical techniques, including pattern recognition, and generating other data visualizations, we were able to validate previously identified findings about overdose deaths, such as the age range and general demographics of affected populations [7]. We found the tools we used to be very useful in helping us to gain a better understanding of our sample set and to generate informative and understandable visuals.

Data Analytics also allowed us to find new information about our sample, particularly when combined with the additional variables added to the base dataset. We were able to generate highly practical visuals and illustrate clear trends from a large dataset with many variables, indicating that there is a possible relationship between drug addiction and lower temperatures and between psychological and behavioral factors and weekend drug use as well as a direct market supply effect on the number of deaths. These findings imply the need for authorities to offer educational workshops to individuals and their families, as well as health practitioners, about the dangers of the current drug epidemic and to design an effective policy for the oversight of drug market supply that includes taking firm action against violators.

We recommend that the authorities spend more time and funds on advertisements to educate individuals and families about the problem as well as investing in approaches to oversee market supply to drug users.

The issue of drug overdose deaths is so pressing that more research by data and population scientists is needed to gain further insight into this epidemic, so that impactful solutions can be found to reduce its harmful effects on communities both in the United States and around the globe.

Results of the current study supported results of the past studies. Thus, the limitation of this research was that there were no new findings obtained. The tools used in this study helped in enabling us to recognize patterns through data visualizations, and we encourage future researchers to leverage the use of multi-dimensional data to find factors that could possibly correlate to the drug addiction epidemic to explore new findings within regional and global communities. It is the hope that this information will be

 ISBN: 978-1-61208-741-2

used in enhancing understanding of readers regarding the subject matter.

REFERENCES

[1] Allegheny County Fatal Accidental Overdoses. (2018, April 11). Retrieved April, 2019, from https://catalog.data.gov/dataset/allegheny-county-fatal-accidental-overdoses

[2] D. Ciccarone, "Fentanyl in the US heroin supply: a rapidly changing risk environment", The International journal on drug policy, vol. 46, pp. 107-111, 2017.

[3] Climate at a Glance. (n.d.). Retrieved April, 2019, from http://www.ncdc.noaa.gov/cag/county/time-series/PA-003/tavg/all/1/2008-2017?base_prd=true&firstbaseyear=2008&lastbaseyear=2017&trend=true&trend_base=10&firsttrendyear=1895&lasttrendyear=2019

[4] B. R. Edlin, "Access to treatment for hepatitis C virus infection: Time to put patients first", The Lancet Infectious Diseases, vol. 16(9). doi:10.1016/s1473-3099(16)30005-6, 2016.

[5] Income Inequality in Allegheny County, PA. (2018, December 7). Retrieved April, 2019, from https://fred.stlouisfed.org/series/2020RATIO042003

[6] Lopez, G. (2019, February 26). A new study shows America's drug overdose crisis is by far the worst among wealthy countries. Retrieved May 20, 2019, from https://www.vox.com/science-and-health/2019/2/26/18234863/drug-overdose-death-america-international-study

[7] Lord, R. (2017, April 06). Allegheny County drug overdose deaths surge to 613 in 2016, breaking record. Retrieved May, 2019, from http://www.post-gazette.com/news/overdosed/2017/04/06/Allegheny-County-drug-overdoses-610-deaths-break-record/stories/201704060199.

[8] Percent Without Health Insurance Data for Allegheny County, PA. (n.d.). Retrieved May, 2019, from http://www.opendatanetwork.com/entity/0500000US42003/Allegheny_County_PA/health.health_insurance.pctui?year=2014&age=18 to 64&race=All races & sex=Both sexes & income=All income levels

[9] B. Saloner, "A public health strategy for the opioid crisis", Public Health Reports, vol. 133, pp. 24S-34S, 2018.

[10] Sederer, L. (2016, February 1). Take Action Against Addiction. Retrieved May 20, 2019, from https://www.usnews.com/opinion/blogs/policy-dose/articles/2016-02-01/10-ways-to-combat-americas-drug-abuse-problem

[11] Unemployment Rate in Allegheny County, PA. (2019, May 31). Retrieved May, 2019, from https://fred.stlouisfed.org/seri

[12] U.S. Census Bureau's Small Area Health Insurance Estimates Program. (N.d.). Retrieved April 2019, from https://www.opendatanetwork.com/entity/0500000US42003/Allegheny_County_PA/health.health_insurance.nui?ref=entityquestion&year=2014&age=18%20to%2064&race=All%20races&sex=Both%20sexes&income=All%20income%20levels

[13] E. Hulsey, and et. al., Opiate-Related Overdose Deaths in Allegheny County – Risks and Opportunities. (2016, July). Retrieved August, 2019, from https://alleghenycounty.us/uploadedFiles/Allegheny_Home/Health_Department/Programs/Special_Initiatives/Overdose_Prevention/Opiate-Related_Overdose_Deaths_in_Allegheny_County.pdf

 ISBN: 978-1-61208-741-2