Published in final edited form as:

Psychopharmacology (Berl). 2020 November ; 237(11): 3369–3381. doi:10.1007/s00213-020-05618-5.

Beyond abstinence and relapse: cluster analysis of drug-use patterns during treatment as an outcome measure for clinical trials

Leigh V. Panlilio^a, Samuel W. Stull^a, Jeremiah W. Bertz^a, Albert J. Burgess-Hull^a, William J. Kowalczyk^{a,1}, Karran A. Phillips^a, David H. Epstein^a, Kenzie L. Preston^a

^aIntramural Research Program, National Institute on Drug Abuse, 251 Bayview Blvd., Suite 200, Baltimore, MD, 21224, United States

Abstract

Rationale—Many people being treated for opioid use disorder continue to use drugs during treatment. This use occurs in patterns that rarely conform to well-defined cycles of abstinence and relapse. Systematic identification and evaluation of these patterns could enhance analysis of clinical trials and provide insight into drug use.

Objectives—To evaluate such an approach, we analyzed patterns of opioid and cocaine use from three randomized clinical trials of contingency management in methadone-treated participants.

Methods—Sequences of drug-test results were analyzed with unsupervised machine-learning techniques, including hierarchical clustering of categorical results (i.e., whether any samples were positive during each week) and *K*-means longitudinal clustering of quantitative results (i.e., the proportion positive each week). The sensitivity of cluster membership as an experimental outcome was assessed based on the effects of contingency management. External validation of clusters was based on drug craving and other symptoms of substance use disorder.

Results—In each clinical trial, we identified four clusters of use patterns, which can be described as opioid use, cocaine use, dual use (opioid and cocaine), and partial/complete abstinence. Different clustering techniques produced substantially similar classifications of individual participants, with strong above-chance agreement. Contingency management increased membership in clusters with lower levels of drug use and fewer symptoms of substance use disorder.

Conclusions—Cluster analysis provides person-level output that is more interpretable and actionable than traditional outcome measures, providing a concrete answer to the question of what clinicians can tell patients about the success rates of new treatments.

Terms of use and reuse: academic research for non-commercial purposes, see here for full terms. <http://www.springer.com/gb/open-access/authors-rights/aam-terms-v1>

¹Present address: Hartwick College, Department of Psychology, Oneonta, NY, 13820, United States

Publisher's Disclaimer: This Author Accepted Manuscript is a PDF file of an unedited peer-reviewed manuscript that has been accepted for publication but has not been copyedited or corrected. The official version of record that is published in the journal is kept up to date and so may therefore differ from this version.

Introduction

Agonist therapy for opioid use disorder is one of the great success stories of behavioral medicine: it improves quality of life and decreases drug use, morbidity and mortality (Bart 2012; Epstein et al. 2018; Mitchell et al. 2015; Salsitz and Wiegand 2016). However, despite the clear benefits and well-deserved first-line status of buprenorphine and methadone as agonist medications (Mattick et al. 2014; Volkow et al. 2014), their success rate can be well under 50% when assessed by strict metrics such as long-term cessation of drug use (Gottheil et al. 1993; Termorshuizen et al. 2005; Zhu et al. 2018).

For years, we and others have primarily reported results of clinical trials in terms of the group-level percentage of drug-negative urine samples or longest duration of abstinence, or in terms of treatment retention, but these measures do not adequately define the proportion of participants who respond well to treatment (Carroll et al. 2014; Donovan et al. 2012; Petry et al. 2005; Stitzer et al. 1993; Walker 2009). Levels of ongoing drug-related symptoms and quality-of-life measures supplement these measures and provide a more complete assessment of outcomes (Epstein et al. 2009; Gardner et al. 2019; Kiluk et al. 2019; Marsden et al. 2019; Stull et al. 2019). However, a thorough evaluation should also include a straightforward yet suitably nuanced assessment of drug use during the course of treatment, because elimination or reduction of this behavior is in itself an important goal (McCann et al. 2015; Volkow et al. 2018).

A seemingly natural approach would be to tally the proportion of patients who relapse. However, if relapse is conceived of as a lengthy descent into problematic drug use after a protracted period of abstinence (Gossop et al. 1989), then we have rarely seen a relapse in the thousands of agonist-treated outpatients who have taken part in our clinical trials—at least not among those who continue to provide data. Instead, we see sporadic lapses, with lapses defined as any instance of drug use during a quit (Gossop et al. 1989).

A promising approach is to systematically identify patterns of use that are associated with a better profile of psychological health. Taking this approach, the therapeutic effectiveness of a treatment could be gauged by how much it shifts people into a healthier use pattern, even if it is not abstinence. This approach aligns with draft guidance from the Food and Drug Administration (Center for Drug Evaluation and Research 2018) indicating that changes in drug-use patterns can be used as a clinical endpoint for assessing the effects of treatment for opioid use disorder, with a focus on the percentage of participants who respond to treatment rather than on group-level measures of use. The guidance states that “patterns other than complete abstinence could be used to define response” to treatment, and that investigators “should specify how the change in drug use pattern will be measured.”

Unfortunately, with the patterns of use typically seen in clinical trials of treatments for opioid use disorder, it is not always obvious where to draw the line between those who respond and those who do not. A logical way to address this difficulty is to categorize patterns of drug use with formal clustering procedures (Hastie et al. 2017; Wang 2014). In preliminary analyses not reported here, we analyzed drug-use patterns from our largest applicable data set, obtained from 309 patients treated with methadone or buprenorphine in a

series of natural-history studies. After finding that we could apply clustering techniques based on sequences of drug-test results to sort patients' outcomes into meaningful categories, we progressed to the current examination of how our approach could be used to report outcomes in randomized clinical trials. Our goal is not to provide a one-size-fits-all recommendation for policy or practice, because we doubt that any such recommendation is possible. Rather, our goals are (1) to give an example of how a clustering approach can be applied in a manner consistent with the FDA's draft guidance, and (2) to demonstrate how this approach can provide insight into the patterns of opioid and cocaine use that occur in people being treated with opioid-agonist therapy and how those patterns relate to other clinical outcomes. This approach has many advantages over traditional analyses, and we suggest that it would be beneficial in almost any longitudinal study of drug use, regardless of whether the research is FDA-related.

The analyses reported here focused on sequences of drug-test results from three of our previously published clinical trials in which methadone-maintained participants were randomly assigned to a contingency-management treatment group or a noncontingent control group, with drug tests three times per week (Epstein et al. 2003; Epstein et al. 2009; Kennedy et al. 2013). We identified basic patterns of opioid and cocaine use by applying unsupervised machine-learning techniques (Hastie et al. 2017; James et al. 2013) that detect clusters of data based only on the patterns themselves (as opposed to supervised techniques that classify patterns based on examples from already-existing data that contain a pre-specified label for each example case, such as "improved" versus "nonimproved"). The procedures we used were hierarchical clustering of categorical sequences (Magura et al. 1998; Milligan and Cooper 1987; Sun et al. 2012) and *K*-means longitudinal clustering of quantitative sequences (Genolini et al. 2015). The cluster assignments obtained with these techniques were then used as an outcome measure to conduct a secondary analysis of our contingency-management results, assessing the potential for these techniques to provide insight that would not be apparent from traditional percentage-positive or longest-abstinence outcome measures. Finally, to externally validate the clusters, we examined between-cluster differences in levels of craving and other symptoms of substance use disorders.

Methods

Data collection and preparation

Our prior publications described the experimental procedures in detail, so the descriptions here will be brief. We will refer to these studies by the protocol numbers assigned at the Intramural Research Program of the National Institute on Drug Abuse: Study 297 (Epstein et al. 2003), Study 326 (Epstein et al. 2009), and Study 390 (Kennedy et al. 2013). These studies also assessed other manipulations (methadone dose, group therapy), but they all were designed and previously analyzed in a way that provided between-group comparisons of contingency management versus a noncontingent control condition. All experiments were performed in accordance with protocols approved by the Addictions Institutional Review Board of the National Institutes of Health. All participants gave informed consent, and the privacy of participants is protected by a Certificate of Confidentiality from the Department of Health and Human Services, limiting the disclosure of identifiable, sensitive information.

Each participant received outpatient treatment with methadone at our Archway Clinic in Baltimore during the study. Based on self-report and urinalysis, all were using opioids and cocaine at the start of the study. Participants were expected to provide urine specimens for drug testing on Mondays, Wednesdays and Fridays. Since the tests detect use within the past 2–4 days, this schedule was designed to detect any use of opioids or cocaine during the week. Each study had a baseline phase (five weeks in Studies 297 and 326 and six weeks in Study 390) and an intervention phase (twelve weeks in Studies 297 and 326 and sixteen weeks in Study 390). During the intervention phase, participants received vouchers exchangeable for goods and services consistent with treatment goals. Participants were randomly assigned to receive vouchers either contingent on testing negative for drug use (abstinence reinforcement), or in a noncontingent fashion (regardless of their test results, but on an unpredictable schedule that simulated the temporal pattern of reward in the contingent group). Each of the three studies had a noncontingent group, and each had a contingent group in which providing cocaine-negative samples was reinforced. Study 326 had an additional “split” contingent group in which providing samples negative for opioids and/or cocaine was reinforced, with a partial reward for being negative for just one and a full reward for being negative for both.

For Studies 297, 326 and 390, respectively, N=166, N=210 and N=50, not including 12, 42 and 4 participants who were excluded from analysis because they were missing all scheduled urine specimens from 20% or more of the weeks from the intervention phase (a level of missingness defined *a priori* as being incompatible with meaningful assessment of use patterns). The number of participants included in the three respective studies were 81, 82 and 32 in the cocaine-contingent groups, 85, 49 and 18 in the noncontingent groups, and 79 in the split contingent group of Study 326. The hierarchical clustering procedure we used handles missing data by essentially treating “missing” as an additional categorical outcome (and ignoring any missing data after the last nonmissing data point in the sequence) (Gabadinho et al. 2011). The *K*-means longitudinal clustering procedure can only handle missing data (in this case, any weeks with all tests missing) by using imputation (Genolini et al. 2015). Various forms of imputation provide viable options for those who wish to apply *K*-means longitudinal clustering (or other procedures) to clinical-trials data, but they all have both advantages and disadvantages (Hedeker et al. 2007). Since there were substantial numbers of complete cases of weekly data in all groups, for the purposes of this demonstration we simply applied the *K*-means procedure only to complete cases: n=146, 192 and 43 for the three respective studies, with n=75, 78 and 27 in the cocaine-contingent groups; n=71, 43 and 16 in the noncontingent groups; and n=71 in the split contingent group. Two participants whose data were collected under the noncontingent condition to maintain blinding to dosing conditions in Study 390 were not included in the original analysis (Kennedy et al. 2013) but were included for the present analysis.

Additional outcome measures were obtained in Study 326 from a semi-structured interview at the end of the intervention phase, assessing DSM-IV symptoms for cocaine dependence, cocaine abuse, opioid dependence and opioid abuse during the previous 30 days (Miele et al. 2000). Levels of opioid and cocaine craving during the previous week were assessed every two weeks by questionnaire during the baseline and intervention phases in Studies 326 and 390, using a Likert scale (0–4) with higher ratings indicating more craving.

Clustering procedures

All cluster analyses were based on the sequences of urinalysis results, with one sequence for each participant. Since participants were randomly assigned to an experimental-treatment group only if they were already being treated with methadone and continuing to use cocaine during the baseline phase, clustering focused only on the intervention phase of each study; we thus avoided the complexity of interpreting sequences that spanned multiple phases, and we extended the generalizability of our cluster findings to trials without baselines. However, we also checked for baseline differences between the groups prior to randomization, by conducting cluster analysis of the baseline data.

For each urine drug test, five kinds of results were possible: positive only for opioids (Opi), positive only for cocaine (Coc), positive for both opioids and cocaine (Both), negative for both opioids and cocaine (Neg), or missing (i.e., no sample provided by the participant, or in rare cases a lab error). The results from the three tests within each week were combined to provide a categorical result for each participant *for each week*: Opi (at least one opioid-positive but no cocaine-positive or both-positive), Coc (at least one cocaine positive but no opioid-positive or both-positive), Both (at least one both-positive or at least one cocaine-positive and one-opioid positive), Neg (none positive), or missing (no urine sample provided for the week). Proportion-positive results were also determined for each week (proportion positive for cocaine and proportion positive for opioids among the non-missing tests). Expressing the data by week had two potential benefits: it removed cyclical effects of the day of the week by including weekdays and weekends in each result, and it reduced the impact of missing urine tests on the clustering algorithms. Although the information was not complete for weeks that had one or two tests missing, we used the partial information (weekly results) rather than discard or impute data missing from individual samples. Historically, the FDA seems to have favored (or at least approved) the practice of assuming that missed tests are drug-positive. However, in our analysis, we need to distinguish among three types of use (Opi, Coc, Both), and arbitrarily scoring missed urine tests as being positive for one of these types would substantially distort the data used to identify patterns. The possibility that tests were missed as a result of use itself is tempered by the fact that participants who repeatedly missed tests were automatically excluded from formal clustering and labeled as dropping out.

Agglomerative hierarchical clustering of the categorical results was conducted with the Ward method (Maechler et al. 2018), using distance scores determined by an optimal-matching algorithm that accounts for sequential patterns (Gabadinho et al. 2011; Studer and Ritschard 2016); unlike most other clustering procedures, it can recognize similar subsequences even if they are shifted in time, which could make it more sensitive to finegrained patterns. *K*-means longitudinal clustering (Genolini et al. 2015) was conducted based simultaneously on the proportion of positive results for cocaine and the proportion of positive results for opioids within each week. These proportions are related to the intensity of use (i.e., frequency and dose) and therefore contain more information than the categorical measure.

Multinomial logistic regression was used to assess whether the randomly assigned contingency conditions in each of the three studies influenced participants' cluster

membership. Cluster membership was modeled as a function of contingency, using the noncontingent group as the reference condition. The results were expressed as odds ratios, which indicate the size and direction of the effects. This logistic analysis also assessed the effects of contingency management on dropping out of the study, defined for this purpose as failing to meet the 20% criterion described above. Cohen's kappa quantified agreement among the three clustering procedures with respect to whether each participant was in a Cocaine, Opioid, Both or Negative cluster. Mixed models were used to analyze craving levels during the baseline and intervention phases. Analysis of variance (ANOVA) was applied to the number of DSM-IV symptoms reported at the end of the intervention phase (referring to the last 30 days of the phase), and χ^2 was used to analyze the number of participants who met DSM-IV criteria for dependence (at least 3 of 7 dependence symptoms) and/or abuse (at least 1 of 4 abuse symptoms). Paired comparisons from these analyses were controlled for familywise error using the Holm procedure ($\alpha=.05$). Effect sizes were assessed with Cramer's V for χ^2 and Cohen's d for ANOVA. Since standard effect-size measures are not established for mixed-model effects with numerator degrees of freedom >1 , the sizes of such effects were conservatively expressed as Cohen's d based on the paired comparison with the highest p value less than .05.

All statistical analyses were conducted with R software (R Core Team 2018), including the packages "TraMineR" (Gabadinho et al. 2011) for optimal matching and cluster exploration, "cluster" (Maechler et al. 2018) for hierarchical clustering, "kml3D" (Genolini et al. 2015) for K-means longitudinal clustering, "ggplot2" (Wickham 2009) for heatmaps, "nnet" (Venables and Ripley 2002) for multinomial logistic regression, and "nlme" for mixed models (Pinheiro et al. 2019).

Results

Clustering results

Clustering procedures—including the unsupervised machine-learning techniques used here, and others such as finite-mixture models (Dong et al. 2019; Genberg et al. 2011; Hser et al. 2017; Jones et al. 2001; Lanza et al. 2010; Liu et al. 2010; Mikolajczyk et al. 2014; Nagin and Odgers 2010) and hidden-Markov models (DeSantis and Bandyopadhyay 2011; DeSantis et al. 2009; Maruotti and Rocci 2012; Shirley et al. 2010) that are based on more strict assumptions about the underlying determinants of the data—require the investigator to choose the number of divisions that makes the most sense for the task at hand (Anderlucci and Hennig 2014; Hastie et al. 2017; McVicar and Anydike-Danes 2002; Nagin and Odgers 2010). This decision is guided by considerations that include interpretability of the clusters, consistency of patterns within clusters, and distinctness of patterns between clusters (James et al. 2013; Milligan and Cooper 1987). Hierarchical clustering produces a tree-like dendrogram that offers solutions at several levels; at each level the clusters are in some respect similar to each other and different from those in the other clusters. For classifying urinalysis results that have four types of outcome (Coc, Opi, Both, Neg), it is natural to prefer a solution that identifies four clusters, with each cluster corresponding to one type of outcome. However, this preference should be outweighed if the four-cluster solution has an

obvious deficiency, such as clusters that are highly similar to each other (indicating too many clusters), or clusters composed of distinct subclusters (indicating not enough clusters).

The clusters identified by the hierarchical and *K*-means procedures are shown in Fig. 1 (expressed as the categorical results that were the basis for hierarchical clustering) and in Fig. 2 (expressed as the use-intensity results that were the basis for *K*-means clustering). Within each clustering technique, the sequences of drug-test results from each study fit well into four naturally interpretable categories that have face validity as types of treatment outcomes. Specifically, within each clustering method and study, each cluster was distinguished by having a higher frequency of Coc, Opi, Both or Neg results compared to the other three clusters.

The clusters identified by the two methods were surprisingly similar, given that they were based on somewhat different kinds of information. Specifically, clustering of categorical results was based only on whether any opioid or cocaine use occurred during each week (Fig. 1), and clustering by the proportion of positive results in each week was based on intensity of use (Fig. 2). Heatmaps of the categorical results highlight the similarity of findings across methods. For example, in Fig. 1: (1) the Both clusters were most internally consistent (with many participants using both drugs during every week of the study); (2) the Cocaine clusters had more cocaine-only weeks relative to the other clusters but also had many weeks in which both cocaine and opioids were used; and (3) all clusters except the Negative clusters had few weeks without either cocaine or opioid use.

The intensity results (Fig. 2) were also similar between clusters obtained with the two different methods. That is, clusters that were based on categorical or intensity results (indicated as “Hierarchical” and “K-means” in the figure, respectively) showed similar intensity profiles (comparing down the pair of columns for each cluster type in Fig. 2), with clear differences between the heatmaps across cluster types (comparing the clusters across each row in Fig. 2). A few sequences were ambiguous, and agreement between the procedures was not perfect. Nonetheless, when agreement was quantified (see Supplementary Material, Fig. S1), the methods mostly agreed on which cluster a participant should be in, with Cohen’s kappa values of 0.76 to 0.93, indicating correspondence above chance and ranging from substantial to strong, with 84% agreement across all three studies.

Collapsing the categorical results across participants within each cluster can provide further insight into cluster-level behavior, including the time course and overall level of drug use during the intervention (Fig. 3). These data clearly show the consistency within clusters, between studies, and between clustering methods. In the Opioid clusters, dual use of cocaine and opioids within the same week tended to decrease over time. In the Negative clusters, the number of participants who were abstinent during the week increased over time.

Clustering of baseline urinalysis data

Clustering of the urinalysis results obtained before participants were randomized to treatment groups (i.e., the sequences shown to the left of the gray vertical lines in each panel of Figs 1 and 2) indicated that the experimental groups did not differ from each other before starting treatment. Specifically, within each of the three experiments, the hierarchical and *K*-

means procedures each identified two clusters during the baseline period, a baseline-Both cluster and a baseline-Cocaine cluster. Under hierarchical clustering, the percentage of participants in the baseline-Both cluster were: 80.2 for the cocaine-contingent group and 76.5 in the noncontingent group in Study 297; 87.8 in the cocaine contingent group, 84.8 in the split-contingency group and 81.6 in the noncontingent group in Study 326; and 90.6 in the cocaine-contingent group and 94.4 in the noncontingent group in Study 390. Under *K*-means clustering, the percentage of participants in the baseline-Both cluster were: 83.3 for the cocaine-contingent group and 86.5 in the noncontingent group in Study 297; 65.4 in the cocaine contingent group, 64.8 in the split-contingency group and 67.4. in the noncontingent group in Study 326; and 66.7 in the cocaine-contingent group and 62.5 in the noncontingent group in Study 390. Logistic regression of these data indicated that the randomized groups did not differ from each other with respect to the relative number of participants in each cluster prior to randomization; for hierarchical clustering in studies 297, 326 and 390, respectively, the p-values for differences between groups were $> .55$, $.33$, and $.64$; for *K*-means clustering, the p-values were $> .59$, $.77$, and $.78$.

Cluster-based analysis of the effects of contingency management

Visual inspection of Fig. 4A indicates that the proportion of participants in the Negative cluster was larger in the contingent conditions than the noncontingent condition in each study, regardless of which method was used to perform the clustering. This increased membership in the Negative cluster appears to have been due to the contingencies' shifting people away from the Both cluster and the Cocaine cluster (since the latter clusters tended to be larger in the noncontingent groups than in the contingent groups).

The data shown in Fig. 4A were used to calculate the number needed to treat (NNT) in order to shift one participant into the Negative cluster. For the cocaine contingency in Studies 297, 326 and 390, respectively, NNT values were 4.6, 25.8, and 5.8 under hierarchical clustering and 4.8, 6.7 and 12.5 under *K*-means clustering; for the split contingency, NNT values were 5.6 under hierarchical clustering and 6.1 under *K*-means clustering. Since all contingencies focused on cocaine, but only the split contingency focused on opioids, NNT values were also calculated for shifting one participant into a *non-cocaine* cluster (i.e., the Neg or Opi cluster); for the cocaine contingency in Studies 297, 326 and 390, respectively, these NNT values were 3.8, 3.7, and 4.1 under hierarchical clustering and 4.2, 4.0 and 5.7 under *K*-means clustering; for the split contingency, they were 4.5 under hierarchical clustering and 4.9 under *K*-means clustering.

The effects of contingency on cluster membership were also assessed with multinomial logistic regression (Fig. 4B), in which odds ratios less than one indicate that contingency management decreased the odds of being in a cluster other than the Negative cluster. (An intercept with an odds ratio higher or lower than one simply indicates that, not considering the contingency condition, the overall likelihood of being in a non-negative cluster was high or low, respectively.) This analysis generally shows that the odds of being in the Both cluster or Cocaine cluster were significantly lower in the contingent groups in all three studies. Contingency management also decreased the odds of dropping out of the study by failing to provide scheduled urine samples. Overall, the *K*-means clusters tended to have the smallest

confidence intervals for the odds ratios, and they also had the overall profile of results that most closely matched the expected effects of the contingency manipulation. The omnibus χ^2 values for these logistic regressions were all significant or marginally significant (see Supplementary Material, Table 1) with medium to large effect sizes. The Supplementary Material (Fig. S2 and “Cluster-based analysis of the effects of methadone dose or group therapy”) provides multinomial logistic regression results for the methadone dose and group-therapy manipulations in the three studies, showing results consistent with the analyses published earlier, but providing person-level, cluster-related details about outcomes that were not readily-apparent from the previous analyses.

Addiction-related symptoms as a function of cluster membership

Under DSM-IV, symptoms were assessed separately for dependence and abuse, and criteria based on these symptoms were used for diagnosis; unlike the current DSM-5, craving was not included as a symptom (but drug-related legal problems were included). The number of symptoms and the proportion of participants meeting the diagnostic criteria were generally lowest in the Negative cluster and highest in the Both cluster (Fig. 5A). The omnibus ANOVA and χ^2 results for the DSM-IV data were all statistically significant (see Supplementary Material, Tables 2 and 3), with medium to large effect sizes.

Results from the craving questionnaires (Fig. 5B) also showed better profiles for the Negative cluster, which had low levels of craving for cocaine and opioids and showed significant improvement during the intervention phase in each panel of the figure. In contrast, the Both cluster had the highest levels of craving during intervention and did not improve relative to baseline. The Opioid and Cocaine clusters showed improvement over time, and they had lower levels of craving than the Both cluster, especially for the drug used less frequently (cocaine in the Opioid cluster, opioids in the Cocaine cluster). In a few instances, there were significant differences between clusters at baseline, suggesting that cluster membership was sometimes influenced by pre-intervention differences between participants; nonetheless, these differences increased during the intervention, consistent with the conclusion that contingencies affected cluster membership. The omnibus F values from the mixed models of craving were all significant (see Supplementary Material, Table 4), with medium to large effect sizes.

Discussion

The finding that there were four distinct patterns of opioid and cocaine use during treatment with methadone provides detailed, person-level information that extends the earlier reports from these three studies. In the original reports, the drug-test data were analyzed and described strictly in terms of the longest duration of abstinence and the group-level percentage or odds of drug-negative tests, showing that *on average* contingency management decreased drug use. In contrast, the current analyses: (1) identify each participant as a member of a specific cluster with a particular pattern of drug use; (2) demonstrate and quantify the influence of contingency management on cluster membership; and (3) reveal clear differences between the clusters with respect to craving and other symptoms experienced during treatment. These results demonstrate the feasibility of applying a

clustering approach to data from clinical trials of addiction therapies, providing a rational, objectively grounded means of drawing the line between one pattern of drug use and another. Specifically, contingency management made it more likely that a person would adopt one of the use patterns that was associated with better psychological health.

The results were similar across the clustering methods we used, providing support for the choice of four outcome clusters for these data. Unsupervised machine-learning techniques and other clustering techniques are used when the “ground truth” cannot be directly assessed. They provide objective input for decision-making, but they are not a substitute for decision-making. The obtained clusters must be evaluated by whether they are useful and externally valid. Accordingly, we do not claim that our procedures or their results are ideal, but rather that they are effective and reliable. We used the optimal-matching hierarchical and *K*-means longitudinal clustering procedures because they are straightforward, nonparametric, appropriate for analysis of sequences, and freely available through open-source software packages, but there are other procedures that could produce results similar or complementary to the ones reported here. A variety of finite-mixture models have been used to identify long-term trajectories of drug use (using data typically aggregated by year or month) (Caudy et al. 2014; Dong et al. 2019; Genberg et al. 2011; Hser et al. 2008; Hser et al. 2007; Hser et al. 2017; Liu et al. 2010; Wojciechowski 2019). These models are based on parametric assumptions that the intensity or categorical data are generated from a mixture of distributions of a certain form (e.g., Gaussian, Poisson, multinomial). Hidden-Markov models (DeSantis and Bandyopadhyay 2011; DeSantis et al. 2009; Maruotti and Rocci 2012; Shirley et al. 2010) and latent-transition models (Lanza et al. 2010; Mikolajczyk et al. 2014) can categorize each individual as being in a certain state at each point in time (e.g., having a certain probability of using drugs, but also having a chance of transitioning to another state associated with a different probability of using drugs). Mixture models and other model-based clustering methods that use expectation-maximization algorithms can handle missing data if they are assumed to be missing at random (Nagin and Odgers 2010). Alternatively, imputation of missing data can be used with any modeling strategy, and sensitivity analysis can be performed to determine how the results are affected by imputing missing drug tests by various means, such as scoring them as positive for drug use, scoring them as negative for drug use, carrying the previous observation forward, or using multiple-imputation methods based on various assumptions concerning probabilistic relationships between missingness and drug use in the available data (Hammon and Zinn 2020; Hedeker et al. 2007; Jackson et al. 2014; Zhang et al. 2018).

The results obtained here demonstrate that the clustering approach is viable for assessing the effects of treatments with both categorical and intensity data. One reason that we chose to analyze categorical results is that the same technique could be applied to data that are not aggregated over time, such as series of individual urinalysis tests (e.g., the raw data that were converted to weekly categorical results and weekly intensity results for this paper). For ease of interpretation, we clustered the baseline and intervention phases separately. However, if applied simultaneously to multiple phases of a study, these same clustering techniques would be capable of identifying and distinguishing between clusters that represent phase-related changes in pattern of use, such as increasing versus decreasing over time — patterns that might not be distinguished by a traditional percent-positive measure. Previous natural-

history studies of long-term drug use based on hierarchical clustering (Dobler-Mikola et al. 2005; Magura et al. 1998) and growth mixture models (Caudy et al. 2014; Dong et al. 2019; Genberg et al. 2011; Hser et al. 2008; Hser et al. 2007; Hser et al. 2017; Liu et al. 2010; Wojciechowski 2019) have identified patterns of increased or decreased use over time, or switching from one drug to another. When clusters exhibiting such patterns are identified, it would be natural to consider a cluster that shifts from more use to less use as showing improvement, and a cluster that shifts from less use to more use as showing deterioration (Stull et al. 2019). Although these various clustering techniques take different approaches, they are all capable of producing valid results. Applying more than one method, as we did here, can (1) increase confidence that the results are robust and (2) help identify cases that are ambiguous (e.g., Supplementary Material, Fig. S1).

Reporting results of clinical trials in terms of cluster membership can facilitate the calculation of clinically useful and FDA-preferred indicators of effect size such as NNT, especially when it is not immediately clear where to draw the line between “responder” and “nonresponder.” Based on the median NNT across all of the studies and clustering methods studied here, 4.1 methadone-treated patients would need to be treated with contingency management rather than a noncontingent control condition for one additional patient to become mostly or completely abstinent from cocaine over a 12–16 week period. Thus, although the clustering methods themselves are more complex than tabulation of group means or percentages, their ultimate output can be more interpretable, concretely answering the question, “What can clinicians expect if they implement a new treatment?” One way to answer this question is to look at the breakdown of outcomes in those who received contingency management: about 19% became largely abstinent, 11% mainly used opioids only, 25% continued using cocaine but decreased their use of opioids, and 45% continued using both cocaine and opioids. DSM symptoms and craving improved in all of these groups except those who continued to use both drugs. For a more graphic and detailed impression of the different use patterns, a heatmap such as Fig. 1 provides a clear, intuitive depiction of what the outcomes look like in individual patients.

The potential advantages and drawbacks to clustering both stem from its exploratory nature. We consider the exploratory aspects of clustering to be highly beneficial, forcing examination of the results at different levels (i.e., numbers of clusters) and providing a coherent summary of whatever kinds of patterns are in the data. The flipside of this is that — no matter what method is used to identify clusters— boundaries are inherently fuzzy, and, as we noted above, the need for judgment in determining the number of clusters is inescapable. Since the boundaries are based only on the outcomes that are observed, there is no guarantee that a set of four clusters obtained in one study will have the same features as four clusters obtained in another study. However, applying the clustering procedures with reasonable care will provide an accurate and detailed reflection of whatever findings are observed and can provide valuable insight. For example, in each of the three studies analyzed here —where participants were only eligible if they were receiving methadone for opioid use and were also using cocaine at the start of the study— the cluster designated as the cocaine cluster had a substantial amount of dual cocaine-opioid use. In studies with a different kind of population (e.g., in our preliminary analyses of natural-history data, mentioned in the Introduction), the cocaine cluster of a four-cluster solution will be more monotonic, with

much less dual use. This is not a problem as long as any comparisons between the results of these studies take note of such differences. In fact, unexpected differences between populations can provide valuable information that might otherwise go unrecognized (Stull et al. 2019).

Although we clustered based on patterns of use, we strongly believe that accompanying symptoms of substance-use disorders or quality-of-life measures should be included as part of a rational, integrated approach to outcome assessment. Ideally, “to support a drug use pattern as a response-defining threshold,” there should be “evidence from clinical trials, longitudinal observational studies, or other sources of information to show that such reduction in drug use predicts clinical benefit (i.e., better health outcomes or psychosocial function)” (Center for Drug Evaluation and Research 2018). We have provided such evidence by showing clear differences between clusters in the levels of craving and other symptoms, and in whether they met diagnostic thresholds for drug abuse and dependence.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding and Disclosure

This study was supported by the Intramural Research Program of NIH, NIDA. The authors report no conflicts of interest.

References

- Anderlucci L, Hennig C (2014) The Clustering of Categorical Data: A Comparison of a Model-based and a Distance-based Approach. *Communications in Statistics - Theory and Methods* 43: 704–721.
- Bart G (2012) Maintenance medication for opiate addiction: the foundation of recovery. *Journal of addictive diseases* 31: 207–25. [PubMed: 22873183]
- Carroll KM, Kiluk BD, Nich C, DeVito EE, Decker S, LaPaglia D, Duffey D, Babuscio TA, Ball SA (2014) Toward empirical identification of a clinically meaningful indicator of treatment outcome: features of candidate indicators and evaluation of sensitivity to treatment effects and relationship to one year follow up cocaine use outcomes. *Drug and alcohol dependence* 137: 3–19. [PubMed: 24556275]
- Caudy MS, Tang L, Wooditch A, Taxman FS (2014) Short-term trajectories of substance use in a sample of drug-involved probationers. *J Subst Abuse Treat* 46: 202–13. [PubMed: 23953166]
- Center for Drug Evaluation and Research (2018) Opioid use disorder: Endpoints for demonstrating effectiveness of drugs for medication-assisted treatment -- Guidance for industry Food and Drug Administration, US Department of Health and Human Services
- DeSantis SM, Bandyopadhyay D (2011) Hidden Markov models for zero-inflated Poisson counts with an application to substance use. *Statistics in medicine* 30: 1678–94. [PubMed: 21538455]
- DeSantis SM, Bandyopadhyay D, Back SE, Brady KT (2009) Non-treatment laboratory stress- and cue-reactivity studies are associated with decreased substance use among drug-dependent individuals. *Drug and alcohol dependence* 105: 227–33. [PubMed: 19717245]
- Dobler-Mikola A, Hattenschwiler J, Meili D, Beck T, Boni E, Modestin J (2005) Patterns of heroin, cocaine, and alcohol abuse during long-term methadone maintenance treatment. *J Subst Abuse Treat* 29: 259–65. [PubMed: 16311178]

- Dong H, Hayashi K, Singer J, Milloy MJ, DeBeck K, Wood E, Kerr T (2019) Trajectories of injection drug use among people who use drugs in Vancouver, Canada, 1996–2017: growth mixture modeling using data from prospective cohort studies. *Addiction* 114: 2173–2186. [PubMed: 31328354]
- Donovan DM, Bigelow GE, Brigham GS, Carroll KM, Cohen AJ, Gardin JG, Hamilton JA, Huestis MA, Hughes JR, Lindblad R, Marlatt GA, Preston KL, Selzer JA, Somoza EC, Wakim PG, Wells EA (2012) Primary outcome indices in illicit drug dependence treatment research: systematic approach to selection and measurement of drug use end-points in clinical trials. *Addiction* 107: 694–708. [PubMed: 21781202]
- Epstein DH, Hawkins WE, Covi L, Umbricht A, Preston KL (2003) Cognitive-behavioral therapy plus contingency management for cocaine use: findings during treatment and across 12-month follow-up. *Psychol Addict Behav* 17: 73–82. [PubMed: 12665084]
- Epstein DH, Heilig M, Shaham Y (2018) Science-Based Actions Can Help Address the Opioid Crisis. *Trends in pharmacological sciences* 39: 911–916. [PubMed: 30343726]
- Epstein DH, Schmittner J, Umbricht A, Schroeder JR, Moolchan ET, Preston KL (2009) Promoting abstinence from cocaine and heroin with a methadone dose increase and a novel contingency. *Drug and alcohol dependence* 101: 92–100. [PubMed: 19101098]
- Gabadinho A, Ritschard G, Müller NS, Studer M (2011) Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software* 40: 1–37.
- Gardner RA, Epstein DH, Preston KL, Phillips KA (2019) Defining and Predicting Opioid and Cocaine Treatment Response. *Subst Use Misuse*: 1–10.
- Genberg BL, Gange SJ, Go VF, Celentano DD, Kirk GD, Mehta SH (2011) Trajectories of injection drug use over 20 years (1988–2008) in Baltimore, Maryland. *Am J Epidemiol* 173: 829–36. [PubMed: 21320867]
- Genolini C, Alacoque X, Sentenac M, Arnaud C (2015) kml and kml3d: R Packages to Cluster Longitudinal Data. 2015 65: 34.
- Gossop M, Green L, Phillips G, Bradley B (1989) Lapse, relapse and survival among opiate addicts after treatment. A prospective follow-up study. *The British journal of psychiatry : the journal of mental science* 154: 348–53. [PubMed: 2597836]
- Gottheil E, Sterling RC, Weinstein SP (1993) Diminished illicit drug use as a consequence of long-term methadone maintenance. *Journal of addictive diseases* 12: 45–57.
- Hammon A, Zinn S (2020) Multiple imputation of binary multilevel missing not at random data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 69: 547–564.
- Hastie T, Tibshirani R, Friedman J (2017) *Unsupervised Learning In: Diggle P, Gather U, Zeger S (eds) The elements of statistical learning: data mining, inference, and prediction.* Springer, New York, pp 485–585
- Hedeker D, Mermelstein RJ, Demirtas H (2007) Analysis of binary outcomes with missing data: missing = smoking, last observation carried forward, and a little multiple imputation. *Addiction* 102: 1564–73. [PubMed: 17854333]
- Hser YI, Huang D, Brecht ML, Li L, Evans E (2008) Contrasting trajectories of heroin, cocaine, and methamphetamine use. *Journal of addictive diseases* 27: 13–21.
- Hser YI, Huang D, Chou CP, Anglin MD (2007) Trajectories of heroin addiction: growth mixture modeling results based on a 33-year follow-up study. *Evaluation review* 31: 548–63. [PubMed: 17986707]
- Hser YI, Huang D, Saxon AJ, Woody G, Moskowitz AL, Matthews AG, 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: 63–69. [PubMed: 27898496]
- Jackson D, White IR, Mason D, Sutton S (2014) A general method for handling missing binary outcome data in randomized controlled trials. *Addiction* 109: 1986–93. [PubMed: 25171441]
- James G, Witten D, Hastie T, Tibshirani R (2013) *Unsupervised learning In: Casella G, Fienberg S, Olkin I (eds) An introduction to statistical learning – with applications in R.* Springer, New York, pp 373–418
- Jones BL, Nagin DS, Roeder K (2001) A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories. *Sociological Methods and Research* 29.

- Kennedy AP, Phillips KA, Epstein DH, Reamer DA, Schmittner J, Preston KL (2013) A randomized investigation of methadone doses at or over 100 mg/day, combined with contingency management. *Drug and alcohol dependence* 130: 77–84. [PubMed: 23195924]
- Kiluk BD, Fitzmaurice GM, Strain EC, Weiss RD (2019) What defines a clinically meaningful outcome in the treatment of substance use disorders: reductions in direct consequences of drug use or improvement in overall functioning? *Addiction* 114: 9–15. [PubMed: 29900624]
- Lanza ST, Patrick ME, Maggs JL (2010) Latent Transition Analysis: Benefits of a Latent Variable Approach to Modeling Transitions in Substance Use. *J Drug Issues* 40: 93–120. [PubMed: 20672019]
- Liu LC, Hedeker D, Segawa E, Flay BR (2010) Evaluation of longitudinal intervention effects : An example of latent growth mixture models for ordinal drug-use outcomes. *Journal of Drug Issues* 40: 27–44.
- Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K (2018) cluster: Cluster Analysis Basics and Extensions, R package version 2.07–1
- Magura S, Kang SY, Nwakeze PC, Demsky S (1998) Temporal patterns of heroin and cocaine use among methadone patients. *Subst Use Misuse* 33: 2441–67. [PubMed: 9781824]
- Marsden J, Tai B, Ali R, Hu L, Rush AJ, Volkow N (2019) Measurement-based care using DSM-5 for opioid use disorder: can we make opioid medication treatment more effective? *Addiction* 114: 1346–1353. [PubMed: 30614096]
- Maruotti A, Rocci R (2012) A mixed non-homogeneous hidden Markov model for categorical data, with application to alcohol consumption. *Statistics in medicine* 31: 871–86. [PubMed: 22302505]
- Mattick RP, Breen C, Kimber J, Davoli M (2014) Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence. *The Cochrane database of systematic reviews*: Cd002207. [PubMed: 24500948]
- McCann DJ, Ramey T, Skolnick P (2015) Outcome Measures in Medication Trials for Substance Use Disorders. *Current Treatment Options in Psychiatry* 2: 113–121.
- McVicar D, Anydike-Danes M (2002) Predicting successful and unsuccessful transitions from school to work by using sequence methods. *Journal of the Royal Statistical Society A* 165: 317–334.
- Miele GM, Carpenter KM, Smith Cockerham M, Trautman KD, Blaine J, Hasin DS (2000) Substance Dependence Severity Scale (SDSS): reliability and validity of a clinician-administered interview for DSM-IV substance use disorders. *Drug and alcohol dependence* 59: 63–75. [PubMed: 10706976]
- Mikolajczyk RT, Horn J, Prins M, Wiessing L, Kretzschmar M (2014) Trajectories of injecting behavior in the Amsterdam Cohort Study among drug users. *Drug and alcohol dependence* 144: 141–7. [PubMed: 25248472]
- Milligan GW, Cooper MC (1987) Methodology Review: Clustering Methods. *Applied Psychological Measurement* 11: 329–354.
- Mitchell SG, Gryczynski J, Schwartz RP, Myers CP, O'Grady KE, Olsen YK, Jaffe JH (2015) Changes in Quality of Life following Buprenorphine Treatment: Relationship with Treatment Retention and Illicit Opioid Use. *Journal of psychoactive drugs* 47: 149–57. [PubMed: 25950595]
- Nagin DS, Odgers CL (2010) Group-based trajectory modeling in clinical research. *Annu Rev Clin Psychol* 6: 109–38. [PubMed: 20192788]
- Petry NM, Alessi SM, Marx J, Austin M, Tardif M (2005) Vouchers versus prizes: contingency management treatment of substance abusers in community settings. *J Consult Clin Psychol* 73: 1005–14. [PubMed: 16392974]
- Pinheiro J, Bates D, DebRoy S, Sarkar D (2019) Linear and Nonlinear Mixed Effects Models, R package version 3.1–143
- R Core Team (2018) R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria
- Salsitz E, Wiegand T (2016) Pharmacotherapy of Opioid Addiction: “Putting a Real Face on a False Demon”. *Journal of medical toxicology : official journal of the American College of Medical Toxicology* 12: 58–63. [PubMed: 26567033]
- Shirley KE, Small DS, Lynch KG, Maisto SA, Oslin DW (2010) Hidden Markov models for alcoholism treatment trial data. *The Annals of Applied Statistics* 4: 366–395.

- Stitzer ML, Iguchi MY, Kidorf M, Bigelow GE (1993) Contingency management in methadone treatment: The case for positive incentives. NIDA Research Monograph Series 137.
- Studer M, Ritschard G (2016) What matters in differences between life trajectories: a comparative review of sequence dissimilarity measures. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 179: 481–511.
- Stull SW, Panlilio LV, Moran LM, Schroeder JR, Bertz JW, Epstein DH, Preston KL, Phillips KA (2019) The chippers, the quitters, and the highly symptomatic: A 12-month longitudinal study of DSM-5 opioid- and cocaine-use problems in a community sample. *Addictive behaviors* 96: 183–191. [PubMed: 31108264]
- Sun J, Bi J, Chan G, Oslin D, Farrer L, Gelernter J, Kranzler HR (2012) Improved methods to identify stable, highly heritable subtypes of opioid use and related behaviors. *Addictive behaviors* 37: 1138–44. [PubMed: 22694982]
- Termorshuizen F, Krol A, Prins M, Geskus R, van den Brink W, van Ameijden EJ (2005) Prediction of relapse to frequent heroin use and the role of methadone prescription: an analysis of the Amsterdam Cohort Study among drug users. *Drug and alcohol dependence* 79: 231–40. [PubMed: 16002032]
- Venables WN, Ripley BD (2002) *Modern Applied Statistics with S* Springer
- Volkow ND, Frieden TR, Hyde PS, Cha SS (2014) Medication-assisted therapies--tackling the opioid-overdose epidemic. *The New England journal of medicine* 370: 2063–6. [PubMed: 24758595]
- Volkow ND, Woodcock J, Compton WM, Throckmorton DC, Skolnick P, Hertz S, Wargo EM (2018) Medication development in opioid addiction: Meaningful clinical end points. *Science translational medicine* 10.
- Walker R (2009) Retention in treatment--indicator or illusion: an essay. *Subst Use Misuse* 44: 18–27. [PubMed: 19137480]
- Wang J (2014) Mixture modeling: a useful analytical approach for drug use studies. *Journal of Substance Abuse & Alcoholism* 2: 1009–1010.
- Wickham H (2009) *ggplot2: elegant graphics for data analysis* Springer, New York
- Wojciechowski TW (2019) Developmental Trajectories of Opioid Use Among Juvenile Offenders: An Epidemiological Examination of Group Characteristics and Criminological Risk Factors. *Subst Use Misuse* 54: 1203–1213. [PubMed: 30767596]
- Zhang Y, Luo X, Le CT, Ahluwalia JS, Thomas JL (2018) Analysis of self-report and biochemically verified tobacco abstinence outcomes with missing data: a sensitivity analysis using two-stage imputation. *BMC Med Res Methodol* 18: 170. [PubMed: 30563473]
- Zhu Y, Evans EA, Mooney LJ, Saxon AJ, Kelleghan A, Yoo C, Hser YI (2018) Correlates of LongTerm Opioid Abstinence After Randomization to Methadone Versus Buprenorphine/Naloxone in a Multi-Site Trial. *Journal of neuroimmune pharmacology : the official journal of the Society on NeuroImmune Pharmacology* 13: 488–497. [PubMed: 30094695]

Fig. 1.

Heatmaps showing categorical results from the four-cluster solutions obtained with each of the two clustering procedures: hierarchical clustering of categorical results and *K*-means clustering of the proportion of results positive for opioids and cocaine. Each small, colored cell shows the result for one participant for one week, and each row of cells shows the entire sequence of results for one participant. Results for each week are defined as being all negative (“Neg”) or as having at least one positive result for opioids but none for cocaine (“Opi”), at least one positive result for cocaine but none for opioids (“Coc”), or at least one positive result for opioids and one positive result for cocaine (“Both”) during the week. White cells indicate that all results were missing for the week. Each contiguous block of cells represents one cluster. Each horizontal row of panels shows the results for one of the three studies. Each vertical column of panels shows clusters that were similar across the three studies, with clusters named by their distinguishing characteristic (Both, Opi, Coc or Neg). Participant numbers (i.e., order on the ordinate axis) are not consistent across methods. Within each hierarchical cluster, the order of participants from bottom to top is determined by the clustering procedure (dendrogram). Within each *K*-means cluster, the

order of participants is determined by when they joined the study. Vertical gray lines mark the division between the baseline phase and the intervention phase. Clustering was based on results from the intervention phase.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 2.

Heatmaps showing use-intensity results for cocaine (“Coc”) and opioids (“Opi”) from the same four-cluster solutions shown in Fig. 1. Intensity is defined as the proportion of results positive for cocaine (“Coc use”) and opioids (“Opi use”) during each week. Details of the maps parallel those of Fig. 1, except that each cluster is represented by a pair of panels (one for cocaine and one for opioids), and the darkness of each cell represents the proportion positive (as indicated in the Scale). The order of participants within each cluster is the same

as in Figure 1. Missing data are not explicitly indicated but correspond to the white cells in Fig. 1.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Categorical results by cluster during each week of treatment

Fig. 3.

Categorical urinalysis results collapsed across participants within each cluster during each week of the intervention phase. Each panel depicts the test results of one cluster from Fig. 1, collapsed to show the proportion of participants with each type of test result in each week.

Panel A: Results from hierarchical clustering of categorical results. *Panel B:* Results from K-means longitudinal clustering of the proportions of positive results.

Fig. 4.

Effects of experimental condition (contingency management) on cluster membership. Vouchers were given contingent on cocaine-negative drug tests, contingent on both opioid-negative and cocaine-negative tests, or regardless of test results (noncontingent). *Panel A:* The proportion of participants in each cluster in each study, as determined by the hierarchical and *K*-means clustering methods. “Dropout” indicates participants who had 20% or more of weeks with no drug-test results (hierarchical clustering) or at least one week with no drug-test results (*K*-means clustering). *Panel B:* Forest plots showing results of

multinomial logistic regressions of cluster based on the experimental condition. Circles indicate odds ratio for a contingency and cluster combination relative to the reference condition (Noncontingent group) and the reference outcome (Negative cluster). Black circles represent odds ratios significantly different ($p < .05$) from 1 (with 1 indicating no effect). Bars represent 95% confidence intervals. See main text for interpretation of negative and positive odds ratios.

Fig. 5.

Craving and other substance use disorder symptoms as a function of cluster. *Panel A:* DSM-IV symptoms and diagnoses for cocaine and opioid dependence and abuse in each cluster in Study 326. The number of symptoms endorsed by participants (two left columns) are shown as mean and standard error, with dark bars representing symptoms related to dependence and light bars representing symptoms related to abuse. Diagnoses (two right columns) are shown as the within-cluster proportion of participants having a symptom count at or above the DSM-defined criterion level. Within each set of four bars, letters to the right of a bar indicate

that the cluster was statistically different ($p < .05$) from the Negative cluster (“*n*”), Cocaine cluster (“*c*”), Opioid cluster (“*o*”) and/or the Both cluster (“*b*”). *Panel B*: Self-reported craving levels for opioids and cocaine from questionnaires given every two weeks during the baseline and intervention phases in each cluster from Studies 326 and 390. Craving levels are shown as mean and standard error in each phase. Within each frame, letters beside a point indicate that (during the same phase of the study) the cluster was statistically different ($p < .05$) from the Negative cluster (“*n*”), Cocaine cluster (“*c*”), Opioid cluster (“*o*”) and/or the Both cluster (“*b*”). An asterisk indicates that a cluster showed a significant decrease in craving from baseline to intervention.