A Latent Class Analysis of Accidental Polysubstance Overdose Deaths in Allegheny County, Pennsylvania

by

Kristina L. Boyd

BS Biology, Massachusetts Institute of Technology, 2013

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Kristina L. Boyd

It was defended on

October 23rd, 2019

and approved by

Jeanine M. Buchanich MEd, MPH, PhD, Research Associate Professor of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Robert Krafty, PhD, Associate Professor of Biostatistics, Graduate School of Public Health, University of Pittsburgh

LuAnn Brink, MPH, PhD, Adjunct Assistant Professor of Epidemiology, Graduate School of Public Health, University of Pittsburgh

Thesis Advisor: Jeanine M. Buchanich MEd, MPH, PhD, Research Associate Professor of Biostatistics, Graduate School of Public Health, University of Pittsburgh

A Latent Class Analysis of Accidental Polysubstance Overdose Deaths in Allegheny County, Pennsylvania

Kristina L. Boyd, MS

University of Pittsburgh, 2019

Background: The overdose mortality rate in Allegheny County has risen from 18.5/100,000 in 2008 to 63.3/100,000 in 2017. Despite most accidental drug poisoning deaths involving more than one drug, patterns in polysubstance overdose deaths are not well understood.

Methods: Polysubstance overdose deaths were identified from Allegheny County Office of the Medical Examiner (ACOME) data. Drugs present at post-mortem toxicology were categorized using a modified list from the Council of State and Territorial Epidemiologists (CSTE) overdose analysis tool. Latent class analysis (LCA) of decedent-level factors including demographics and drug categories present at post-mortem toxicology were used to identify common patterns within polysubstance overdose deaths.

Results: Among n = 3,749 toxicology reports from the Allegheny County Office of the Medical Examiner (ACOME) covering case years (cy) 2008 to 2018, n=2,864 involved more than one drug. LCA identified five latent class (LC) decedent groups associated with polysubstance overdose deaths: featured co-occurring fentanyl and fentanyl analogs (LC1); benzodiazepine presence with other opioid presence (LC2); ages 25-34 years and heroin presence (LC3); female sex, antidepressant presence, and amphetamine presence (LC4); and ages 45-54, black race, cocaine presence, alcohol presence, and fentanyl presence (LC5). LC1 was the largest class, representing 38.7% of all polysubstance overdose deaths.

Conclusion: Among polysubstance overdose deaths, five sub-populations were identified. These findings match findings in the literature and suggest surveillance targets.

Public health significance: This analysis is among the first to characterize accidental polysubstance overdose death in Allegheny County, PA. The composition of the sub-populations identified in this analysis will inform future overdose surveillance policy and outreach.

Table of Contents

List of Tables

List of Figures

Preface

To the members of my committee, for taking the time to guide me through the writing of this thesis. To my advisor Dr. Jeanine Buchanich for holding me accountable to my October timeline. To Dr. Robert Krafty for feedback on project scope and results. To Dr. LuAnn Brink for allowing me to continue this work at ACHD. To all three committee members especially for keeping my analysis relevant and readable.

To Dr. Abdus Wahed, for all of the extra feedback during my preparation for the comprehensive exam retake.

To Dr. Ada Youk, for excellent advising, patient replies to my many questions, no matter how trivial, and belief in my abilities when I had none.

To Dr. Dianne Finkelstein and Nora Horick, MS, and everyone at the Massachusetts General Hospital Biostatistics Center, for instilling in me a love of statistical practice and scientific inquiry.

To my friends Qi and Dominika. Thank you so much for all of your support over the one of the hardest years of my life.

To my partners, whose love and support illuminate my life in innumerable ways. To my partner, Stentor, for sending me excellent cat GIFs, laughing at my math jokes, and wandering both near and afar with me. To my spouse, Kevin, for making me a much more savvy Excel user, holding down the domestic fort during my many late nights over the last two years, and for infinite patience. I love you both so much.

Para mi mama. Todo lo que he logrado en esta vida nunca hubiera sido posible sin tu amor y apoyo. Gracias por escuchando sin juzgarme. Hicimos esto juntos.

To my son, Garrett. In the words of John Nash:

> I've always believed in numbers and the equations and logics that lead to reason. But after a lifetime of such pursuits, I ask, "What truly is logic? Who decides reason?" My quest has taken me through the physical, the metaphysical, the delusional – and back. And I have made the most important discovery of my career, the most important discovery of my life: it is only in the mysterious equations of love that any logic or reasons can be found. I'm only here tonight because of you. You are the only reason I am. You are all of my reasons.

Thank you all so much.

1.0 Introduction

Drug overdose is the number one cause of accidental death in the United States [8]. As of 2017, Pennsylvania ranks third among all states in overdose death rates (44.3 per 100,000 persons), outranked only by its neighboring states, West Virginia (57.8 per 100,000 persons) and Ohio (46.3 per 100,000 persons) [7]. Among Pennsylvania counties, Allegheny County ranks among the top ten in terms of overdose death rates (49 deaths per 100,000 persons) [14]. Despite a recent 40% year-over-year decrease in Allegheny County overdose death counts from 2017 to 2018[1], accidental drug overdose death remains a significant public health concern in southwestern Pennsylvania.

Identifying the drugs involved in accidental drug overdose is key to stopping the epidemic. Analysis of death certificate International Classification of Diseases 10th edition (ICD-10) codes is convenient [28], but can be lacking in accuracy [6] due to the variation in reporting practices amongst county coroners and medical examiners. Analysis of post-mortem toxicology results, when available, can provide more detailed insights into patterns of drug co-occurrence in overdose deaths [5].

Coupled with the finding by Jalal and colleagues that the overdose epidemic is composed of smaller sub-epidemics, and that the composition of these sub-epidemics differ greatly by demographic, geographic, and temporal groupings [11], analysis of drug groupings in polysubstance deaths provides a potential way to frame the characteristics of the epidemic. Though polysubstance deaths are the rule rather than the exception among fatal drug overdoses [5], the small number of deaths attributed to each drug combination makes analysis difficult. Drug groupings quickly lead to dimensionality and data sparseness issues in analysis.

Latent class analysis (LCA) of polysubstance overdose deaths offers a powerful way to gain insight into the diversity of drug co-occurrence in post-mortem toxicology [19]. LCA is a measurement model in which individuals from a heterogenous population can be classified into mutually exclusive and exhaustive latent classes, based on a set of categorical indicator variables. In LCA, true class membership is unknown for each individual. Recent studies have used LCA to examine classes of overdose-related hospitalizations in Pennsylvania [16]

and substance use patterns in various populations of polysubstance users [24], but to date the same techniques have not been applied to post-mortem toxicology data.

This thesis explores latent class analysis in a set of toxicology-confirmed accidental overdose deaths from the Allegheny County Office of the Medical Examiner (ACOME) during the years 2008 to 2018. The two goals of this analysis are to (1) determine the number of latent classes best suitable for modeling patterns of drug co-occurrence at post-mortem toxicology in polysubstance deaths in Allegheny County and (2) explore the composition of those latent classes.

To give context to the methods and models, section 2 describes the data management and statistical analyses carried out. Section 3 contains results of the analysis. Section 4 contains a discussion. The R code used to perform this analysis is contained within appendix A.

2.0 Methods

2.1 Data Management

2.1.1 Data sources

Data from n = 3,749 toxicology reports from the Allegheny County Office of the Medical Examiner (ACOME) covering case years (cy) 2008 to 2018 were assessed for presence of more than one drug at toxicology. N = 1,105 records were removed for listing only a single drug. The final analytic sample (n=2,861) included all overdose deaths in Allegheny County, PA that occurred between 2008 and 2018 and involved more than one drug except three records for which race was missing. Because all individuals included in the final analytic sample were deceased, Institutional Review Board approval was not required.

2.1.2 Drug classification

Toxicology data was stored as a comma-separated string of drugs the decedent tested positive for at post-mortem toxicology. The string-to-column function in Excel (Microsoft Corporation) was used to separate drug names into separate columns. The presence or absence of each drug was noted with a 0 or a 1. Drugs were cross-referenced to a list of drug categories derived from a literal text analysis tool developed by the Council of State and Territorial Epidemiologists (CSTE) Overdose Subcommittee. [23] Drug category prevalences were calculated. The sedative and barbiturate markers were combined into one, and all drug categories that did not meet minimum prevalence (1% of the final analytic sample) were combined into an "other" category.

2.2 Statistical Analyses

2.2.1 Covariates

Demographics available from ACOME data included patient age, sex, and race. Age was categorized into the following groups : 15-19, 20-24, 25-34, 35-44, 44-54, 55-59, and 65+ years. Race was categorized as White, Black, and Other. Drugs were grouped as follows: alcohol, antidepressants, antihistamines, antipsychotics, amphetamines, benzodiazepines, cocaine, fentanyl, fentanyl analogues, heroin, muscle relaxants, non-opioid analgesics, other opioids, an aggregate other category, and a combination sedative/barbiturate category.

2.2.2 Latent class analysis

2.2.2.1 Model specification, estimation, and fit First introduced by Lazarsfeld and Henry in 1968, latent class analysis (LCA) focuses on the identification of latent class variables through a number of categorical observed response variables [21].

Formally the model can be defined as follows. Let there be:

- p manifest categorical variables $x_1 \dots x_p$ with the j-th variable having m_j levels.
- a latent categorical variable y with $K \geq 2$ levels.

Let each level of y define a population of subjects. Conditioning on the level of the latent variable, the manifest variables are independent. That is: $(x_1 \perp x_2 \perp \dots x_p) \mid y$. Suppose that for an individual in the population, the prior probability of class membership is $P(y = j) = \gamma_j$ with $\sum \gamma_j = 1$. Let π_{ij} be the probability that a person in latent category j gives a response on manifest variable i such that $x_i = 1$ corresponds to the presence of a state and $x_i = 0$ corresponds to the absence of a state – for instance, the presence or absence of a drug on a toxicology screen. Then according to this:

$$f(\mathbf{x} \mid y = j) = \prod_{i=1}^{p} \pi_{ij}^{x_i} \tag{2.1}$$

. The unconditional distribution of 2.1 is given by

$$f(\mathbf{x}) = \sum_{j=1}^{K} \{\gamma_j \prod_{i=1}^{p} \pi_{ij}^{x_i} (1 - \pi_{ij})^{1-x_i}\}. \tag{2.2}$$

Suppose also a sample from a random population. This implies that the average fraction of individuals in class j estimates γ_j. The parameters can then be estimated via maximum likelihood estimation (MLE) via an expectation-maximization (EM) algorithm. The EM algorithm is a way to find maximum likelihood estimates in models with dependence on unobserved latent variables [2]. Using an initial guess, this iterative process generates a sequence of improving approximations in which the n-th approximation is derived from preceding approximations. As the EM algorithm name suggests, there is alternation between an expectation step in which a function for the expectation (E) of the log-likelihood using guesses for parameter values and a maximization (M) step which computes parameters maximizing the expected log-likelihood found in the (E) step. These steps are repeated until a termination criteria is reached, typically some pre-set number of iterations or until convergence[2].

Formally, given a sample of n observations from the aforementioned random sample from a population, the log-likelihood is

$$\log L(\theta) = \log(\prod_{h=1}^{n} f(\mathbf{x}_h)) \tag{2.3}$$

where θ consists of the $K-1$ free parameter γs , also called the latent class factor mean estimates and the $\{\pi_{ij}\}$, also called the class membership probabilities. Differentiating 2.3 with respect to the parameters in the usual way yields the following estimates for γ_j and π_{ij}

$$\hat{\gamma_j} = \sum_{h=1}^{n} h(j \mid \mathbf{x}_h)/n, i = 1, \ldots, K \tag{2.4}$$

$$\hat{\pi_{ij}} = \frac{1}{n\hat{\gamma_j}} \sum_{h=1}^{n} x_{ih} h(j \mid \mathbf{x}_h), i = 1, \ldots, K, j = 1, \ldots, p \tag{2.5}$$

. Equation 2.4is the estimated fraction of individuals allocated to latent class j. Equation 2.5 is the fraction of those with manifest variable i in latent class j.

A review of model selection criteria by McLachlan and Rathnayke (2014) [18] favored the Bayesian Information Criterion (BIC) as an acceptable model selection criteria for mixture models such as LCA. Formally, the BIC is defined as follows

$$BIC = \ln(n)k - 2\ln\hat{L} \tag{2.6}$$

where k is the number of parameters being estimated, n is the sample size, and $\hat{L}$ is 2.3 evaluated at the MLEs specified in equation 2.4 and equation 2.5. Other model fit statistics include the adjusted Bayesian Information Criterion (aBIC), which is a sample-size adjusted BIC, Akaike's Information Criterion (AIC), and consistent Akaike's Information Criterion (cAIC). They are defined as follows

$$aBIC = -2LL + \log \frac{n+2}{24} \tag{2.7}$$

$$AIC = 2k - 2\ln \hat{L} \tag{2.8}$$

$$cAIC = -2LL + k(1 + \log(n)) \tag{2.9}$$

where LL refers to the log-likelihood of the model in question.

For this thesis, LCA was performed using the R package poLCA created by Drew Linzer and Jeffrey Lewis.[15]. Models from 1 to 10 classes were fit and their fit assessed by calculating and comparing BIC, aBIC, AIC, and cAIC. Following the suggestions of Nylund and colleagues[20], the model with lowest BIC was determined to be the best fitting model. Conditional item response probabilities were examined and classes labeled.

3.0 Results

3.1 Identification of polysubstance overdose deaths

Polysubstance overdose deaths were derived using the literal text overdose analysis tool developed by the CSTE overdose subcommittee. The original data set included a single comma-separated string of all drugs a decedent tested positive for on toxicology screen. First, the string was stripped of delimiters and individual drug names separated into columns using Excel (Microsoft Corporation). Binary indicator variables were derived for each individual drug and the number of drugs present at post-mortem toxicology enumerated. Decedents with two or more unique drugs present at post-mortem toxicology and without missing values for race were entered into the final analytic sample. In this final analytic sample, the binary indicators derived in the previous data cleaning step were cross-referenced to the drug category list provided with the CSTE tool. In consultation with local public health officials, the list was modified to reflect local overdose trends. For example, heroin, fentanyl, and fentanyl analogs were given their own categories instead of labeled opioids. Cocaine was separated from other stimulants as were all amphetamines.

A total of 137 unique substances were found across all case years occurring in 1,472 unique combinations. An average of 3.1 drugs were found in polysubstance overdose deaths and an average of three deaths were attributed to each combination. A full listing of substances identified is available in Appendix B.

3.2 Sample characteristics

Among 3,749 toxicology reports, 2,861 (73.7%) involved more than one drug. Decedents in this sample were more likely to be male (68.8%), white (86.3%). More than three quarters of decedents were aged 25 to 54 years (76.5%) and no decedents were younger than 15 years old (Table 1). Fentanyl was the single most prevalent drug category (46.8%).

3.3 Defining latent classes

The best-fitting LCA model identified 5 distinct groups of polysubstance overdose deaths (Table 3). The five-class model had the lowest BIC and cAIC, and second-lowest aBIC and AIC (Figure 1). The six-class model had marginally lower aBIC and AIC compared to the five-class model, but the five-class model classes were more easily interpreted.

Latent class 1 (LC1) was defined by a 100% probability of fentanyl presence and a 29% probability of fentanyl analog presence. Latent class 2 (LC2) was defined by benzodiazepine presence (56%) and other opioid presence (69%). Latent class 3 (LC3) was defined by age 25-34 years (35%) and 100% chance of heroin presence. Latent class 4 (LC4) included decedents with a higher than average probability of being female (58%), antidepressant presence (90%), and amphetamine presence (23%). Latent class 5 (LC5) was defined by black race (50%), cocaine presence (91%),alcohol presence (62%), and fentanyl presence (Table 4).

Among these five classes, LC1 was the largest class, representing 38.7% of all polysubstance overdose deaths handled by ACOME. LC5 was the smallest, representing 9.8% of deaths in this sample (Table 4).

Table 1: Demographic characteristics of n = 2,861 polysubstance overdose deaths in Allegheny County, PA 2008-2018

Sample characteristics	**Count (%)**
Age Group	
15 to 19 yrs	26 (0.9%)
20 to 24 yrs	159 (5.6%)
25 to 34 yrs	727 (25.4%)
35-44 yrs	693 (24.2%)
44-54 yrs	769 (26.9%)
55-59 yrs	289 (10.1%)
60-64 yrs	136 (4.8%)
65+ yrs	62 (2.2%)
Sex	
Female	893 (31.2%)
Male	1968 (68.8%)
Race	
White	2469 (86.3%)
Black	375 (13.1%)
Other	17 (0.6%)

Table 2: Drug categories present at toxicology in n = 2,861 polysubstance overdose deaths in Allegheny County, PA 2008-2018

Drug categories present at toxicology	**Count (%)**
Fentanyl	1338 (46.8%)
Cocaine	1064 (37.2%)
Benzodiazepines	933 (32.6%)
Other opioids	927 (32.4%)
Heroin	870 (30.4%)
Alcohol	837 (29.3%)
Antidepressants	410 (14.3%)
Fentanyl analogs	312 (10.9%)
Antihistamines	119 (4.2%)
Amphetamines	105 (3.7%)
Sedatives and barbiturates	96 (3.4%)
Other drugs	77 (2.7%)
Muscle relaxants	76 (2.7%)
Non-opioid analgesics	50 (1.7%)
Antipsychotics	46 (1.4%)

Table 3: Model fit statistics for 1-10 latent classes.

Number of latent classes	Model	Log-likelihood	resid. df	BIC	aBIC	AIC	cAIC	Likelihood Ratio
1	Model A	-23142.36	2836.00	46483.70	46404.26	46334.72	46508.70	8913.61
2	Model B	-22637.89	2810.00	45681.68	45519.64	45377.78	45732.68	7904.66
3	Model C	-22379.56	2784.00	45371.96	45127.30	44913.12	45448.96	7388.00
4	Model D	-22170.19	2758.00	45160.15	44832.88	44546.38	45263.15	6969.27
5	Model E	-22015.64	2732.00	45057.97	44648.10	44289.27	45186.97	6660.16
6	Model F	-21923.61	2706.00	45080.86	44588.37	44157.23	45235.86	6476.11
7	Model G	-21836.43	2680.00	45113.42	44712.69	44034.85	45294.42	6301.74
8	Model H	-21763.63	2654.00	45174.76	44837.01	43941.26	45381.76	6156.14
9	Model I	-21620.70	2628.00	45095.83	44961.33	43707.40	45328.83	5870.28
10	Model J	-21529.19	2602.00	45119.73	45085.65	43576.37	45378.73	5687.26

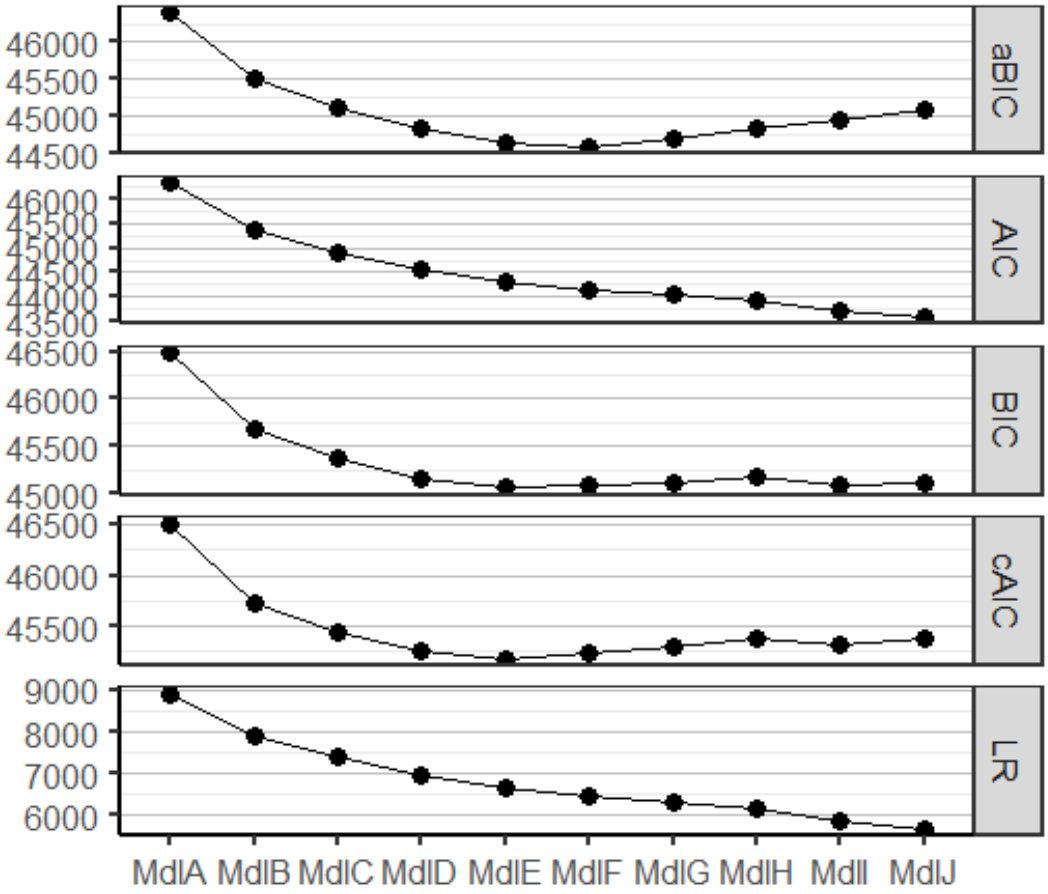

Figure 1: Graph of model fit statistics for 1-10 latent classes.

Table 4: Class membership probabilities (%) for the 5-class model.

		Latent classes (percentage of analytic sample, n = 2,681)				
Variable	Outcome	LC1 (38.7%)	LC2 (25.7%)	LC3 (14.7%)	LC4 (11.1%)	LC5 (9.8%)
Age group	15-19 yrs	0.8	1.8	0.6	0.3	0.0
	20-24 yrs	5.9	6.8	6.4	2.7	2.7
	25-34 yrs	32.9	19.5	35.4	20.0	8.6
	35-44 yrs	26.4	23.9	23.6	26.4	17.1
	45-54 yrs	20.0	31.1	20.0	29.4	44.4
	55-59 yrs	7.1	10.9	7.0	11.6	20.1
	60-64 yrs	4.5	3.6	5.0	6.3	6.5
	65+ yrs	2.2	2.4	2.0	3.2	0.6
Sex	Female	24.1	35.3	25.7	58.5	24.1
	Male	75.8	64.7	74.3	41.6	75.9
Race	White	90.5	92.3	90.4	90.0	48.9
	Black	8.8	7.4	8.6	9.5	50.3
	Other	0.7	0.3	1.0	0.5	0.7
Drugs present at post-mortem toxicology	Alcohol	20.8	30.6	30.5	16.8	62.4
	Amphetamines	3.1	0.0	0.5	22.8	0.8
	Antidepressants	5.4	6.3	4.1	90.2	3.6
	Antihistamines	2.7	7.4	0.9	8.9	0.4
	Antipsychotics	0.3	2.1	0.5	7.9	0.0
	Benzodiazepines	24.5	55.7	33.7	30.2	0.5
	Cocaine	35.2	29.8	29.9	16.9	91.3
	Fentanyl	100.0	11.6	8.1	25.6	43.2
	Fentanyl analogs	29.4	0.5	0.2	1.3	3.7
	Heroin	36.0	0.7	100.0	10.2	12.2
	Muscle relaxants	1.2	4.7	1.1	5.6	1.4
	Non-opioid analgesics	0.5	3.7	0.3	4.5	0.0
	Other drugs	0.5	3.9	1.0	9.5	2.3
	Other opioids	16.0	69.0	14.0	37.1	11.7
	Sedatives and barbiturates	0.0	0.1	0.0	0.0	0.0

4.0 Discussion

Drug poisoning and overdose has reached epidemic proportions in the United States [27], and presents a major public health concern in southwestern Pennsylvania ([6]). The problem of overdose is well-characterized in the literature via analysis of single-substance cause-specific mortality. Despite the majority of overdoses involving more than one drug, polysubstance mortality burden is not as well understood [5]. Most studies that examine polysubstance use focus on risk group identification of living subjects through the use of hospitalization [17] and other government service data [19, 5]. This analysis provides one of the first characterizations of polysubstance mortality by identifying five distinct classes of polysubstance overdose deaths in Allegheny County from 2008-2018.

Polysubstance overdose was identified via enumeration of positive drug results at post-mortem toxicology. The individual substances were then categorized using the documentation accompanying an overdose analysis tool developed by the CSTE overdose subcommittee. Though the CSTE tool provided a comprehensive listing of overdose-related terms, including substance names and common misspellings, there were a small subset of terms that required manual cleaning and identification. The process was time-consuming and not scalable for larger data sets. More automated methods of text cleaning should be considered for future work.

It was determined that polysubstance overdose commonly featured co-occurring fentanyl and fentanyl analogs (LC1); benzodiazepine presence with other opioid presence (LC2); ages 25-34 years and heroin presence (LC3); female sex, antidepressant presence, and amphetamine presence (LC4); and black race, cocaine presence, and alcohol presence (LC5). Identifying decedent-level characteristics associated with polysubstance overdose death has implications for public health surveillance and prevention strategies. While risk-stratification is best inferred from LCA of data from living subjects, characterization of mortality trends over time via LCA can be used to monitor the effect of prevention strategies.

LC1 consisted of fentanyl and fentanyl analog presence. The presence of fentanyl in Allegheny County is well documented. Since 2013, the number of overdose deaths in Al-

legheny County due to illicit fentanyl has rapidly increased [9]. Creppage and colleagues [9] found the proportion of fentanyl-containing stamp bag evidence in illicit drug evidence increased from 2.1% in 2014 to 17.1% in 2016. Additionally, the rise in local fentanyl related overdoses, both fatal and non-fatal, has been documented in analyses focusing on naloxone distribution and other surveillance efforts [4, 6]. Because fentanyl is prevalent across single and polysubstance overdoses, it should come as no surprise LC1 was the largest of the latent classes in this analysis.

LC2 was characterized by benzodiazepine and other opioid presence. Taking these medications in combination is discouraged due to both drugs having depressive effects on the respiratory system [3]. This grouping is possibly a local manifestation of national overdose trends that have been noted since 2013; more than 30% of overdoses involving an opioid also involve a benzodiazepine of some kind [3].

LC3 was characterized by the presence of heroin in decedents between the ages of 25 to 34 years. As recently as 2018, a study showed that prevalence of lifetime heroin use increased from 2% in 25 year-olds to 4% in 35 year-olds, the upper and lower ends of the age group that predominates in this grouping [25]. The relatively younger age in this class suggests a possible target for future non-medical substance use education and intervention efforts. The surgeon general's report on addiction in the United States has shown that early intervention in a young non-medical substance consuming population has similar response rates to similar intervention for non-substance-use related chronic illnesses such as diabetes and asthma [10]. Moreover, the same report suggests that individuals with possible substance use disorder are easily identified through screening. Because substance use disorder increases in complexity and severity with continued substance misuse, intervention in a young population is key to reducing mortality burden [10].

LC4 was characterized by female sex, antidepressant presence, and amphetamine presence. This is the only class for which female sex and antidepressants had a conditional probability greater than 50%. The replacement of tricyclic antidepressants with selective serotonin re-uptake inhibitors has reduced the role of antidepressants in fatal overdose [22]; the presence of the antidepressant signature in this group might reflect treatment for mental

and behavioral health co-morbidities, which are known to occur in an amphetamine-using population [13].

LC5 was characterized black race, ages 45-54, cocaine presence, alcohol presence, and fentanyl presence. This is the only grouping for which black race was preferred over other racial categories. The crack-cocaine epidemic of the 1980s leveled off in the 1990s, but recent data suggest a resurgence in cocaine use in this population [27, 12]. Cocaine and alcohol together produce the cardio-toxic metabolite cocaethylene [29], lending a plausible biological underpinning to the variables clustered in this class.

These groupings should be interpreted with the following limitations in mind. First, this analysis used toxicology data from the Allegheny County Office of the Medical Examiner. The appearance of a positive result on a drug screen might be from intentional ingestion of that substance or from adulteration of one substance with another. Positive toxicology results reflect only the presence of drugs or metabolites in the tissue and sera. Toxicology screens do not capture all drugs in use in a population, so this analysis is limited to those drugs which are captured on standard toxicology screening tools[26]. LCA also has inherent limitations. The output of the latent classes is nominal, and might not necessarily have scientific meaning. The labeling of these classes was subjective and heterogeneity in class membership remains. Changing the number of classes would change the composition of those classes.

Despite these limitations, identification of these latent classes provides an important first step towards characterizing polysubstance mortality burden in Allegheny County.

5.0 Conclusion

This thesis identified five distinct classes within polysubstance overdose deaths in Allegheny County from 2008 to 2018. Future work could include the change in class membership over time, both in the years studied, and with the addition from data corresponding to future case years as it becomes available. Additionally, this analysis could be improved by incorporating other data sources that provide more information about decedent co-morbidities, which could provide more nuanced groupings and further insight for public health surveillance.

Appendix A

R code

```
#Same analysis as previous versions but drops year and location

#install packages required for this analysis
#install.packages("ggparallel") #"igraph", "tidyr","knitr")
#install.packages("igraph")
#install.packages("tidyr")
#install.packages("knitr")
#install.packages("reshape2")
#install.packages("dplyr")
#install.packages("Rcpp")
#install.packages("reshape2")
#install.packages("getRcppVersion")

#load packages required for this analysis
library("reshape2")
library("plyr")
library("dplyr")
library("poLCA")
library("ggplot2")
library("ggparallel")
library("igraph")
library("tidyr")
library("knitr")



# these are the defaults of the poLCA command
#poLCA(formula, data, nclass=2,
\\ maxiter=1000, graphs=FALSE, \\ tol=1e-10, na.rm=TRUE,
probs.start=NULL, nrep=1, verbose=TRUE, calc.se=TRUE)

#estimate the model with k-classes
#k<-3

#lc<-poLCA(f, data, nclass=k, nrep=30, na.rm=FALSE, graph = TRUE)
```

```
#fake data frame for troubleshooting

data3 = data.frame(replicate(10, sample(1:3,1000, rep = TRUE)))
# select variables

mydata3 <- data3 %>% dplyr::select(X1,X2,X3,X4,X5,X6,X7,X8,X9,X10)


# define function
f3<-with(mydata3, cbind(X1,X2,X3,X4,X5,X6,X7,X8,X9,X10)~1)


k<-3

lc<-poLCA(f3, mydata3, nclass=k, nrep=10, na.rm=FALSE, graph = TRUE)

#??poLCA


#real data now

#read in data to "data" object
data <- read.csv(file= "C:/Users/krist/
OneDrive/␣␣Documents/Thesis/Code/polysubstace_ODs_v6.csv",
header =TRUE, sep =",")

#rename variables for ease of graphical presentation

colnames(data)[colnames(data)=="AgeGroup"] <- "X1"
colnames(data)[colnames(data)=="Race"] <- "X2"
#colnames(data)[colnames(data)=="Location"] <- "X3"
colnames(data)[colnames(data)=="Year"] <- "X4"
colnames(data)[colnames(data)=="Cocaine"] <- "X5"
colnames(data)[colnames(data)=="FentanylAnalog"] <- "X6"
colnames(data)[colnames(data)=="Amphetamine"] <- "X7"
colnames(data)[colnames(data)=="OtherAggregate"] <- "X8"
colnames(data)[colnames(data)=="Antidepressant"] <- "X9"
colnames(data)[colnames(data)=="Antihistamine"] <- "X10"
colnames(data)[colnames(data)=="Antipsychotic"] <- "X11"
colnames(data)[colnames(data)=="Benzo"] <- "X12"
colnames(data)[colnames(data)=="Heroin"] <- "X13"
colnames(data)[colnames(data)=="MuscleRelaxant"] <- "X14"
colnames(data)[colnames(data)=="Alcohol"] <- "X15"
colnames(data)[colnames(data)=="NonOpioidAnalgesic"] <- "X16"
```

```
colnames(data)[colnames(data)== "OtherOpioid"] <- "X17"
colnames(data)[colnames(data)=="SedativeBarbituateCombo"] <- "X18"
colnames(data)[colnames(data)=="Fentanyl"] <- "X19"
colnames(data)[colnames(data)=="Gender"] <- "X20"


#select variables

mydata <- data %>% dplyr::select(X1,X2,X5,X6,X7,X8,X9,X10,X11,
X12,X13,X14,X15,X16,X17,X18,X19, X20)

#define formula
f <- with(mydata, cbind(X1,X2,X5,X6,X7,X8,X9,X10,X11,X12,
X13,X14,X15,
X16,X17,X18,X19, X20)~1)

#lc<-poLCA(f, mydata, nclass=k, nrep=10, na.rm=FALSE, graph = TRUE)

#------ run a sequence of models with 1-10 classes and print out \\
the model with the lowest BIC

#code modified from statistics.ohlsen-
web.de/latent-class-analysis-polca/

#set seed for reproducibility
set.seed(0932019)
#define values
#max_II <- -100000
min_bic <- 100000
for(i in 2:10){
  lc <- poLCA(f, mydata, nclass=i, maxiter=3000,
              tol=1e-5, na.rm=FALSE,
              nrep=10, verbose=TRUE, calc.se=TRUE)
  if(lc$bic < min_bic){
    min_bic <- lc$bic
    LCA_best_model<-lc
  }
}
LCA_best_model


#models with different numbers of groups without covariates

lc1<-poLCA(f, data=mydata, nclass=1,
na.rm = FALSE, nrep=10, maxiter=3000)
lc2<-poLCA(f, data=mydata, nclass=2,
na.rm = FALSE, nrep=10, maxiter=3000)
```

```
lc3<-poLCA(f, data=mydata, nclass=3,
na.rm = FALSE, nrep=10, maxiter=3000)
lc4<-poLCA(f, data=mydata, nclass=4,
na.rm = FALSE, nrep=10, maxiter=3000)
lc5<-poLCA(f, data=mydata, nclass=5,
na.rm = FALSE, nrep=10, maxiter=3000)
lc6<-poLCA(f, data=mydata, nclass=6,
na.rm = FALSE, nrep=10, maxiter=3000)
lc7<-poLCA(f, data=mydata, nclass=7,
na.rm = FALSE, nrep=10, maxiter=3000)
lc8<-poLCA(f, data=mydata, nclass=8,
na.rm = FALSE, nrep=10, maxiter=3000)
lc9<-poLCA(f, data=mydata, nclass=9,
na.rm = FALSE, nrep=10, maxiter=3000)
lc10<-poLCA(f, data=mydata, nclass=10,
na.rm = FALSE, nrep=10, maxiter=3000)

# generate dataframe with fit-values

    results <- data.frame(Model=c("Model␣A"),
log_likelihood=lc1$llik,
df = lc1$resid.df,
BIC=lc1$bic,
ABIC=  (-2*lc1$llik) + ((log((lc1$N + 2)/24)) * lc1$npar),
AIC = lc1$aic,
CAIC = (-2*lc1$llik)
+ lc1$npar * (1 + log(lc1$N)).
 likelihood_ratio=lc1$Gsq)

    results$Model<-as.integer(results$Model)
    results[1,1]<-c("Model␣A")
    results[2,1]<-c("Model␣B")
    results[3,1]<-c("Model␣C")
    results[4,1]<-c("Model␣D")
    results[5,1]<-c("Model␣E")
    results[6,1]<-c("Model␣F")
    results[7,1]<-c("Model␣G")
    results[8,1]<-c("Model␣H")
    results[9,1]<-c("Model␣I")
    results[10,1]<-c("Model␣J")

    results[2,2]<-lc2$llik
    results[3,2]<-lc3$llik
    results[4,2]<-lc4$llik
    results[5,2]<-lc5$llik
    results[6,2]<-lc6$llik
    results[7,2]<-lc7$llik
```

```
results[8,2]<-lc8$llik
results[9,2]<-lc9$llik
results[10,2]<-lc10$llik

results[2,3]<-lc2$resid.df
results[3,3]<-lc3$resid.df
results[4,3]<-lc4$resid.df
results[5,3]<-lc5$resid.df
results[6,3]<-lc6$resid.df
results[7,3]<-lc7$resid.df
results[8,3]<-lc8$resid.df
results[9,3]<-lc9$resid.df
results[10,3]<-lc10$resid.df

results[2,4]<-lc2$bic
results[3,4]<-lc3$bic
results[4,4]<-lc4$bic
results[5,4]<-lc5$bic
results[6,4]<-lc6$bic
results[7,4]<-lc7$bic
results[8,4]<-lc8$bic
results[9,4]<-lc9$bic
results[10,4]<-lc10$bic

results[2,5]<-(-2*lc2$llik) +
((log((lc2$N + 2)/24)) * lc2$npar) #abic
results[3,5]<-(-2*lc3$llik) + ((log((lc3$N + 2)/24)) * lc3$npar)
results[4,5]<-(-2*lc4$llik) + ((log((lc4$N + 2)/24)) * lc4$npar)
results[5,5]<-(-2*lc5$llik) + ((log((lc5$N + 2)/24)) * lc5$npar)
results[6,5]<-(-2*lc6$llik) + ((log((lc6$N + 2)/24)) * lc6$npar)
results[7,5]<-(-2*lc6$llik) + ((log((lc7$N + 2)/24)) * lc7$npar)
results[8,5]<-(-2*lc6$llik) + ((log((lc8$N + 2)/24)) * lc8$npar)
results[9,5]<-(-2*lc6$llik) + ((log((lc9$N + 2)/24)) * lc9$npar)
results[10,5]<-(-2*lc6$llik) +
((log((lc10$N + 2)/24)) * lc10$npar)

#aic
results[2,6]<-lc2$aic
results[3,6]<-lc3$aic
results[4,6]<-lc4$aic
results[5,6]<-lc5$aic
results[6,6]<-lc6$aic
results[7,6]<-lc7$aic
results[8,6]<-lc8$aic
results[9,6]<-lc9$aic
results[10,6]<-lc10$aic
```

```
    results[2,7]<- (-2*lc2$llik) + lc2$npar * (1 + log(lc2$N)) #caic
    results[3,7]<- (-2*lc3$llik) + lc3$npar * (1 + log(lc3$N))
    results[4,7]<- (-2*lc4$llik) + lc4$npar * (1 + log(lc4$N))
    results[5,7]<- (-2*lc5$llik) + lc5$npar * (1 + log(lc5$N))
    results[6,7]<- (-2*lc6$llik) + lc6$npar * (1 + log(lc6$N))
    results[7,7]<- (-2*lc7$llik) + lc7$npar * (1 + log(lc7$N))
    results[8,7]<- (-2*lc8$llik) + lc8$npar * (1 + log(lc8$N))
    results[9,7]<- (-2*lc9$llik) + lc9$npar * (1 + log(lc9$N))
    results[10,7]<- (-2*lc10$llik) + lc10$npar * (1 + log(lc10$N))

    results[2,8]<-lc2$Gsq
    results[3,8]<-lc3$Gsq
    results[4,8]<-lc4$Gsq
    results[5,8]<-lc5$Gsq
    results[6,8]<-lc6$Gsq
    results[7,8]<-lc7$Gsq
    results[8,8]<-lc8$Gsq
    results[9,8]<-lc9$Gsq
    results[10,8]<-lc10$Gsq

    results


# combining results \\ to a dataframe
colnames(results)<-c("Model","Log-likelihood",
"resid. df","BIC","aBIC", "AIC","cAIC","LR")
lca_results<-results



# output table for copy and paste
#install.packages("ztable")
ztable::ztable(lca_results)

#output table to console for LaTeX
#install.packages("xtable")
library(xtable)
xtable(lca_results)

# plot 1


#convert to long format
results2<-tidyr::gather(results,criteria,Guete,4:8)
results2

#plot
```

```
fit.plot<-ggplot(results2) +
  geom_point(aes(x=Model,y=Guete),size=3) +
  geom_line(aes(Model, Guete, group = 1)) +
  scale_x_discrete(labels = abbreviate) +
  theme_bw()+
  labs(x = "", y="", title = "") +
  facet_grid(criteria ~. ,scales = "free") +
  theme_bw(base_size = 16, base_family = "") +
  theme(panel.grid.major.x = element_blank() ,
        panel.grid.major.y = element_line(colour="grey", size=0.5),
        legend.title = element_text(size = 12, face = 'bold'),
        axis.text = element_text(size = 12),
        axis.title = element_text(size = 12),
        legend.text=  element_text(size=12),
        axis.line = element_line(colour = "black"))


# save 650 x 800
fit.plot #need to fix model order
```

Appendix B

Alphabetical listing of drugs identified in polysubstance overdose death toxicology reports

- 1,1-Difluoroethane
- 2-Fluoro Deschloroketamine
- 3-Methylfentanyl
- 4-Methoxy-Butyryl Fentanyl
- 7-Aminoclonazepam
- Acetaminophen
- Acetone
- Acetyl Fentanyl
- Alcohol
- Alprazolam
- Amitriptyline
- Amphetamines
- Benzodiazepines
- Benzofuran
- Benzoylecgonine
- Benzyl fentanyl
- Buprenorphine
- Bupropion
- Butalbital
- Butylone
- Butyryl Fentanyl/Isobutyryl Fentanyl
- Carbamazepine
- Carfentanil
- Carisoprodol
- Chlordiazepoxide
- Chlorodiazepam
- Chlorpheniramine
- Chlorpromazine
- Cis-3-Methylfentanyl
- Citalopram
- Citalopram/Escitalopram
- Clomipramine
- Clonazepam
- Clozapine
- Cocaine
- Codeine
- Cyclobenzaprine
- Cyclopropyl Fentanyl
- Delorazepam
- Demoxepam
- Desipramine
- Despropionyl Fentanyl 4-ANPP
- Dextromethorphan
- Diazepam
- Diclazepam
- Difluoroethane
- Dihydrocodeine
- Diltiazem
- Diphenhyramine
- Doxepin
- Doxylamine
- Duloxetine
- EDDP
- Ephedrine
- Ethyl Chloride
- Ethylene Glycol
- Ethylone
- Etizolam
- Fentanyl
- Fluoxetine
- Flurazepam
- Fluvoxamine

- Furanyl Fentanyl
- Gabapentin
- Gama-Hydroxybutyric Acid
- Glucophage
- Heroin
- Hydrocodone
- Hydromorphone
- Hydroxyzine
- Imipramine
- Isobutanol
- Isoflurane
- Ivermectin
- Ketamine
- Lamotrigine
- Lithium
- Loperamide
- Lorazepam
- MDMA
- Mephobarbital
- Meprobamate
- Metaxalone
- Methadone
- Methamphetamine
- Methanol
- Methoxyacetyl Fentanyl
- Methylphenidate
- Midazolam
- Mirtazapine
- Mitragynine
- Modafinil
- Morphine
- Naproxen
- N-Ethylpentylone
- Nordiazepam
- Norpropoxyphene
- Nortriptyline
- Olanzapine
- Opiates
- Oxazepam
- Oxycodone
- Oxymorphone
- Para-Fluorobutyryl Fentanyl/FIBF
- Paroxetine
- Phencyclidine
- Phendimetrazine
- Phenethylamine
- Phenobarbital
- Phentermine
- Phenylpropanolamine
- Phenytoin
- Primidone
- Promethazine
- Propofol
- Propoxyphene
- Pseuodoephedrine
- Quetiapine
- Quinine
- Remifentanil
- Salicylate
- Sertraline
- Temazepam
- Topiramate
- Tramadol
- Trans-3-Methylfentanyl
- Trazodone
- U-47700 Synthetic Opioid
- U-48800 Synthetic Opioid
- Valeryl Fentanyl
- Valproic Acid
- Venlafaxine
- Verapamil
- Zolpidem
- Zopiclone

Bibliography

[1] Data on Opioid Use and Overdose — Health Department — Allegheny County, 2019.

[2] Robert Andersen, Jacques A. Hagenaars, and Allan L. McCutcheon. Applied Latent Class Analysis. *Canadian Journal of Sociology / Cahiers canadiens de sociologie*, 2003.

[3] Marcus A. Bachhuber, Sean Hennessy, Chinazo O. Cunningham, and Joanna L. Starrels. Increasing benzodiazepine prescriptions and overdose mortality in the United States, 1996-2013, 4 2016.

[4] Lauren C. Balmert, Jeanine M. Buchanich, Janice L. Pringle, Karl E. Williams, Donald S. Burke, and Gary M. Marsh. Patterns and Trends in Accidental Poisoning Deaths: Pennsylvanias Experience 1979-2014. *PLOS ONE*, 11(3):e0151655, 3 2016.

[5] Joshua A. Barocas, Jianing Wang, Brandon D.L. Marshall, Marc R. LaRochelle, Amy Bettano, Dana Bernson, Curt G. Beckwith, Benjamin P. Linas, and Alexander Y. Walley. Sociodemographic factors and social determinants associated with toxicology confirmed polysubstance opioid-related deaths. *Drug and Alcohol Dependence*, 2019.

[6] Jeanine M. Buchanich, Lauren C. Balmert, Karl E. Williams, and Donald S. Burke. The effect of incomplete death certificates on estimates of unintentional opioid-related overdose deaths in the united states, 1999-2015. *Public Health Reports*, 2018.

[7] Center for Disease Control and Prevention. Drug Overdose Death Data. *Centers for Disease Control and Prevention*, 2017.

[8] Centers for Disease Control and Prevention. Understanding the Epidemic — Drug Overdose — CDC Injury Center, 2018.

[9] Kathleen E. Creppage, Joshua Yohannanms, Karl Williams, Jeanine M. Buchanich, Thomas J. Songer, Stephen R. Wisniewski, and Anthony Fabio. The rapid escalation of fentanyl in illicit drug evidence in allegheny county, pennsylvania, 2010-2016. *Public Health Reports*, 2018.

[10] Ed Day. Facing addiction in America: The Surgeon General's Report on Alcohol, Drugs, and Health U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES, OFFICE OF THE SURGEON GENERAL Washington, DC, USA: U.S. Department of Health and Human Services, 2016 382 pp. Online (grey literature): https://addiction.Surgeongeneral.Gov/. *Drug and alcohol review*, 37(2):283–284, 2 2018.

[11] Hawre Jalal, Jeanine M. Buchanich, Mark S. Roberts, Lauren C. Balmert, Kun Zhang, and Donald S. Burke. Changing dynamics of the drug overdose epidemic in the United States from 1979 through 2016. *Science*, 2018.

[12] William S. John and Li Tzy Wu. Trends and correlates of cocaine use and cocaine use disorder in the United States from 2011 to 2015. *Drug and Alcohol Dependence*, 180:376–384, 11 2017.

[13] Mansour Khoramizadeh, Mohammad Effatpanah, Alireza Mostaghimi, Mehdi Rezaei, Alireza Mahjoub, and Sara Shishehgar. Treatment of amphetamine abuse/use disorder: a systematic review of a recent health concern. *Daru : journal of Faculty of Pharmacy, Tehran University of Medical Sciences*, 6 2019.

[14] Kate L. Kingery. County Health Rankings & Roadmaps. *Journal of Youth Development*, 2018.

[15] Drew A. Linzer and Jeffrey B. Lewis. poLCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software*, 2011.

[16] Stephen J. Liu, Christina Mair, Thomas J. Songer, Elizabeth E. Krans, Abdus Wahed, and Evelyn Talbott. Opioid-related hospitalizations in Pennsylvania: A latent class analysis. *Drug and Alcohol Dependence*, 2019.

[17] Yiyang Liu, Amy L. Elliott, Mirsada Serdarevic, Robert F. Leeman, and Linda B. Cottler. A latent class analysis of the past-30-day substance use patterns among lifetime cocaine users: Findings from a community sample in North Central Florida. *Addictive Behaviors Reports*, 2019.

[18] Geoffrey J. Mclachlan and Suren Rathnayake. On the number of components in a Gaussian mixture model. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2014.

[19] Neerav Monga, Jrgen Rehm, Benedikt Fischer, Suzanne Brissette, Julie Bruneau, Nady El-Guebaly, Lina Noël, Mark Tyndall, Cameron Wild, Francesco Leri, Jean Sbastien Fallu, and Saira Bahl. Using latent class analysis (LCA) to analyze patterns of drug use in a population of illegal opioid users. *Drug and Alcohol Dependence*, 2007.

[20] Karen L Nylund, Tihomir Asparouhov, and Bengt O Muthén. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. Technical Report 4, 2007.

[21] Mariano Porcu and Francesca Giambona. Introduction to Latent Class Analysis With Applications. *Journal of Early Adolescence*, 2016.

[22] Bin Qin, Yuqing Zhang, Xinyu Zhou, Pengfei Cheng, Yiyun Liu, Jin Chen, Yuying Fu, Qinghua Luo, and Peng Xie. Selective serotonin reuptake inhibitors versus tricyclic antidepressants in young patients: a meta-analysis of efficacy and acceptability. *Clinical therapeutics*, 36(7):1087–1095, 7 2014.

[23] Jennifer Sabel, Jim Davis, Dagan Wright, and Svetla Slavova. Epi Tool to Analyze Overdose Death Data - Council of State and Territorial Epidemiologists.

[24] Kristin E. Schneider, Ju Nyeong Park, Sean T. Allen, Brian W. Weir, and Susan G. Sherman. Patterns of polysubstance use and overdose among people who inject drugs in Baltimore, Maryland: A latent class analysis. *Drug and Alcohol Dependence*, 2019.

[25] John E Schulenberg, Lloyd D Johnston, Patrick M O'malley, Jerald G Bachman, Richard A Miech, and Megan E Patrick. MONITORING FUTURE NATIONAL SURVEY RESULTS ON DRUG USE. Technical report, 1975.

[26] Svetla Slavova, Chris Delcher, Jeannine M. Buchanich, Terry L. Bunn, Bruce A. Goldberger, and Julia F. Costich. Methodological Complexities in Quantifying Rates of Fatal Opioid-Related Overdose. *Current Epidemiology Reports*, 6(2):263–274, 6 2019.

[27] University of Maryland Center for Substance Abuse Research. National Drug Early Warning System NFLIS Drug Report Dashboards, 2019.

[28] Diane K. Wysowski. Surveillance of prescription drug-related mortality using death certificate data. *Drug Safety*, 2007.

[29] Xirong Zheng, Linyue Shang, Chang-Guo Zhan, and Fang Zheng. In vivo characterization of toxicity of norcocaethylene and norcocaine identified as the most toxic cocaine metabolites in male mice. *Drug and alcohol dependence*, 204:107462, 8 2019.