Specific polysubstance use patterns predict relapse among patients entering opioid use disorder treatment

Yue Pan [a],[*], Daniel J. Feaster [a], Gabriel Odom [c], Laura Brandt [d], Mei-Chen Hu [b], Roger D. Weiss [e], John Rotrosen [f], Andrew J. Saxon [g], Sean X. Luo [b], Raymond R. Balise [a]

[a] *Department of Public Health Sciences, Division of Biostatistics, University of Miami, Office 1023, 1120 NW 14th St, Miami, FL 33136, USA*
[b] *Department of Psychiatry, Columbia University, USA*
[c] *Department of Biostatistics, Florida International University, USA*
[d] *Department of Psychology, The City College of New York, USA*
[e] *Department of Psychiatry, Harvard Medical School, Boston, and McLean Hospital, USA*
[f] *Department of Psychiatry, NYU Grossman School of Medicine, USA*
[g] *Department of Psychiatry and Behavioral Sciences, University of Washington, USA*

A R T I C L E   I N F O

A B S T R A C T

Introduction: While polysubstance use has consistently been associated with higher rates of relapse, few studies have examined subgroups with specific combinations and time course of polysubstance use (i.e., polysubstance use patterns). This study aimed to classify and compare polysubstance use patterns, and their associations with relapse to regular opioid use in 2637 participants in three large opioid use disorder (OUD) treatment trials.
Methods: We explored the daily patterns of self-reported substance use in the 28 days prior to treatment entry. Market basket analysis (MBA) and repeated measure latent class analysis (RMLCA) were used to examine the subgroups of polysubstance use patterns, and multiple logistic regression was used to examine associations between identified classes and relapse.
Results: MBA and RMLCA identified 34 "associations rules" and 6 classes, respectively. Specific combinations of polysubstance use and time course (high baseline use and rapid decrease of use prior to initiation) predicts a worse relapse outcome. MBA showed individuals who co-used cocaine, heroin, prescription opioids, and cannabis had a higher risk for relapse (OR = 2.82, 95%CI = 1.13, 7.03). In RMLCA, higher risk of relapse was observed in individuals who presented with high baseline prescription opioid (OR = 1.9, 95% CI = 1.3, 2.76) or heroin use (OR = 3.54, 95%CI = 1.86, 6.72), although use decreased in both cases prior to treatment initiation.
Conclusions: Our analyses identified subgroups with distinct patterns of polysubstance use. Different patterns of polysubstance use differentially predict relapse outcomes. Interventions tailored to these individuals with specific polysubstance use patterns prior to treatment initiation may increase the effectiveness of relapse prevention.

1. Introduction

The staggering impact of opioid use disorder (OUD) is compounded by polysubstance use (Cicero et al., 2020; Compton et al., 2021). While estimates on the percentage of people with OUD who use multiple substances vary, the most recent estimates suggest that polysubstance use is the norm in people with OUD (Cicero et al., 2020; Makarenko et al., 2018; Soyka et al., 2017). Rates of polysubstance use among people in treatment for OUD range between 65% (Jarlenski et al., 2017) to 85% (Raffa et al., 2007). While some overdoses occur in the context of unintentional ingestion of multiple substances, such as carfentanil mixed with heroin, other people actively seek out multiple drugs (Ataiants et al., 2020). However, little is known about the patterns in

which various types of drugs are used. Some combinations, such as opioids with benzodiazepines, are frequently used together and likely contribute to overdose events (Hernandez et al., 2018; Seth et al., 2018). Less is known about other combinations, such as opioids with marijuana or opioids with alcohol, and their impact on treatment response is unclear (Hassan and Le Foll, 2019; Wagner et al., 2018). There is a need to better characterize which substances are used together and whether the patterns of use change prior to engagement in treatment.

Polysubstance use researchers focusing on drugs other than opioids, such as cannabis (Connor et al., 2013, 2014) and tranquilizers (Votaw et al., 2020), have used a classic statistical method, Latent Class Analysis (LCA) (Collins and Lanza, 2009), to identify drug use patterns. An extension of LCA, repeated measures latent class analysis (RMLCA),

---

considers repeated drug exposure patterns and how they cluster through time (Collins and Lanza, 2009). Knowing which drugs are used together and how this does or does not change over time allows for identifying patterns in multiple comorbid conditions, such as depression and anxiety frequently co-occurring in "wide-range substance users." Understanding these patterns may, in turn, lead to more targeted treatments (Carlsen et al., 2020).

While methods like LCA are useful for finding such patterns, particularly when dealing with a small number of grouping features (e.g., 6 to 12 drugs or classes of drugs), modern machine learning methods such as market basket analyses (MBA) are adept at identifying patterns when there are many features (Lantz, 2019). The method has been applied to identify co-occurring patterns in a host of domains, such as detecting patterns of genes associated with disease, co-occurring food allergies, and even detecting tactical patterns in elite beach volleyball (Aguinis et al., 2013; Wenninger et al., 2019). This ability to detect patterns, even in the presence of rarely used drugs out of a pool of many possible substances, makes MBA a useful tool for studying polysubstance use. MBA, as opposed to RMLCA, focuses on polysubstance use over short periods (e.g., in the same day).

Although studies have documented the rate of polysubstance use and/or reported on drugs frequently used together with opioids (Hser et al., 2017; Northrup et al., 2015; Ruglass et al., 2019), few have looked at the fine details of drug use patterns and how substance use changes through time in people with OUD before entering treatment. Here we explore the daily patterns of substance use self-reported in the "Timeline Followback" in the 28 days before the onset of Medication for Opioid Use Disorder (MOUD) in three National Institute on Drug Abuse (NIDA) Clinical Trials Network (CTN) studies. Data from these three CTN studies were harmonized as part of a fourth CTN study (CTN-0094) to allow more extensive secondary analyses of OUD treatment. Different from the previous studies that model the trajectories of outcomes (Hser et al., 2017; Northrup et al., 2015; Ruglass et al., 2019), we focused on modeling trajectories of substance use prior to treatment initiation as predictors for relapse. We report patterns of drug use using MBA and RMLCA. MBA is useful to identify patterns of all individual drugs used on any given day, while RMLCA allows exploring patterns over longer periods of time. We examine daily reports of use across 12 broad classes of substances over the 28 days before starting MOUD. Associations between selected RMLCA classes with relapse outcomes were also examined. We hypothesized that subgroups of polysubstance use, with different prevalence, can be identified by MBA and RMLCA, and that some of these patterns are more likely than others to be associated with relapse to regular opioid use during treatment.

2. Material and methods

2.1. Study population

Full protocols for the individual studies were previously published (Lee et al., 2016; Saxon et al., 2013; Weiss et al., 2011). Briefly, all three studies enrolled individuals who met the criteria of DSM-IV-TR for opioid dependence (CTN-0027/CTN-0030) or DSM-5 diagnosis of OUD (CTN-0051). The trials enrolled adult participants over 18, with very broad pragmatic inclusion and few exclusions except for major medical and unstable psychiatric comorbidities. CTN-0027 was the most inclusive, whereas CTN-0030 excluded individuals with OUD who only used heroin, and CTN-0051 excluded individuals currently receiving methadone treatment. CTN-0027 randomized individuals to buprenorphine and methadone for outpatient treatment for 24 weeks. CTN-0030 randomized individuals to two different types of psychotherapy programs, and all participants received identical medication treatment in two phases: in the initial phase, participants received a buprenorphine taper, and those who relapsed (about 90% of all participants), were treated with buprenorphine maintenance over 12 weeks in the

Table 1

Structured drug use variables from three harmonized studies and how they were categorized for repeated measure latent class analyses.

Name	CTN-0027	CTN-0030	CTN-0051	Drug category
Alcohol	Yes	Yes	Yes	Alcohol
Heavy Drinking	*	**	Yes	
Cannabinoids	Yes	Yes	Yes	Cannabis
Cocaine	Yes	Yes	Yes	Cocaine
Crack	*	No	Yes	Cocaine
Amphetamine	Yes	Yes	Yes	Amphetamine
Methamphetamine	Yes	Yes	No	Amphetamine
Opiates	Yes	Yes	Yes	Opioid
Benzodiazepines	Yes	Yes	Yes	Benzodiazepines
Propoxyphene	Yes	Yes	No	Opioid
Oxycodone	Yes	Yes	No	Opioid
Other	Yes	***	Yes	
Morphine	No	Yes	No	Heroin
Hydromorphone	No	Yes	No	Heroin
Codeine	No	Yes	No	Opioid
Hydrocodone	No	Yes	No	Opioid
Sedatives Not Benzo	No	Yes	Yes	Depressants
Heroin/Opium	No	Yes	Yes	Heroin
Hallucinogens/MDMA	No	No	Yes	Hallucinogens
Buprenorphine	No	No	Yes	
Inhalants	No	No	Yes	

* Can be calculated.
** Only a count of days.
*** If yes to opiates, then follow up for which drug.

second phase. CTN-0051 randomized patients from inpatient treatment facilities and they either received buprenorphine or extended-release naltrexone after discharge and were followed as outpatients for 24 weeks.

2.2. Substance use measures

We focused our analysis on polysubstance use prior to treatment initiation as a baseline predictor of treatment outcomes. For all participants who enrolled in the above three studies, substance use history in the 28 days prior to treatment initiation was surveyed with the Timeline Followback method (Sobell and Sobell, 1992). The free text drug names were harmonized to account for spelling errors and differences in names (generic, brand versus street) and to remove superfluous information (e.g., mg/capsule/oz), resulting in a set of 44 distinct substances. These data included both high-frequency drugs of use which were also identified in structured questions and low-frequency substances of use (e.g., cathinones $N = 5$ events), drugs of concern (e.g., gabapentin $N = 21$), as well as other drugs such as antiemetics ($N = 18$), non-opioid pain killers (e.g., acetaminophen $N = 5$) and antipsychotics ($N = 3$). These values were used for MBA. To allow LCA to operate on a more tractable set of options, these harmonized drug names were grouped into twelve "drug categories." Table 1 lists the drugs that were surveyed as part of structured questions in the three studies and how they were grouped for LCA. Other repeatedly occurring drugs were: fentanyl and opium (grouped-with/classified-as Heroin for LCA); GHB (Gamma hydroxybutyrate) (Depressants); K2 (Synthetic cannabinoids); merperidine, tramadol, oxymorphone (Opioids); barbiturate, sedative-hypnotic (Depressants); MDMA (3,4-Methyl enedioxy methamphetamine), hallucinogen unspecified (Hallucinogens); muscle relaxant unspecified, soma, methocarbamol, flexeril, baclofen, carisoprodol (Relaxant).

2.3. Relapse

Relapse was defined by four consecutive opioid use weeks between 21 days after treatment initiation and the end of the 12-week treatment period. An opioid use week was defined as having either a urine drug screen (UDS) positive for any non-prescribed opioid or a miss-

ing/refused UDS in that week. We defined relapse as a categorical variable representing three possible outcomes: 0 indicated that the definition of relapse was not met (50% of the sample), 1 indicated at least one opioid-positive UDS during four consecutive weeks with positive or missing UDSs (indicating that the participant, albeit positive for nonprescribed opioids, showed up to the clinic at least once; 29% of the sample), and 2 indicated missing UDS data for four consecutive weeks (21% of the sample). This definition closely followed a more complex definition from CTN-0051, but only required UDS results and not the self-reported drug use information gathered on the Timeline Followback.

2.4. Statistical analysis

MBA was originally designed to guide business decisions using massive datasets, such as all the transactions in a supermarket for a month. In these scenarios, traditional hypothesis testing with p-values becomes meaningless because anything is statistically significant with a big enough sample size. Instead of focusing on p-values, MBA methods rely on three statistics to assess the importance of an association: lift, support, and confidence. These measures assess the chances of seeing drug A and drug B (or a set of drugs which we will label A and a different set that we will label as B) appearing on a drug screening on the same day. Lift, which assesses if an association exists, is calculated as $\frac{P(A \cap B)}{P(A)*P(B)}$. It is the ratio of the actual probability of drug A and drug B occurring on a given day divided by the probability of seeing drug A on any day times the probability of seeing drug B on any day. In other words, it is the increased or decreased probability of seeing both drugs relative to their overall chance of either one appearing (assuming their co-occurrence is random). Support is synonymous with probability of seeing the various sets of drugs being examined. MBA practitioners focus on the support for A and B, $P(A \cap B)$, that is, the probability of seeing both items on the same day. Support is of decreasing utility as the number of possible drugs in the set increases. Confidence, which describes the probability that a set of items appears given that one has already occurred, is calculated as $\frac{P(A \cap B)}{P(A)}$. In other words, it is the conditional probability of seeing drug B given that you see A, $P(B|A)$.

Using these concepts, it is possible to think of predictive association rules where A leads to B (which is written as A -> B). It is useful to think of the relationship between lift and confidence: $lift(A \rightarrow B) = \frac{P(B|A)}{P(B)}$. Importantly, A and B are not interchangeable in the confidence equations. That is, the chances of seeing chips after seeing guacamole (guacamole -> chips) is not the same as the chances of seeing guacamole after seeing chips (chips -> guacamole). This "directional" property allows MBA to make predictions that can describe the greatly increased risk of drug B if the person has used drug A, but seeing drug B may not increase the chances of seeing drug A.

For MBA, the a priori algorithm was used after excluding drugs with a support of less than 0.001. That is, drugs that occurred with a frequency of less than twice a month per transaction (28*2/56,000) were excluded. The algorithm was set to extract all rules with a confidence of at least 0.5. We then selected the top ten rules based on the lift (i.e., ratio of the support of the antecedent drugs co-occurring with the consequent drugs, divided by the probability that the antecedent and consequent drugs co-occur if the two are independent) and created a binary class to indicate the particular basket of polydrug use. Participants who reported using all substances in the basket rule during the study period were identified and categorized into different polydrug baskets.

RMLCA for the drugs used in the 28 days leading up to MOUD initiation was conducted using the drug groups described above. First, class enumeration was done without covariates by estimating models with increasing numbers of classes until the sample size in each latent class was considered to be too small for practical interpretation (less than 5% of the total sample size) and/or information criteria showed worse model fit. To ensure the models converged to the

global maximum, 1000 random starts and 100 replicated likelihoods were used for each model. Model fit statistics for each of the models were used to determine the model that best fit the data, including entropy (Celeux and Soromenho, 1996) and penalized information criteria (Bayesian Information Criteria (BIC), Akaike's Information Criteria (AIC) (Vrieze, 2012), Vuong-Lo-Mendell-Rubin, Lo-Mendell-Rubin adjusted LRT tests (Nylund et al., 2007), and bootstrapped likelihood ratio tests (BLRT) (Asparouhov and Muthén, 2014; Feng and McCulloch, 1996). Second, once the best-fitting class structures were determined, a 3-step approach using the R3STEP (Asparouhov and Muthén, 2014) procedure was used to examine between-class differences in the relapse covariate using multinomial logistic regression. The 3-step procedure ensures that the inclusion of the covariates does not change the class structure.

Logistic regressions were used to test for the effects of RMLCA classes and MBA top lift baskets on relapse controlling for which of the three trials, and treatment arms within the trials. Odds ratios (OR) and 95% confidence intervals (CIs) were reported. P-values < 0.05 for two-sided tests were considered statistically significant. Demographic and substance use data from CTN-0027, 0030, and 0051 were harmonized using SAS 9.4. Exploratory data analyses and MBA were conducted with R 3.6.2 with packages including haven (Wickham and Miller, 2021) (version 2.2), tidyverse (Wickham et al., 2019) (version 1.3), arules (Hahsler et al., 2022, 2011; Hornik et al., 2005) (version 1.6–5), and arulesViz (Hahsler, 2017; Hahsler and Chelluboina, 2021) (version 1.3–3). Haven was used to export the data from SAS, and tidyverse was used for data cleaning. Arules and arulesViz were used for MBA. RMLCA analyses were conducted in Mplus (version 8.3).

3. Results

3.1. Parent study and participants

All people who provided drug use Timeline Followback from CTN-0027 (*N* = 1300), CTN-0030 (*N* = 661) and CTN-0051 (*N* = 676) were included in this study. Participants typically provided full drug use information for the 28 days prior to initiation of MOUD, and 97.9% provided more than three weeks of history. Despite different participant selection criteria and differences in available data across the three trials, Table 2 shows similar patterns in terms of baseline demographic and psychiatric co-morbidities.

3.2. Market basket

The (*N* = 2637) participants reported using at least one drug on 66,098 (89%) days in the 28 days prior to MOUD initiation and randomization. Fig. 1 shows the top 10 frequently used substances. The most commonly occurring substances were heroin (reported on *N* = 40,740 days), oxycodone (*N* = 12,289 days), cannabis (*N* = 9183 days), hydrocodone (*N* = 7336 days), and methadone (*N* = 3663 days). Almost two-thirds (64%) of drug use days involved using a single substance (mean drugs used per day = 1.45). Participants used two substances on 29%, three substances on 6%, and four substances on 6% of drug use days. One person reported using 13 substances on a single day.

The a priori MBA algorithm identified 34 association rules. As can be seen in Table 3, these rules included extremely strong predictions for rare events, for example, a 16-fold increase (with 71% confidence) in the chance of seeing heroin with the use of opium (lift = 16, count = 69). Predictions were also strong for relatively more common events, such as using heroin with crack (lift = 1.46, count = 2884) or cocaine (lift = 1.35, count = 2598). The algorithm noted several large effect-size combinations, such as the 13-fold increase in the chances of reported cocaine use when a combination of heavy drinking, opioid, and cannabis was observed, and a 6.13-fold increase in the chances of cannabis in a combination of cocaine, heavy drinking, and opioid.

Table 2
Demographic details on 2637 participants from three clinical trials, who self-reported drug use in the 28 days before initiation of medication for opioid use disorder and randomization.

	CTN-0027 (N = 1300)	CTN-0030 (N = 661)	CTN-0051 (N = 676)	Overall (N = 2637)
Age				
Mean (SD)	36.9 (11.1)	32.6 (10.2)	33.9 (9.63)	35.1 (10.7)
Median [Min, Max]	35.5 [18.0, 67.0]	30.0 [18.0, 77.0]	31.0 [19.0, 67.0]	33.0 [18.0, 77.0]
Missing	0 (0%)	0 (0%)	106 (15.7%)	106 (4.0%)
Sex				
Female	420 (32.3%)	265 (40.1%)	199 (29.4%)	884 (33.5%)
Male	880 (67.7%)	396 (59.9%)	477 (70.6%)	1753 (66.5%)
Race				
Black	131 (10.1%)	22 (3.3%)	88 (13.0%)	241 (9.1%)
Other	236 (18.2%)	35 (5.3%)	80 (11.8%)	351 (13.3%)
Refused/missing	6 (0.5%)	1 (0.2%)	8 (1.2%)	15 (0.6%)
White	927 (71.3%)	603 (91.2%)	500 (74.0%)	2030 (77.0%)
Ethnicity				
Hispanic	212 (16.3%)	33 (5.0%)	118 (17.5%)	363 (13.8%)
Not Hispanic	1088 (83.7%)	628 (95.0%)	558 (82.5%)	2274 (86.2%)
Education				
Less than HS	46 (3.5%)	100 (15.1%)	130 (19.2%)	276 (10.5%)
HS/GED	61 (4.7%)	261 (39.5%)	251 (37.1%)	573 (21.7%)
More than HS	58 (4.5%)	298 (45.1%)	232 (34.3%)	588 (22.3%)
Missing	1135 (87.3%)	2 (0.3%)	63 (9.3%)	1200 (45.5%)
Marital Status				
Married or Partnered	26 (2.0%)	186 (28.1%)	86 (12.7%)	298 (11.3%)
Never married	87 (6.7%)	332 (50.2%)	447 (66.1%)	866 (32.8%)
Separated/Divorced/Widowed	52 (4.0%)	139 (21.0%)	140 (20.7%)	331 (12.6%)
Unknown	1135 (87.3%)	4 (0.6%)	3 (0.4%)	1142 (43.3%)
Bipolar				
No	1147 (88.2%)	622 (94.1%)	548 (81.1%)	2317 (87.9%)
Yes	147 (11.3%)	38 (5.7%)	88 (13.0%)	273 (10.4%)
Missing	6 (0.5%)	1 (0.2%)	40 (5.9%)	47 (1.8%)
Schizophrenia				
No	1260 (96.9%)	653 (98.8%)	627 (92.8%)	2540 (96.3%)
Yes	33 (2.5%)	8 (1.2%)	9 (1.3%)	50 (1.9%)
Missing	7 (0.5%)	0 (0%)	40 (5.9%)	47 (1.8%)
Major Depression				
No	941 (72.4%)	440 (66.6%)	439 (64.9%)	1820 (69.0%)
Yes	354 (27.2%)	220 (33.3%)	197 (29.1%)	771 (29.2%)
Missing	5 (0.4%)	1 (0.2%)	40 (5.9%)	46 (1.7%)
Anxiety/Panic Disorder				
No	902 (69.4%)	447 (67.6%)	350 (51.8%)	1699 (64.4%)
Yes	392 (30.2%)	213 (32.2%)	286 (42.3%)	891 (33.8%)
Missing	6 (0.5%)	1 (0.2%)	40 (5.9%)	47 (1.8%)
Brain Damage				
No	1172 (90.2%)	543 (82.1%)	560 (82.8%)	2275 (86.3%)
Yes	120 (9.2%)	118 (17.9%)	76 (11.2%)	314 (11.9%)
Missing	8 (0.6%)	0 (0%)	40 (5.9%)	48 (1.8%)
Epilepsy				
No	1254 (96.5%)	639 (96.7%)	584 (86.4%)	2477 (93.9%)
Yes	41 (3.2%)	22 (3.3%)	52 (7.7%)	115 (4.4%)
Missing	5 (0.4%)	0 (0%)	40 (5.9%)	45 (1.7%)

While the majority of the rules (76%) predicted heroin use after consuming other drugs, 18% of the rules were associated with cannabis use.

3.3. RMLCA

RMLCA with three to eight classes was fit to all the harmonized data. Fit indices for each LCA model are presented in Appendix Table A1. The Vuong-Lo-Mendell-Rubin test and Lo-Mendell-Rubin adjusted LRT test, both with p-values of 0.811, suggested that a six-class solution was an adequate fit and that seven classes were not needed. However, the bootstrapped parametric likelihood ratio test, with a p-value < 0.0001, suggested seven classes. Given the large improvement in adjusted BIC between the five and the six-class solutions and the fact that the seven-class solution has some classes with smaller size groups, we selected the six-class solution (Table A1).

The daily probability and prevalence of drug use by latent class for the six-class solution are shown in Fig. 2 (by days) and Fig. 3 (by drug use). More comprehensive figures by days, and by drug use, stratified by the three studies are provided as in the appendix (Figs. A1–A6). The largest class was C1-"All time low" (n = 829, 31.4%), which describes a group of patients who used few substances, including any opioids, in the 28 days prior to treatment initiation. This group primarily included participants from CTN-0030 (n = 466), CTN-0051 (n = 208) and CTN-0027 (n = 155). The second largest class was C2-High opioid decreasing (n = 824, 31.3%). This class was comprised of individuals who had a self-initiated decrease of prescription opioids in the 28 days prior to the date of assessment on treatment entry. This group likely represents individuals who voluntarily decreased opioid intake in preparation for either inpatient (CTN-0051, n = 13) or outpatient treatment (CTN-0027 n = 811); no CTN-0030 participants were represented in this class. As can be seen in the corresponding figures, across all classes except the low substance use group, there was a precipitous drop-off in substance use in the two weeks leading up to treatment initiation, particularly heroin, methadone, opioid, and cannabis use. However, the degree of reduction differed by drug.

Table 3
Market basket analysis result, sorted by lift, shows which drug is predicted (C is consequent) given the presence of other drugs (+ shows antecedents) which were used on the same day.

Amphetamine	Benzodiazepine	Cocaine	Crack	Heavy Drinking	Heroin	Light Drinking	Methamphetamine	Opioid	Opium	Cannabis	Lift	Confidence	Coverage	Support	Count
		+		+				C		+	16	0.71	0.001	0.001	69
		C		+				+		+	13	0.61	0.002	0.001	69
		+		+				+		C	6.1	0.85	0.001	0.001	69
	+				+			+		C	4.8	0.67	0.002	0.001	85
		+			+			+		C	4.8	0.66	0.002	0.001	71
	+	+			+					C	4.7	0.65	0.002	0.001	89
		+						+		C	4.3	0.6	0.004	0.002	147
	+	+								C	4.1	0.56	0.002	0.001	90
					C				+		1.6	1	0.001	0.001	97
	+	+			C					+	1.6	0.99	0.001	0.001	89
	+				C	+					1.5	0.92	0.001	0.001	71
	+			+	C					+	1.5	0.92	0.001	0.001	71
	+		+		C						1.5	0.92	0.002	0.002	130
			+		C						1.5	0.9	0.048	0.044	2884
		+			C	+					1.4	0.88	0.002	0.001	93
	+	+			C						1.4	0.85	0.002	0.002	136
			+		C					+	1.4	0.85	0.005	0.005	301
		+			C						1.4	0.83	0.047	0.039	2598
					C	+					1.3	0.81	0.027	0.022	1431
					C		+				1.3	0.81	0.008	0.007	433
					C	+				+	1.3	0.8	0.004	0.003	212
+					C						1.3	0.8	0.016	0.013	851
			+		C	+					1.3	0.78	0.002	0.002	116
			+	+	C						1.3	0.77	0.002	0.002	122
+					C					+	1.2	0.72	0.003	0.002	129
		+			C					+	1.2	0.72	0.008	0.005	357
	+		+		C						1.1	0.69	0.004	0.002	163
			+		C						1.1	0.67	0.032	0.021	1415
		+	+		C						1.1	0.66	0.003	0.002	140
			+		C					+	1.1	0.65	0.006	0.004	272
					C						1	0.62	1.000	0.616	40,740
	+				C			+		+	1	0.59	0.002	0.001	85
					C					+	0.9	0.55	0.139	0.076	5011
	+				C					+	0.9	0.53	0.009	0.005	317

* +: Antecedent; C: Consequent.
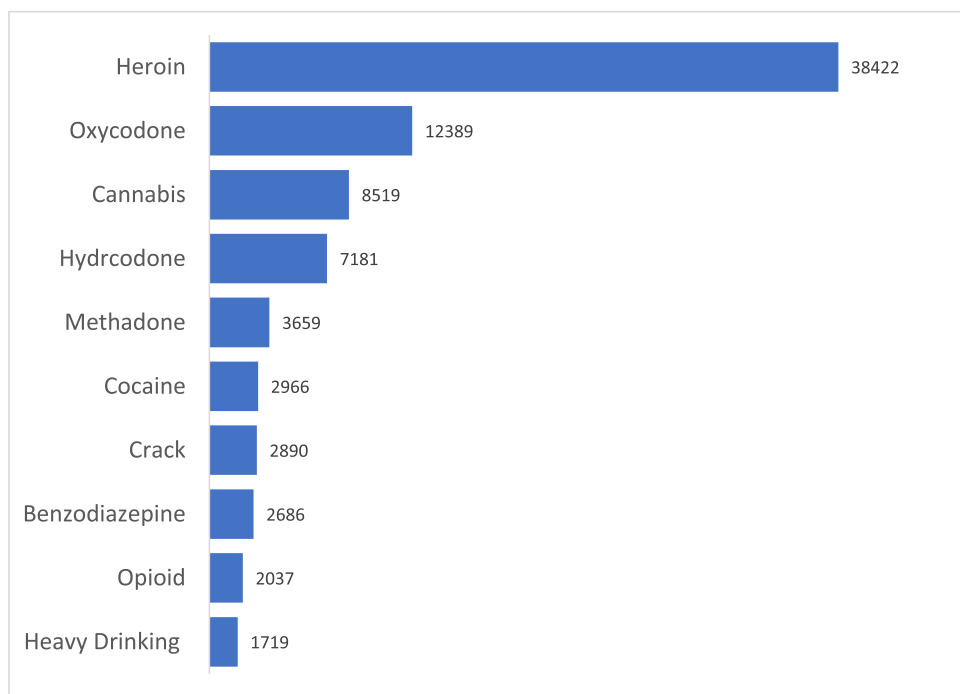
Fig. 1. Substance use frequency (Top 10).

Table 4
Association between relapse and Market Basket top 10 lift baskets.

OR (95% CI)	Relapse* Yes vs. No	Missing vs. No
MB Top 10 Lift Baskets		
Cocaine, Heavy Drinking, Cannabis, Opioid ($n = 218$, 8.6%)	0.76 (0.53, 1.09)	**0.44 (0.25, 0.78)**
Benzodiazepine, Heroin, Opioid, Cannabis ($n = 624$, 24.6%)	1.43 (0.98, 2.09)	**0.58 (0.39, 0.86)**
Cocaine, Heroin, Opioid, Cannabis ($n = 694$, 27.4%)	**2.82 (1.13, 7.03)**	**5.13 (1.44, 18.22)**
Benzodiazepine, Cocaine, Heroin, Cannabis ($n = 487$, 19.2%)	1.21 (0.39, 3.69)	0.54 (0.14, 2.17)
Cocaine, Opioid, Cannabis ($n = 806$, 31.8%)	1.3 (0.53, 3.21)	0.34 (0.09, 1.29)
Benzodiazepine, Cocaine, Cannabis ($n = 581$, 22.9%)	0.64 (0.23, 1.77)	1.37 (0.38, 4.94)
Opium, Heroin ($n = 84$, 3.3%)	1.16 (0.58, 2.31)	1.82 (0.68, 4.84)

* Relapse defined as Yes=at least one positive Urine Drug Screen (UDS) during four consecutive weeks with positive or missing UDS; No=No consecutive missing or positive for four weeks; Missing=Missing UDS for four consecutive weeks. Odds ratios (OR) and 95% confidence intervals (CIs) were reported. P-values < 0.05 for two-sided tests were considered statistically significant and are in bold font.

3.4. Association between relapse and top 10 lift baskets

We selected the top 10 lift rules based on the MBA (Table 4). Individuals co-using cocaine, heroin, opioids, and cannabis had a higher risk of having a relapse event defined by the presence of at least one opioid-positive UDS (OR = 2.82, 95% CI = 1.13, 7.03) or exclusively missing UDS for four consecutive weeks (OR = 5.13, 95% CI = 1.44, 18.22) compared to individuals who did not report use of this particular drug combination, controlling for other polydrug baskets and trial. Individuals who reported polydrug use of cocaine, heavy drinking, cannabis, and opioids had a lower risk of relapse defined by four consecutive missing UDS (OR = 0.44, 95% CI = 0.25, 0.78). Similarly, we identified *decreased* odds of relapse for individuals who reported polydrug use of benzodiazepine, heroin, opioid, and cannabis (OR = 0.58, 95% CI = 0. 39, 0.86). The number of participants with polydrug patterns represented by the top 10 lift baskets is listed in Appendix Table 2.

3.5. Association between relapse and RMLCA classes

Table 5 shows the association between relapse and the six latent classes. Compared to the C1 "All-time low" class, individuals in the C2

Table 5
Association between relapse and classes.

OR (95% CI)	Relapse* Yes vs. No	Missing vs. No
Classes		
1. All-Time Low	ref	ref
2. High Opioid Decrease	**1.9 (1.3, 2.76)**	1.05 (0.71, 1.56)
3. High Heroin Decrease	**3.54 (1.86, 6.72)**	1.5 (0.97, 2.31)
4. Mid Opioid, Alcohol Decrease	0.92 (0.6, 1.43)	0.74 (0.47, 1.17)
5. High Cannabis, Mid Opioid Decrease	0.88 (0.54, 1.42)	1.1 (0.68, 1.79)
6. High Methadone Decrease	0.54 (0.29, 1.03)	1.54 (0.9, 2.65)

* Relapse defined as Yes=at least one positive Urine Drug Screen (UDS) during four consecutive weeks with positive or missing UDS; No=No consecutive missing or positive for four weeks; Missing=Missing UDS for four consecutive weeks. Odds ratios (OR) and 95% confidence intervals (CIs) were reported. P-values < 0.05 for two-sided tests were considered statistically significant and are in bold font.

"High opioid decreasing" class had a higher risk of relapse defined by the presence of at least one opioid-positive UDS (OR = 1.9, 95% CI = 1.3, 2.76). Similarly, C3 "High Heroin decreasing" had a higher risk of relapse (OR = 3.54, 95% CI = 1.86, 6.72) compared to the C1 class. How-
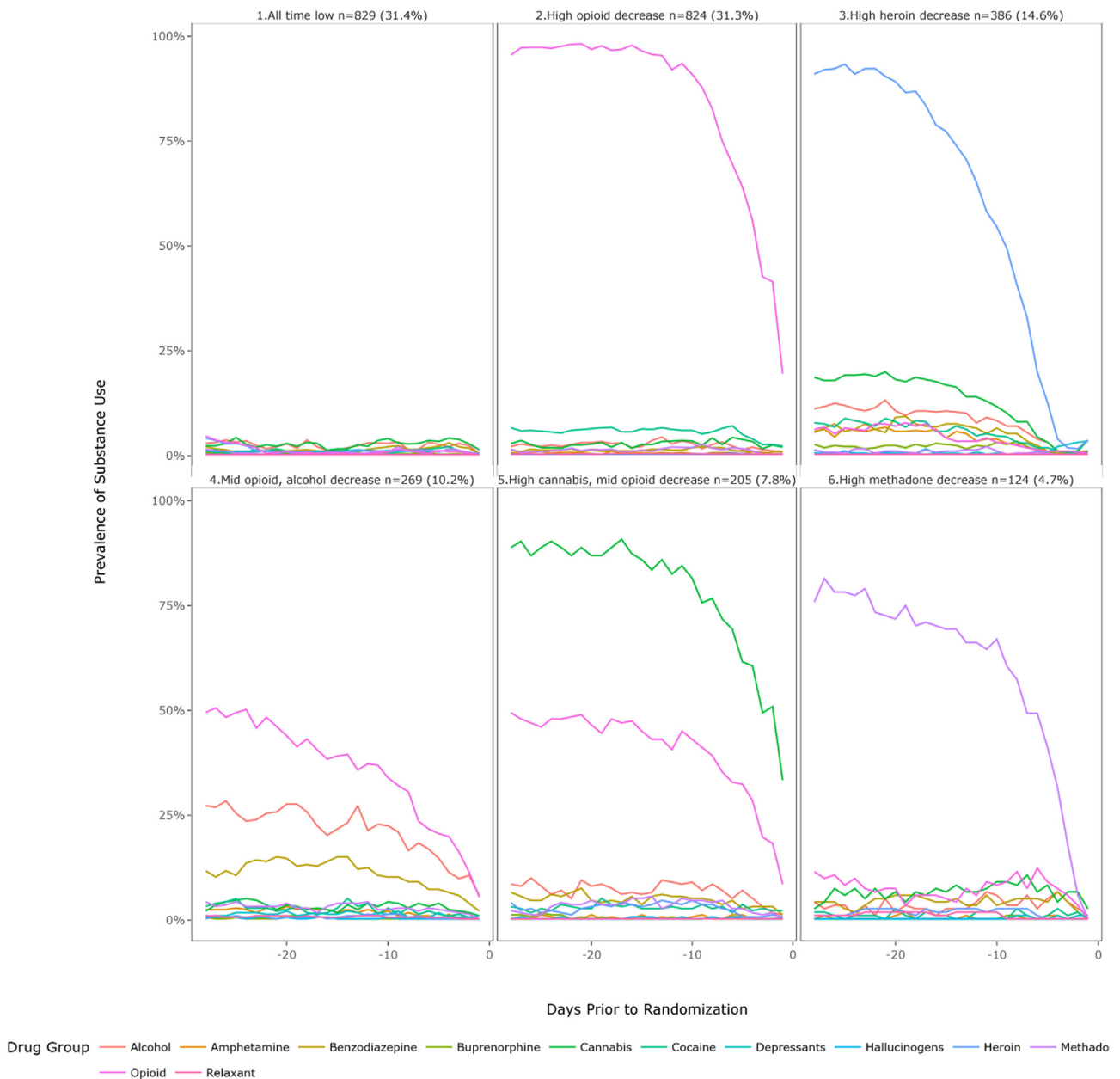
**Fig. 2.** Six class solution by RMLCA class (Overall).

ever, there were no significant associations between C4-C6 classes and relapse risk.

4. Discussion

This study provides novel evidence on patterns of polysubstance use in the immediate period prior to treatment initiation among individuals with OUD. Polysubstance use was common, and both RMLCA and MBA offer valuable insights into the specific patterns. RMLCA is a repeated measures extension of LCA and person-centered data analytic technique. Therefore, this approach helps identify latent patterns of responses to categorical items with varying probabilities of endorsement. Compared with other data segmentation methods, such as hierarchical clustering, RMLCA derives clusters using a formal probabilistic approach and can be used in conjunction with multivariate methods to estimate parameters. It helps determine how many patterns of responses/behavior are present in the data, how prevalent each pattern is, and how likely item endorse-

ment is in each latent class. The optimal number of classes minimizes the degree of relationship among cases belonging to different classes. To select the optimal number, methods such as the Bayesian Information Criterion are used, which capitalize on the value of the negative log-likelihood function, a well-established measure of the goodness of fit of a statistical model. MBA was used to reveal the most common combinations of substances used together on the same day and it affords additional insights that are otherwise masked by broad drug classes needed for traditional LCA.

Our results indicate, for example, that opium users were also using heroin, alcohol, and cannabis, and there were several unidirectional relationships where the use of specific drugs (e.g., benzodiazepine, cocaine, and cannabis) was likely to lead to heroin use. RMLCA identified six broad classes of polysubstance use patterns prior to treatment initiation and revealed that a large number of patients either had shown low use of all substances, including opioids or substantially reduced use prior to treatment initiation. Further, both methods used in our anal-
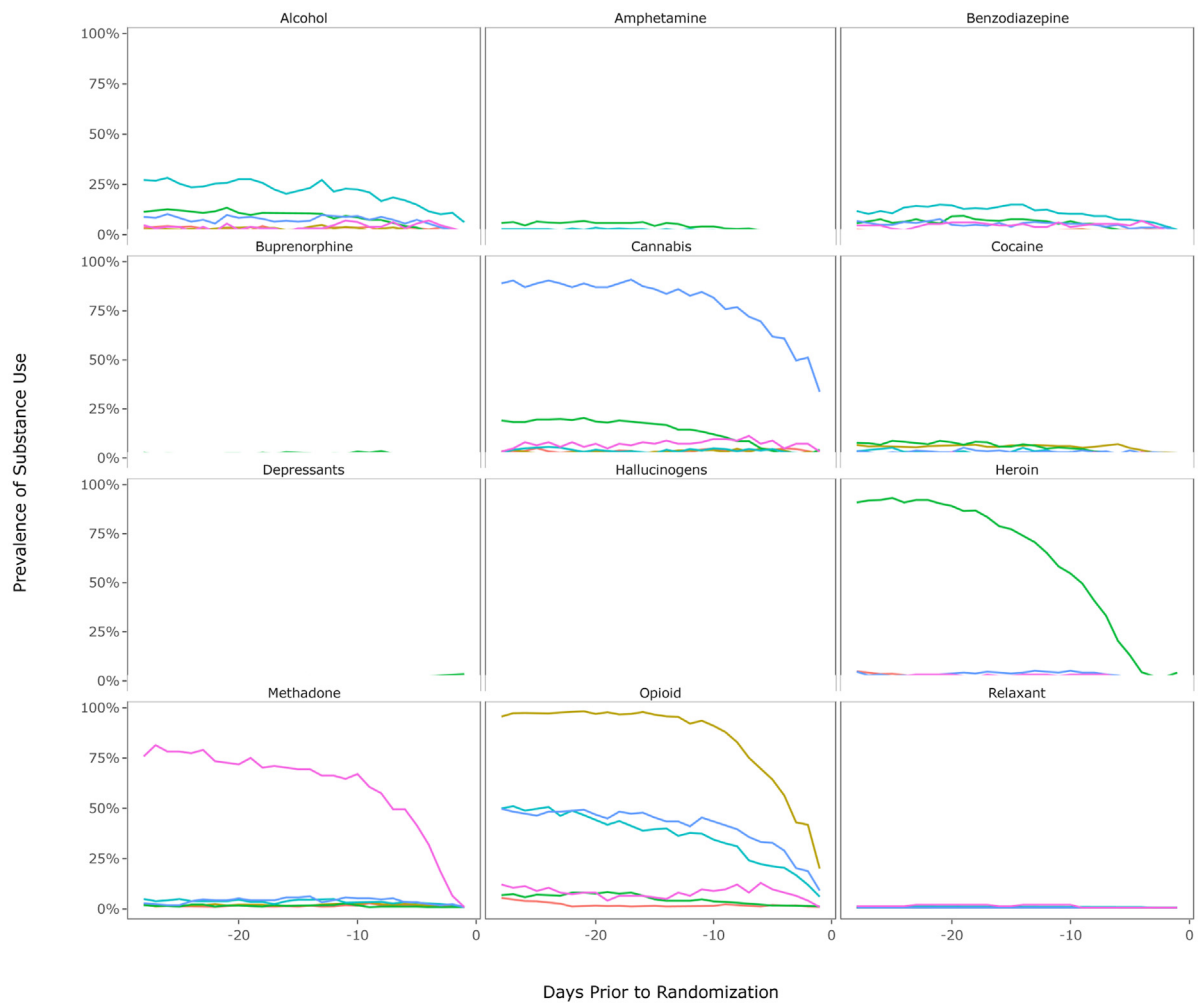
Fig. 3. Six class solution by drug use (Overall).

yses revealed subgroups of polysubstance use patterns associated with relapse outcomes.

We found some unexpected associations between specific MBA baskets and relapse by week 12: the observance of a combination of cocaine, cannabis, and heavy drinking with prescription opioids did not negatively impact the odds of eventual relapse but instead was correlated with decreased odds of relapse. In contrast, the appearance of the same combination with heroin predicted a marked increase in the odds of relapse. This suggests that a polysubstance "binge" is not automatically associated with relapse to regular opioid use or non-attendance at scheduled clinic visits (i.e., missed UDS). In fact, the occurrence of a polysubstance heroin binge portends a worse prognosis, and a polysubstance binge with prescription opioids much less so.

Most importantly, the different RMLCA classes and MBA polydrug use groups had different relationships with relapse. For example, RMLCA revealed that people in both the "high opioid decreasing" and "high heroin decreasing" classes were more likely to relapse to regular

opioid use. The implication is that even if patients decrease their use of opioids prior to treatment engagement, their risk of relapse is still substantially higher compared to those who start with lower levels of use or who have more sporadic usage patterns. MBA also showed that there were subgroups among opioid users and one of the subgroups who also consumed heroin, cocaine, and cannabis, was more likely to have a potentially worse outcome—consecutive non-attendance at scheduled clinic visits (i.e., missed UDSs) for four weeks between 21 days after randomization and the end of the 12-week treatment period. These results demonstrate that there are quantifiably different subgroups among polydrug users associated with clinically meaningful differences in treatment outcomes.

These results should be considered in light of several limitations. Measures of substance use are self-reported and may be subject to error because of social desirability and/or recall bias. Comparing the two models, MBA is a marginal model across time and therefore does not allow conclusions about the timing of the use of different drug

combinations. In contrast, RMLCA specifically clusters the data on time patterns. This may be useful in evaluating predictors of treatment success due to the importance of a period of abstinence prior to induction on extended-release naltrexone and to a lesser extent, buprenorphine. Strengths of our study include the use of geographically diverse sites and large samples from three randomized clinical trials. In addition, we were able to establish a temporal relationship between identified RMLCA classes and MBA groups before treatment initiation and relapse outcomes after treatment initiation. Observing the pattern of polysubstance use prior to initiation of treatment may facilitate early identification of patients with unique needs and different probabilities of responding to treatment. Our ongoing work, using both traditional and modern machine learning methods to predict response to MOUD, will use these features to help identify who is best suited for different kinds of treatment.

Our data harmonization efforts highlight the need to standardize Timeline Followback questions. The fact that CTN-0030 did not gather data on unspecified, non-opioid substances was a missed opportunity to assess details of how other unexpected drugs may impact OUD. Differences in the breadth of drug categories, grouping hallucinogens with MDMA versus listing individual drugs as structured questions versus free text, may also limit the interpretability and clinical usefulness of our results. Furthermore, inconsistencies in the way in which substances are described or entered as free text (leading to more than a dozen spellings of "street buprenorphine" and half a dozen spellings of Adderall), and instances where respondents reported a set of drugs as a single free text "event" suggests a need for additional tools to support uniform Timeline Followback data collection. Future directions include tools that would automatically map clinical datasets, such as those obtained from medical records, to structured datasets such as Timeline Followback, providing a venue to replicate our findings on larger observational datasets.

In summary, this study suggests the presence of subgroups with distinct patterns of polysubstance use among individuals with OUD prior to MOUD treatment initiation. We reported the most prevalent polysubstance use combinations and identified six subgroups with different substance use behaviors during the 28 days leading up to MOUD initiation as well as associations between the effects of RMLCA classes and MBA top lift baskets to predict relapse in treatment. Our results suggest that patterns of substance use prior to treatment initiation may be useful in tailoring OUD interventions. Future studies should replicate these findings in larger datasets and different treatment contexts.

Role of funding source

Funding

Declaration of Competing Interest

No conflict declared.

CRediT authorship contribution statement

Yue Pan: Visualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Daniel J. Feaster:** Visualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Gabriel Odom:** Writing – review & editing, Data curation. **Laura Brandt:** Writing – review & editing. **Mei-Chen Hu:** Writing – review & editing, Data curation. **Roger D. Weiss:** Writing – review & editing. **John Rotrosen:** Writing – review & editing. **Andrew J. Saxon:** Writing – review & editing. **Sean X. Luo:** Visualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Raymond R. Balise:** Visualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing.

Acknowledgments

Supplementary materials

References

Aguinis, H., Forcum, L.E., Joo, H., 2013. Using market basket analysis in management research. J. Manag. 39 (7), 1799–1824.

Asparouhov, T., Muthén, B., 2014. Auxiliary variables in mixture modeling: using the BCH method in Mplus to estimate a distal outcome model and an arbitrary secondary model. Mplus Web Notes 21 (2), 1–22.

Ataiants, J., Roth, A.M., Mazzella, S., Lankenau, S.E., 2020. Circumstances of overdose among street-involved, opioid-injecting women: drug, set, and setting. Int. J. Drug Policy 78, 102691.

Carlsen, S.L., Lunde, L.H., Torsheim, T., 2020. Opioid and polydrug use among patients in opioid maintenance treatment. Subst. Abus. Rehabil. 11, 9–18.

Celeux, G., Soromenho, G., 1996. An entropy criterion for assessing the number of clusters in a mixture model. J. Classif. 13 (2), 195–212.

Cicero, T.J., Ellis, M.S., Kasper, Z.A., 2020a. Polysubstance use: a broader understanding of substance use during the opioid crisis. Am. J. Public Health 110 (2), 244–250.

Cicero, T.J., Ellis, M.S., Kasper, Z.A., 2020b. Polysubstance use: a broader understanding of substance use during the opioid crisis. Am. J. Public Health 110 (2), 244–250.

Collins, L.M., Lanza, S.T., 2009. Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences. Wiley, Hoboken, New Jersey.

Compton, W.M., Valentino, R.J., DuPont, R.L., 2021. Polysubstance use in the US opioid crisis. Mol. Psychiatry 26 (1), 41–50.

Connor, J.P., Gullo, M.J., Chan, G., Young, R.M., Hall, W.D., Feeney, G.F., 2013. Polysubstance use in cannabis users referred for treatment: drug use profiles, psychiatric comorbidity and cannabis-related beliefs. Front. Psychiatry 4, 79.

Connor, J.P., Gullo, M.J., White, A., Kelly, A.B., 2014. Polysubstance use: diagnostic challenges, patterns of use and health. Curr Opin. Psychiatry 27 (4), 269–275.

Feng, Z.D., McCulloch, C.E., 1996. Using bootstrap likelihood ratios in finite mixture models. J. R. Stat. Soc. Ser. B Methodol. 58 (3), 609–617.

Hahsler, M., 2017. ArulesViz: interactive visualization of association rules with R. R. J. 9 (2), 163.

Hahsler, M., Buchta, C., Gruen, B., Hornik, K., Hahsler, M.M., 2022. Package 'arules'.

Hahsler, M., Chelluboina, S., 2021. arulesViz: visualizing association rules and frequent itemsets. R package version 1.5-1.

Hahsler, M., Chelluboina, S., Hornik, K., Buchta, C., 2011. The arules R-package ecosystem: analyzing interesting patterns from large transaction data sets. J. Mach. Learn. Res. 12, 2021–2025.

Hassan, A.N., Le Foll, B., 2019. Polydrug use disorders in individuals with opioid use disorder. Drug Alcohol Depend. 198, 28–33.

Hernandez, I., He, M., Brooks, M.M., Zhang, Y., 2018. Exposure-response association between concurrent opioid and benzodiazepine use and risk of opioid-related overdose in medicare part D beneficiaries. JAMA Netw. Open 1 (2), e180919.

Hornik, K., Grün, B., Hahsler, M., 2005. arules-A computational environment for mining association rules and frequent item sets. J. Stat. Softw. 14 (15), 1–25.

Hser, Y.I., Huang, D., Saxon, A.J., Woody, G., Moskowitz, A.G., Ling, W., 2017. Distinctive trajectories of opioid use over an extended follow-up of patients in a multisite trial on buprenorphine + naloxone and methadone. J. Addict. Med. 11 (1), 63–69.

Jarlenski, M., Barry, C.L., Gollust, S., Graves, A.J., Kennedy-Hendricks, A., Kozhimannil, K., 2017. Polysubstance use among us women of reproductive age who use opioids for nonmedical reasons. Am. J. Public Health 107 (8), 1308–1310.

Lantz, B., 2019. Machine learning with r: expert techniques for predictive modeling, Machine Learning with R: Expert techniques For Predictive Modeling.

Lee, J.D., Friedmann, P.D., Kinlock, T.W., Nunes, E.V., Boney, T.Y., Hoskinson Jr, R.A., Wilson, D., McDonald, R., Rotrosen, J., Gourevitch, M.N., 2016. Extended-release naltrexone to prevent opioid relapse in criminal justice offenders. N. Engl. J. Med. 374 (13), 1232–1242.

Makarenko, I., Mazhnaya, A., Marcus, R., Pykalo, I., Madden, L., Filippovich, S., Dvoriak, S., Altice, F.L., 2018. Concurrent drug injection during opioid agonist treatment among people who inject drugs in Ukraine. J. Subst. Abus. Treat. 87, 1–8.

Northrup, T.F., Stotts, A.L., Green, C., Potter, J.S., Marino, E.N., Walker, R., Weiss, R.D., Trivedi, M., 2015. Opioid withdrawal, craving, and use during and after outpatient buprenorphine stabilization and taper: a discrete survival and growth mixture model. Addict. Behav. 41, 20–28.

Nylund, K.L., Asparouhov, T., Muthén, B.O., 2007. Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. Structural Equation Modeling: A Multidisciplinary Journal 14, 535–569.

Raffa, J.D., Grebely, J., Tossonian, H., Wong, T., Viljoen, M., Khara, M., Mead, A., McLean, M., Duncan, F., Petkau, A.J., DeVlaming, S., Conway, B., 2007. The impact of ongoing illicit drug use on methadone adherence in illicit drug users receiving treatment for HIV in a directly observed therapy program. Drug Alcohol Depend. 89 (2–3), 306–309.

Ruglass, L.M., Scodes, J., Pavlicova, M., Campbell, A.N.C., Fitzpatrick, S., Barbosa-Leiker, C., Burlew, K., Greenfield, S.F., Rotrosen, J., Nunes Jr., E.V., 2019. Trajectory classes of opioid use among individuals in a randomized controlled trial comparing extended-release naltrexone and buprenorphine-naloxone. Drug Alcohol Depend. 205, 107649.

Saxon, A.J., Ling, W., Hillhouse, M., Thomas, C., Hasson, A., Ang, A., Doraimani, G., Tasissa, G., Lokhnygina, Y., Leimberger, J., 2013. Buprenorphine/naloxone and methadone effects on laboratory indices of liver health: a randomized trial. Drug Alcohol Depend. 128 (1–2), 71–76.

Seth, P., Scholl, L., Rudd, R.A., Bacon, S., 2018. Overdose deaths involving opioids, cocaine, and psychostimulants - United States, 2015-2016. MMWR Morb. Mortal Wkly Rep. 67 (12), 349–358.

Sobell, L.C., Sobell, M.B., 1992. Timeline Follow-Back, Measuring Alcohol Consumption. Springer, pp. 41–72.

Soyka, M., Strehle, J., Rehm, J., Buhringer, G., Wittchen, H.U., 2017. Six-year outcome of opioid maintenance treatment in heroin-dependent patients: results from a naturalistic study in a nationally representative sample. Eur. Addict. Res. 23 (2), 97–105.

Votaw, V.R., McHugh, R.K., Vowles, K.E., Witkiewitz, K., 2020. Patterns of polysubstance use among adults with tranquilizer misuse. Subst. Use Misuse 1–10.

Vrieze, S.I., 2012. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Psychol. Methods 17 (2), 228.

Wagner, F., Martin, G., Winter, C., Plorer, D., Krause, D.L., Adorjan, K., Koller, G., Pogarell, O., 2018. Concomitant drug abuse of opioid dependent patients in maintenance treatment detected with a multi-target screening of oral fluid. The American journal on addictions 27 (5), 407–412.

Weiss, R.D., Potter, J.S., Fiellin, D.A., Byrne, M., Connery, H.S., Dickinson, W., Gardin, J., Griffin, M.L., Gourevitch, M.N., Haller, D.L., 2011. Adjunctive counseling during brief and extended buprenorphine-naloxone treatment for prescription opioid dependence: a 2-phase randomized controlled trial. Arch. Gen. Psychiatry 68 (12), 1238–1246.

Wenninger, S., Link, D., Lames, M., 2019. Data mining in elite beach volleyball – detecting tactical patterns using market basket analysis. Int. J. Comput. Sci. Sport 18 (2), 1–19.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D.A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., 2019. Welcome to the tidyverse. J. Open Source Softw. 4 (43), 1686.

Wickham, H., Miller, E., 2021. haven: import and Export 'SPSS', 'Stata' and 'SAS' Files. R package version 2.4.3.